THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Fitting parton distribution data with multiplicative normalization uncertainties

OPEN ACCESS

Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties

The NNPDF Collaboration:

Richard D. Ball[1], Luigi Del Debbio[1], Stefano Forte[2], Alberto Guffanti[3],
José I. Latorre[4], Juan Rojo[2] and Maria Ubiali[1,5].

[1] *School of Physics and Astronomy, University of Edinburgh,*
JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland
[2] *Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano,*
Via Celoria 16, I-20133 Milano, Italy
[3] *Physikalisches Institut, Albert-Ludwigs-Universität Freiburg,*
Hermann-Herder-Straße 3, D-79104 Freiburg i. B., Germany
[4] *Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona,*
Diagonal 647, E-08028 Barcelona, Spain
[5] *Center for Particle Physics and Phenomenology (CP3),*
Université Catholique de Louvain,
Chemin du Cyclotron, B-1348 Louvain-la-Neuve, Belgium

Abstract

The extraction of robust parton distribution functions with faithful errors requires a careful treatment of the uncertainties in the experimental results. In particular, the data sets used in current analyses each have a different overall multiplicative normalization uncertainty that needs to be properly accounted for in the fitting procedure. Here we consider the generic problem of performing a global fit to many independent data sets each with a different overall multiplicative normalization uncertainty. We show that the methods in common use to treat multiplicative uncertainties lead to systematic biases. We develop a method which is unbiased, based on a self–consistent iterative procedure. We then apply our generic method to the determination of parton distribution functions with the NNPDF methodology, which uses a Monte Carlo method for uncertainty estimation.

November 2009

Contents

1 Introduction

The interpretation of forthcoming experiments at the Large Hadron Collider requires the development of precision statistical analysis tools. One context where this is especially apparent is the determination of the parton distribution functions of the proton, which are obtained by global analysis of existing data sets (see [1] for a review). These PDFs, together with their associated uncertainties, should be without theoretical prejudice, and should be associated with a genuine statistical confidence level. In a series of recent papers [2–7] the NNPDF collaboration has adopted a method based on a Monte Carlo estimate of uncertainties [8] that allows one to propagate the uncertainty on the experimental data to the fitted PDFs, and to any other quantity which depends on them. The PDFs are parametrized by neural networks, containing large numbers of free parameters, sufficient to ensure that the resulting ensembles of fitted PDFs are free from any bias due to assumptions about the underlying functional form. Within this context, in which one aims typically at

accuracies at the percent level in physical observables, more subtle problems in data analysis become relevant. One such problem is the treatment of normalization uncertainties.

When combining data sets from independent experiments it is necessary to take account of the overall normalization uncertainty associated with each experiment: an experiment with large normalization uncertainty should contribute less to the fit than one with a small uncertainty. Normalization uncertainties are usually multiplicative, in the sense that each data point within the set has a normalization uncertainty proportional to the measurement at that point. All these normalization uncertainties are however correlated across the whole set of data points. Fitting the data with the usual Hessian method, using the complete covariance matrix, leads to a substantial bias in the fitted value due to the fact that smaller data points are assigned a smaller uncertainty than larger ones [9]. This problem is usually avoided by using the "penalty trick": the normalization of each data set is treated as a free parameter to be determined during the fitting, within a range restricted by the quoted experimental uncertainty. While this method indeed gives correct results when fitting data from a single experiment, we will show below that it remains biased when used in fits which combine several different data sets.

In this paper we develop a treatment of normalization uncertainties which is always free from bias, even when fitting to many different data sets. While the original motivation of this study comes from PDF determination, all our results are of general relevance whenever a quantity has to be determined from data which are affected by multiplicative uncertainties. Hence most of our discussion will be completely general, and we will only consider the issue of PDF determination in the end, as an example of the application of our proposed method.

The paper is constructed as follows: in Sect. 2 we review the Hessian and Monte Carlo methods for the extraction of a quantity from a set of experimental measurements, and how they might be implemented when there are normalization uncertainties. In Sect. 3 we review the well–known fact that using the full covariance matrix for the treatment of normalization uncertainties leads to a bias for a single experiment, and we show that further biases arise when combining several experiments. In Sect. 4 we show that the so–called "penalty trick" method commonly used to overcome this bias, while working well for a single experiment, still leads to biased results when used to combine results from several experiments. In Sect. 5 we attempt to construct a self–consistent covariance matrix, but find that this too leads to biased results when used to combine several different experiments. In Sect. 6 we cure the defect in this method by determining the self–consistent covariance matrix through an iterative procedure using as a starting point the result of a previous fit: we show that this method (the "t_0-method") is completely unbiased and rapidly convergent. Finally in Sect. 7 we discuss the treatment of normalization errors in PDF fitting: after a brief summary of the methods used currently, we discuss the effect of our new t_0-method on the NNPDF1.2 parton fit. Specifically, we demonstrate explicitly the practicality and convergence of the method, and use it to quantify the effect of including the normalization uncertainties in the determination of PDFs.

The language of the paper is that of a particle physicist (and thus similar in approach to e.g. Ref. [9]), not a professional statistician. We thus make no pretence of mathematical rigour, but hope nonetheless that our results will be of practical use to physicists interested in fitting large datasets.

2 Hessian and Monte Carlo

2.1 Hessian Methods

Consider a simple but typical experimental situation in which we have n measurements m_i of a single theoretical quantity t, with experimental uncertainties given by a covariance matrix $(\text{cov})_{ij}$: typically this takes the form

$$(\text{cov})_{ij} = \delta_{ij}\sigma_i^2 + \sum_{k=1}^{n} \bar{\sigma}_{ik}\bar{\sigma}_{kj}, \tag{1}$$

where the σ_i are uncorrelated uncertainties (typically obtained as the sum in quadrature of statistical and uncorrelated systematic uncertainties) and $\bar{\sigma}_{ik}$ are correlated (typically systematic) uncertainties. All experimental uncertainties are generally assumed to be Gaussian. Then the least squares estimate for t is given by minimizing the χ^2 function

$$\chi^2(t) = \sum_{i,j=1}^{n} (t - m_i)(\text{cov}^{-1})_{ij}(t - m_j), \tag{2}$$

and thus by

$$t = \frac{\sum_{i,j=1}^{n}(\text{cov}^{-1})_{ij}m_j}{\sum_{i,j=1}^{n}(\text{cov}^{-1})_{ij}}. \tag{3}$$

The variance of t, V_{tt} is found through

$$V_{tt} = \left(\tfrac{1}{2}\frac{\partial^2\chi^2}{\partial t^2}\right)^{-1} = \frac{1}{\sum_{i,j=1}^{n}(\text{cov}^{-1})_{ij}}. \tag{4}$$

Of course when there are several quantities t to be determined, the variances are given by the diagonal elements V_{tt} of the matrix determined by inversion of the Hessian matrix of second derivatives of χ^2: hence the name of the method. Here the simplest case of only one quantity t will be sufficient to illustrate the points we wish to make. The situation is yet more complicated when fitting parton distribution functions: then, t is actually some nontrivial but calculable function (such as a cross section or structure function) of the (many) fitted parameters which describe the shape of the underlying parton distribution functions which is being determined. Again this complication is irrelevant to the issues to be discussed here, so we will ignore it.

In the very simplest case in which the measurements m_i have completely uncorrelated (e.g. purely statistical) uncertainties σ_i,

$$(\text{cov})_{ij} = (\text{cov}_0)_{ij} = \sigma_i^2\delta_{ij}, \tag{5}$$

so

$$\chi^2(t) = \sum_{i=1}^{n} \frac{(t - m_i)^2}{\sigma_i^2}, \tag{6}$$

whence at the minimum $t = w$ and $V_{tt} = \Sigma^2$, where

$$w \equiv \Sigma^2 \sum_{i=1}^{n} \frac{m_i}{\sigma_i^2}, \tag{7}$$

3

and Σ^2 is given by

$$\frac{1}{\Sigma^2} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}. \tag{8}$$

Thus the theoretical quantity t is given by the average value of the measurements m_i weighted by the inverses of the variances σ_i^2, and the inverse of the variance of t is likewise the average of the inverses of the variances σ_i^2.

When all the variances are equal, $\sigma_i = \sigma$, and $w = \bar{m}$, where

$$\bar{m} = \frac{1}{n} \sum_{i=1}^{n} m_i, \tag{9}$$

i.e. the unweighted average of the measurements. In this limiting case, the determination Eq. (3) of t is manifestly seen to be unbiased (and in particular it tends to the true value in the limit of large number of measurements): no measurement is preferred, and $\Sigma^2 = \sigma^2/n$, so the variance is reduced by a factor of $1/n$, due to there being n independent measurements of the same quantity. In what follows we will use the words "biased" and "unbiased" to describe estimates of the mean which pass or fail this simple test.

In the opposite extreme, if one of the variances, say σ_i^2, becomes very large compared to the others, the contribution of the measurement m_i to w and Σ^2 becomes very small, so this measurement decouples from the rest as it must. In what follows we shall use freedom from bias (as defined above) and decoupling as two criteria to assess the usefulness of a particular method of determining t and its variance.[1]

2.2   The Monte Carlo Method

A different way of determining a theoretical quantity from a set of measurements is to construct a Monte Carlo representation of the data. First, the n data points are associated to n random variables M_i, normally distributed around the averages m_i according to the covariance matrix $(\text{cov})_{ij}$. Then, an ensemble $\{M_i\}$ of replicas of the data is constructed: these by construction satisfy

$$\langle M_i \rangle = m_i, \qquad \langle M_i M_j \rangle = m_i m_j + (\text{cov})_{ij}, \tag{10}$$

where $\langle \rangle$ denotes averaging over the set of replicas. Finally an ensemble $\{T\}$ of replicas of the theoretical quantity t is determined from the data replicas $\{M_i\}$, by minimizing a suitable error function $E_{\text{MC}}(T)$. The mean value and variance of t (and indeed any other function of t) may then be found simply by averaging over the replicas:

$$\text{E}[t] = \langle T \rangle, \quad \text{Var}[t] = \langle T^2 \rangle - \langle T \rangle^2. \tag{11}$$

This method becomes advantageous when the determination of some function of t is called for: once the ensemble of replicas $\{T\}$ has been found, error propagation to any function of t, no matter how complicated, may be performed by simply averaging over

---

[1]Note that the terms "bias" and "decoupling" as used here are not quite the same as the technical definitions of consistency and bias used by statisticians, which distinguish more carefully between results obtained with finite size samples and those when the sample size becomes infinite.

replicas. This is especially useful in situations in which t is multidimensional, or a nontrivial function of some underlying theoretical quantity (such as a PDF).

Clearly, the features of the result obtained with this method depend on the choice of error function $E_{MC}(T)$ which determines the the ensemble of replicas $\{T\}$ from the data replicas $\{M_i\}$. In the simple situation discussed in the previous section we may choose $E_{MC}(T) = \chi^2(T)$, where the χ^2 function is given by Eq. (2). Then

$$T = \frac{\sum_{i,j=1}^{n}(\mathrm{cov}^{-1})_{ij}M_j}{\sum_{i,j=1}^{n}(\mathrm{cov}^{-1})_{ij}}, \tag{12}$$

so that, using Eq. (11),

$$\mathrm{E}[t] = \frac{\sum_{i,j=1}^{n}(\mathrm{cov}^{-1})_{ij}m_j}{\sum_{i,j=1}^{n}(\mathrm{cov}^{-1})_{ij}}, \tag{13}$$

while

$$\begin{aligned}
\mathrm{Var}[t] &= \frac{\sum_{i,j,k,l=1}^{n}(\mathrm{cov}^{-1})_{ij}(\mathrm{cov}^{-1})_{kl}(\langle M_j M_l\rangle - \langle M_j\rangle\langle M_l\rangle)}{\left(\sum_{i,j=1}^{n}(\mathrm{cov}^{-1})_{ij}\right)^2} \\
&= \frac{1}{\sum_{i,j=1}^{n}(\mathrm{cov}^{-1})_{ij}}, \tag{14}
\end{aligned}$$

in agreement with Eq. (3) and Eq. (4) found using the Hessian method.

