



On the consistent use of scale variations in PDF fits and predictions

L. A. Harland-Lang^{1,a}, R. S. Thorne²

¹ Rudolf Peierls Centre, Beecroft Building, Parks Road, Oxford OX1 3PU, UK

² Department of Physics and Astronomy, University College London, London WC1E 6BT, UK

Received: 30 November 2018 / Accepted: 27 February 2019 / Published online: 12 March 2019
© The Author(s) 2019

Abstract We present an investigation of the theoretical uncertainties in parton distribution functions (PDFs) due to missing higher-order corrections in the perturbative predictions used in the fit, and their relationship to the uncertainties in subsequent predictions made using the PDFs. We consider in particular the standard approach of factorization and renormalization scale variation, and derive general results for the consistent application of these at the PDF fit stage. To do this, we use the fact that a PDF fit may be recast in a physical basis, where the PDFs themselves are bypassed entirely, and one instead relates measured observables to predicted ones. In the case of factorization scale variation we find that in various situations there is a high degree of effective correlation between the variation in the fit and in predicted observables. In particular, including such a variation in both cases can lead to an exaggerated theoretical uncertainty. More generally, a careful treatment of this correlation appears mandatory, at least within the standard scale variation paradigm. For the renormalization scale, the situation is less straightforward, but again we highlight the potential for correlations between related processes in the fit and predictions to enter at the same level as between processes in the fit or prediction alone.

1 Introduction

The history of the determination of parton distribution functions (PDFs) from comparison to data goes back many decades, see [1] for a recent review. For some years the precision in both the data and theory was such that no systematic uncertainty estimate on the PDF at all was warranted or required. If some estimate of uncertainty were needed, then a comparison of different PDFs from different groups, or using different assumptions for one group, were thought to be sufficient. The situation changed in the first years of the new millennium. This was largely driven by the very precise

measurements of structure function data over a wide range of both x and Q^2 by the HERA experiment (see [2] for the final Run I + II combination). In addition, various apparent observed excesses over the Standard Model predictions, such as in high E_{\perp} inclusive jet production at CDF [3] were subsequently explained by a suitable modification of the PDFs [4], rather than being due to new physics. A systematic evaluation of PDF uncertainties therefore became essential, and the first global PDFs with an estimate of the uncertainty due to the experimental precision were released by CTEQ [5,6] and MRST [7], building on earlier DIS—only fits [8–11].

In these first fits PDF uncertainties were a few percent at best, and much larger for many PDF flavours and x regions. Soon after this the full calculation of the next-to-next-to leading order (NNLO) splitting functions for the evolution of PDFs were presented in full [12,13] and NNLO extractions of PDFs became possible, as in e.g. [14,15] (with minor approximations for some data sets). At this point it was assumed that the theoretical precision on the PDFs was rather better than the uncertainties from the data, as well as being more difficult to quantify. Hence PDF uncertainties were always interpreted as being an experimental uncertainty due to the statistical and systematic uncertainties of the data included in the fit. In recent years the understanding of the experimental uncertainties on PDFs has improved, and as well as the Hessian-based approach of the original PDF fitters an alternative approach based on neural networks and generation of statistically distributed replicas of PDFs has reached full maturity, as implemented in the NNPDF sets (see [16] for the most recent fit). Good agreement between the two approaches, both for central values and uncertainties is seen [17]. Indeed PDFs can be combined and those generated via the Hessian procedure can be converted to replicas and vice versa [18,19].

Any systematic consideration of theoretical uncertainties in PDF fits was limited to variations due to changes in the strong coupling $\alpha_S(M_Z^2)$, quark masses and sometimes possible higher twist terms. In addition to this, the convergence of different variants of variable flavour number schemes has

^a e-mail: lucian.harland-lang@physics.ox.ac.uk

been demonstrated [20], and some agreement on the influence of use of variable flavour as opposed to fixed flavour schemes, and on fitting rather than imposing cuts for higher twist effects has been demonstrated [21–23]. However, these omit a potentially significant source of uncertainty, namely due to the fact that fixed-order perturbative predictions are used for the theoretical input to PDF fits. The uncertainty due to the approximate form of these, namely due to missing higher order (MHO) corrections, has until very recently not been considered at all, even on a semi-quantitative basis. We should note that introduction of a tolerance in PDF fits, such as the dynamic tolerance procedure for $\Delta\chi^2$ determination introduced in [24], is largely introduced to take account of tension between data sets in the fit, as justified in for example [18]. However, some of the apparent tension between data sets in a fixed order fit may in fact be due to these missing higher order corrections. Such an effect is clearly seen in a LO PDF fit, for example, where the NLO corrections to DIS and Drell Yan cross sections are very different. Hence some part of the tolerance can likely be attributed to this, although in a NNLO fit this is probably a small component, and it is certainly challenging to quantify this precisely. Nonetheless, it would be interesting to pursue a quantitative study of the degree to which observed tensions between data sets in the fit reduce as additional higher order corrections and/or uncertainties due to the missing higher orders are included.

A more focussed study of the theoretical uncertainties in PDFs has just begun in some quarters, and preliminary results have been presented in [25]. This is based on a variation of factorization and renormalization scales by a fixed factor of two in the theory input for the fit, a method that is frequently taken as the standard means of estimating MHO corrections in QCD. The purpose of this article is to critically examine a potentially important and quite general issue with taking such an approach, based on the fact that the PDFs are not themselves physical quantities. In particular, we argue that this type of straightforward scale variation does not provide a particularly obvious definition of what one means by ‘theoretical uncertainty’ for PDFs.

In more detail, in practice one obtains the PDFs by fitting to data for one physical quantity (or more generally a set of them), and then predicts another physical quantity from these, using perturbatively calculated partonic cross sections for these quantities. Ultimately it is the uncertainty on the predicted quantity that is required. This clearly has a contribution from the experimental uncertainty on the PDFs, and this is included as standard. There is then also a theoretical uncertainty on the prediction arising from the finite order of the calculation for both the predicted cross section and for the cross sections entering the PDF extraction. The former of these is included as standard (normally using scale variations) while the latter is not. However, we will argue in this article that for factorisation scale variation there is a highly

non-trivial interplay between the scale variation when obtaining the PDF and when obtaining the required prediction, and this can potentially lead to a misinterpretation of the ‘theory uncertainty’ on the prediction. In short, the aim of the process is to measure one physical quantity, and in terms of this predict another quantity. If, as seems natural, one interprets the ‘theory uncertainty’ as that inherent in expressing the predicted quantity in terms of the measured quantity due to MHOs in the relationship between the two, then we will show that varying the factorisation scale by a set amount in both the PDF extraction and also in the prediction in terms of PDFs can lead to an effectively exaggerated factorisation scale variation when determining the full theoretical uncertainty. Our arguments rely on the fact that it is possible, and sometimes preferable (in principle at least), to bypass the intermediate PDFs entirely, instead working purely at the level of physical observables (structure functions and so on) and the relationships between them. This was behind the original proposal of the DIS factorisation scale [26], and has subsequently been developed in for example [27–32] under the general name of ‘physical schemes’.

We also consider the case of the renormalization scale variation, which we find to be less amenable to this treatment, implying that the conventional approach should be reliable here. Nonetheless, one basic implication of working in this physical basis where one considers a PDF fit to be a (complex) relation between physical observables does follow in this case. Namely, any correlation between renormalization scale variations that one assumes is present in related physical processes entering the fit, should also in principle be included at the same level between fit and predicted processes.

The outline of this paper is as follows. In Sect. 2 we consider the simplest possible case of fitting to a non-singlet structure function observable, and then predicting a second such structure function, and a non-singlet Drell–Yan cross section, in Sects. 2.1 and 2.2, respectively. In Sect. 3 we generalise this to the case of structure functions involving both quark and gluon contributions, and comment on the corresponding high and low x limits. In Sect. 4 we discuss the case of renormalization scale variation. In Sect. 5 we discuss the implications of our findings and conclude. In Appendix A we briefly summarise the case of DGLAP evolution in the diagonal basis at NLO.

2 A simple example: the non-singlet quark

2.1 Structure functions

To illustrate the key issue, we start with the simplified case of a fit to a non-singlet structure function, and corresponding prediction of a second (distinct) non-singlet structure function. This represents the simplest example of our general

argument, and having done this we will show how the case of a predicted hadronic observables, namely the non-singlet Drell–Yan cross section, follows straightforwardly.

