Testing momentum dependence of the nonperturbative hadron structure in a global QCD analysis

Aurore Courtoy[1,*] and Pavel Nadolsky[2,†]

[1]*Instituto de Física, Universidad Nacional Autónoma de México Apartado Postal 20-364, 01000 Ciudad de México, Mexico*
[2]*Department of Physics, Southern Methodist University, Dallas, Texas 75275-0181, USA*

We discuss strategies for comparisons of nonperturbative QCD predictions for parton distribution functions (PDFs) with high-energy experiments in the region of large partonic momentum fractions x. Analytic functional forms for PDFs cannot be uniquely determined solely based on discrete experimental measurements because of a mathematical property of mimicry of PDF parametrizations that we prove using a representation based on Bézier curves. Predictions of nonperturbative QCD approaches for the x dependence of PDFs instead should be cast in a form that enables decisive comparisons against experimental measurements. Predictions for effective power laws of $(1 - x)$ dependence of PDFs may play this role. Expectations for PDFs in a proton based on quark counting rules are compared against the effective power laws of $(1 - x)$ dependence satisfied by CT18 next-to-next-to-leading-order parton distributions. We comment on implications for studies of PDFs in a pion, in particular on the comparison of nonperturbative approaches with phenomenological PDFs.

I. INTRODUCTION

Quantum chromodynamics governs interactions of strongly interacting particles and predicts the existence of bound states of hadronic matter. The scale dependence of the QCD coupling constant leads to the existence of two regimes of the strong interaction—the nonperturbative and perturbative—that result in formation of hadronic bound states at low energies and in quasifree interactions of QCD partonic degrees of freedom at high energies. While both regimes require complex theoretical treatments, the perturbative regime of QCD has the advantage of rendering calculable predictions using the expansions in small parameters such as the inverse hard energy scale of the process and small QCD coupling constant. Experiments at the Large Hadron Collider and other facilities test the perturbative phase of QCD to high accuracy.

As for the nonperturbative regime, powerful approaches characterize internal dynamics of hadrons based on the models of the hadronic wave function (or the bound-state amplitude) and the effective Lagrangian approaches

incorporating emergent low-energy symmetries. The pioneering studies of the hadron structure in nonperturbative models (see, e.g., Refs. [1,2]) have paved the way for recent rapid advancements, driven particularly by discretized (lattice) QCD [3,4] as well as by other approaches such as analytical representations (e.g., Ref. [5]) or Schwinger-Dyson formalism (e.g., Ref. [6]). Of particular interest to these studies are parton distribution functions (PDFs)—universal functions quantifying probabilities for finding partons in a fast-moving hadron probed at a factorization scale $\mu \gg 1$ GeV.

The highly challenging computation of a PDF for an arbitrary parton becomes more amenable when the parton carries a significant fraction x of the hadron's momentum, of order 0.1 or more. Recent nonperturbative/lattice computations provide many predictions for PDFs at $x \to 0.1$, as well as of the Mellin moments dominated by large-x PDFs [4]. At $x \gtrsim 0.5$, the flavor dependence of the proton wave function further simplifies, with only the up and down quarks having the appreciable PDFs. Furthermore, under specific conditions outlined in Sec. II, notably requiring that the parton entering the hard scattering carries nearly all of the hadron's momentum (i.e., when $x \to 1$), the proton wave function right before the hard scattering may reduce to a small number of simplest quasifree partonic Fock states, which in turn may allow one to predict the $x \to 1$ asymptotics of the PDFs probed at sufficiently high μ. This physical picture gives rise to the famous quark counting rules [7–9]. The x dependence of the PDFs at

*aurore@fisica.unam.mx
†nadolsky@smu.edu

$x > 0.5$, the topic of interest for this paper, may open avenues for appraising the predictiveness of nonperturbative approaches.

On the phenomenological side, the PDFs are used to predict long-distance contributions to hadronic cross sections, when combined with perturbative parton scattering cross sections. Precise phenomenological parametrizations of PDFs [10–17] for unpolarized protons are determined by performing the global QCD analysis of experimental measurements. A global QCD analysis is a large-scale study involving fits of parametrized PDFs to various experimental datasets in the framework of perturbative QCD (PQCD). Flexible functional forms for these PDFs are fitted to measured cross sections in diverse high-energy processes, such as deeply inelastic scattering and production of vector bosons and jets. We will focus primarily on unpolarized proton PDFs, as they are best constrained experimentally, although baryons are not the simplest particles from the nonperturbative point of view.