2.3 Normalization Uncertainties

The two methods discussed in the previous two sections cover most situations of uncorrelated and correlated errors found in combining experimental data. However problems arise when data sets have overall multiplicative uncertainties, such as normalization uncertainties: this is due to biases arising from the rescaling of errors [9–11]. The effect of these biases in the naive application of the Hessian method can be very severe, as will be discussed in the following section. Here we consider normalization uncertainties in the Monte Carlo method, which is possibly more straightforward.

First consider the situation in which all the data come from a single experiment, with a single overall normalization uncertainty s, assumed to be Gaussian. In the Monte Carlo method the normalization uncertainty is taken into account by multiplying the data by a random factor N, which is normally distributed around 1 with variance s. Assuming N to be uncorrelated with M_i

$$\langle N\rangle = 1, \qquad \langle N^2\rangle = 1 + s^2, \qquad \langle f(N)g(M_i)\rangle = \langle f(N)\rangle\langle g(M_i)\rangle. \tag{15}$$

When fitting to the replicas, the error function will depend on NM_i, but the weights of the different measurements will be unchanged, since an overall rescaling of the data should not affect the relative weight of the measurements in the fit. Thus we now take

$$E_{MC}(T) = \sum_{i,j=1}^{n}(T - NM_i)(\mathrm{cov}^{-1})_{ij}(T - NM_j) \tag{16}$$

5

where $(\text{cov})_{ij}$ is the same covariance matrix used when there was no normalization uncertainty. The result of the minimization of Eq. (16) is the same as Eq. (12), but with an overall factor of N

$$T = N\frac{\sum_{i,j=1}^{n}(\text{cov}^{-1})_{ij}M_j}{\sum_{i,j=1}^{n}(\text{cov}^{-1})_{ij}}, \tag{17}$$

so $\text{E}[t]$ is the same as in Eq. (13), while now

$$\begin{aligned}\text{Var}[t] &= \frac{\sum_{i,j,k,l=1}^{n}(\text{cov}^{-1})_{ij}(\text{cov}^{-1})_{kl}(\langle N^2\rangle\langle M_j M_l\rangle - \langle N\rangle^2\langle M_j\rangle\langle M_l\rangle)}{\left(\sum_{i,j=1}^{n}(\text{cov}^{-1})_{ij}\right)^2} \\ &= \frac{1+s^2}{\sum_{i,j=1}^{n}(\text{cov}^{-1})_{ij}} + s^2\text{E}[t]^2.\end{aligned} \tag{18}$$

The first term on the right–hand side of Eq. (18) is the same as was found previously in Eq. (14), but with an extra factor of $1 + s^2$, while the second term is the contribution to the variance from the normalization uncertainty. This is as expected: indeed, the variance of a product of random variables is

$$\text{Var}[NM] = \text{E}[N]^2\text{Var}[M] + \text{E}[M]^2\text{Var}[N] + \text{Var}[N]\text{Var}[M], \tag{19}$$

where the last term is usually neglected because, being a product of two variances, it corresponds to a higher order moment of the probability distribution for NM.

For the simple case of uncorrelated statistical measurement errors Eq. (5),

$$E_{\text{MC}}(T) = \sum_{i=1}^{n}\frac{(T - NM_i)^2}{\sigma_i^2}, \tag{20}$$

whence

$$T = N\Sigma^2\sum_{i=1}^{n}\frac{M_i}{\sigma_i^2}, \tag{21}$$

so $\text{E}[t] = w$, Eq. (7), while

$$\text{Var}[t] = \Sigma^2 + s^2w^2 + s^2\Sigma^2. \tag{22}$$

Clearly then this method gives correct unbiased results for the case of a single experiment with normalization uncertainty.

Let us now consider a slightly more complex situation, where each of the measurements m_i comes from a different experiment, and also has an independent normalization uncertainty s_i. Here and henceforth when discussing this situation we will neglect possible correlations between these measurements with independent uncertainties, which are usually absent if the measurements are obtained from independent experiments; however, the inclusion of such correlations is straightforward and it does not affect our subsequent results. Following the same line of reasoning as above, we want to derive the fitted value for t and its variance in the Monte Carlo approach. We thus introduce independent normally distributed random variables N_i to represent the normalization uncertainties, each with mean one, variance s_i^2, and uncorrelated to each other and to the M_i:

$$\langle N_i\rangle = 1, \qquad \langle N_i N_j\rangle = 1 + s_i^2\delta_{ij}, \qquad \langle f(N_i)g(M_j)\rangle = \langle f(N_i)\rangle\langle g(M_j)\rangle. \tag{23}$$

6

The difficulty now is to choose an appropriate error function. The simplest choice would be

$$E_{\mathrm{MC}}(T) = \sum_{i=1}^{n} \frac{(T - N_i M_i)^2}{\sigma_i^2},$$ (24)

Minimizing with respect to T and averaging over the replicas then gives $\mathrm{E}[t] = w$ as before, but

$$
\begin{aligned}
\mathrm{Var}[t] &= \Sigma^4 \sum_{i,j=1}^{n} (\langle N_i N_j \rangle \langle M_i M_j \rangle - \langle N_i \rangle \langle N_j \rangle \langle M_i \rangle \langle M_j \rangle)/\sigma_i^2 \sigma_j^2, \\
&= \Sigma^2 + \Sigma^4 \sum_{i=1}^{n} s_i^2 (m_i^2 + \sigma_i^2)/\sigma_i^4.
\end{aligned}
$$ (25)

When all the experiments have the same normalization uncertainty, $s_i = s$, these results are not so unreasonable: $\mathrm{E}[t] = w$, while

$$\mathrm{Var}[t] = (1 + s^2)\Sigma^2 + s^2 \Sigma^4 \sum_{i=1}^{n} \frac{m_i^2}{\sigma_i^4}.$$ (26)

However Eq. (24) is clearly incorrect in general because differences in the normalization uncertainties s_i are not taken account of in the weighting of the different measurements. In particular if one of the experiments has a relatively large normalization uncertainty, it still contributes to the mean, but spoils the measurement by giving a very large contribution to the variance.

Therefore, results found using the Monte Carlo method with the error function Eq. (24) are unbiased when the normalization uncertainties are equal, but do not satisfy the criterion of decoupling. It follows that when we have more than one experiment, and in particular when we wish to include experiments with a large overall normalization uncertainty, we need to choose a better error function than Eq. (24), which incorporates differences in the normalization uncertainties s_i. We are thus led to consider error functions built using the full covariance matrix, including normalization uncertainties, and thus rather closer to the χ^2-function.

3   The d'Agostini Bias

We saw in the previous section that when we are combining different experiments with independent and different normalization uncertainties, it is necessary to incorporate these differences into the χ^2-function or error function used in the fitting procedure. This is true both for the Hessian method and for the Monte Carlo method. In the previous section we have shown that this is easily done in the case of a single experiment, but when several experiments must be combined the simplest choice of error function Eq. (24) leads to results which do not satisfy the decoupling criterion. In this section we will see that in the Hessian case the simplest choice of error function leads to results which are severely biased even in the case of a single experiment.

Specifically, we consider the case in which normalization uncertainties are included by using as an error function the χ^2-function computed using the full covariance matrix

including normalizations:

$$\chi_m^2(t) = \sum_{i,j=1}^{n} (t - m_i)(\mathrm{cov}_m^{-1})_{ij}(t - m_j). \tag{27}$$

As in the previous section we consider two cases separately: when we have n measurements all made within a single experiment, and thus with a common normalization uncertainty, so

$$(\mathrm{cov}_m)_{ij} = (\mathrm{cov})_{ij} + s^2 m_i m_j, \tag{28}$$

and then when each of the n measurements is made in an independent experiment, all uncorrelated, and in particular with different normalization uncertainties, so

$$(\mathrm{cov}_m)_{ij} = (\sigma_i^2 + s_i^2 m_i^2)\delta_{ij}. \tag{29}$$

Note that the more realistic case in which there are n_{exp} experiments each with n_{dat}^i measurements can be built from these two simpler examples by first combining the many measurements in each individual experiment together into one measurement, using Eq. (28), and then combining the experiments using Eq. (29), so these two cases should suffice to illustrate all the issues involved.

3.1   One experiment

Consider first a very simple model of a single experiment with only two data points. The covariance matrix Eq. (28) is then simply

$$(\mathrm{cov}_m)_{ij} = \begin{pmatrix} \sigma_1^2 + s^2 m_1^2 & s^2 m_1 m_2 \\ s^2 m_1 m_2 & \sigma_2^2 + s^2 m_2^2 \end{pmatrix} \tag{30}$$

so the χ^2 Eq. (27) is

$$\chi_m^2(t) = \frac{(t - m_1)^2(\sigma_2^2 + m_2^2 s^2) + (t - m_2)^2(\sigma_1^2 + m_1^2 s^2) - 2(t - m_1)(t - m_2)m_1 m_2 s^2}{\sigma_1^2 \sigma_2^2 + (m_1^2 \sigma_2^2 + m_2^2 \sigma_1^2)s^2}. \tag{31}$$

Minimizing this χ^2-function with respect to t gives after a straightforward calculation the result

$$t = \frac{m_1/\sigma_1^2 + m_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2 + (m_1 - m_2)^2 s^2/\sigma_1^2 \sigma_2^2} = \frac{w}{1 + (m_1 - m_2)^2 s^2/\Sigma^2}, \tag{32}$$

where w is the weighted mean Eq. (7) with $n = 2$.

It follows that when $m_1 \neq m_2$ and $s \neq 0$ the result for t has a downward shift. That this shift is clearly a bias can be seen for instance by considering the simple case $\sigma_1 = \sigma_2 = \sigma$. Then Eq. (32) gives

$$t = \frac{\bar{m}}{1 + 2r^2 s^2 \bar{m}^2/\sigma^2} = \bar{m}(1 - 2r^2 s^2 \bar{m}^2/\sigma^2 + O(r^4)). \tag{33}$$

where we have defined

$$\bar{m} \equiv \tfrac{1}{2}(m_1 + m_2), \qquad r \equiv \frac{m_1 - m_2}{m_1 + m_2}. \tag{34}$$

Thus simply minimizing the χ^2 derived from the correlated covariance matrix Eq. (28) leads to a central value which is shifted downwards: for a sufficiently large s^2/σ^2 one can get an average which is lower than either of the two values which are being averaged, so one must conclude that the result is biased.

This bias gets worse as the number of data points increases [9, 11, 12]. For n data m_i, with statistical uncertainties σ_i, the covariance matrix given by Eq. (28) and Eq. (5) has inverse

$$(\text{cov}_m^{-1})_{ij} = \frac{\delta_{ij}}{\sigma_i^2} - \frac{m_i m_j}{\sigma_i^2 \sigma_j^2} \frac{s^2}{1 + s^2 m^2/\Sigma^2}. \tag{35}$$

where

$$m^2 \equiv \Sigma^2 \sum_{i=1}^{n} \frac{m_i^2}{\sigma_i^2}. \tag{36}$$

The χ^2-function Eq. (27) is then minimized when

$$t = \frac{\sum_{i,j=1}^{n} (\text{cov}_m^{-1})_{ij} m_j}{\sum_{i,j=1}^{n} (\text{cov}_m^{-1})_{ij}} = \frac{w}{1 + r^2 s^2 w^2/\Sigma^2} \tag{37}$$

where w is defined in Eq. (7) while r is defined through

$$m^2 - w^2 = \Sigma^2 \sum_{i=1}^{n} \frac{(m_i - w)^2}{\sigma_i^2} \equiv r^2 w^2. \tag{38}$$

So again we have a downwards bias unless $s^2 = 0$ or all the measurements m_i are equal. Note further that when the data are consistent and n is large, r^2 is simply given by the variance Σ^2 of the measurements: $r^2 w^2 \simeq n\Sigma^2$. The bias is thus by a factor $1/(1 + ns^2)$, which will become arbitrarily large as the number of data points increases.

The origin of the bias is clear: smaller values of m_i have a smaller normalization uncertainty $m_i s$, and are thus preferred in the fit. Several examples of situations where this leads to absurd results may be found in Ref. [9]. The variance of t is afflicted by the same downward bias:

$$V_{tt} = \frac{\Sigma^2 + s^2 w^2 (1 + r^2)}{1 + r^2 s^2 w^2/\Sigma^2}. \tag{39}$$

We will henceforth refer to this as the "d'Agostini bias", after Ref. [9, 11] where it was studied and explained.