The structure function we consider is given purely in terms of the quark non-singlet distribution. A concrete example of this is the neutral current structure function F_3^{NC} due to Z exchange, but we will for the sake of generality refer to the observable as F_{NS} . We write this to NLO as

$$F_{\text{NS}}(x, Q^2) = xq_{\text{NS}}(x, \mu^2) + \tilde{\alpha}_S C_q^{(1)} \otimes xq_{\text{NS}}(x, \mu^2) + \tilde{\alpha}_S \ln\left(\frac{Q^2}{\mu^2}\right) P_{qq}^{(0)} \otimes xq_{\text{NS}}(x, \mu^2), \tag{1}$$

where $\tilde{\alpha}_S \equiv \alpha_S/2\pi$, μ is the factorization scale, $C_q^{(n)}$ is the order n coefficient function, and we consider for simplicity only one quark flavour. Note that for now we assume a fixed renormalization scale, the argument of which is suppressed for simplicity, and do not consider any issues related to its variation.¹ To keep the expressions which follow simpler we have also taken the leading order coefficient function as $C_q^{(0)} = 1$, which can always be achieved by a suitable redefinition of the normalisation of F_{NS} . We use the shorthand

$$g \otimes f(x) = \int_x^1 \frac{dz}{z} f(z)g\left(\frac{x}{z}\right), \tag{2}$$

throughout. The non-singlet quark combination $q_{\text{NS}} = q - \bar{q}$ obeys the usual NLO DGLAP evolution

$$\frac{\partial q_{\text{NS}}(x, \mu^2)}{\partial \ln \mu^2} = \tilde{\alpha}_S \left(P_{qq}^{(0)} + \tilde{\alpha}_S P_{qq}^{(1)} \right) \otimes q_{\text{NS}}(x, \mu^2). \tag{3}$$

We now consider an idealised PDF fit of q_{NS} to the structure function, that is we assume that F_{NS} has been measured to arbitrary accuracy over the x region we are interested in. We are free to do this as we are only considering the impact of theoretical uncertainties on the fit, and the inclusion of the (unrelated) experimental sources of uncertainty will not qualitatively effect the argument which follows. Indeed, we are precisely most interested in the case where the former dominates over the latter. Defining the ratio $a_i = \mu^2/Q^2$, we can rewrite (1) as

¹ We are implicitly using the ‘‘standard’’ convention, see e.g. [33], that in the PDF evolution the scale of the coupling is taken to be the same as the factorization scale, i.e the PDFs depends on only one scale. However, the arguments all remain the same if this scale of the coupling in the evolution is related to the factorization scale by $\mu_R = c\mu_F$ if c is the same for all physical quantities, i.e. the scale choice in the coupling for PDF evolution is not process-dependent.

$$\begin{aligned} xq_{\text{NS}}(x, \mu^2) &= F_{\text{NS}}(x, \mu^2/a_i) - \tilde{\alpha}_S C_q^{(1)} \otimes xq_{\text{NS}}(x, \mu^2) \\ &\quad + \tilde{\alpha}_S \ln a_i P_{qq}^{(0)} \otimes xq_{\text{NS}}(x, \mu^2), \tag{4} \\ &= F_{\text{NS}}(x, \mu^2/a_i) - \tilde{\alpha}_S C_q^{(1)} \otimes F_{\text{NS}}(x, \mu^2/a_i) \\ &\quad + \tilde{\alpha}_S \ln a_i P_{qq}^{(0)} \otimes F_{\text{NS}}(x, \mu^2/a_i), \tag{5} \end{aligned}$$

where in the second line we consistently drop terms of $O(\alpha_S^2)$. Note a ‘standard’ fit, i.e. where one does not consider any scale variation and takes the conventional choice of $\mu^2 = Q^2$, simply corresponds to

$$xq_{\text{NS}}(x, \mu^2) = F_{\text{NS}}(x, \mu^2) - \tilde{\alpha}_S C_q^{(1)} \otimes F_{\text{NS}}(x, \mu^2), \tag{6}$$

while for example a standard factor of 2 scale variation about the central scale $\mu = Q$ would correspond to taking $a_i \in (\frac{1}{4}, 4)$.

We now use this to predict a second, distinct, non-singlet structure function, F'_{NS} :

$$\begin{aligned} F'_{\text{NS}}(x, Q^2) &= xq_{\text{NS}}(x, \mu^2) + \tilde{\alpha}_S C_q^{(1)} \otimes xq_{\text{NS}}(x, \mu^2) \\ &\quad + \tilde{\alpha}_S \ln\left(\frac{Q^2}{\mu^2}\right) P_{qq}^{(0)} \otimes xq_{\text{NS}}(x, \mu^2), \\ &= xq_{\text{NS}}(x, a_f Q^2) + \tilde{\alpha}_S C_q^{(1)} \otimes xq_{\text{NS}}(x, a_f Q^2) \\ &\quad - \tilde{\alpha}_S \ln a_f P_{qq}^{(0)} \otimes xq_{\text{NS}}(x, a_f Q^2), \tag{7} \end{aligned}$$

where we have defined $a_f = \mu^2/Q^2$. We now consider the standard fit approach, that is using (6) to express the quark distribution in terms of the structure function, giving

$$\begin{aligned} F'_{\text{NS}}(x, Q^2) &= F_{\text{NS}}(x, a_f Q^2) \\ &\quad + \tilde{\alpha}_S \left(C_q^{(1)} - C_q^{(1)} \right) \otimes F_{\text{NS}}(x, a_f Q^2) \\ &\quad - \tilde{\alpha}_S \ln a_f P_{qq}^{(0)} \otimes xF_{\text{NS}}(x, a_f Q^2), \tag{8} \end{aligned}$$

Thus we can rewrite our prediction so that no reference is made to the intermediate PDFs, and instead we express one observable quantity, F'_{NS} , in terms of the another, F_{NS} . This expression is accurate to $O(\alpha_S)$ and following the standard rule of thumb approach we can then evaluate the theoretical uncertainty from MHOs by performing a scale variation with $a_f \in (\frac{1}{4}, 4)$.

We now consider the case that the scale is allowed to vary in the fit as well, which simply corresponds to keeping the a_i dependence as in (5). We find

$$\begin{aligned} F'_{\text{NS}}(x, Q^2) &= F_{\text{NS}}(x, a_{fi} Q^2) \\ &\quad + \tilde{\alpha}_S \left(C_q^{(1)} - C_q^{(1)} \right) \otimes F_{\text{NS}}(x, a_{fi} Q^2) \\ &\quad - \tilde{\alpha}_S \ln a_{fi} P_{qq}^{(0)} \otimes xF_{\text{NS}}(x, a_{fi} Q^2), \tag{9} \end{aligned}$$

where we have defined $a_{fi} \equiv a_f/a_i$. Thus the a_f and a_i dependence is entirely contained in the ratio a_f/a_i , such that (9) is identical to (8) upon the replacement $a_f \rightarrow a_{fi}$. Of course, if we expand out about e.g. fixed scale Q^2 , then

$$F_{NS}(x, a_{fi} Q^2) = F_{NS}(x, Q^2) + \tilde{\alpha}_S \ln a_{fi} P_{qq}^{(0)} \otimes x F_{NS}(x, Q^2), \tag{10}$$

and the dependence on $\ln(a_{fi})$ in (9) vanishes at $O(\alpha_S)$ due to cancellations between the first and third terms, and only appears at $O(\alpha_S^2)$, where higher-order terms beyond those shown in (9) would be required for further cancellations. We note also, that at this stage the distinction between a_{fi} and a_f is really entirely artificial. There is only one physical relation between the two structure functions

$$F'_{NS}(x, Q^2) = F_{NS}(x, a Q^2) + \tilde{\alpha}_S \left(C_q^{(1)} - C_q^{(1)} \right) \otimes F_{NS}(x, a Q^2) - \tilde{\alpha}_S \ln a P_{qq}^{(0)} \otimes x F_{NS}(x, a Q^2), \tag{11}$$

where a corresponds to the relative difference in scale at which we evaluate F_{NS} and F'_{NS} . In other words, we can see that the effect of varying the scale in the fit (5) is the same as the previous approach, but with a larger range of variation, $a = a_{fi} \in (\frac{1}{16}, 16)$.

How should we interpret this result? The rule of thumb variation is applied to a broad category of observables, under the expectation that this will provide an estimate of the MHO uncertainty. Concretely, one varies the logarithms in a within a reasonable range, in order to track of the decreasing dependence on these with increasing perturbative order, but nonetheless keeping the argument a to be $O(1)$ in order to avoid spoiling the overall perturbative convergence. The precise choice of $a \in (\frac{1}{4}, 4)$ above is of course arbitrary, but is nonetheless guided by these principles.

We have seen in the above scenario that the intermediate PDFs themselves can be bypassed entirely in favour of a straightforward and arguably more fundamental relation between the physical observables F_{NS} and F'_{NS} . This simply reflects the fact that the PDFs are not themselves observables, and follows in a similar way to the physical factorization approach discussed elsewhere [27–29,32]. In terms of this relation, there is only one degree of freedom for scale variation, namely a in (11). Within the context of the standard rule of thumb variation, the only reasonable and consistent choice appears to be to take (11) and vary $a \in (\frac{1}{4}, 4)$.

Now of course from a practical point of view one will not in general work explicitly in this physical framework, but rather in terms of the PDFs. The aim should therefore be to be consistent with the above results when doing so and evaluating a theoretical uncertainty on the PDFs themselves. We have seen above that in our example one should either vary the factorization scale in the prediction by the canonical factor of 2, or equivalently in the fit, but not in both. One may clearly call into question the reliability of such simple scale variations, but nonetheless at least under the assumption that this $a \in (\frac{1}{4}, 4)$ variation provides an accurate estimate of

the theoretical uncertainty for general observables, this result will hold.