Phenomenological PDFs can provide (and have provided) useful guidance for models of nonperturbative dynamics, e.g., by identifying the energy scale where the model is applicable [18,19]. However, comparisons of model and phenomenological PDFs must not be done uncritically. On the one hand, the most common $\overline{MS}$ PDFs enter a factorized *approximation* for the hadronic scattering cross section that is valid up to process-dependent power-suppressed terms. While the nonperturbative computations predict the PDFs in a free hadron, the initial-state hadrons in high-energy scattering processes are not truly free. They interact with other participating particles through soft QCD interactions. Scale-dependent PDFs provide a logarithmic approximation to the process of collinear parton showering in the initial state. The full radiation pattern also depends on particle masses and kinematic constraints. The QCD factorization formulas approximate the hadronic cross sections in simple inclusive processes in a way that accounts for soft and collinear contributions, and neglects numerically small mass terms. The relation of these formulas to the nonperturbative PDFs includes power-suppressed terms that are not controlled to the necessary extent.

On the other hand, the PDFs enter the fitted cross sections through elaborate, flavor-dependent convolution integrals and are determined from complex experimental measurements. The global analysis of PDFs relies on the critical assumption of universality, that the PDFs do not depend on the hard-scattering process. Multiple factors contribute to the final PDF uncertainty, as detailed, for instance, in the recent reviews [20,21]. The question then arises to which extent the phenomenological PDF analyses can genuinely reproduce the features reflecting nonperturbative dynamics.

To illustrate these issues, we will revisit a classical problem in the PDF analysis, determination of the power laws that govern the falloff of PDFs as x approaches 1. Quark counting rules (QCRs), reviewed in Sec. II, are one of the earliest predictions that the valence PDFs in the proton fall off roughly like $(1 - x)^3$ for both the up and down quarks. Remarkably, the power law predicted by the QCRs as well as the asymptotic behavior of the $d(x)/u(x)$ at $x \to 1$—a consequence of the spin-flavor extension of the original QCRs [22]—are consistent with the behavior of the actual phenomenological PDFs, although alternative behaviors are not ruled out.

We will examine this $(1 - x)$ falloff in the recent CT18 next-to-next-to-leading-order global analysis [23]. (For an informative study of the empirical small-x and large-x power laws in the neural network PDF (NNPDF) fit, which follows a different methodology, see Ref. [24].) We start in Sec. II A by briefly reviewing the rationale for the QCRs and by stating the conditions under which the QCRs are expected to hold. Next, we revisit the connection between the PDFs and factorized formulas for fitted hadronic cross sections in Sec. II B.

Can the power laws predicted by the QCRs be tested by experimental hadron measurements? We address this question by presenting a mathematical argument in Sec. III A to show that the polynomial functional form of the structure functions or PDFs cannot be uniquely determined from experimental observations. This conclusion follows from basic properties of polynomial interpolation. As an alternative to the reconstruction of the analytic form, we introduce an empirical power-law exponent for the proton PDFs, as defined in Eq. (27). The empirical exponent allows a phenomenologist to reliably confront theoretical models with the observed behavior of hadronic cross sections. We address the dependence of the empirical power-law exponent on the functional form of the PDFs, factorization scale, and the type of the scattering process in the remainder of Sec. III.

Functional forms of phenomenological PDFs incorporate various assumptions about the asymptotic behaviors of the PDFs at $x \to 0$ and 1, including flavor dependence. It is important to understand compatibility of these assumptions with the experimental data. This study also serves as a sandbox problem illustrating broad aspects of comparisons of phenomenological PDFs with the large-x predictions. In particular, investigations of the x dependence of the PDFs in pions and kaons have been proposed as a powerful test to understand mechanisms for the emergence of the hadronic mass [25]. Yet modern pion PDF analyses [26,27] arrive at varied conclusions about the validity of QCRs for the pion, or they may even appear to be at odds with expectations from nonperturbative approaches [28]. In Sec. IV, we comment on the reconciliation of the physical pictures arising from the perturbative and nonperturbative descriptions of QCD and on implications for pion PDF studies.

II. QUARK COUNTING RULES AND QCD FACTORIZATION

A. Weakly coupled gauge theory

1. QCRs for structure functions

The QCRs for a structure function $F(x_B, Q^2)$ in lepton-hadron deeply inelastic scattering arise from the parton model in gauge theories with small quark-boson coupling constants, such as QED or asymptotically free QCD. Consider a Feynman diagram in Fig. 1(a) in such a weakly coupled theory with massless quarks. The diagram corresponds to scattering of a virtual photon $\gamma^*(q)$ on a highly boosted "proton" $p(P)$ whose lowest Fock state entering the hard scattering (at momentum resolution scales somewhat below $Q^2 \equiv -q^2$) consists of three weakly bound quarks. (Alternatively, we could consider scattering on a "meson" consisting of a quark and an antiquark.) ϕ is the low-energy (long-distance) part of the hadronic wave function, describing the binding of quarks into the hadron at virtualities much less than Q^2. H, the hard-scattering subgraph of the diagram, can be approximated by the quark-photon bag diagram (the squared tree-level amplitude of the $\gamma^* q$ scattering) if all couplings are small. The diagram in Fig. 1(a) dominates the cross section when the $\gamma^* p$ scattering energy $W^2 = Q^2(1/x_B - 1) + m_p^2$ barely exceeds the mass m_p^2 of the initial proton. This regime corresponds to the maximal Bjorken variable, $x_B \equiv Q^2/(2P \cdot q) \to 1$. The contribution of this diagram to the structure function at $x_B \to 1$ behaves as