3.2   More than one experiment

In the second example with n distinct experiments the d'Agostini bias is much milder. With the covariance matrix Eq. (29), the χ^2-function is

$$\chi_m^2(t) = \sum_{i=1}^{n} \frac{(t - m_i)^2}{\sigma_i^2 + m_i^2 s_i^2}, \tag{40}$$

which is minimized when

$$t = \frac{\sum_{i=1}^{n} \frac{m_i}{\sigma_i^2 + s^2 m_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + s^2 m_i^2}}. \tag{41}$$

9

To exhibit the bias, consider again the case when all statistical uncertainties are equal, $\sigma_i = \sigma$: then on expanding in powers of

$$r^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{(m_i - \bar{m})^2}{\bar{m}^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{m_i^2 - \bar{m}^2}{\bar{m}^2}, \tag{42}$$

with $\bar{m}$ given by Eq. (9) we find

$$t = \bar{m}\left(1 - 2r^2\frac{s^2\bar{m}^2}{\sigma^2 + s^2\bar{m}^2} + O(r^4)\right). \tag{43}$$

The origin of the bias is the same as in Eqs. (33,37), and indeed for two data points the biases are approximately the same: however for n independent experiments, the bias Eq. (43) is stable for large n. In this case the variance is given by (again expanding in r)

$$V_{tt} = \tfrac{1}{n}(\sigma^2 + s^2\bar{m}^2(1 + r^2)) + O(r^4), \tag{44}$$

which is as expected, since $\bar{m}^2(1 + r^2) = \frac{1}{n}\sum_{i=1}^{n}m_i^2$.

4   The Penalty Trick

The standard [9,11,12] way to include normalization uncertainties in the Hessian approach while avoiding the d'Agostini bias consists of including the normalizations of the data n_i as parameters in the fit, with penalty terms to fix their estimated value close to one with variance s_i^2. In this section we will discuss this widely used method, which we will refer to as the "penalty trick": we will see that while it gives correct results for a single experiment, when used to combine results from several experiments it is actually still biased.

4.1   One Experiment

We first consider a single experiment, with covariance matrix Eq. (5), but now with an overall normalization uncertainty with variance s^2. The value of t is then obtained by minimizing the error function

$$E_{\text{Hess}}(t, n) = \sum_{i=1}^{n}\frac{(t/n - m_i)^2}{\sigma_i^2} + \frac{(n-1)^2}{s^2}. \tag{45}$$

where the last term is called the penalty term. The parameters t and n are then determined by minimizing this error function: minimizing with respect to t gives $t = nw$, with w as defined in Eq. (7), while minimization with respect to n fixes $n = 1$, so for the central values the result is the same as in the Monte Carlo approach.[2]

---

[2] Note that to obtain an unbiased result from this approach, the factor n must rescale the theory, not the data: if instead of Eq. (45) we took

$$E_{\text{Hess}} = \sum_{i=1}^{n}\frac{(t - nm_i)^2}{\sigma_i^2} + \frac{(n-1)^2}{s^2}.$$

we would get a result with a strong downward bias similar to that in Eq. (37) [9,11].

10

In order to compute the error on the fitted quantity in this approach, we need to evaluate the Hessian matrix:

$$V^{-1} = \frac{1}{2} \begin{pmatrix} \frac{\partial^2 \chi^2}{\partial t^2} & \frac{\partial^2 \chi^2}{\partial t \partial n} \\ \frac{\partial^2 \chi^2}{\partial n \partial t} & \frac{\partial^2 \chi^2}{\partial n^2} \end{pmatrix} = \frac{1}{\Sigma^2} \begin{pmatrix} 1 & -t \\ -t & \Sigma^2/s^2 + t^2 \end{pmatrix} . \tag{46}$$

The covariance matrix is obtained by inverting V^{-1}. In this way one recovers

$$V_{tt} = \Sigma^2 + s^2 w^2 . \tag{47}$$

This is the same result as Eq. (22) obtained with the Monte Carlo approach, apart from the cross-correlation term between the variances of the measurements and the variances of the normalization, akin to the last term in Eq. (19).

For the case of a single experiment, we have thus recovered within the Hessian approach the Monte Carlo result of Sect. 2.3: in this simple case the Hessian and Monte Carlo methods are (almost) equivalent. It is straightforward to generalize this equivalence to a general covariance matrix Eq. (1).

4.2 More than one experiment

Let us now turn to the more complex situation where we have several data points from different experiments, and thus with independent normalizations, $n_i = 1 \pm s_i$. In this case, the Monte Carlo result Eq. (25) is unbiased (in the sense that it gives an unbiased average over the data when all uncertainties are equal), but it does not satisfy the requirement of decoupling. We will now show that the penalty trick leads to a result which does satisfy decoupling, but is biased when all uncertainties are equal.

Following the same line of reasoning as for the single experiment above, we set up the error function

$$E_{\text{Hess}}(t, n_i) = \sum_{i=1}^{n} \frac{(t/n_i - m_i)^2}{\sigma_i^2} + \sum_{i=1}^{n} \frac{(n_i - 1)^2}{s_i^2} , \tag{48}$$

where now we have a separate penalty term for each of the normalizations to be fitted. The minimum is obtained for

$$t = \frac{\sum_{i=1}^{n} \frac{m_i}{n_i \sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{n_i^2 \sigma_i^2}}, \tag{49}$$

$$n_i = 1 + \frac{s_i^2 t}{n_i^2 \sigma_i^2} \left(\frac{t}{n_i} - m_i \right) . \tag{50}$$

These $n + 1$ equations are now complicated nonlinear relations which must be solved for t and n_i. A general analytic solution is probably impossible, and it seems very likely that for a large number of experiments the number of solutions will grow rapidly, making it difficult to select the correct one. However it is possible to find solutions for certain special cases, which are sufficient to show that the approach is biased.

First we specialize to the case of only two experiments: to explore the bias we again assume that $\sigma_1 = \sigma_2 \equiv \sigma$ and $s_1 = s_2 = s$, as in Sect. 3.1. Adding and subtracting the two equations for n_1 and n_2, and substituting the equation for t, we find

$$n_1^2 + n_2^2 = n_1 + n_2 , \qquad n_1^2 - n_2^2 = -2\Delta n_1 n_2 , \tag{51}$$
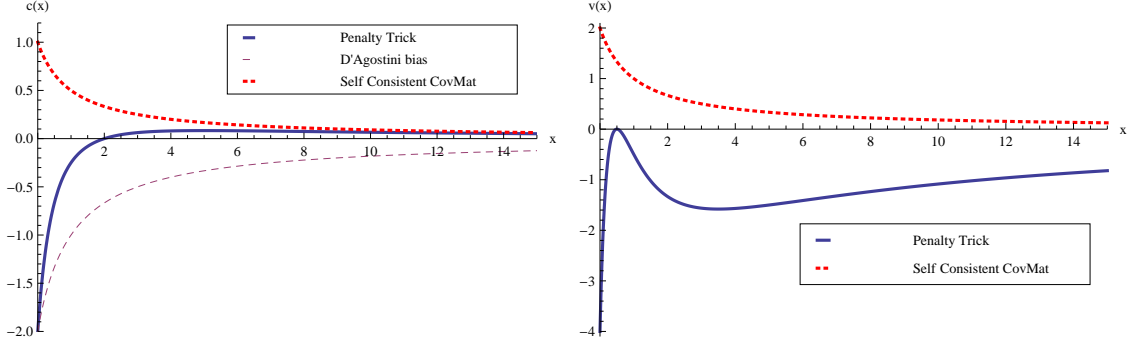
Figure 1: The "bias" functions $c(x)$ for the central value (left plot) and $v(x)$ for the variance (right plot), as defined in Eqs. (57-58), corresponding to the results obtained when the d'Agostini bias is present, Eqs. (43-44), using the penalty trick, Eqs. (55-56), and using the self-consistent covariance matrix method, Eqs. (75-76). The unbiased result corresponds to $c = v = 0$.

where

$$\Delta = \frac{s^2}{\sigma^2 + m_1 m_2 s^2} \tfrac{1}{2}(m_1^2 - m_2^2) \,. \tag{52}$$

Solving these for n_1 and n_2, and choosing the solution close to one (note that Δ will generally be rather small) gives

$$n_i = \tfrac{1}{2}\left(1 + \frac{1 \mp \Delta}{\sqrt{1 + \Delta^2}}\right), \tag{53}$$

$$t = \bar{m}\frac{\tfrac{1}{2}(1 + \sqrt{1 + \Delta^2} + r\Delta)}{1 + \Delta^2}, \tag{54}$$

where $\bar{m}$ and r are defined in Eq. (34).

The variance may now be computed as before by inversion of the three by three Hessian matrix. The result is not very enlightening, however it simplifies if we expand in powers of r: the Hessian method then gives

$$t = \bar{m}\left(1 + \frac{s^2 \bar{m}^2(\sigma^2 - 2s^2 \bar{m}^2)}{(\sigma^2 + s^2 \bar{m}^2)^2}r^2 + O(r^4)\right), \tag{55}$$

$$V_{tt} = \tfrac{1}{2}\sigma^2 + \tfrac{1}{2}s^2 \bar{m}^2\left(1 + r^2 - 4\frac{(2\sigma^2 + s^2 \bar{m}^2)^2}{(\sigma^2 + s^2 \bar{m}^2)^3}s^2 \bar{m}^2 r^2 + O(r^4)\right). \tag{56}$$

Clearly, both the central value and variance Eqs. (55,56) are biased: the normalization uncertainties leads to data being weighed differently based on their central values, even when their normalization uncertainties are the same. It is not difficult to show from the structure of Eqs. (49,50) that the results Eqs. (55,56) remain true for any number of data points, i.e. for all $n \geq 2$, provided that $\bar{m}$ and r are defined as in Eqs. (9,38), and an overall factor of $2/n$ is included in the variance.

This bias is more subtle than the d'Agostini bias Eq. (43), in that it is caused by nonlinearities in the error function rather than by a consistent bias in the variances. It can thus have either sign, depending on the relative weight of statistical and normalization
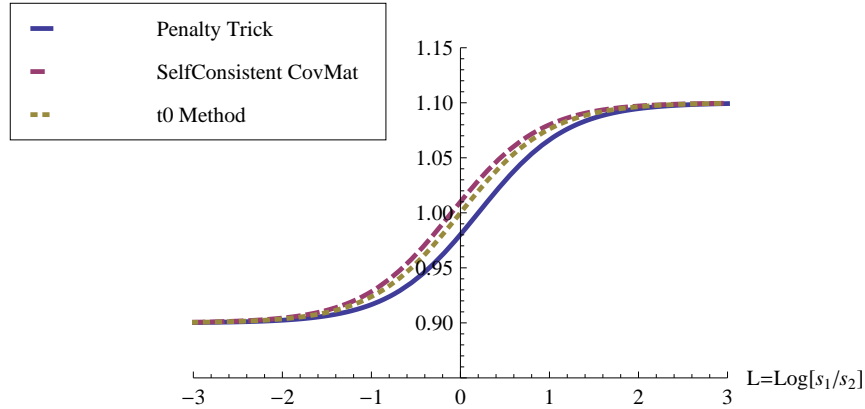
Figure 2: Dependence of the central value t on the ratio of normalization uncertainties s_1 and s_2 for a pair of measurements with central values $m_1 = 0.9$, $m_2 = 1.1$ and negligible uncertainties σ_i. The curves correspond to the cases of the penalty trick method Eq. (62), to the self–consistent covariance matrix method Eq. (78), and to the t_0 method Eq. (98). The unbiased result must be symmetric about the point $L = 0$ (i.e. $s_1 = s_2$): only the t_0 curve is unbiased. Decoupling of the data point with the larger error is clearly visible in the graph for $L \to \pm\infty$.

uncertainties: we can rewrite Eqs. (55-56) in the form

$$t = \bar{m}\big(1 + c(\tfrac{\sigma^2}{s^2\bar{m}^2})r^2 + O(r^4)\big), \tag{57}$$

$$V_{tt} = \tfrac{1}{n}\sigma^2 + \tfrac{1}{n}s^2\bar{m}^2\big(1 + r^2 + v(\tfrac{\sigma^2}{s^2\bar{m}^2})r^2 + O(r^4)\big), \tag{58}$$

with the functions c and v given by

$$c(x) = \frac{x - 2}{(x + 1)^2}, \qquad v(x) = -4\frac{(2x - 1)^2}{(x + 1)^3}. \tag{59}$$

The unbiased results would correspond to $c = v = 0$. The d'Agostini–biased results Eqs. (43-44) can also be cast in the form of Eq. (57-58), but now with $c(x) = -2/(x + 1)$, $v(x) = 0$ (since in this case the variance is unbiased at $O(r^2)$).