While the above result is demonstrated at NLO for simplicity, this remains true at arbitrary order. This becomes clearest when we work in Mellin space, where we can write the DGLAP evolution (3) in the simple form

$$q_{NS}(j, \mu^2) = q_{NS}(j, Q^2) \left(\frac{\mu^2}{Q^2} \right)^{\tilde{\alpha}_S \gamma_{qq}(j, \tilde{\alpha}_S)}, \tag{12}$$

where j denotes the Mellin moment. Then, our expression for F_{NS} , at a scale Q , can be written in the form

$$F_{NS}(j, Q^2) = q_{NS}(j + 1, \mu^2) \left(\frac{Q^2}{\mu^2} \right)^{\tilde{\alpha}_S \gamma_{qq}(j, \tilde{\alpha}_S)} c_q(j, \tilde{\alpha}_S), \\ = q_{NS}(j + 1, \mu^2) a_i^{-\tilde{\alpha}_S \gamma_{qq}(j, \tilde{\alpha}_S)} c_q(j, \tilde{\alpha}_S), \tag{13}$$

where we define $a_i = \mu^2/Q^2$ as before. In the above expression, and in what follows, it is understood that the result at any arbitrary order in perturbation theory is given by expanding this out to the desired order in α_S . Here

$$\gamma_{qq}(j, \tilde{\alpha}_S) = \sum_{k=0}^n (\tilde{\alpha}_S)^k \gamma_{qq}^{(k)}(j), \quad c_q(j, \tilde{\alpha}_S) = \sum_{k=0}^n (\tilde{\alpha}_S)^k c_q^{(k)}(j), \tag{14}$$

correspond to the non-singlet qq anomalous dimension and Mellin transform of the coefficient function at order n , respectively. Note that we have defined $c_q^{(0)} \equiv 1$ in the structure function case above, but we leave the expression completely general here. Now if we consider the second structure function at the same scale Q , we have

$$F'_{NS}(j, Q^2) = q_{NS}(j + 1, \mu^2) a_i^{\tilde{\alpha}_S \gamma_{qq}(j, \tilde{\alpha}_S)} c'_q(j, \tilde{\alpha}_S), \tag{15}$$

and so expressing $F'_{NS}(j, Q^2)$ in terms of $F_{NS}(j, Q^2)$ we obtain the very simple result

$$F'_{NS}(j, Q^2) = \frac{c'_q(j, \tilde{\alpha}_S)}{c_q(j, \tilde{\alpha}_S)} F_{NS}(j, Q^2). \tag{16}$$

We can see that there is now no explicit dependence on the factorization scale, μ , at all. The above situation is however defined for the quite specific case that we wish to relate the two observables at exactly the same scale. More generally, we have the freedom to express $F'_{NS}(j - 1, Q_f^2)$ in terms of $F_{NS}(j - 1, Q_i^2)$, that is choose to express the predicted quantity a different physical scale, Q_f , to the scale at which we express the measured quantity, Q_i . In this case we have

$$F'_{NS}(j, Q_f^2) = \frac{c'_q(j, \tilde{\alpha}_S)}{c_q(j, \tilde{\alpha}_S)} \left(\frac{Q_f^2}{Q_i^2} \right)^{\tilde{\alpha}_S \gamma_{qq}(j, \tilde{\alpha}_S)} F_{NS}(j, Q_i^2). \tag{17}$$

Hence, we see that to any arbitrary order the relationship between the prediction and the measurement can be expressed as a ratio of scales, and that the relationship between the factorization scale chosen when fitting the PDF to that chosen when making the prediction becomes equivalent to the relationship between the scale at which one predicts the physical quantity and that at which one evaluates the physical quantity entering the fit. To be completely explicit, we can suggestively define $Q^2 \equiv Q_f^2$ and $\mu^2 \equiv Q_i^2$, in terms of which we have

$$F'_{NS}(j, Q^2) = \frac{c'_q(j, \tilde{\alpha}_s)}{c_q(j, \tilde{\alpha}_s)} \left(\frac{Q^2}{\mu^2}\right)^{\tilde{\alpha}_s \gamma_{qq}(j, \tilde{\alpha}_s)} F_{NS}(j, \mu^2), \tag{18}$$

where F_{NS} obeys the same DGLAP evolution Eq. (12) as q_{NS} , up to terms in the ratio of the coefficient function which depend on the running of the coupling; indeed, if we more correctly re-introduce the scale dependence of α_s into the above results, these would follow in precisely the same way, up to this difference in evolution. Thus, we can express this at any arbitrary order, as is done in (1) at NLO, and we have an analogous freedom to vary the scale μ at which one evaluates F_{NS} as in the case of the standard factorization scale. However, as we have only one ratio of scales involved in this relation, there is only one such degree of freedom, and not the two implied by varying the factorization scale independently in the fit and prediction.

Finally, we note that while it might seem most direct in the above expression to choose the scale of F_{NS} ($\mu = Q_i$) equal to the scale at which the measurement is made, this is, of course, not mandatory. If the scale at which one structure function is fit is significantly different to that at which the second is to be predicted ($Q = Q_f$) it would normally be assumed to be more sensible to express the measured quantity at a scale similar to the predicted quantity, relying on the validity of the evolution equation and avoiding obvious large logarithms in the expression relating the two physical quantities.

2.2 Drell–Yan cross section

As a second simple example of the above argument, and to demonstrate how the above result can readily be generalised to the case of hadronic observables, we can use the same fit as above to predict the non-singlet Drell–Yan production cross section, i.e. $q_{NS}(x_1)q_{NS}(x_2) \rightarrow l^+l^-$. We can write this at NLO as

$$\frac{d\sigma_{NS}^{DY}}{dQ^2} = \int_0^1 dx_1 dx_2 dz \delta(x_1 x_2 z - \tau) q_{NS}(x_1, \mu^2) q_{NS}(x_2, \mu^2) \cdot \left[\delta(1 - z) + 2\tilde{\alpha}_S \left(P_{qq}^0(z) \ln \frac{Q^2}{\mu^2} + C_{DY,q}^{(1)}(z) \right) \right], \tag{19}$$

where Q^2 is the dilepton invariant mass and $\tau = Q^2/s$. As before, for simplicity we normalise our cross section so that the LO coefficient function is unity. Defining $a_f = \mu^2/Q^2$, to the order we are calculating we can rewrite this as

$$\begin{aligned} \frac{d\sigma_{NS}^{DY}}{dQ^2} = & \int_{\tau}^1 \frac{dx_1}{x_1} \left(q_{NS}(x_1, a_f Q^2) - \tilde{\alpha}_S \ln a_f P_{qq}^0 \right. \\ & \left. \otimes q_{NS}(x_1, a_f Q^2) + C_{DY,q}^{(1)} \otimes q_{NS}(x_1, a_f Q^2) \right) \\ & \cdot \left(q_{NS}(x_2, a_f Q^2) - \tilde{\alpha}_S \ln a_f P_{qq}^0 \right. \\ & \left. \otimes q_{NS}(x_2, a_f Q^2) + C_{DY,q}^{(1)} \otimes q_{NS}(x_2, a_f Q^2) \right). \end{aligned}$$

where $x_2 = \tau/x_1$. Proceeding as before, and expressing the quark distribution in terms of the non-singlet structure function, we have

$$\begin{aligned} \frac{d\sigma_{NS}^{DY}}{dQ^2} = & \int_{\tau}^1 \frac{dx_1}{x_1} \left(F_{NS}(x_1, a Q^2) - \tilde{\alpha}_S C_q^1 \otimes F_{NS}(x_1, a Q^2) \right. \\ & \left. - \tilde{\alpha}_S \ln a P_{qq}^0 \otimes F_{NS}(x_1, a Q^2) \right. \\ & \left. + C_{DY,q}^{(1)} \otimes F_{NS}(x_1, a Q^2) \right) \left(F_{NS}(x_2, a Q^2) \right. \\ & \left. - \tilde{\alpha}_S C_q^1 \otimes F_{NS}(x_2, a Q^2) \right. \\ & \left. - \tilde{\alpha}_S \ln a P_{qq}^0 \otimes F_{NS}(x_2, a Q^2) + C_{DY,q}^{(1)} \otimes F_{NS}(x_2, a Q^2) \right), \tag{20} \end{aligned}$$

where $a = a_f$ for the case that we vary the scale only in the prediction, and $a = a_{fi}$ if we vary in both cases. In other words, the argument follows through in exactly the same way.