$$F_2(x_B, Q^2) \xrightarrow[x_B \to 1]{} (1 - x_B)^{2n_s - 1 + 2|\lambda_q - \lambda_A|} \cdot \{\text{const} + \mathcal{O}(1 - x_B)\},$$

$$(1)$$

with n_s being the number of spectator partons (two for a baryon and one for a meson) and λ_A and λ_q denoting helicities of the parent hadron and active (struck) quark. For spin-averaged proton and pion structure functions, we obtain the limits

$$\lim_{x_B \to 1} F_2^p(x_B, Q^2) \propto (1 - x_B)^3, \qquad \lim_{x_B \to 1} F_2^\pi(x_B, Q^2) \propto (1 - x_B)^2.$$

$$(2)$$

The $(1 - x_B)$ power law for $F_2(x_B, Q^2)$ thus arises when the $(n_s + 1)$-quark Fock state dominates in the $x_B \to 1$ limit. In this picture, the $(1 - x)$ falloff is driven primarily by semihard gluon propagators binding the $(n_s + 1)$ quarks before the hard scattering, on the top of long-distance binding effects included in the nonperturbative wave function ϕ. The QCRs were initially demonstrated based on the examination of leading perturbative diagrams [22,29,30] as well as analyticity of partial-wave amplitudes [31] and including helicity dependence as in Eq. (1) [22]. They are also expected to apply in various nonperturbative approaches; see examples in Sec. II B. Adding even more gluon propagators to the graph in Fig. 1(a) suppresses the rate both by additional powers of $(1 - x_B)$ and by additional factors of the coupling constant. The term of order
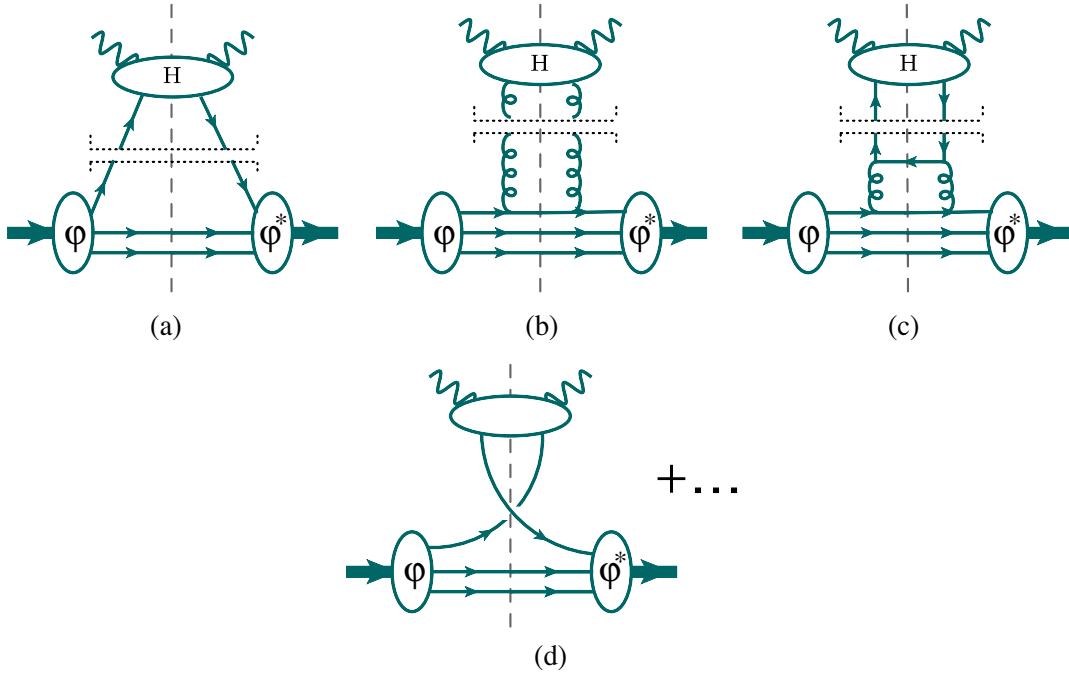

FIG. 1. Leading radiative contributions giving rise to the counting rules for (a) valence quarks, (b) gluons, and (c) sea quarks. We assume $x \to 1$ and a very small coupling constant. (d) A "resolved photon" diagram that is non-negligible for small virtualities of the photon.