The "bias" functions $c(x)$ and $v(x)$ for these two cases, as well as for a further biased case to be discussed below in Sect. 5, are compared in Fig. 1. Thanks to the penalty trick, the bias in the central value is generally less severe, but the variance is now also biased.

It is also interesting to explore decoupling in this approach. To do this we again consider two experiments, in the special case in which the normalization errors dominate, so $\sigma_1^2, \sigma_2^2 \ll s_1^2 m_1^2, s_2^2 m_2^2$. Then Eqs.(49,50) can again be solved: $n_i = t/m_i$, with

$$t = m_1 m_2 \frac{m_1 s_1^2 + m_2 s_2^2}{m_1^2 s_1^2 + m_2^2 s_2^2}. \tag{60}$$

The variance is now

$$V_{tt} = m_1^2 m_2^2 \frac{s_1^2 s_2^2}{m_1^2 s_1^2 + m_2^2 s_2^2}. \tag{61}$$

When $s_1 \ll s_2$, we recover $t = m_1$ and $V_{tt} = m_1^2 s_1^2$, as we should. However the result Eq. (60) looks rather strange: one would expect that in this limit the measurements m_1

13

and m_2 to be simply weighted by $1/s_1^2$ and $1/s_2^2$, in analogy to the weighting Eq. (7) in the single experiment case, and the variance to be a similar weighted average of m_1^2 and m_2^2 (see Eqs. (98–99) below). Instead the weighting is more complicated, reflecting the bias in this method for general $s_1 \neq s_2$. Note for example that when $s_1 = s_2$, $t = m_1 m_2 (m_1 + m_2)/(m_1^2 + m_2^2) \neq \frac{1}{2}(m_1 + m_2)$, and $V_{tt} = s^2 m_1^2 m_2^2/(m_1^2 + m_2^2) \neq \frac{1}{4}s^2(m_1^2 + m_2^2)$. The result Eq. (60) is shown in Fig. 2 for the special case $m_1 = 0.9$ and $m_2 = 1.1$. The bias is readily apparent in the asymmetry of the curve.

The generalization to n independent experiments is straightforward in this limit: Eq. (60) becomes

$$t = \frac{\sum_{i=1}^{n} \frac{1}{m_i s_i^2}}{\sum_{i=1}^{n} \frac{1}{m_i^2 s_i^2}},$$
(62)

so if $s_i = s$, when there should be no bias, we have instead $t = \sum_{i=1}^{n} m_i^{-1} / \sum_{i=1}^{n} m_i^{-2} \neq \bar{m}$ unless $m_i = m$.

5 A Self–Consistent Covariance Matrix

We saw in Sect. 3 that minimizing the χ^2 Eq. (27) constructed using a covariance matrix of the form Eq. (28) gives the so–called d'Agostini bias. This bias comes from the dependence of the normalization term in the covariance matrix on m_i and m_j: indeed the bias is proportional to the differences $m_i - m_j$. A possible way out, alternative to the penalty trick of the previous section and based on a covariance matrix approach, was suggested by d'Agostini in Ref. [11]. Namely, one could choose to use

$$(\text{cov}_t)_{ij} = (\text{cov})_{ij} + s^2 t^2 \,,$$
(63)

since t is, by construction, a more precise estimator of the observable than m_i, and has already averaged out the differences in central value of the different measurements. We will now show that this method leads to results which are similar to those found using the penalty trick: for one experiment there is no bias, but for several experiments a bias arises. There is also a problem with multiple solutions.

5.1 One experiment

With two measurements within a single experiment, we now have a covariance matrix

$$(\text{cov}_t)_{ij} = \begin{pmatrix} \sigma_1^2 + s^2 t^2 & s^2 t^2 \\ s^2 t^2 & \sigma_2^2 + s^2 t^2 \end{pmatrix}$$
(64)

so the χ^2 is

$$\chi_t^2(t) = \frac{(t - m_1)^2(\sigma_2^2 + t^2 s^2) + (t - m_2)^2(\sigma_1^2 + s^2 t^2) - 2(t - m_1)(t - m_2)s^2 t^2}{\sigma_1^2 \sigma_2^2 + (\sigma_1^2 + \sigma_2^2)s^2 t^2} .$$
(65)

It is easy to check that minimizing this χ^2 with respect to t gives again $t = w$, where w is the weighted average Eq. (7). The variance V_{tt} is now simply the inverse of the second derivative of the χ^2 Eq. (31) at the minimum: a straightforward but tedious calculation then leads back to Eq. (47).

14

It is not difficult to show that everything works out for a general covariance matrix Eq. (63) for a single experiment with n data:

$$
\begin{aligned}
\tfrac{1}{2}\frac{\partial \chi_t^2}{\partial t} &= \sum_{i,j=1}^{n} (\mathrm{cov}_t^{-1})_{ij}(t - m_j) + \tfrac{1}{2}\sum_{i,j=1}^{n}(t - m_i)\left(\frac{\partial \mathrm{cov}_t^{-1}}{\partial t}\right)_{ij}(t - m_j) \\
&= \sum_{i,j=1}^{n}(\mathrm{cov}_t^{-1})_{ij}(t - m_j) - \tfrac{1}{2}\sum_{i,j,k,l=1}^{n}(t - m_i)(\mathrm{cov}_t^{-1})_{ik}\left(\frac{\partial \mathrm{cov}_t}{\partial t}\right)_{kl}(\mathrm{cov}_t^{-1})_{lj}(t - m_j) \\
&= \sum_{i,j=1}^{n}(\mathrm{cov}_t^{-1})_{ij}(t - m_j)\left[1 - s^2 t \sum_{k,l=1}^{n}(\mathrm{cov}_t^{-1})_{lk}(t - m_k)\right],
\end{aligned}
\tag{66}
$$

which vanishes when

$$
t = \frac{\sum_{i,j=1}^{n}(\mathrm{cov}_t^{-1})_{ij}m_j}{\sum_{i,j=1}^{n}(\mathrm{cov}_t^{-1})_{ij}}.
\tag{67}
$$

The variance of t is given by evaluating $\frac{\partial^2 \chi_t^2}{\partial t^2}$ at the minimum: this yields

$$
V_{tt} = \frac{1}{\sum_{i,j=1}^{n}(\mathrm{cov}_t^{-1})_{ij}}.
\tag{68}
$$

When

$$
(\mathrm{cov}_t)_{ij} = \delta_{ij}\sigma_i^2 + s^2 t^2,
\tag{69}
$$

the inverse is

$$
(\mathrm{cov}_t^{-1})_{ij} = \frac{\delta_{ij}}{\sigma_i^2} - \frac{s^2 t^2}{\sigma_i^2 \sigma_j^2}\frac{\Sigma^2}{\Sigma^2 + s^2 t^2},
\tag{70}
$$

where Σ is as defined in Eq. (8). Thus Eq. (67) and Eq. (68) simplify to the familiar results $t = w$ Eq. (7) and $V_{tt} = \Sigma^2 + s^2 w^2$ Eq. (47).

Note however that since χ_t^2 Eq. (31) is no longer quadratic in t, there is also a spurious solution: the term in square brackets in Eq. (66) vanishes when $t = \Sigma^2/ws^2$ which might be troublesome since it may lie close to the correct solution $t = w$ whenever $s^2 w^2 \sim \Sigma^2$.

5.2 More than one experiment

Consider now the case of n independent experiments. The covariance matrix is then

$$
(\mathrm{cov}_t)_{ij} = (\sigma_i^2 + s_i^2 t^2)\delta_{ij},
\tag{71}
$$

so the χ^2 is

$$
\chi_t^2(t) = \sum_{i=1}^{n}\frac{(t - m_i)^2}{\sigma_i^2 + s_i^2 t^2}.
\tag{72}
$$

The minimum of χ_t^2 is found by solving the system of nonlinear equations

$$
\sum_{i=1}^{n}\frac{(t - m_i)(t m_i s_i^2 + \sigma_i^2)}{(t^2 s_i^2 + \sigma_i^2)^2} = 0.
\tag{73}
$$

15

In general these equations will have $4n - 2$ solutions, so for a large number of experiments finding the correct solution might be difficult.

To explore the bias we consider as usual the symmetric situation $\sigma_i = \sigma$, $s_i = s$, as we did in Sect. 4.2. Then t is the (upper) solution to the quadratic equation

$$(\bar{m}s^2 t + \sigma^2)(t - \bar{m}) = r^2 \bar{m}^2 s^2 t, \tag{74}$$

so again when r is small we have a solution close to $t = \bar{m}$:

$$t = \bar{m}\left(1 + \frac{\bar{m}^2 s^2}{\sigma^2 + s^2 \bar{m}^2} r^2 + O(r^4)\right). \tag{75}$$

Thus as might be expected from the shape of the χ^2 the minimum is pushed upwards for $m_1 \neq m_2$. The variance is also biased:

$$V_{tt} = \frac{1}{n}\sigma^2 + \frac{1}{n}s^2 \bar{m}^2\left(1 + r^2 + \frac{2s^2 \bar{m}^2}{\sigma^2 + s^2 \bar{m}^2} r^2 + O(r^4)\right). \tag{76}$$

The biases can again be cast in the form of Eq. (57) and Eq. (58) but now with

$$c(x) = \frac{1}{x+1}, \quad v(x) = \frac{2}{x+1}. \tag{77}$$

They are compared to those of the penalty trick method in Fig. 1.

Another interesting special case is when the normalization errors are dominant, so (for n experiments) $s_i m_i \gg \sigma_i$. The solution for t then reduces to

$$t = \frac{\sum_{i=1}^n \frac{m_i^2}{s_i^2}}{\sum_{i=1}^n \frac{m_i}{s_i^2}}. \tag{78}$$

So when one of the s_i is very large, this experiment decouples as expected. However the weighting is still biased: when $s_i = s$, $t = \sum_i m_i^2 / \sum_i m_i \neq \bar{m}$ unless all $m_i = m$. The result Eq. (78) is also plotted in Fig. 2: it is instructive to compare it with the superficially similar result Eq. (62) obtained with the penalty trick, which was also biased, but in the opposite direction.

Thus the self–consistent covariance matrix discussed here is biased when used for more than one experiment, like the penalty trick method of Sect. 4. Since these methods are biased in the Hessian approach, they would also fail to provide a suitably unbiased error function for fitting to Monte Carlo replicas. Moreover these methods have multiple solutions, which also make them difficult to implement in a Monte Carlo, since if some of the replicas are fitted to the wrong solution, these will clearly also lead to an incorrect final result. In the next section, we will present a new method which is free of multiple solutions, is unbiased when uncertainties are equal, but which unlike the method of Sect. 2.3 also correctly weights the different experiments according to their normalization uncertainties when these are unequal.

16

6 Unbiased Fitting

The biases and multiple solutions found in the previous section come from the fact that the $\chi^2(t)$ function used in the fitting is no longer a quadratic function of the observable t being fitted, and thus the distribution of $\exp(-\frac{1}{2}\chi_t^2(t))$ is no longer Gaussian. The dependence of the covariances of the data on t^2 distorts the shape of the χ^2, and thus introduces a bias. The only way to avoid this is to hold the covariance matrix fixed when performing the fitting. We can do this by evaluating the covariance matrix using some fixed value t_0 rather than t. The value of t_0 can then be tuned independently to be consistent with the value of t obtained from the fit. The basic idea is then to determine t_0 self–consistently in an iterative way. The use of theoretical estimates to avoid d'Agostini bias through an iterative procedure can also be found in the treatment of multiplicative systematic errors by the H1 Collaboration [13].