Exactly as in the previous example, we note that the above results can be more simply expressed in Mellin space. In particular, taking the Mellin transform with respect to τ the DY cross section simply becomes

$$\begin{aligned} \frac{d\sigma_{NS}^{DY}(j, Q^2)}{dQ^2} = & \left(q_{NS}(j, \mu^2) a_f^{\tilde{\alpha}_s \gamma_{qq}(j, \tilde{\alpha}_s)} \right)^2 \left[1 + 2\tilde{\alpha}_S c_q^{DY}(j) + \dots \right], \\ = & \left(F_{NS}(j - 1, a_{fi} Q^2) a_{fi}^{\tilde{\alpha}_s \gamma_{qq}(j, \tilde{\alpha}_s)} \right)^2 \\ & \times \left[1 + 2\tilde{\alpha}_S c_q^{DY}(j) + \dots \right], \tag{21} \end{aligned}$$

where the anomalous dimension can be calculated at any arbitrary order, and the ‘...’ indicate these higher order contributions to the coefficient function.

3 A more general example: the quark singlet and gluon

3.1 Set-up

The above examples considered the special case of observables given in terms of a single non-singlet quark distribution. This leads to a simple and transparent result, but it is

not immediately clear how it will generalise to the case that includes both quark and gluon partons, which obey the fully coupled DGLAP equation.

For the sake of generality and demonstration, we consider a PDF fit to a pair of arbitrary ‘structure function’ observables F and H , which are then used to predict a third such observable, K . The generalisation to the case of hadronic observables would render the corresponding analysis a great deal more complex in practice, but in principle should not change the basic argument. We will also work in Mellin space, as this will simplify the calculation, although all the results which follow hold analogously in x space as well. We write

$$\begin{aligned}
 F(j, Q^2) &= \Sigma_q(j + 1, \mu^2) F_q \left(j, \frac{Q^2}{\mu^2} \right) \\
 &\quad + g(j + 1, \mu^2) F_g \left(j, \frac{Q^2}{\mu^2} \right), \\
 H(j, Q^2) &= \Sigma_q(j + 1, \mu^2) H_q \left(j, \frac{Q^2}{\mu^2} \right) \\
 &\quad + g(j + 1, \mu^2) H_g \left(j, \frac{Q^2}{\mu^2} \right). \tag{22}
 \end{aligned}$$

Thus, we assume that these observables depend on the gluon (g) and total quark singlet (Σ_q) PDFs only. In other words, any dependence on non-singlet quark combinations, which would be introduced by e.g. including a quark flavour-dependence (due typically to the quark EW charges), is omitted to limit the observables we need to consider, although the set-up can readily be generalised to include these.

We will drop the Mellin moment argument j for brevity in what follows. We can write the coefficient functions at NLO as

$$\begin{aligned}
 F_q \left(\frac{Q^2}{\mu^2} \right) &= c_q^{(0)} + \tilde{\alpha}_S c_q^{(1)} + \tilde{\alpha}_S c_q^{(0)} \ln \left(\frac{Q^2}{\mu^2} \right) \gamma_{qq}^{(0)} \\
 &\quad + \tilde{\alpha}_S c_g^{(0)} \ln \left(\frac{Q^2}{\mu^2} \right) \gamma_{gq}^{(0)}, \tag{23}
 \end{aligned}$$

$$\begin{aligned}
 F_g \left(\frac{Q^2}{\mu^2} \right) &= c_g^{(0)} + \tilde{\alpha}_S c_g^{(1)} + \tilde{\alpha}_S c_q^{(0)} \ln \left(\frac{Q^2}{\mu^2} \right) \gamma_{qg}^{(0)} \\
 &\quad + \tilde{\alpha}_S c_g^{(0)} \ln \left(\frac{Q^2}{\mu^2} \right) \gamma_{gg}^{(0)}, \tag{24}
 \end{aligned}$$

and similarly for H . Note that the $q \leftrightarrow g$ mixing introduces a corresponding mixing in the coefficients $c_{q,g}$ of the expansions of the $F_{q,g}$, and similarly for $H_{q,g}$. This simplifies if we instead use the basis of eigenvectors of the DGLAP equation, which we denote Σ_{\pm} . In terms of these we can write

$$F(Q^2) = \Sigma_+(\mu^2) \left(\frac{Q^2}{\mu^2} \right)^{\tilde{\alpha}_S \gamma_+} F_+ + \Sigma_-(\mu^2) \left(\frac{Q^2}{\mu^2} \right)^{\tilde{\alpha}_S \gamma_-} F_-, \tag{25}$$

$$H(Q^2) = \Sigma_+(\mu^2) \left(\frac{Q^2}{\mu^2} \right)^{\tilde{\alpha}_S \gamma_+} H_+ + \Sigma_-(\mu^2) \left(\frac{Q^2}{\mu^2} \right)^{\tilde{\alpha}_S \gamma_-} H_-, \tag{26}$$

at arbitrary order. Here the diagonal anomalous dimensions γ_{\pm} and the PDF eigenvectors are given explicitly in Appendix A. These then define the coefficients F_{\pm} and H_{\pm} above, which are given in terms of F_q and F_g , and similarly for H .

We now consider a PDF fit to these observables, for which we take the factorization scales $\mu^2 = a_f Q^2$ and $\mu^2 = a_h Q^2$. In this case we have (see 47)

$$F \left(\frac{\mu^2}{a_f} \right) = \Sigma_+(\mu^2) (a_f)^{-\tilde{\alpha}_S \gamma_+} F_+ + \Sigma_-(\mu^2) (a_f)^{-\tilde{\alpha}_S \gamma_-} F_-, \tag{27}$$

$$H \left(\frac{\mu^2}{a_h} \right) = \Sigma_+(\mu^2) (a_h)^{-\tilde{\alpha}_S \gamma_+} H_+ + \Sigma_-(\mu^2) (a_h)^{-\tilde{\alpha}_S \gamma_-} H_-, \tag{28}$$

which we can then invert to give

$$\Sigma_+(\mu^2) = \frac{H_- F \left(\frac{\mu^2}{a_f} \right) (a_f)^{\tilde{\alpha}_S \gamma_-} - F_- H \left(\frac{\mu^2}{a_h} \right) (a_h)^{\tilde{\alpha}_S \gamma_-}}{F_+ H_- (a_f)^{\tilde{\alpha}_S (\gamma_- - \gamma_+)} - F_- H_+ (a_h)^{\tilde{\alpha}_S (\gamma_- - \gamma_+)}} \tag{29}$$

$$\Sigma_-(\mu^2) = \frac{H_+ F \left(\frac{\mu^2}{a_f} \right) (a_f)^{\tilde{\alpha}_S \gamma_+} - F_+ H \left(\frac{\mu^2}{a_h} \right) (a_h)^{\tilde{\alpha}_S \gamma_+}}{F_- H_+ (a_f)^{-\tilde{\alpha}_S (\gamma_- - \gamma_+)} - F_+ H_- (a_h)^{-\tilde{\alpha}_S (\gamma_- - \gamma_+)}} \tag{30}$$

Substituting these into the third structure function K at scale $\mu^2 = a_k Q^2$, at NLO we find

$$\begin{aligned}
 K(Q^2) &= \frac{1}{F_+ H_- - F_- H_+} \left\{ K_+ H_- - H_+ K_- \right. \\
 &\quad + \tilde{\alpha}_S \ln \left(\frac{a_f}{a_k} \right) [\gamma_+ K_+ H_- - \gamma_- K_- H_+] \\
 &\quad + \tilde{\alpha}_S (\gamma_- - \gamma_+) \ln \left(\frac{a_h}{a_f} \right) \frac{H_+ H_-}{F_+ H_- - F_- H_+} \\
 &\quad \times (K_+ F_- - K_- F_+) \left. \right\} F \left(\frac{a_k}{a_f} Q^2 \right) \\
 &\quad + \frac{1}{H_+ F_- - H_- F_+} \left\{ K_+ F_- - F_+ K_- + \tilde{\alpha}_S \right. \\
 &\quad \times \ln \left(\frac{a_h}{a_k} \right) [\gamma_+ K_+ F_- - \gamma_- K_- F_+] \\
 &\quad - \tilde{\alpha}_S (\gamma_- - \gamma_+) \ln \left(\frac{a_h}{a_f} \right) \frac{F_+ F_-}{H_+ F_- - H_- F_+} \\
 &\quad \times (K_+ H_- - K_- H_+) \left. \right\} H \left(\frac{a_k}{a_h} Q^2 \right), \tag{31}
 \end{aligned}$$

where it is understood that the fixed-order coefficients, H_{\pm} and so on, should also be expanded out to the appropriate order, but we omit this for clarity. For comparison, if we do not vary the scales in the fit, i.e. we take $a_f = a_h = 1$, then we have

$$\begin{aligned}
 K(Q^2) = & F \left(a_k Q^2 \right) \frac{1}{F_+ H_- - F_- H_+} \left\{ K_+ H_- - H_+ K_- \right. \\
 & \left. - \tilde{\alpha}_S \ln a_k \left[\gamma_+ K_+ H_- - \gamma_- K_- H_+ \right] \right\} \\
 & + H \left(a_k Q^2 \right) \frac{1}{H_+ F_- - H_- F_+} \left\{ K_+ F_- - F_+ K_- \right. \\
 & \left. - \tilde{\alpha}_S \ln a_k \left[\gamma_+ K_+ F_- - \gamma_- K_- F_+ \right] \right\}. \quad (32)
 \end{aligned}$$

Thus we can immediately see that the situation is somewhat more complex than in the simplified purely non-singlet case considered in the previous section. In particular, our general expression (31) contains the two ratios $a_{f,h}/a_k$ corresponding to the arguments of the F, H , similar to the non-singlet case we considered before. However in addition we can see that the result includes a contribution that depends on the ratio a_h/a_f , which is purely due to the scale variation in the fit stage, and is completely absent in the prediction without this variation. We note that while the above expression is written in terms of three ratios, only two of these are independent, exactly as we would expect following the discussion towards the end of Sect. 2.1. In particular, we are now expressing one predicted physical quantity defined at one scale in terms of two measured physical quantities, each of which may be evaluated at a different scale. There are therefore two independent scale ratios, and two physically meaningful ratios of factorization scales. On the other hand, while the ratio a_h/a_f can be written in terms of the independent ratios $a_{f,h}/a_k$, we can see that this results in a mixing of these two ratios which does not immediately reduce to the simple situation we had in the case of the non-singlet structure functions.