$(1 - x_B)$ in curly brackets in Eq. (1) arises from these higher-order radiative contributions. The extra suppression power can be found by counting the added propagators. Such contributions also introduce anomalous dimensions that make the $(1 - x_B)$ exponent dependent on the renormalization scale Q [31,32].

We thus see that the QCRs are mostly directly formulated for the structure functions; QCRs for the PDFs are discussed below. Section III A argues on general grounds that the power law introduced by the QCRs in Eq. (2) cannot be directly tested. Instead, we construct an effective power-law exponent that can be compared against experimental measurements, as discussed in Sec. III B. In the case of proton structure functions, Sec. III D demonstrates that the effective power-law exponent predicted by the QCRs is compatible with the global QCD analysis of hadronic scattering data. See, in particular, the discussion of Fig. 3.

2. QCRs for electromagnetic form factors

The described physics picture also applies to exclusive processes [7–9]. Drell-Yan–West duality [33,34] relates deep inelastic structure functions near the threshold, $x_B \to 1$, to elastic electromagnetic form factors at large momentum transfer, $Q^2 \gg 1$ GeV^2. The scattering contributions that dominate the large-x_B limit of the deep inelastic scattering (DIS) structure function are thus expected to drive the $1/Q^2$ falloff of the elastic form factor [29,31]. For example, in the parton model, the inelastic structure function $\nu W_2(x_B)$ and electromagnetic form factor $F_1(Q^2)$, when both represented in terms of the target wave function, are related through position-dependent parton distributions which have an explicit parton density interpretation [31], implying the interdependent power laws for $\nu W_2(x_B)$ and $F_1(Q^2)$. More generally, the structure functions and form factors are related via impact-parameter generalized parton distributions for the x region for which DGLAP evolution applies (see, e.g., Ref. [35]).

3. QCRs for parton distribution functions: DIS scheme

In real life QCD, computation of high-energy hadronic cross sections involves factorization of long-distance and short-distance QCD radiative contributions. The power-law falloff of factorized structure functions translates into the falloff of PDFs $f_a(x, Q^2)$ at large light-cone momentum fractions x (which differ from the Bjorken variable x_B starting at the next-to-leading order).

The connection is most transparent in *the DIS factorization scheme* [36], where the neutral-current DIS structure function is given by the charge-weighted sum of quark parton distributions to all orders in the QCD coupling constant α_s:

$$F_2(x_B, Q^2)|_{\text{DIS scheme}}$$
$$= \text{const} \cdot \sum_{i=u,d,\dots} e_i^2 (f_i^{\text{DIS}}(x_B, Q^2) + f_{\bar{i}}^{\text{DIS}}(x_B, Q^2))$$
$$+ \mathcal{O}(M/Q). \tag{3}$$

For valence-quark DIS PDFs in the proton, this implies the same falloff power as in Eqs. (1) and (2),

$$\lim_{x\to 1} f_{i/p}^{\text{DIS}}(x, Q^2) = (1-x)^{A_{2i}} \cdot \{\text{const} + \mathcal{O}(1-x)\}, \tag{4}$$

where

$$A_{2i} = 3 \quad \text{for } i = u \quad \text{and} \quad d, \tag{5}$$

if the diagram in Fig. 1(a) dominates. Note that the predicted falloff power is the same for valence up and down quarks.

Starting at the next order in α_s, the cross sections receive significant contributions from QCD radiation. DGLAP evolution equations [37–40] implement a collinear approximation for initial-state QCD radiation, valid when Q^2 is much larger than 1 GeV^2. The gluon and sea (anti)quark PDFs are generated by collinear radiation off the valence-quark lines, with the respective lowest-order diagrams shown in Figs. 1(b) and 1(c). The splitting functions for the $q \to g$ and $g \to q$ splittings in these diagrams are

$$P_{g\leftarrow q}(x) = \frac{\alpha_s}{2\pi} C_F \left(\frac{1 + (1-x)^2}{x} \right) + \cdots,$$
$$P_{q\leftarrow g}(x) = \frac{\alpha_s}{2\pi} \frac{1}{2} (x^2 + (1-x)^2) + \cdots. \tag{6}$$

By computing the convolutions of $P_{g\leftarrow q}(x)$ and $P_{q\leftarrow g}(x)$ with the leading term of the valence PDF, Eq. (4), one determines the lowest-order estimates for the falloff exponents for sea (anti)quark and gluon PDFs in the proton:

$$A_{2i} = 4 \quad \text{for } i = g, \qquad \text{Fig.1(b)}; \tag{7}$$

$$A_{2i} = 5 \quad \text{for } \bar{i} = \bar{u}, \bar{d}, \dots, \qquad \text{Fig.1(c)}. \tag{8}$$

If, instead, we solve the DGLAP differential equations for the scale evolution, we exponentiate the cumulative effect of collinear splittings from all α_s orders. The solution introduces anomalous dimensions for the leading asymptotic powers $A_{2i}(Q^2)$. In QCD, the respective anomalous dimensions are positive [24]; $A_{2i}(Q^2)$ grow with Q^2. For an arbitrary Q^2, the QCRs thus predict

$$A_{2i}(Q^2) \geq 3, 4, \quad \text{and} \quad 5 \tag{9}$$

for the valence, gluon, and sea quark PDFs in the proton, respectively.