The necessity for such a procedure is particularly clear in the Monte Carlo approach. When we fit to replicas, we do so on the assumption that each replica is generated with a given distribution of the data central values according to their experimental uncertainties. Clearly this uncertainty should be fixed once and for all, not vary from replica to replica. We will now show that this procedure indeed gives unbiased results for equal normalization uncertainties and that it also satisfies the decoupling criterion, for both the Hessian and Monte Carlo methods. Finally we also show that the iterative determination of t_0 converges very rapidly, and thus that the method is also practical.

6.1 One experiment

For a single experiment, in place of Eq. (63) the covariance matrix is chosen to be

$$(\text{cov}_{t_0})_{ij} = (\text{cov})_{ij} + t_0^2 s^2, \tag{79}$$

where t_0 should be viewed as a guess for t, to be fixed beforehand. In a Hessian approach the χ^2 is then

$$\chi_{t_0}^2(t) = \sum_{i,j=1}^{n} (t - m_i)(\text{cov}_{t_0}^{-1})_{ij}(t - m_j), \tag{80}$$

and minimization is trivial:

$$t = \frac{\sum_{i,j=1}^{n}(\text{cov}_{t_0}^{-1})_{ij}m_j}{\sum_{i,j=1}^{n}(\text{cov}_{t_0}^{-1})_{ij}}, \tag{81}$$

while

$$V_{tt} = \frac{1}{\sum_{i,j=1}^{n}(\text{cov}_{t_0}^{-1})_{ij}}. \tag{82}$$

For the special case in which the data have uncorrelated statistical errors only, Eq. (5), using the inversion Eq. (70) (with t replaced by t_0) we thus recover $t = w$ Eq. (7), independent of the value chosen for t_0, while

$$V_{tt} = \Sigma^2 + s^2 t_0^2. \tag{83}$$

This reduces to Eq. (47), but only if we tune $t_0 = t$.

17

In a Monte Carlo approach the χ^2 for each replica is

$$\chi^2_{t_0}(T) = \sum_{i,j=1}^{n} (T - NM_i)(\text{cov}_{t_0}^{-1})_{ij}(T - NM_j), \tag{84}$$

and minimization is again trivial:

$$t = \frac{\sum_{i,j=1}^{n}(\text{cov}_{t_0}^{-1})_{ij}NM_j}{\sum_{i,j=1}^{n}(\text{cov}_{t_0}^{-1})_{ij}}. \tag{85}$$

The expectation value and variance of t can now be straightforwardly evaluated:

$$\text{E}[t] = \frac{\sum_{ij=1}^{n}(\text{cov}_{t_0}^{-1})_{ij}m_j}{\sum_{i,j=1}^{n}(\text{cov}_{t_0}^{-1})_{ij}}, \tag{86}$$

$$\text{Var}[t] = (1 + s^2)\left(\frac{1}{\sum_{i,j=1}^{n}(\text{cov}_{t_0}^{-1})_{ij}} - s^2 t_0^2\right) + s^2\text{E}[t]^2. \tag{87}$$

This coincides with the Hessian result Eqs. (81-82) when $t_0 = \text{E}[t]$. Also it reduces again to the old results $\text{E}[t] = w$ Eq. (7) and $\text{Var}[t] = (1 + s^2)\Sigma^2 + w^2$ Eq. (22) for uncorrelated statistical errors Eq. (5). These results are then quite independent of the value of t_0, explaining the success of the old error function Eq. (16). Note however that the value of the χ^2 Eq. (80) at the minimum for each replica will still depend on t_0, and will indeed be quite different from that of the error function Eq. (16), to which it reduces only when $t_0 = 0$.

6.2 More than one experiment

When we have n independent experiments (each with one data point), we now choose in place of Eq. (71)

$$(\text{cov}_t)_{ij} = (\sigma_i^2 + s_i^2 t_0^2)\delta_{ij}, \tag{88}$$

so the χ^2 is now

$$\chi^2 = \sum_{i=1}^{n} \frac{(t - m_i)^2}{\sigma_i^2 + s_i^2 t_0^2}, \tag{89}$$

whence we have the single solution

$$t = \frac{\sum_{i=1}^{n} \frac{m_i}{\sigma_i^2 + s_i^2 t_0^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + s_i^2 t_0^2}}, \tag{90}$$

and

$$V_{tt} = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + s_i^2 t_0^2}}. \tag{91}$$

For the special case $s_i = s$, $\sigma_i = \sigma$, these reduce to $t = \bar{m}$ and $V_{tt} = \frac{1}{n}(\sigma^2 + s^2 t_0^2)$ as they should: the fit is unbiased, and the variance is correctly estimated provided only that $t_0 = t$. When instead the normalization uncertainties dominate,

$$t = \frac{\sum_{i=1}^{n} \frac{m_i}{s_i^2}}{\sum_{i=1}^{n} \frac{1}{s_i^2}}, \qquad V_{tt} = \frac{t_0^2}{\sum_{i=1}^{n} \frac{1}{s_i^2}}, \tag{92}$$

18

and thus the central value exhibits decoupling (if some s_i is much larger than the others, t becomes independent of the corresponding measurement m_i), and the variance is again correctly estimated whenever $t_0 = t$.

In the Monte Carlo approach the χ^2 for a given replica is now

$$\chi^2 = \sum_{i=1}^{n} \frac{(t - N_i M_i)^2}{\sigma_i^2 + s_i^2 t_0^2} \,. \tag{93}$$

Minimization is again straightforward:

$$t = \frac{\sum_{i=1}^{n} \frac{N_i M_i}{\sigma_i^2 + s_i^2 t_0^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + s_i^2 t_0^2}}, \tag{94}$$

whence on averaging over replicas

$$\mathrm{E}[t] = \frac{\sum_{i=1}^{n} \frac{m_i}{\sigma_i^2 + s_i^2 t_0^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + s_i^2 t_0^2}}, \tag{95}$$

$$\mathrm{Var}[t] = \frac{\sum_{i=1}^{n} \frac{\sigma_i^2 + s_i^2(m_i^2 + \sigma_i^2)}{(\sigma_i^2 + s_i^2 t_0^2)^2}}{\left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + s_i^2 t_0^2}\right)^2}. \tag{96}$$

The central value coincides with the Hessian result Eq. (90), while the variance is now determined more accurately: Eq.(96) reduces to Eq. (91) whenever $m_i^2 + \sigma_i^2 \sim t_0^2$.

In the special cases considered previously, in the symmetric case $s_i = s$, $\sigma_i = \sigma$ we have

$$\mathrm{E}[t] = \bar{m}, \qquad \mathrm{Var}[t] = \tfrac{1}{n}(\sigma^2(1 + s^2) + s^2 \bar{m}^2(1 + r^2)), \tag{97}$$

which is indeed unbiased. Note that unlike in the Hessian method, the variance cross term is now properly included (see Eq. (19)), the spread of values m_i now contributes to the variance, as it should (see Eq. (44)), and both the expectation value and variance are actually independent of t_0, just as they were for the case of a single experiment.

When the normalization uncertainties dominate, $\sigma_i^2 \ll s_i^2 t_0^2$

$$\mathrm{E}[t] = \frac{\sum_{i=1}^{n} \frac{m_i}{s_i^2}}{\sum_{i=1}^{n} \frac{1}{s_i^2}}, \tag{98}$$

$$\mathrm{Var}[t] = \frac{\sum_{i=1}^{n} \frac{m_i^2}{s_i^2}}{\left(\sum_{i=1}^{n} \frac{1}{s_i^2}\right)^2}. \tag{99}$$

The result Eq. (98) is compared in Fig. 2 to the biased results Eq. (78) and Eq. (60) obtained previously. It is manifestly unbiased. Furthermore, when the normalization uncertainty of one of the experiments is particularly large, this experiment now decouples from both the mean and the variance just as it should. Note again that the results Eq. (98) and Eq. (99) are entirely independent of t_0: in the Monte Carlo method t_0 only controls the relative balance between the statistical and normalization errors, and then only when these are different amongst themselves.

6.3 Determining t_0

The results of Sect. 6.1-6.2 imply that the t_0-covariance matrix Eq. (79) gives a χ^2 function that can be used to give unbiased fits to the individual replicas, with t_0 controlling the relative balance between statistical and normalization errors, both in the Hessian and Monte Carlo method. The remaining difficulty with this approach is that t_0 is not determined self–consistently within the minimization, but rather must be fixed beforehand. Clearly if the value chosen is incorrect, this may itself lead to an incorrect fit. Indeed, we have seen that the Monte Carlo and Hessian results with the t_0–method lead to the same central prediction only when $t_0 = E[t]$.

However, the dependence on t_0 is rather weak. That this is the case is qualitatively clear: firstly t_0 only determines the uncertainties, so an error we make in t_0 is a second order effect; furthermore all dependence on t_0 cancels when all the m_i are equal, when σ_i and s_i are equal, or when normalization errors dominate over statistical, or indeed vice versa. To make this more precise, consider a small shift $t_0 \to t_0 + \delta t_0$. The corresponding shift in $E[t]$ Eq. (95) (or its Hessian counterpart Eq. (90) is then given by

$$\delta E[t] = \delta t_0 \frac{\sum_{i,j=1}^{n} t_0 \frac{(m_i - m_j)(s_i^2 \sigma_j^2 - \sigma_i^2 s_j^2)}{(\sigma_i^2 + s_i^2 t_0^2)^2 (\sigma_j^2 + s_j^2 t_0^2)^2}}{\left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + s_i^2 t_0^2}\right)^2}. \tag{100}$$

As expected this vanishes when all the m_i are equal, in the symmetric case $s_i = s$, $\sigma_i = \sigma$, and in the limits when either statistical or normalization uncertainties dominate. Elsewhere, we expect it to be very small.

To quantify this, we first note that when combining a large number of experiments, so that n in Eq. (100) is large, the result is essentially independent of n (since both numerator and denominator grow as n^2), so a typical contribution to the sums may be taken as indicative of the overall result. Now for any pair of measurements, $2t_0(m_i - m_j) \sim m_i^2 - m_j^2$ will typically be of the same size as the uncertainties in the measurements, so if all percentage uncertainties are of the same order Δ, i.e. $s_i \sim \Delta$, $\sigma_i \sim \Delta t_0$, then Eq. (100) gives $\delta E[t] \sim \Delta^2 \delta t_0$. So since Δ is always rather less than one, we always expect $\delta E[t] \ll \delta t_0$.

It follows that t_0 can be determined iteratively: a first determination of t is performed with a zeroth–order guess for t_0, such as, for example, $t_0 = 0$, which in the Monte Carlo approach corresponds to the simple choice Eq. (24) in which normalization uncertainties are not included in the error function. The result for $E[t]$ thus obtained is used as t_0 for a second iteration, and so on. Since $\delta E[t] \sim \Delta^2 \delta t_0$, and $\Delta \ll 1$, it is clear that these iterations will converge rapidly to a result within the envelope of the purely statistical uncertainties on $E[t]$. Moreover, since as we saw in the previous section $E[t]$ is unbiased for any t_0, the final result (which has $t_0 = E[t]$) will also be unbiased. Note incidentally that this also implies that the t_0-method must give a result different in general to that obtained with the penalty trick, since the latter is biased (see Fig.2).

The estimate Eq. (100) then suggests that already the first iteration should be sufficient to provide a determination with a relative uncertainty of Δ^2 for data affected by typical relative uncertainties Δ, i.e. the procedure is expected to converge at the first iteration for all practical purposes.