We note that in [25] it was advocated that, as all physical quantities share common PDFs, the factorization scale should be varied in a fully correlated way across all processes entering the fit. In the above analysis, we can see that this corresponds to taking $a_f = a_h$, and we are left with (32), after replacing $a_k \rightarrow a_k/a_f = a_f/a_h$. In other words, our situation and conclusions are exactly the same as for the simpler non-singlet case, i.e. variations of factorization scale in the predictions are entirely equivalent to those in the fitting and vice versa. We fully expect this to hold in the more general case appropriate to a global fit, as here too we will only have one independent ratio of scales for any given predicted process. Thus, if one makes this assumption, one could bypass the complication of including these variations at the fit stage entirely and simply include them in the prediction, with the

assumption of full correlations implying that this should be done in the same way for different predictions at the same time. On the other hand, varying the factorization scale in both the fit and prediction would be a type of double counting, i.e. varying the scale by a factor of two more than may naively be expected.

However, as discussed further below, in general this appears to be an overly constraining assumption, given there is the question of the central choice of scale to consider and, potentially more significantly, the fact that the partons and x range probed by the fit processes can be rather different. With this in mind, how do we interpret our above result if we do not make this simplifying assumption of fully correlated factorization scales for quantities in the PDF fit? We will first consider this result in various kinematic limits, before discussing the more general implications.

3.2 The low and high x limits

One can simplify the full result (31) by for example assuming that $F_- = 0$ and $H_+ = 0$, in other words that F, H are only sensitive to the contribution from either the negative or positive eigenvectors. In this case we have

$$\begin{aligned}
 K(Q^2) = & \frac{K_+}{F_+} \left\{ 1 + \tilde{\alpha}_S \ln \left(\frac{a_f}{a_k} \right) \gamma_+ \right\} F \left(\frac{a_k}{a_f} Q^2 \right) \\
 & + \frac{K_-}{H_-} \left\{ 1 + \tilde{\alpha}_S \ln \left(\frac{a_h}{a_k} \right) \gamma_- \right\} H \left(\frac{a_k}{a_h} Q^2 \right), \quad (33)
 \end{aligned}$$

and the terms proportional to the ratio a_f/a_h vanish. As an aside, if in addition the prediction is only sensitive to one eigenvector, then this will reduce to a single ratio, in precisely the same way we saw for the non-singlet distribution.

While at first glance it may appear somewhat arbitrary to consider these limits, in fact in the low and high x regions this can be precisely the situation we find ourselves in. As discussed further in Appendix A, if we take the high x limit, we have the well known result that

$$\Sigma_+(j, \mu^2) = g(j, \mu^2), \quad \Sigma_-(j, \mu^2) = \Sigma_q(j, \mu^2), \quad (34)$$

that is the quark distribution ($\sim \Sigma_q, q_{NS}$) and gluon are independent eigenvectors of the DGLAP evolution. In the alternative low x ($j \sim 1$) limit we have

$$g(j, \mu^2) \sim q(j, \mu^2) \sim \Sigma_+(j, \mu^2), \quad (35)$$

for sufficiently high scale μ . That is, the positive eigenvector is dominant.

These regimes play a direct role in PDF phenomenology at the LHC and elsewhere. For example, a topical case is the high x gluon, which is relatively poorly determined, and in which there is currently a great deal of interest in placing further constraints. This typically involves the use of LHC observables such as inclusive jet and $t\bar{t}$ production, and the

Z boson p_{\perp} distribution, for which the high x gluon plays a dominant role. Although a global fit includes of course a wider dataset, the extracted high x gluon will to a significant extent be driven by these. One can then take the result of this fit and predict the gluon-initiated production of e.g. a high mass BSM object. In such a scenario the gluon evolution will be effectively decoupled and both the fit and predicted observables will be dominated by the positive eigenvector Σ_+ . In other words, we are in an analogous situation to Sect. 2, where the factorization scales for the fit and prediction are fully correlated, and varying the factorization scale in both will lead to an effective double counting of the theoretical uncertainty. The corresponding situation for the high x quark, where both the singlet and non-singlet are decoupled from the gluon, is similar.

At low x , as we increase the scale of the observed process we find that the quark and gluon contribution are completely correlated by evolution, and only the positive eigenvector contributes. This is equally true for observables such as scaling violations of the structure function, $dF_2/d \ln Q^2$, which only depends on the positive eigenvector at low x , for all scales. Thus, for fit processes such as a $dF_2/d \ln Q^2$ and predicted processes such as Drell–Yan production at the LHC (in particular in the lower mass region), we will expect a large degree of correlation.

We note that in a realistic PDF fit we will in general include multiple observables at the fit stage which may be dominated by a particular PDF eigenvector. This will therefore introduce a common set of factorization scale dependent logarithms in to the corresponding predictions, or equivalently the scale evolution of these observables will be same up to running coupling effects. It therefore seems natural in such a case to vary the factorization scale about some fixed central value for each data set in a correlated way across these observables, either in the fit or prediction stage (but not both). However, the choice of central/best fit is not obvious, e.g. $\mu = Q^2$ may be more appropriate for DIS data and $\mu = M/2$ may be more appropriate for Drell Yan production. Further to this, while one might argue that if one varies the scale for one type of structure function from a central value of $\mu = Q$ to $\mu = 2Q$, then one does it for all structure functions, it does not seem so clear that e.g. for some jet data related observable one should simultaneously vary the scale from a central value of e.g. $\mu = p_T$ to $\mu = 2p_T$.² Such scale allocations are therefore not in any clear sense correlated, and certainly the degree of variation in the cross section when applying the rule of thumb variation of the scale will depend on the central scale, the precise ‘preferred’ value of which is not necessarily clear.

² Indeed, for some quantities the choice between $\mu = p_T$, $\mu = p_{T,\max}$ or $\mu = \hat{H}_T$ is also open in principle, see e.g. [34] for a recent discussion for the case of inclusive jets. For dijets one has additionally has the choice of a p_T -based scale or the invariant mass m_{jj} .

We do not advocate strictly fitting the best scale for each type of process, but advise that some note, based on experience, should probably be taken to use central scales that provide good fits for a given process (or at the very least, to avoid those known not to be optimal).

As described above, even for processes that depend dominantly on the same eigenvector, the correlation of the scale variation between these between processes is not entirely trivial. However, if two quantities (either fit and predicted, or both fit) are dominated by alternative PDF eigenvectors and therefore completely independent, then clearly the variation of scales, in either fit or prediction, is not correlated. In this case imposing correlation between scale variations in these processes can result in artificial correlations between the predicted processes (depending on how they depend on the initial PDFs), or in the context of a fit, the PDFs themselves. In general, most predictions will depend on combinations of fit PDFs where full correlation is, to a lesser or greater degree, an overly restrictive assumption.