The errors in these estimates critically depend on the size of radiative corrections to the lowest Feynman diagrams in Figs. 1(a)–1(c). These come from the higher powers in the coupling constant as well as from the power-suppressed contributions of order M/Q, as indicated in Eq. (3), where M is a nonperturbative scale of order 1 GeV.

At $x \to 1$, the main channel of QCD radiation is due to emission of gluons off valence quarks, as described by the nonsinglet DGLAP equation for the valence quark PDFs. Contributions with radiation off initial-state sea (anti) quarks and gluons are strongly suppressed by smallness of their respective PDFs. The Q^2 dependence of $A_{2i}(Q^2)$ computed based on the DGLAP equations thus reflects the magnitude of higher-order corrections beyond the simplest QCR picture.

Figure 1(d) presents an example of a contribution that is normally not discussed in the derivations of the quark counting rules, but may be an important part of the inclusive $F(x_B, Q^2)$ at relatively low Q^2. In this diagram, the photon splits into a $q\bar{q}$ dipole that interacts with the hadronic state. For highly virtual photons, this contribution can be estimated perturbatively. At Q^2 of order 1 GeV^2 or less, the resolved photon contribution is independent from Fig. 1(a) and is not small. It requires an independent "resolved photon" PDF.

4. QCRs for parton distribution functions: $\overline{MS}$ scheme

Modern phenomenological PDFs are provided predominantly in the $\overline{MS}$ factorization scheme, which offers a number of advantages compared to the DIS factorization scheme. The $\overline{MS}$ PDFs are defined in a process-independent way as summarized in the next subsection. In the $\overline{MS}$ scheme, the leading-power Feynman integrals for DIS inclusive cross sections are given by convolutions of perturbative Wilson coefficients H_a and nonperturbative PDFs $f_{a/A}$ (where $A = p$ or π) according to Eq. (10). In Fig. 1, factorization of Feynman subgraphs is indicated by the horizontal double dotted lines separating the hard and PDF parts. We expect roughly the same falloff powers for the $\overline{MS}$ PDFs as in Eq. (9). Differences between the DIS and $\overline{MS}$ factorization schemes start at the next-to-leading order in α_s.

B. Large-x behavior of QCD processes

We see that the QCRs reflect a simplified picture of hadron scattering, in which cross sections near the elastic limit are dominated by the lowest-order diagrams like those in Fig. 1. Is this simple depiction tethered to realistic measurements?

Our view is that an arbitrary hadron scattering process is likely to include substantial violations of the QCRs. Thus, the QCRs need not be precisely obeyed by all processes included in global PDF fits. However, there may be processes where the kinematics favors the dominance of the lowest diagrams, and the QCRs are more closely followed. Specifically, when the initial hadron is "minimally perturbed" by the hard scattering, the higher-order Fock states may be better suppressed in the elastic limit. Several factors may indicate the minimally perturbed regime, including the smallness of the QCD coupling constant and vanishing Q^2 dependence of the effective power laws preferred by the experimental measurement.

The key assumption of the QCRs, that the $(1 - x)$ dependence is determined mostly by scattering off a few quarks knocked out of the parent hadron, suggests two possible conditions under which the QCRs may hold. First, a weakly bound incoming state may be required, so that only the diagram with the minimal number of semihard propagators gives an appreciable rate in the elastic limit. In that case, each additional perturbative vertex introduces a large suppression factor into the scattering rate.

Second, the hadron-parton vertex described by a bound-state amplitude ϕ reflects the low-energy dynamics; i.e., the long-distance interaction that cannot be approximated by a few (semi)hard gluons. The respective part of the hadronic correlator function can be evaluated consistently in a fully nonperturbative approach to hadron binding, such as MIT bag (e.g., Ref. [1]), Isgur-Karl (e.g., Ref. [41]), or chiral quark soliton models (e.g., Ref. [42]) for the proton and Nambu–Jona-Lasinio or chiral quark models (e.g., Ref. [2]) as well as Schwinger-Dyson equations (e.g., Ref. [43]) for the pion. In such approaches, the proton correlator is computed starting with the lowest-energy bound states consisting of three quark fields. The coupling here is large, but the analyticity of partial-wave amplitudes [31] indicates that the power-law falloff may be realized in a variety of theories that lead to asymptotic freedom at short distances. The dominance of final Fock states with lowest parton multiplicities is essential for realizing the QCRs in both cases. In the latter case, the QCRs may be more evident in a subsample of DIS events in which the final-state hadron multiplicity is low.