In table 1 we summarise the features of our new t_0–method compared to those of all the other procedures discussed in this paper, based on the two criteria of freedom from bias

method	E or χ^2	Bias for $s_i = s$, $\sigma_i = \sigma$		Decoupling $s_i^2 m_i^2 \gg \sigma_i^2$
		1 expt, n data	n expts	2 expts
MC with $t_0 = 0$	(16),(24)	1	1	$\frac{1}{2}(m_1 + m_2)$
Hessian with d'Agostini bias	(31),(40)	$\frac{1}{1+ns^2}$	$1 - \frac{2s^2\sigma^2}{\sigma^2+s^2\bar{m}^2} + \ldots$	$\frac{1/m_1 s_1^2 + 1/m_2 s_2^2}{1/m_1^2 s_1^2 + 1/m_2^2 s_2^2}$
Hessian with penalty trick	(45),(48)	1	$1 - \frac{s^2(\sigma^2-2s^2\bar{m}^2)}{(\sigma^2+s^2\bar{m}^2)^2} + \ldots$	$\frac{1/m_1 s_1^2 + 1/m_2 s_2^2}{1/m_1^2 s_1^2 + 1/m_2^2 s_2^2}$
t-cov. mat.	(65),(72)	1	$1 + \frac{s^2\sigma^2}{\sigma^2+s^2\bar{m}^2} + \ldots$	$\frac{m_1^2/s_1^2 + m_2^2/s_2^2}{m_1/s_1^2 + m_2/s_2^2}$
t_0-cov. mat.	(84),(93)	1	1	$\frac{m_1/s_1^2 + m_2/s_2^2}{1/s_1^2 + 1/s_2^2}$

Table 1: Summary of the various methods for the inclusion of normalization uncertainties. The second column provides the reference to the pair of E– or χ^2–functions whose minimization determines the results for one experiment and many experiments in each case. Results for the bias (defined in the text) are given assuming the data asymmetry $r^2 \approx \sigma^2/\bar{m}^2$.

when uncertainties are all equal, and decoupling when the normalization uncertainty of one experiment is very large, as presented in the end of Sect. 2.1. Specifically, we consider the Monte Carlo method with normalization uncertainties not included in the error function discussed in Sect. 2.3; the d'Agostini–biased Hessian method of Sect. 3; the Hessian method with penalty trick presented in Sect. 4; the self–consistent covariance matrix method (t-cov) discussed in Sect. 5, and finally the t_0–method discussed in this section. The two central columns in the table give the ratio of the central value to the unbiased results Eq. (86) and Eq. (95) in the one–experiment and n–experiment cases respectively, thereby exposing the bias whenever the ratio is not equal to one. The result is given for uncorrelated systematic uncertainties, all equal to σ, and normalization uncertainties all equal to s, and assuming that the difference between measurements r Eq. (42) is $r^2 \approx \sigma^2/\bar{m}^2$. The last column gives the central value for a fit to two experiments, in the limit where normalization uncertainties are dominant and other uncertainties can be neglected, in order to show how (or if) an experiment decouples when its normalization uncertainty becomes very large.

6.4 Likelihood

So far in this paper we have taken a naive approach, in which we constructed least squares estimators, and minimised them with respect to the theoretical prediction t. It is interesting to consider instead how one might construct a likelihood function for the measurements and normalizations, and thus whether any of our estimators are maximum likelihood estimators.

Consider for definiteness the case of several experiments, with measurements m_i and variances σ_i. In presenting the measurements in this form, there is an underlying assumption that the measurements are Gaussian. The likelihood is then simply defined as the probability that the measurements take the observed values given a certain theoretical value t for such measurements:

$$P(m|t) = N \exp\left(-\tfrac{1}{2}\sum_{i=1}^{n}\frac{(m_i - t)^2}{\sigma_i^2}\right), \tag{101}$$

where N is some overall normalisation factor for the probability, dependent on σ_i but not on m_i or t. Of course this probability is just $N\exp(-\tfrac{1}{2}\chi^2(t))$, so the maximum likelihood estimator is found by minimising the χ^2, i.e. it is the same as the least squares estimator.

Now consider what happens when there are also normalization uncertainties n_i with variances s_i. Again in the absence of further information it is natural to assume these are also Gaussian. Furthermore they are clearly entirely independent of the measurement uncertainties, since the physics involved in determining the normalization is generally quite independent of that related to the measurements m_i or indeed the theoretical value t. Thus the total likelihood should factorise: $P(m, n|t) = P(m|t)P(n)$. The maximum likelihood estimators obtained from $P(m, n|t)$ and $P(m|t)$ should thus be the same.

Now the Hessian method of Sec. 3, by adopting the χ^2-function Eq.(40), assumes for the likelihood

$$P_m(m|t) = N \exp\left(-\tfrac{1}{2} \sum_{i=1}^{n} \frac{(m_i - t)^2}{\sigma_i^2 + s_i^2 m_i^2} \right). \tag{102}$$

This is incorrect, because it is no longer Gaussian in m_i, and indeed not even properly normalised (to normalise it, N must depend on t, and then maximising $P_m(m|t)$ is no longer the same as minimising the χ^2). Thus the χ^2-function Eq.(40) is not a maximum likelihood estimator, principally because the probability distribution it assumes is skewed by the normalization uncertainties.

Similarly the Hessian method with penalty trick presented in Sec. 4, by adopting the error function Eq.(48), assumes for the likelihood

$$P_n(m, n|t) = N \exp\left(-\tfrac{1}{2} \sum_{i=1}^{n} \left[\frac{(m_i - t/n_i)^2}{\sigma_i^2} + \frac{(n_i - 1)^2}{s_i^2} \right] \right). \tag{103}$$

This is also incorrect, because now the likelihood, while Gaussian in m_i, is not Gaussian in n_i, and furthermore cannot be factorised into a product $P(m|t)P(n)$. Once again it is not properly normalised: N must depend on t. So this too does not give us a maximum likelihood estimator. The main problem here is that since the model for the likelihood does not factorise, the assumption of a common theoretical result t introduces artificial correlations between the normalization measurements n_i; this leads to biases since these measurements are in principle completely independent. This is why the penalty trick, while giving correct results for a single experiment with only one overall normalization uncertainty, fails when applied to several independent experiments.

The self–consistent covariance matrix method (t-cov) discussed in Sec. 5, with χ^2-function Eq.(72), assumes the likelihood to be

$$P_t(m|t) = N \exp\left(-\tfrac{1}{2} \sum_{i=1}^{n} \frac{(m_i - t)^2}{\sigma_i^2 + t^2 s_i^2} \right). \tag{104}$$

This is better, because it is now Gaussian in m_i. However the normalization N still depends on t, and thus the maximum likelihood estimator no longer minimises the χ^2, but has a correction from $\ln N(t)$. It may thus be possible to construct the maximum likelihood estimator by this route, but it is difficult because everything is so nonlinear in t.

Finally consider the t_0–method discussed in this section, which takes as its starting point the χ^2-function Eq.(89). Here the assumption for the likelihood is

$$P_{t_0}(m|t) = N \exp\left(-\tfrac{1}{2} \sum_{i=1}^{n} \frac{(m_i - t)^2}{\sigma_i^2 + t_0^2 s_i^2} \right) \tag{105}$$

This is now correct: it is Gaussian in m_i, properly normalised, and all we have to do to make sure we get the correct answer is choose t_0 consistently. Note that it can alternatively be formulated as

$$P_{t_0}(m, n|t) = N \exp\left(-\tfrac{1}{2} \sum_{i=1}^{n} \left[\frac{(m_i - t/n_0)^2}{\sigma_i^2 + t_0^2 s_i^2} + \frac{(n_i - n_0)^2}{s_i^2}\right]\right). \tag{106}$$

This is rather like the penalty trick Eq.(103), but with the n_i in the first term replaced with its best estimate n_0, just as in going from m-cov Eq.(102) to t_0-cov Eq.(105) we replace m_i in the denominator with our best estimate t_0. Note of course that actually $n_0 = 1$: this is the natural choice for presenting the data that all experimentalists choose. Eq.(106) is also a good definition of the likelihood: it is Gaussian in both m_i and n_i, the normalization is determined quite independently of t, and it factorises correctly into the product $P_{t_0}(m|t)P(n)$. Thus when we minimise wrt t, $P_{t_0}(m|t)$ and $P_{t_0}(m, n|t)$ give the same maximum likelihood estimator for t, as they should.

Since the t_0-method yields the maximum likelihood estimator, it possesses all the nice asymptotic properties of maximum likelihood estimators: in particular it is consistent and unbiased, and asymptotically efficient (see for example Ref. [14]).

7 Application to PDF determination

As mentioned in the introduction, in PDF determinations an accurate result is sought while combining many different sources of uncertainty for a large number of experiments. Typical uncertainties in a global parton fit are summarized in Table 2 for deep–inelastic scattering (DIS) experiments and Table 3 for hadronic experiments. Specifically, we list in Tables 2 and 3 the uncertainties of the DIS data included in the NNPDF1.2 [7] analysis, and those of the Drell-Yan, weak boson production and jet data included in the NNPDF2.0 [15] global fit. The total number of data points included in these sets is about 3000 for the NNPDF1.2 analysis, and about 3500 for NNPDF2.0. The datasets used of other recent global parton fits [16, 17] are similar. For DIS data, all uncertainties including normalizations are of the order of a few percent, while for hadronic data uncertainties vary widely and can be as large as 20%. We expect therefore the impact of a full treatment of normalization uncertainties to be moderate when fitting DIS data, and more dramatic in truly global fits which include hadronic data.

7.1 Impact of normalization uncertainties

Normalization uncertainties have been included in recent PDF determinations either by using the penalty trick with the Hessian method in recent MSTW [16] and CTEQ [17, 18] fits, or using the Monte Carlo method with the $t_0 = 0$ error function Eq. (24) in recent NNPDF fits[3] [5, 7].

As we have seen, neither of these procedures is entirely satisfactory. However, the bias in the penalty trick method used by MSTW and CTEQ, with uncertainties such as those

---

[3]In fact, the error functions used by NNPDF differs from Eq. (24) by a factor of N_i^2 in the denominator. It is not difficult to see that the only effect of this extra factor is to introduce a small downward bias of relative order s_i^2 (so $\sim 0.01\%$ for the data in NNPDF1.2) in E[t] and Var[t].

DIS data (NNPDF1.2 [7] data set)					
Experiment	Set	STAT (%)	SYS (%)	NORM (%)	TOT (%)
NMC-pd					
	NMC-pd	2.0	0.4	0.0	2.1
NMC					
	NMC	3.7	2.3	2.0	5.0
SLAC					
	SLACp	2.7	0.0	2.2	3.6
	SLACd	2.5	0.0	1.8	3.1
BCDMS					
	BCDMSp	3.2	2.0	3.2	5.5
	BCDMSd	4.5	2.3	3.2	6.6
ZEUS					
	Z97lowQ2	2.5	3.0	2.1	4.7
	Z97NC	6.2	3.1	2.1	7.8
	Z97CC	33.6	5.6	2.0	34.2
	Z02NC	12.2	2.1	1.8	12.7
	Z02CC	38.6	6.2	1.8	39.3
	Z03NC	6.9	3.2	2.0	8.3
	Z03CC	29.1	5.6	2.0	29.8
H1					
	H197mb	2.8	2.0	1.7	3.9
	H197lwQ2	2.7	2.5	1.7	4.2
	H197NC	12.5	3.2	1.5	13.3
	H197CC	27.5	4.6	1.5	28.1
	H199NC	14.7	2.8	1.8	15.2
	H199CC	25.5	3.8	1.8	25.9
	H199NChy	7.2	1.7	1.8	7.7
	H100NC	9.4	3.2	1.5	10.4
	H100CC	20.4	3.8	1.5	20.9
CHORUS					
	CHORUSnu	4.2	6.4	7.9	11.2
	CHORUSnb	13.8	7.8	8.7	18.7
FLH108					
	FLH108	47.2	53.3	5.0	71.9
NTVDMN					
	NTVnuDMN	16.2	0.0	2.1	16.3
	NTVnbDMN	26.6	0.0	2.1	26.7
ZEUS-H2					
	Z06NC	3.8	3.7	2.6	6.4
	Z06CC	25.5	14.3	2.6	31.9

Table 2: Different sources of uncertainty for data included in the NNPDF1.2 [7] analysis, which is representative of the deep-inelastic scattering data included in typical parton fits. All uncertainties are given as a percentages, obtained as averages over all points of the percentage values in the corresponding set; "stat" denotes all uncorrelated uncertainties (mostly statistical but also uncorrelated systematics); "sys" uncertainties which are correlated between all experiments in a set; "norm" multiplicative (normalization) uncertainties which are also correlated between all experiments in a set; and "tot" the average of the sum in quadrature of all uncertainties. Normalization uncertainties are fully correlated between the NMCp and NMCd datasets and cancel in the NMC-pd ratio.