4 Renormalization scale variation

In this section we will discuss how the issues presented above for factorization scale variation apply to the separate question of renormalization scale variation. We define a physical quantity beginning at $\mathcal{O}(\alpha_S)$:

$$A(Q^2) = \tilde{\alpha}_S(\mu_f^2)A_1 + \tilde{\alpha}_S^2(\mu_f^2) \left(A_2 - \ln \beta^{(0)} A_1 \left(\frac{Q^2}{\mu_f^2} \right) \right), \tag{36}$$

where μ_i is the renormalization scale. We first consider this to be the quantity we measure in order to determine the value of the coupling. With this, we wish to predict the second quantity

$$B(Q^2) = \tilde{\alpha}_S(\mu_f^2)B_1 + \tilde{\alpha}_S^2(\mu_f^2) \left(B_2 - \beta^{(0)} B_1 \ln \left(\frac{Q^2}{\mu_f^2} \right) \right). \tag{37}$$

As with the PDFs and structure functions one can invert (36) to obtain

$$\begin{aligned} \tilde{\alpha}_S(\mu^2) &= \frac{A(\mu^2/a_i)}{A_1} \left(1 - \frac{\alpha_S(\mu^2)}{A_1} \left(A_2 + \beta^{(0)} A_1 \ln a_i \right) \right), \\ &= \frac{A(\mu^2/a_i)}{A_1} \left(1 - \frac{A(\mu^2/a_i)}{A_1^2} \left(A_2 + \beta^{(0)} A_1 \ln a_i \right) \right), \end{aligned} \tag{38}$$

where as usual $a_i = \mu_i^2/Q^2$ and we drop terms of $\mathcal{O}(\alpha_S^2)$; note that while the physical quantities begin at $\mathcal{O}(\alpha_S)$, the relation between them is one order lower, and is accurate to

$O(\alpha_S)$. For $B(Q^2)$, taking $a_f = \mu_f^2/Q^2$ and substituting in our expression for $A(Q^2)$, we get

$$B(Q^2) = \frac{A(a_{fi} Q^2)}{A_1} \left(B_1 + \frac{A(a_{fi} Q^2)}{A_1} \times \left[B_2 - \frac{B_1 A_2}{A_1} + \beta^{(0)} B_1 \ln a_{fi} \right] \right), \tag{39}$$

Hence, we can indeed derive a final result which is expressed only in terms of a single ratio $a_{fi} = a_f/a_i$, as in the structure function case.

In the above example we have used our initial physical quantity in order to determine the coupling constant, not the PDFs. This is really the most natural thing to do when thinking about renormalization scale variation, since the scale is that used for the definition of the coupling, while the factorization scale is that used for the definition of the PDFs. However, in a PDF fit one instead uses the physical quantity $A(Q^2)$ to determine the PDFs. To see how this changes the result, we now assume our toy observables depend on a single PDF q (the non-singlet quark, say, although this is not essential). Implicitly we work in Mellin space to avoid complications with convolutions, but as before this does not change the basic argument. We consider the case of a fixed factorization scale $\mu_F^2 = Q^2$, while setting $\mu_i^2 = a_i Q^2$ for the renormalization scale. We have

$$A(Q^2) = \tilde{\alpha}_S(\mu_i^2) A_1 \left[1 + \frac{\tilde{\alpha}_S(\mu_i^2)}{A_1} \times \left(A_2 + \beta^{(0)} A_1 \ln a_i \right) \right] q(Q^2). \tag{40}$$

In terms of this the PDF can be written as

$$q(Q^2) = \frac{A(Q^2)}{A_1 \tilde{\alpha}_S(\mu_i^2)} \left[1 - \frac{\tilde{\alpha}_S(\mu_i^2)}{A_1} \left(A_2 + \beta^{(0)} A_1 \ln a_i \right) \right], \tag{41}$$

Inserting this into the expression for $B(Q^2)$, we obtain

$$B(Q^2) = \tilde{\alpha}_S(\mu_f^2) B_1 \left[1 + \frac{\tilde{\alpha}_S(\mu_f^2)}{B_1} \left(B_2 + \beta^{(0)} B_1 \ln a_f \right) \right] q(Q^2) \\ = \frac{B_1 A(Q^2)}{A_1} \frac{\tilde{\alpha}_S(\mu_f^2)}{\tilde{\alpha}_S(\mu_i^2)} \left[1 + \beta^{(0)} \left(\tilde{\alpha}_S(\mu_f^2) \ln a_f - \tilde{\alpha}_S(\mu_i^2) \ln a_i \right) + \frac{B_2}{B_1} \tilde{\alpha}_S(\mu_f^2) - \frac{A_2}{A_1} \tilde{\alpha}_S(\mu_i^2) \right]. \tag{42}$$

We see that now the expression is certainly not just a function of the ratio of scales $\mu_f^2/\mu_i^2 = a_{fi}$. Let us examine the explicit consequence of this. For example, in our earlier case of factorisation scale variation, the choice of $\mu_f^2 = \mu_i^2$ i.e. $a_{fi} = 1$ resulted in no change in the expression for the prediction compared to choosing both $a_i = a_f = 1$. In fact

for the first term in (42) this equivalent result appears, and similarly for the sum of the second and third terms. However, for the fourth and fifth terms, i.e. those dependent on the coefficients of the scale-independent parts of the NLO expressions for $A(Q^2)$ and $B(Q^2)$, this is not the case. In this limit each becomes proportional to $\tilde{\alpha}_S(\mu_f^2)$, but depends on the absolute value of the scale. This term depends on the difference in the relative size of these NLO corrections (compared to the LO contributions to each quantity), so the violation of the dependence on ratios is violated by the scale independent NLO corrections. If we consider other types of scale variation, e.g. multiplying μ_f by 2 but dividing μ_i by 2 we see that even though the effect in the combination of the first, second and third terms is close to effect of either multiplying μ_f by 4 or dividing μ_i by 4, it is not identical, and the discrepancy is larger in the fourth and fifth terms.

The fact that the expression of the predicted physical quantity in terms of the measured physical quantity does not break down into an expression depending on the ratio of the renormalization scales used for each calculation is a consequence of the fact that the renormalization scale is fundamentally associated to the scale of the coupling, but here we do not directly relate the physical quantities to the coupling constant, but to the PDF. It is also the case that different physical quantities depend on the coupling in different ways, i.e. the perturbative order starts at zeroth, first or second order for very standard quantities (and higher order for more exclusive quantities). Here we have given perhaps the simplest example of two quantities which each start at first order. However, the common input in PDF fits of the $F_{2,3}$ structure functions starts at zeroth order, so at lowest order has no renormalization scale dependence in the hard cross section. The renormalization scale dependence of F_2 will therefore be suppressed by a power of α_S relative to the case of e.g. top-pair production in hadron-hadron scattering, which begins at $O(\alpha_S^2)$. In contrast, all cross sections are linear in the PDF of any of the hadrons participating in the scattering.

Finally, we can also consider the case of two related physical observables. This could be for example, jet production at ATLAS and CMS, or more generally W and Z boson production, for which the LO results are of course uncorrelated in normalization, but the effect of higher-order QCD corrections is similar. Considering the latter example, for our toy observables above we would have

$$\frac{A_2}{A_1} \sim \frac{B_2}{B_1} \equiv C_{\text{NLO}}, \tag{43}$$

and so (42) can be written as

$$\frac{B(Q^2)}{A(Q^2)} = \frac{B_1}{A_1} \left[1 + \beta^{(0)} \left(\tilde{\alpha}_S(\mu_f^2) \ln a_f - \tilde{\alpha}_S(\mu_i^2) \ln a_i \right) + C_{\text{NLO}} \left(\tilde{\alpha}_S(\mu_f^2) - \tilde{\alpha}_S(\mu_i^2) \right) \right]. \tag{44}$$

Now to maintain consistency with our requirement that this ratio should be approximately constant under inclusion of higher-order QCD corrections, we can see that we must take $\mu_f = \mu_i$ ($a_f = a_i$), i.e. vary the renormalization scale in the fit and prediction in a correlated way. It is of course a well-known procedure to vary QCD renormalization scales in such a way when predicting this type of ratio (of e.g. the W to Z boson cross sections.³) In [25] this argument is extended to hold between processes at the fit stage. It is perhaps not particularly surprising to find an equivalent requirement between the fit and prediction here, and clearly the inclusion of this in a global fit would be intractable. Nonetheless, we can see that this correlation enters in principle at the same level as that between processes entering the fit, and so the question of whether it is necessary or sensible to include one without the other requires further investigation. Certainly, the relative importance of the correlation between processes in the fit stage and between the fit and prediction will in general depend on the specific data sets being considered.

5 Summary and conclusions

In this paper we have discussed the inclusion of theoretical uncertainties in PDFs due to missing higher-order terms in the pQCD results for the processes entering the fit. Such uncertainties, while routinely included in the predictions, have previously not been explicitly included in the PDF fit itself. We are now firmly in the high precision LHC era, both in terms of the available data for PDF fitting and the standard for phenomenology which applies these PDFs. Therefore, such an approach may be increasingly called into question, and certainly requires careful consideration.

As a first step towards this, we have considered the standard approach to evaluating MHO uncertainties, namely due to QCD factorization and renormalization scale variation around a central value by some set factor. Focussing on the case of the factorization scale, we have in particular shown that if we take this standard criterion seriously and apply it consistently to both a PDF fit and the predicted observables resulting from that fit, then in general there is a strong overlap between the variation in the fit and prediction stage. To demonstrate this, we have considered in Sect. 2 the simplest possible case of a fit to a non-singlet structure function, before generalising to include coupled quark and gluon contributions in Sect. 3. We have shown how the explicit dependence on the PDFs can be removed entirely, and the outcome of the fit recast instead in terms of observable quantities only. We

have then found that written in this way, scale variation in the fit corresponds to precisely a scale variation in the prediction in certain regimes, in particular at low or high enough x . Our results have relied on the basic fact that it is possible, and sometimes preferable, to bypass the intermediate PDFs entirely, instead working purely at the level of physical observables (structure functions and so on) and the relationships between them. This idea of working in such a ‘physical basis’ is in fact quite an old one; here we simply derive the implications of this for scale variation uncertainties in PDF fits.