Exclusive processes like the deeply virtual Compton scattering with a photon that minimally perturbs the proton, and with the QCD coupling constant and mass terms tuned down, should satisfy the QCRs, as shown in the original derivation of Brodsky and Farrar. In particular, we must assume that the QCD radiation is weak enough so that the excited intermediate Fock states with five or more partons (or equivalently, the correlator contributions with sea partons, or with disconnected topologies) are negligible. The QCRs do not directly hold if there are excited Fock states.

Neither picture—a weak coupling or a minimally perturbed hadron—applies automatically in typical experimental measurements used in the proton PDF fits. Indeed, a typical inclusive hadronic observable used to determine the phenomenological PDFs includes high-multiplicity events.

But we may spot the trace of the QCRs in some kinematic regime, when final-state multiplicities are small, and the impact of other corrections is minimal.

We will further argue in Sec. IV B that the conditions supporting QCRs may be easier to achieve in pion scattering than in nucleon scattering.

We take neutral-current DIS on a proton as an example. In this process, two scales control the QCD radiation, the photon-proton center-of-mass energy W^2 (equal to the invariant mass squared of the hadronic final state) and the photon virtuality Q^2. For any reasonable choice of W^2 and Q^2—with the Bjorken regime limited by $W^2 > m_p^2$ with $W^2 = m_p^2 + (1 - x_{\rm B})/x_{\rm B} Q^2$—the proton bound state is not minimally perturbed. There is no region of W^2 and Q^2 where both the QCD coupling $\alpha_s(Q^2)$ is small and initial-state radiation into final states with more than three partons can be neglected. Now, consider three relevant kinematic regions of DIS:

  (i) In the elastic limit, i.e., when $W^2 \to m_p^2 \sim 1$ GeV^2, the proton mass m_p^2 is not negligible. The relevant three quark degrees of freedom are not massless and free, and some modification of the original Brodsky-Farrar motivation is necessary.

 (ii) When W^2 increases up to about 4 GeV^2, the proton mass terms eventually become negligible, but the behavior of the DIS cross section is initially complex in this region because of the resonant contributions. Globally, one may expect that the picture based on scattering of quasifree partons approximates the DIS cross section on average because of the Bloom-Gilman parton-hadron duality [44]. However, over small intervals of W^2, the cross section can exhibit very complex resonant behavior that does not satisfy the QCRs [45,46].

(iii) At even higher W^2, the power-suppressed terms become small. The *leading-power* contribution dominates a DIS structure function $F(x_{\rm B}, Q^2)$ and can be factorized in terms of the PDFs $f_{a/p}$ and coefficient functions H_a as

$$F(x_{\rm B}, Q^2) = \sum_a \int_{x_{\rm B}}^1 \frac{dx}{x} f_{a/p}(x, \mu^2) H_a\left(\frac{x_{\rm B}}{x}, \frac{\mu^2}{Q^2}\right)$$
$$+ \mathcal{O}(M/Q), \qquad (10)$$

where H_a consists of a delta function for quark a at the zeroth order of α_s and of respective higher-order radiative contributions for $a = q, g$ at higher orders. The quark PDFs $f_{a/p}$, with their dependence on the partonic momentum fraction x and factorization scale μ, are defined in the $\overline{MS}$ scheme as

$$f_{a/p}(x, \mu^2) = \frac{1}{4\pi} \int dy^- e^{-ixP^+ y^-} \langle P|\bar{\psi}_a(0, y^-, \mathbf{0})$$
$$\times \gamma^+ W(y^-, 0)\psi_a(0)|P\rangle, \qquad (11)$$

where

$$W(y^-, 0) = \mathcal{P} \exp\left(-ig \int_0^{y^-} d\bar{y}^- \hat{A}^+(0^+, \bar{y}^-, \vec{0}_T)\right)$$
$$(12)$$

is the Wilson eikonal line and we have used the light-cone coordinates; see, e.g., Ref. [21]. At these W^2 and Q^2, we can finally talk about scattering on *nearly* independent initial-state partons, which nevertheless feel some long-distance interaction with other particles mediated by long-wavelength gluon fields. The factorization formula captures this interaction in two places, through the insertion of the eikonal line $W(y^-, 0)$ in $f_{a/A}(x, \mu^2)$ to approximate the interaction of the initial-state quark field with the soft gluon field $\hat{A}(y)$ connecting to the other particles, and through nonfactorizable terms in the power-suppressed correction $\mathcal{O}(M/Q)$.