Hadronic data (NNPDF2.0 [15] data set)					
Experiment	Set	STAT (%)	SYS (%)	NORM (%)	TOT (%)
DYE605					
	DYE605	16.6	0.0	15.0	22.6
DYE886					
	DYE886p	20.4	0.0	6.5	22.1
	DYE886d	18.3	0.0	6.5	20.6
	DYE886r	3.6	1.0	0.0	3.8
CDFWASY					
	CDFWASY	4.2	4.2	0.0	6.0
CDFZRAP					
	CDFZRAP	5.1	6.0	6.0	11.5
D0ZRAP					
	D0ZRAP	7.6	0.0	6.1	10.2
CDFR2KT					
	CDFR2KT	4.5	21.1	5.8	23.0
D0R2CON					
	D0R2CON	4.4	14.3	6.1	16.8

Table 3: Same as Table 2 but for the hadronic data which is included, in addition to the DIS data of Table 2, in the NNPDF2.0 [15] analysis. Again, this dataset is representative of the hadronic data included in typical parton fits. Normalization uncertainties are fully correlated between the DYE886p and DYE886d datasets and cancel in the DYE866r ratio.

of Tables 2 and 3 is likely to be at or below the percent level for DIS data, though it could be non-negligible for some hadronic (in particular Drell-Yan) data. Similarly, the error in using the Monte Carlo method with $t_0 = 0$ can be estimated using Eq. (100) which for DIS experiments gives $\delta E[t] \sim 0.01 \, \delta t_0$. Hence, the NNPDF1.0 and NNPDF1.2 fits performed in Ref. [5, 7] to DIS data are expected to be within a few percent of the true result. The deviation is more significant for hadronic data.

7.2   Implementation of the t_0 method

In order to test our general arguments, and also as a practical illustration of the new t_0 method we have implemented it in the NNPDF framework and repeated the NNPDF1.2 fit [7] with the t_0 method. As a starting point we take the old NNPDF1.2 fit, and thus $t_0 = 0$. We then performed several iterations in each one taking the central value of the previous fit to determine t_0, and thus the covariance matrix Eq. (79): specifically ite1 thus takes t_0 to be the central value of NNPDF1.2 (referred to hereafter as ite0), ite2 takes t_0 to be the central value of the ite1 fit, and so on. In this way we can assess the convergence of the method. Three iterations proves to be more than sufficient. At each iteration we produce one thousand replicas: the uncertainty on t_0 is then the overall pdf uncertainty divided by $\sqrt{1000}$, which is sufficiently small that random fluctuations are kept under control.

In Fig. 3 we show the results for the central pdfs (i.e. t_0) in the various iterations, ite0, ite1, ite2 and ite3 normalized with respect to ite3. We show, besides the overall one-sigma PDF uncertainty of ite3 (the very broad band in the plots), also the expected range for the fluctuations of the t_0 iterations in the convergence regime (the relatively narrow band). As expected, convergence is reached very quickly, essentially at the first iteration: while the original central value (ite0) often lies some distance from the central narrow band, all subsequent iterations lie more or less within it. Note however that even ite0 lies essentially within the broad PDF uncertainty band: the shift in central values is generally within one
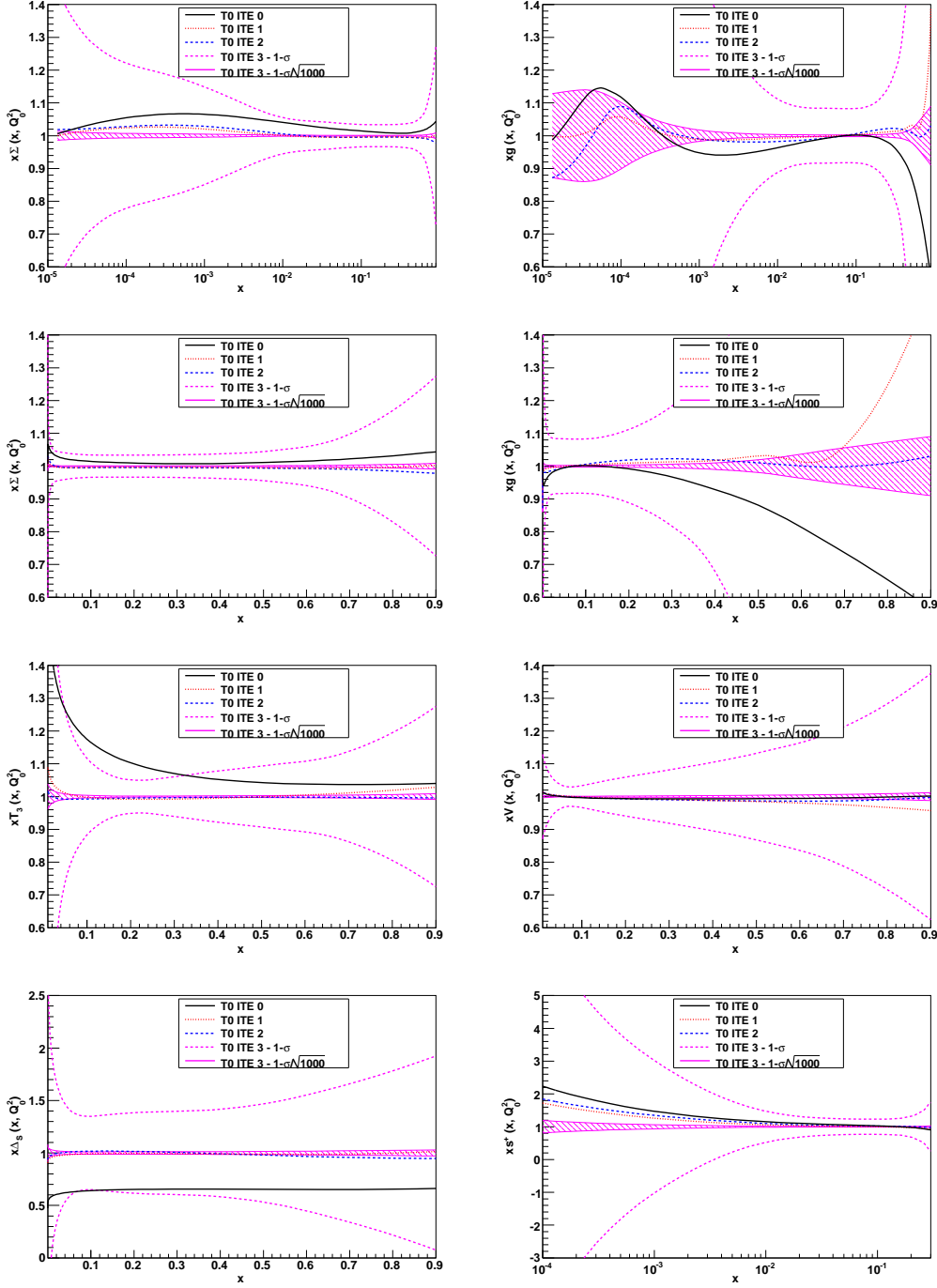
Figure 3: Comparison between NNPDF1.2 PDFs and the results of the three first t_0–iterations. In each case the PDFs shown have been used to compute the t_0 covariance matrix for the next iteration. Each iterations has been computed with 1000 Monte Carlo replicas. We show the one-sigma PDF uncertainty band for ite3, as well as the same band divided by $\sqrt{1000}$, which corresponds to the expected statistical fluctuations of the PDF central values (hatched area). For the singlet and the gluon PDFs plots displaying both the small x and large x regions are shown.

26

	ite1	ite2	ite3
$\Sigma(x, Q_0^2)$	9.2	2.6	2.8
$g(x, Q_0^2)$	2.9	1.9	2.1
$T_3(x, Q_0^2)$	7.5	0.9	0.7
$V(x, Q_0^2)$	1.1	0.8	2.8
$\Delta_S(x, Q_0^2)$	14.4	1.6	1.2
$s_+(x, Q_0^2)$	2.8	3.5	3.3
$s_-(x, Q_0^2)$	1.9	1.2	2.6

Table 4: The stability distances for various PDFs for the various iterations of the NNPDF1.2 t_0 fits, all computed from 1000 replicas. For each fit, distances are computed with respect to the fit from the previous column (ite1 being with respect to NNPDF1.2). The distances are averaged over the data regions, as defined for each PDF in [7].

Experiment	ite0	ite1	ite2	ite3
Total	1.32	1.25	1.25	1.25
SLAC	1.33	1.30	1.31	1.29
BCDMS	1.57	1.44	1.42	1.44
NMC	1.71	1.66	1.67	1.66
NMC-pd	1.73	1.30	1.28	1.29
ZEUS	1.05	1.02	1.03	1.03
H1	1.02	0.99	0.99	1.00
CHORUS	1.38	1.31	1.32	1.33
FLH108	1.65	1.67	1.67	1.67
NuTeV Dimuon	0.65	0.66	0.65	0.68
ZEUS HERA-II	1.53	1.49	1.49	1.49

Table 5: The χ^2 per degree of freedom, both total and for individual experiments, for the various iterations of the NNPDF1.2 t_0 fits, always with 1000 replicas; "ite0" denotes the starting NNPDF1.2 fit.

sigma of the overall uncertainty. The only exception is the triplet T_3 in the valence region.

These statements may be made more quantitative by looking at the distances between the various curves, in units of their standard deviations combined in quadrature (as defined in Appendix B of [4]). Distance equal to one means that two curves are within one sigma of each other, so the average distance between a random sample of curves should tend roughly to one in the limit of large samples. The distances between ite1 and ite0, ite2 and ite1, and ite3 and ite2 are shown as a function of x in Fig. 4 for the various PDFs considered, and in Table 4 averaged over x (the average is performed by sampling the distance at ten values of x in the data region, see Ref. [7]). Again the convergence at ite1 is apparent: only the distance between ite1 and ite0 is ever appreciably larger than unity.

It is clear from this analysis that the PDFs which are most affected by the inclusion of normalization uncertainties in the fits are the singlet Σ in the region $0.001 < x < 0.1$ (where there is tension between NMC and HERA), and the triplet T_3 and sea asymmetry Δ_S in the region $0.04 < x < 0.4$ (due mainly to tension between BCDMS, NMC and CHORUS).