We have also briefly considered the case of renormalization scale variation, finding that the situation is not as straightforward. This is unsurprising, given the quite different roles that the the QCD coupling and PDFs play in fits. However, one basic implication of working in a physical basis is that the motivation for including correlated renormalization scale variations between related processes in the fit or prediction state, is equally present between processes entering both the fit and prediction. While including such correlations in a realistic global fit would certainly in practice be impossible, clearly this raises questions if for example one wishes to include such correlations at the fit stage.

Now, the true situation in a global fit is certainly significantly more complicated than the examples we have considered explicitly in this paper. Here, we fit a very wide array of structure function and hadron collider data, sensitive to a range of different (and overlapping) x values and scales. Indeed, while in the simple non-singlet case of Sect. 2 we find a complete overlap between the fit and prediction, we have seen that in the somewhat more general (although still simplified) scenario of Sect. 3 the situation is not as straightforward. Nonetheless, as mentioned above in certain (e.g. low and high x) regimes the same conclusion holds, and more generally these considerations serve as clear guidance for the case of a genuine PDF fit. In particular, a naive variation of factorization scales in both the PDF fit and prediction will certainly correspond to a degree of overestimation in the total theoretical error, and should be avoided. On the other hand, the considerations of Sec. 3 also suggest that variation of the overall factorization scale in the prediction alone does not capture the full degree of uncertainty due to MHOs in the problem. This suggests that the correct approach, maintaining generality while avoiding overlap in the regimes where it may occur, would be to consider scale variations only in the fit and not in the prediction.

Further to this, we have also seen that if one considers factorization scales between all quantities to be fully correlated in the fit, then the factorization scale variation can equivalently be performed entirely in the calculation of the predictions, with the first assumption implicitly leading to full correlation also being maintained between the factorization scale across different predictions. As discussed earlier,

³ See for example [35] for the example of vector boson plus jets. In this study an additional, conservative, process dependent uncertainty is also introduced to account for the difference between the K -factors of the different quantities.

this seems to be a overly strong assumption in general, but as the arguments in Section 3 suggest it may, in practice, not be such a bad assumption for certain specific physical quantities. Therefore, for factorisation scale variations, the current approach of only varying the scale in the prediction is certainly an underestimate of the full uncertainty from this source, but probably not as significant an underestimate as might naively be expected. The arguments in this article then suggest it is more reliable to consider factorization scale in the fit alone, but this should in general include a variation which is only correlated for physical processes which depend on the same independent PDF combinations. This type of procedure would clearly need significant compromise in practice, as few physical processes depend on exactly the same PDFs, so many sets of processes will either be weakly correlated, strongly correlated, or somewhere in between. We note that in all cases the choices of central scale is largely a separate issue: this should be taken independent for different quantities, allowing for something close to the best possible fit at a fixed theoretical order, and possibly relieving some tensions between data sets in a fit. Indeed, such an approach also has the likely benefit of reducing the sensitivity of the fit quality, χ^2 , to MHOs, which may confuse the interpretation of PDF fits.

The eventual interpretation of these results has the potential to be a matter of some debate, given the known issues with the ‘rule of thumb’ scale variation approach and availability of alternative, potentially superior, approaches (see for example [36–38]). Nonetheless the initial investigations of the inclusion of theoretical uncertainties in PDF fits currently apply the scale variation paradigm [25, 33, 39], and so this result is certainly directly relevant to these studies. Thus, one is free to apply a potentially more complete and reliable approach than scale variation to evaluating the theoretical uncertainty due to missing higher orders in the PDF fit. Indeed, our result may be taken as further evidence that such an approach is preferable. In such a case, the analysis above will not directly apply, although some element of the basic approach, namely expression of the predicted observables directly in terms of the fit observables, will certainly be relevant. If, on the other hand, one does apply the standard factorization scale variation approach, then clearly considerable care is necessary to maintain consistency with the requirements demonstrated in this paper. Future work will consider the impact of such variations, consistently performed, within the context of the global MMHT fit and its interplay with the tolerance criteria to evaluate the PDF uncertainties.

Acknowledgements LHL thanks the Science and Technology Facilities Council (STFC) for support via Grant award ST/P004547/1. RST thanks the Science and Technology Facilities Council (STFC) for support via Grant award ST/P000274/1. We would like to thank James Stirling for many illuminating discussions on this topic in particular, and many others besides over the years.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors’ comment: There is no data to deposit.]

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. Funded by SCOAP³.

A DGLAP evolution in the diagonal basis

In this appendix, we summarise a few standard formulae for the DGLAP evolution in a diagonal basis. The coupled DGLAP equation for the quark singlet Σ_q and gluon g PDF reads

$$\frac{\partial}{\partial \ln \mu^2} \begin{pmatrix} g \\ \Sigma_q \end{pmatrix} = \tilde{\alpha}_S \begin{pmatrix} \gamma_{gg} & \gamma_{gq} \\ 2n_f \gamma_{qg} & \gamma_{qq} \end{pmatrix} \begin{pmatrix} g \\ \Sigma_q \end{pmatrix}, \tag{45}$$

in Mellin space, where we leave the scale and moment arguments implicit. This can be diagonalized to give

$$\frac{\partial}{\partial \ln \mu^2} \begin{pmatrix} \Sigma_+ \\ \Sigma_- \end{pmatrix} = \tilde{\alpha}_S \begin{pmatrix} \gamma_+ & 0 \\ 0 & \gamma_- \end{pmatrix} \begin{pmatrix} \Sigma_+ \\ \Sigma_- \end{pmatrix}, \tag{46}$$

allowing us to write

$$\Sigma_{\pm}(j, \mu^2) = \Sigma_{\pm}(j, Q^2) \left(\frac{\mu^2}{Q^2} \right)^{\tilde{\alpha}_S \gamma_{\pm}}, \tag{47}$$

where

$$\Sigma_+(j, \mu^2) = \frac{1}{r_- - r_+} \left[r_- g(j, \mu^2) - \Sigma_q(j, \mu^2) \right], \tag{48}$$

$$\Sigma_-(j, \mu^2) = \frac{1}{r_- - r_+} \left[-r_+ g(j, \mu^2) + \Sigma_q(j, \mu^2) \right], \tag{49}$$

with

$$\gamma_{\pm} = \frac{1}{2} \left[\gamma_{gg} + \gamma_{qq} \pm \sqrt{(\gamma_{gg} - \gamma_{qq})^2 + 8n_f \gamma_{qg} \gamma_{gq}} \right],$$

$$r_{\pm} = \frac{2n_f \gamma_{qg}}{\gamma_{\pm} - \gamma_{qq}}. \tag{50}$$

The LO splitting functions are given by

$$\gamma_{qq}(j) = C_F \left[-\frac{1}{2} + \frac{1}{j(j+1)} - 2 \sum_{k=2}^j \frac{1}{k} \right], \tag{51}$$

$$\gamma_{qg}(j) = T_R \left[\frac{(2+j+j^2)}{j(j+1)(j+2)} \right], \tag{52}$$

$$\gamma_{gq}(j) = C_F \left[\frac{(2+j+j^2)}{j(j+1)(j-1)} \right], \tag{53}$$

$$\gamma_{gg}(j) = 2N_C \left[-\frac{1}{12} + \frac{1}{j(j-1)} + \frac{1}{(j+1)(j+2)} - \sum_{k=2}^j \frac{1}{k} \right] - \frac{2}{3} n_f T_R, \tag{54}$$

while the NLO case is given in [40]. In the high x ($j \gg 1$) limit, we then have at LO

$$\gamma_{qq} = -2C_F \ln j, \quad \gamma_{gg} = -2N_C \ln j, \quad \gamma_{qg}, \gamma_{gq} = 0, \tag{55}$$

with the same scaling present at NLO. In other words, (45) is already diagonal and we can simply write

$$\Sigma_+(j, \mu^2) = g(j, \mu^2), \quad \Sigma_-(j, \mu^2) = \Sigma_q(j, \mu^2), \tag{56}$$

up to overall normalisation factors. In the low x ($j \sim 1$) limit, (see [41] for further discussion) we have at LO

$$\gamma_{gq} = \frac{2C_F}{j-1}, \quad \gamma_{gg} = \frac{2N_C}{j-1}, \quad \gamma_{qq} = 0, \quad \gamma_{qg} = \frac{2}{3} T_R \tag{57}$$

so that

$$\gamma_+ = \gamma_{gg} \sim \frac{1}{j-1}, \quad \gamma_- = -2n_f \frac{\gamma_{qg}\gamma_{gq}}{\gamma_{gg}} \sim \text{const.}, \tag{58}$$

and

$$r_+ = \frac{1}{9} n_f (j-1), \quad r_- = -\frac{9}{4}. \tag{59}$$

Thus, if we consider the evolution of the quark singlet and gluon at the low input scale $\mu_0 \sim 1$ GeV, and a larger scale μ , e.g. characteristic of an LHC process, we have