The inelastic cross section grows quickly in this region of W^2, indicating that final states with multiple partons are now easily produced. This effect is captured in the leading-power logarithmic approximation by the scale dependence of $f_{a/A}(x, \mu^2)$. These multiparton final states violate the naive prediction of the QCRs. One indication of this violation is significant Q^2 dependence of the effective power law.

1. Threshold resummation

The collinear factorization formula (10) is based on a highly nontrivial proof [47,48] that separates the leading-power convolution integral from power-suppressed terms $\mathcal{O}(M/Q)$ such as target-mass corrections. The collinear formula is perturbatively stable when W^2 is of order Q^2. When $x \to 1$, the inclusive DIS cross section becomes sensitive to soft interactions among various particles that are not necessarily associated with the PDF(s). A different factorization formula, including a soft exponential factor, replaces the collinear factorization (10) in this limit. Soft radiation can be reliably approximated by a resummed all-order series of large logarithms if Q is much larger than 1 GeV. At Q of a few GeV, when the perturbative logarithms are not large, the threshold behavior is most sensitive to the nonperturbative part of the soft factor that should be fitted together with the PDFs. In either case, radiation of *multiple* soft partons modifies the x dependence of the DIS and Drell-Yan (DY) cross sections at $x \to 1$ as compared to the QCR-based estimates.

2. QCD factorization for other processes

To determine *phenomenological* functional forms for $\overline{MS}$ PDFs of various flavors, a global QCD analysis includes a comprehensive combination of experimental

measurements in DIS, production of lepton pairs, jets, $t\bar{t}$ pairs, and other processes. As in the case of DIS, the connection between PDFs and inclusive cross sections relies on factorization theorems, and those are known with less confidence for more complex measurements. The cross sections used in the PDF fits are usually evaluated at a fixed order in α_s and often neglecting power-suppressed terms.

The QCRs demonstrated for inclusive DIS cross sections do not translate automatically to the other processes. For example, while the leading-power collinear factorization for the DY pair production cross section,

$$\sigma = \sum_{a,b} \int dx_a \int dx_b f_{a/A}(x_a, \mu_{\mathrm{F}}^2) f_{b/B}(x_b, \mu_{\mathrm{F}}^2) H_{a,b,x_a,x_b,\mu_{\mathrm{F}}^2}$$
$$+ \mathcal{O}(M/Q), \tag{13}$$

is structurally similar to that in inclusive DIS, as given in Eq. (10), the underlying scattering processes and factorization proofs are drastically different between two processes. In the Drell-Yan process, the underlying hadronic activity from multiperipheral scattering of *two* parent hadron remnants plays a far more prominent role and creates difficulties in proving the factorization [47]. The parent hadrons are more perturbed by soft interactions in hadron-hadron scattering than in DIS. Any factorization formula holds up to power-suppressed terms which are different in the collinear and threshold factorization formalisms and which are different at some level in DIS and DY, or between different hadron and heavy nuclei parent species.[1]

To summarize, in realistic QCD processes that determine the PDFs, the large-x behavior is modified compared to the predictions of the massless parton model. The scope of modifications in the large-x power laws introduced by higher Fock states, mass terms, resonant contributions, and nuclear effects varies by the scattering process. It is reasonable to expect large modifications of parton-model predictions in events with high final-state parton multiplicities. Still, there can be situations that are close to realizing the assumptions that underlie the QCRs, such as when one tests the internal structure of a meson or looks at a subsample of DIS events with low final-state hadronic multiplicities.

III. TESTING LARGE-x PDFS IN EXPERIMENTAL MEASUREMENTS

A. Bézier curves as polynomial interpolations of discrete data

Models of the hadron structure make concrete predictions for the x dependence of the structure functions and

PDFs. One can straightforwardly check the agreement of a given model with an experimental observation within the uncertainties. A stronger assertion, that the experiment *demands* the $1 - x$ dependence of the PDFs to follow a specific power law, is difficult to demonstrate since the functional forms of the PDFs are not known exactly. This is clearly not possible in the presence of local or resonant structures that disagree with the global trend. Even when the PDF functional forms are restricted to be polynomial, the discrete experimental data can be compatible with multiple functional forms.

To illustrate why, consider an idealized example, in which we seek a polynomial function $f^{(n)}(x)$ of degree n to interpolate $k + 1$ data points $\{x_0, p_0\}, \{x_1, p_1\},..., \{x_k, p_k\}$ that have no uncertainty. Our points satisfy $0 \leq x_i \leq 1$. From mathematics, we know that the existence and number of the interpolating solutions depend on the degree n of the polynomial.