In Table 5 we show the value of the χ^2 per degree of freedom in the reference fit ite0 and in the various iterations of the t_0 fit. We note the improvement in the fit quality from the better handling of normalization uncertainties, particularly in BCDMS, NMC and
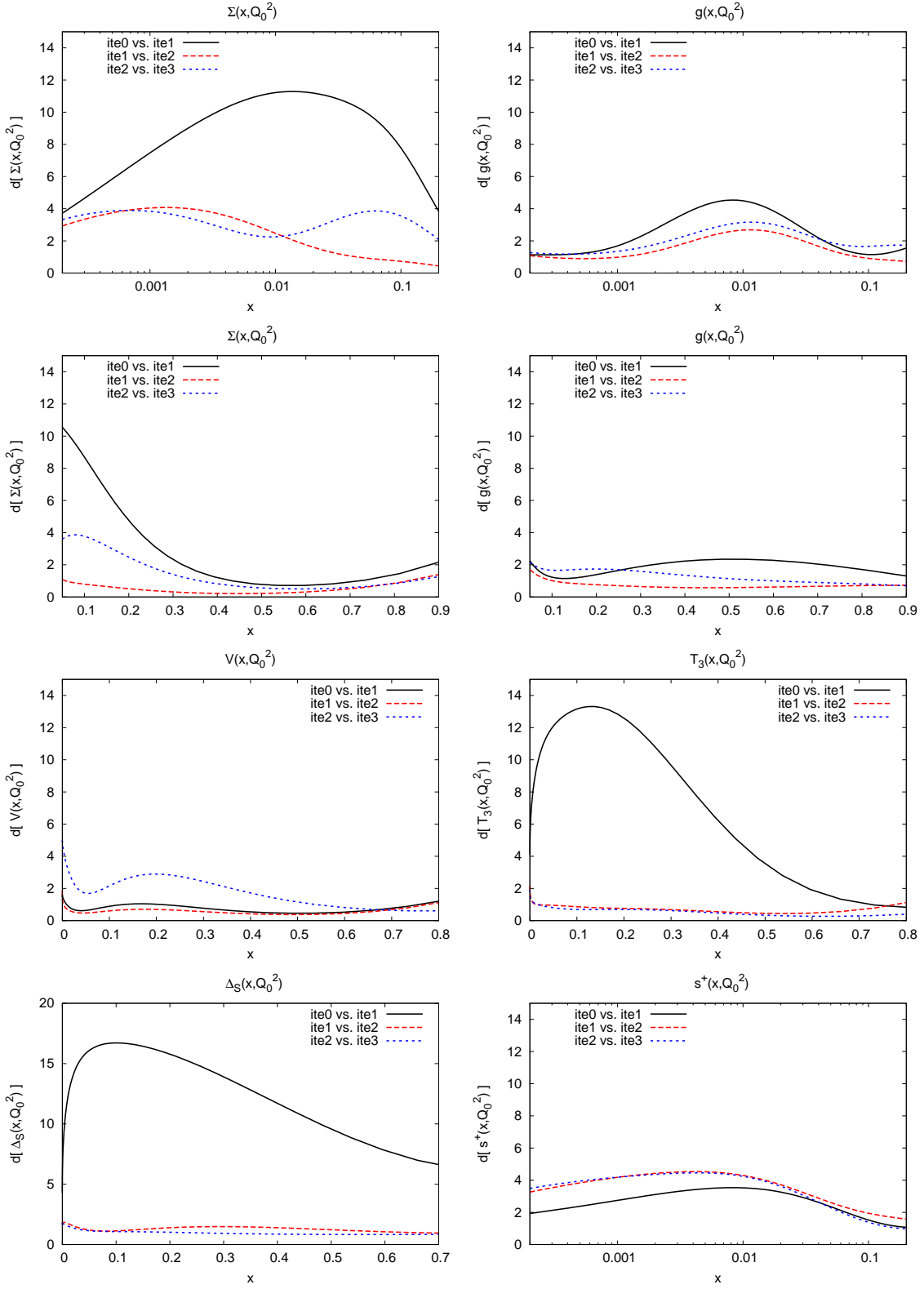
Figure 4: Distance between pairs of PDFs in subsequent iterations of the t_0 method. For the singlet and the gluon PDFs the distances are shown both at small x and at large x.

CHORUS. The table also shows that the χ^2 does not improve after the first iteration, which provides further evidence for the convergence of the t_0 method at the first iteration.

Finally, the PDFs from the starting $t_0 = 0$ NNPDF1.2 fit and those of the final iteration (ite3) of the t_0 method, together with their respective uncertainties, are compared in Fig. 5. As expected from the distance analysis the most important shifts in central values due to the inclusion of normalization uncertainties may be seen in the singlet, the triplet, and the sea asymmetry, though even these are generally small (ie around one-sigma). It is interesting to observe that these changes in central values of the triplet in the valence region, and of the sea asymmetry, are accompanied by a reduction in the overall uncertainty, due to the improved compatibility between the various datasets once normalization uncertainties are properly taken into account.

8   Conclusions

We have studied various methods for the inclusion of multiplicative normalization uncertainties in combined fits to multiple data sets, using both Hessian and Monte Carlo methods, specifically but not necessarily in view of applications to PDF determination. We reviewed how the simplest approach of including the normalization uncertainties in the covariance matrix leads to the well–known "d'Agostini bias" [9], and showed that the commonly used penalty trick method, while fine for the analysis of a single set of experimental data, can lead to biases when combining several independent data sets. We then developed a new technique, the t_0-method, in which normalization uncertainties are introduced into the covariance matrix in such a way that the results are unbiased. While this technique requires iteration to self–consistency, we showed that in practice the convergence is very fast, so that only one iteration is generally required.

To demonstrate the practical application of the t_0-method, we implemented it in the most recent published parton fit by the NNPDF collaboration, NNPDF1.2 [7]. This confirmed the rapid convergence of the technique, showed that the inclusion of normalization uncertainties can lead to a small improvement in the quality of the fit through the resolution of tensions between datasets, and moreover that where these tensions are significant this can lead to a subsequent reduction in PDF uncertainties.

We note that the t_0-method, while very well suited to the determination of PDF uncertainties by Monte Carlo methods, could also be used in the more traditional Hessian fitting methods, where it would lead to faster minimization (since in the t_0-method the dataset normalizations are not fitted), and more reliable central values (since unlike the penalty trick the t_0-method is free from systematic bias). However, the Hessian estimate of uncertainties is still a little less reliable than that from the Monte Carlo method, since quadratic cross-variance terms (such as in Eq. (19)) are always missing, and Gaussian distributions of PDF parameters are always implicitly assumed.

The t_0-method has now been used in the global NNPDF fits [15], where the compatibility of deep–inelastic and hadronic data is a relevant issue. As expected, here the inclusion of the normalization uncertainties results in a significant improvement in the quality of the fit to the hadronic data sets.

Note that the d'Agostini and penalty trick biases discussed in this paper affect all multiplicative errors, not only overall normalizations. Many of the systematic errors in cross-
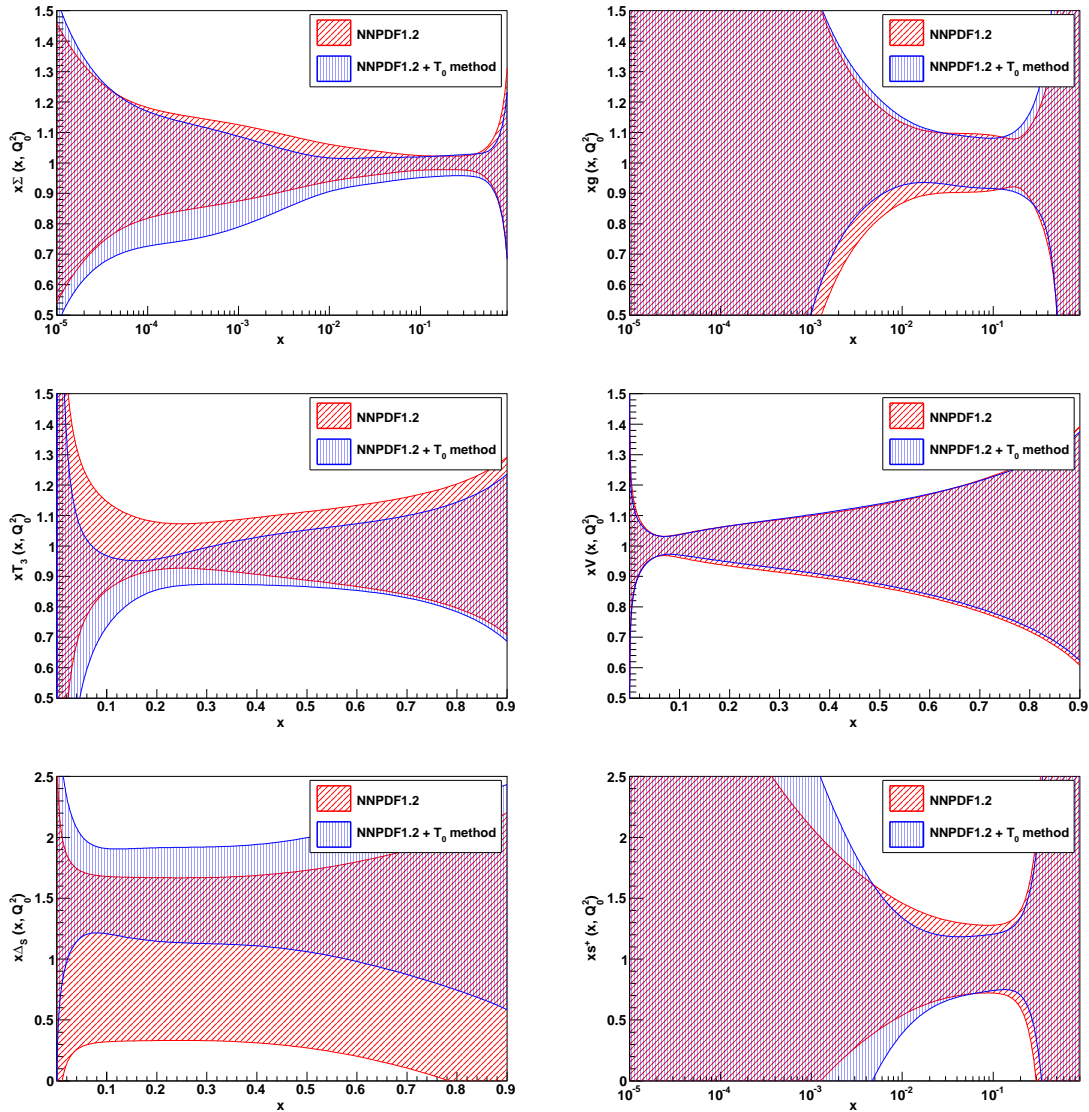
Figure 5: Comparison between the PDFs obtained from the **ite3** NNPDF1.2$-t_0$ fit with the standard ($t_0 = 0$) NNPDF1.2 [7] fit. The quantity plotted is the ratio of the difference between the NNPDF1.2$-t_0$ and NNPDF1.2 results to the NNPDF1.2 itself.

section measurements are closer to multiplicative than additive (see for example Ref. [13]). The t_0-method might thus be developed into a general technique to obtain bias free fits to data sets with a variety of multiplicative systematic uncertainties.

Acknowledgments

References

[1] M. Dittmar *et al.*, arXiv:hep-ph/0511119.

   M. Dittmar *et al.*, arXiv:0901.2504 [hep-ph].

[2] S. Forte, L. Garrido, J. I. Latorre and A. Piccione, JHEP **0205** (2002) 062 [arXiv:hep-ph/0204232].

[3] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione and J. Rojo [NNPDF Collaboration], JHEP **0503** (2005) 080 [arXiv:hep-ph/0501067].

[4] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione and J. Rojo [NNPDF Collaboration], JHEP **0703** (2007) 039 [arXiv:hep-ph/0701127].

[5] R. D. Ball *et al.* [NNPDF Collaboration], Nucl. Phys. B **809** (2009) 1 [Erratum-ibid. B **816** (2009) 293] [arXiv:0808.1231 [hep-ph]].

[6] J. Rojo *et al.* [NNPDF Collaboration], arXiv:0811.2288 [hep-ph].

[7] R. D. Ball *et al.* [The NNPDF Collaboration], Nucl. Phys. B **823**, 195 (2009) [arXiv:0906.1958 [hep-ph]].

[8] W. T. Giele, S. A. Keller and D. A. Kosower, arXiv:hep-ph/0104052.

[9] G. D'Agostini, "Bayesian reasoning in data analysis: A critical introduction", World Scientific, 2003.

[10] H.J. Behrend et al [CELLO Collaboration], Phys. Lett. **B183** (1987) 400.

[11] G. D'Agostini, Nucl. Instrum. Meth. **A346** (1994) 306.

[12] T. Takeuchi, Prog. Theor. Phys. Suppl. **123** (1996) 247.

[13] F.D. Aaron et al, Eur. Phys. J. **C63** (2009) 625; JHEP **1001** (2010) 109 [arXiv:0911.0884 [hep-ex]].

[14] W.T. Eadie, D. Dryard, F.E. James, M. Roos and B. Sadoulet, "Statistical Methods in Experimental Physics", North Holland (1971).

[15] R. D. Ball *et al.* [The NNPDF Collaboration], [arXiv:1002.4407 [hep-ph]].

[16] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, Eur. Phys. J. C **63**, 189 (2009) [arXiv:0901.0002 [hep-ph]].

[17] P. M. Nadolsky *et al.*, Phys. Rev. D **78** (2008) 013004 [arXiv:0802.0007 [hep-ph]].

[18] P. M. Nadolsky, *private communication*

[19] L. Lyons, A.J. Martin and D.H. Saxon, Phys. Rev. D**41** (1990) 982.