$$g(j, \mu^2) = \Sigma_+(j, \mu_0^2) \left(\frac{\mu^2}{\mu_0^2} \right)^{\tilde{\alpha}_{S\gamma_+}} + \Sigma_-(j, \mu_0^2) \mu^2 \gg \mu_0^2 \rightarrow \Sigma_+(j, \mu_0^2) \left(\frac{\mu^2}{\mu_0^2} \right)^{\tilde{\alpha}_{S\gamma_+}}, \tag{60}$$

and

$$\begin{aligned} \Sigma_q(j, \mu^2) &= -\frac{9}{4} \Sigma_-(j, \mu_0^2) + \frac{1}{9} n_f (j-1) \Sigma_+(j, \mu_0^2) \left(\frac{\mu^2}{\mu_0^2} \right)^{\tilde{\alpha}_{S\gamma_+}}, \\ &= \Sigma_q(j, \mu_0^2) + \frac{1}{9} n_f (j-1) \left(g(j, \mu_0^2) + \frac{4}{9} \Sigma_q(j, \mu_0^2) \right) \\ &\quad \times \left[\left(\frac{\mu^2}{\mu_0^2} \right)^{\tilde{\alpha}_{S\gamma_+}} - 1 \right], \\ &\xrightarrow{\mu^2 \gg \mu_0^2} \frac{1}{9} n_f (j-1) \left(g(j, \mu_0^2) + \frac{4}{9} \Sigma_q(j, \mu_0^2) \right) \left(\frac{\mu^2}{\mu_0^2} \right)^{\tilde{\alpha}_{S\gamma_+}}. \end{aligned} \tag{61}$$

Thus at high scale the positive eigenvector dominates for both the gluon and quark singlet, in the latter case overcoming the normalization due to the power of $j - 1$ compared to the

negative eigenvector input. At NLO, the γ_- eigenvector picks up a $\sim \alpha_S/(j - 1)$ pole, but this is nonetheless suppressed by a power of α_S relative to γ_+ and thus the leading high μ^2 scaling is still driven by the latter eigenvector, with the $O(\alpha_S)$ coefficient of the Σ_+ contribution Σ_q being non-vanishing as $j \rightarrow 1$.

We therefore have that for sufficiently high scales an observable will be given dominantly by the Σ_+ eigenvector, with $g \sim \Sigma_q$ due to the low- x PDF evolution. Alternatively, if the our observable is given by the rate of change with scale Q^2 , e.g. dF_2/dQ^2 , then the input Σ_- component will give no contribution, and the positive eigenvector will be dominant at all scales.

References

1. J. Gao, L. Harland-Lang, J. Rojo, Phys. Rep. **742**, 1 (2018). [arXiv:1709.04922](https://arxiv.org/abs/1709.04922)
2. H1. ZEUS, H. Abramowicz et al., Eur. Phys. J. **C75**, 580 (2015)
3. CDF, F. Abe et al., Phys. Rev. Lett. **77**, 438 (1996) [hep-ex/9601008](https://arxiv.org/abs/hep-ex/9601008)
4. J. Huston et al., Phys. Rev. Lett. **77**, 444 (1996). [hep-ph/9511386](https://arxiv.org/abs/hep-ph/9511386)
5. J. Pumplin et al., Phys. Rev. D **65**, 014013 (2001). [hep-ph/0101032](https://arxiv.org/abs/hep-ph/0101032)
6. J. Pumplin et al., JHEP **0207**, 012 (2002). [hep-ph/0201195](https://arxiv.org/abs/hep-ph/0201195)
7. A.D. Martin, R.G. Roberts, W.J. Stirling, R.S. Thorne, Eur. Phys. J. C **28**, 455 (2003). [hep-ph/0211080](https://arxiv.org/abs/hep-ph/0211080)
8. S. Alekhin, Eur. Phys. J. C **10**, 395 (1999). [hep-ph/9611213](https://arxiv.org/abs/hep-ph/9611213)
9. M. Botje, Eur. Phys. J. C **14**, 285 (2000). [hep-ph/9912439](https://arxiv.org/abs/hep-ph/9912439)
10. V. Barone, C. Pascaud, F. Zomer, Eur. Phys. J. C **12**, 243 (2000). [hep-ph/9907512](https://arxiv.org/abs/hep-ph/9907512)
11. W.T. Giele, S.A. Keller, D.A. Kosower (2001) [hep-ph/0104052](https://arxiv.org/abs/hep-ph/0104052)
12. S. Moch, J.A.M. Vermaseren, A. Vogt, Nucl. Phys. B **688**, 101 (2004). [hep-ph/0403192](https://arxiv.org/abs/hep-ph/0403192)
13. A. Vogt, S. Moch, J.A.M. Vermaseren, Nucl. Phys. B **691**, 129 (2004). [hep-ph/0404111](https://arxiv.org/abs/hep-ph/0404111)
14. A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, Phys. Lett. B **652**, 292 (2007). [arXiv:0706.0459](https://arxiv.org/abs/0706.0459)
15. S. Alekhin, J. Blumlein, S. Klein, S. Moch, Phys. Rev. D **81**, 014032 (2010). [arXiv:0908.2766](https://arxiv.org/abs/0908.2766)
16. NNPDF, R.D. Ball et al., Eur. Phys. J. **C77**, 663 (2017) [arXiv:1706.00428](https://arxiv.org/abs/1706.00428)
17. J. Butterworth et al., J. Phys. **G43**, 023001 (2016). [arXiv:1510.03865](https://arxiv.org/abs/1510.03865)
18. G. Watt, R.S. Thorne, JHEP **08**, 052 (2012). [arXiv:1205.4024](https://arxiv.org/abs/1205.4024)
19. S. Carrazza, S. Forte, Z. Kassabov, J.I. Latorre, J. Rojo, Eur. Phys. J. C **75**, 369 (2015). [arXiv:1505.06736](https://arxiv.org/abs/1505.06736)
20. J.R. Andersen et al. (2014) [arXiv:1405.1067](https://arxiv.org/abs/1405.1067)
21. R.S. Thorne, Phys. Rev. D **86**, 074017 (2012). [arXiv:1201.6180](https://arxiv.org/abs/1201.6180)
22. NNPDF, R.D. Ball et al., Phys. Lett. **B723**, 330 (2013) [arXiv:1303.1189](https://arxiv.org/abs/1303.1189)
23. R.S. Thorne, Eur. Phys. J. C **74**, 2958 (2014). [arXiv:1402.3536](https://arxiv.org/abs/1402.3536)
24. A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, Eur. Phys. J. C **63**, 189 (2009). [arXiv:0901.0002](https://arxiv.org/abs/0901.0002)
25. R.L. Pearson, C. Voisey, Towards parton distribution functions with theoretical uncertainties. **1810**, 01996 (2018)
26. G. Altarelli, R.K. Ellis, G. Martinelli, Nucl. Phys. B **157**, 461 (1979)
27. G. Grunberg, Phys. Rev. D **29**, 2315 (1984)
28. S. Catani, Z. Phys. C **75**, 665 (1997). [hep-ph/9609263](https://arxiv.org/abs/hep-ph/9609263)
29. R.S. Thorne, Nucl. Phys. B **512**, 323 (1998). [hep-ph/9710541](https://arxiv.org/abs/hep-ph/9710541)
30. J. Blumlein, V. Ravindran, W.L. van Neerven, Nucl. Phys. B **586**, 349 (2000). [hep-ph/0004172](https://arxiv.org/abs/hep-ph/0004172)
31. M. Hentschinski, M. Stratmann (2013) [arXiv:1311.2825](https://arxiv.org/abs/1311.2825)

32. J. Davies, A. Vogt, Phys. Lett. B **776**, 189 (2018). [arXiv:1711.05267](#)
33. Stefano Forte. http://nnpdf.mi.infn.it/wp-content/uploads/2018/07/SF_cernth18.pdf
34. J. Currie et al., Submitted to: JHEP (2018) [arXiv:1807.03692](#)
35. J.M. Lindert et al., Eur. Phys. J. C **77**, 829 (2017). [arXiv:1705.04664](#)
36. M. Cacciari, N. Houdeau, JHEP **09**, 039 (2011). [arXiv:1105.5152](#)
37. A. David, G. Passarino, Phys. Lett. B **726**, 266 (2013). [arXiv:1307.1843](#)
38. E. Bagnaschi, M. Cacciari, A. Guffanti, L. Jenniches, JHEP **02**, 133 (2015). [arXiv:1409.5036](#)
39. J. Gao et al., Phys. Rev. D **89**, 033009 (2014). [arXiv:1302.6246](#)
40. A. Gonzalez-Arroyo, C. Lopez, Nucl. Phys. B **166**, 429 (1980)
41. J. Bluemlein, V. Andreas, Phys. Rev. D **58**, 014020 (1997). <https://doi.org/10.1103/PhysRevD.58.014020>