If $n = k$, the unisolvence theorem guarantees that there exists a unique interpolating polynomial going through all points: $f^{(n)}(x_i) = p_i$. Two equivalent closed-form solutions for the interpolating polynomial are given by the Lagrange polynomial,

$$\mathcal{L}^{(n)}(x) \equiv \sum_{i=0}^{k} p_i \prod_{\substack{m=1 \\ m \neq i}}^{k} \frac{x - x_m}{x_i - x_m} \quad \text{for } n = k, \tag{14}$$

and by a Bézier curve of degree n,

$$\mathcal{B}^{(n)}(x) = \sum_{l=0}^{n} c_l B_{n,l}(x), \tag{15}$$

constructed from Bernstein basis polynomials

$$B_{n,l}(x) \equiv \binom{l}{n} x^l (1 - x)^{n-l}. \tag{16}$$

Denote the vector $\mathcal{B}^{(n)}(x_i)$ as B. This vector can be written in a matrix form [50,51],

$$B = T \cdot M \cdot C, \tag{17}$$

where $C \equiv \|c_l\|$;

$$M \equiv \|m_{lp}\| \quad \text{with}$$

$$m_{lp} = \begin{cases} (-1)^{p-l} \binom{l}{n} \binom{n-p}{n-l}, & l \leq p \\ 0, & l > p \end{cases}; \tag{18}$$

and $T \equiv \|t_{ip}\|$ with $t_{ip} = x_i^p$. Here, i runs from 0 to k, and l, p run from 0 to n.

Given the matrix $P \equiv \|p_i\|$ of data values, the matrix C for the Bézier curve $\mathcal{B}^{(n)}(x)$ going through all points satisfies [51]

---

[1]A well-known example of loss of universality of factorization is T-odd distributions in transverse momentum distributions factorization, which have opposite signs in DIS and DY process [49].

$$C = M^{-1} \cdot T^{-1} \cdot P \quad \text{for } n = k. \qquad (19)$$

This equation shows that $k + 1$ data points uniquely determine the polynomial of order $n = k$, assuming no experimental errors.

If $n < k$, an interpolating solution that goes through all points may not exist. Rather, there is a Bézier curve that minimizes the total squared distance to p_i,

$$\chi^2(P, B) = \sum_{i=0}^{k} (\mathcal{B}^{(n)}(x_i) - p_i)^2$$
$$= (P - T \cdot M \cdot C)^T \cdot (P - T \cdot M \cdot C). \qquad (20)$$

The matrix of the coefficients of this Bézier curve is

$$C = M^{-1} \cdot (T^T T)^{-1} \cdot T^T \cdot P \quad \text{for } n < k. \qquad (21)$$

The total squared distance from this special curve to p_i is

$$\min \chi^2(P, B) = P^T \cdot K^T \cdot K \cdot P \quad \text{for } n < k, \qquad (22)$$

where

$$K \equiv I_{(k+1) \times (k+1)} - T \cdot (T^T \cdot T)^{-1} \cdot T^T. \qquad (23)$$

If we set $n = k$ in Eq. (21), it reduces to Eq. (19). We also get $K = 0$ and $\min \chi^2(P, B) = 0$.

Finally, if $n > k$, an infinite number of polynomial solutions have $\min \chi^2(P, B) = 0$. They can be constructed by adding $n - k$ arbitrary points to the Lagrange polynomial (14) found for $n = k$.

Equations (19) and (21) for the coefficients of the Bézier curve are readily solvable and can be used to explore strategies for experimental determination of the x dependence of the PDFs. The numerical solution for the interpolating polynomial is generally unstable for large n. The Bézier form is more stable compared to the other forms, sometimes allowing us to get stable interpolation for n as large as 10 or 15. In our equations, numerical instabilities may arise from the inversion of matrix T if T is ill conditioned when n is large (especially if $k > n$), or when some points are spaced too closely in x.

Once found from the data, the Bézier curve (15) can be expanded into the monomial (Taylor) power series of $(1 - x)$,

$$\mathcal{B}^{(n)}(x) = \sum_{l=0}^{n} \bar{c}_l (1 - x)^l, \qquad (24)$$

which in turn can be compared against the predictions of the quark counting rules. The QCRs discussed in Sec. II might predict that the empirically found coefficients $\bar{c}_l$ vanish when $l \leq A_2(Q)$. Alternatively the whole set of c_l or $\bar{c}_l$ can be compared against predictions of a given model.

This comparison is impeded, however, by large cancellations between the terms with highest powers l in the monomial expansion (24) when p_i values are sampled from a realistic PDF shape. The high-l monomial terms tend to have alternating signs when interpolating such p_i. The monomial components with low l, signifying the cutoff by the QCRs and sensitive to the high-l cancellations, vary significantly depending on the range and spacing of x_i.

Figure 2 illustrates this feature by interpolating nine points that are sampled from the functions $f(x)$ specified inside the subfigures. The interpolation is constructed using the Bézier curve $\mathcal{B}^{(n)}(x)$ according to Eqs. (19) and (21). We also show truncated monomial approximations for this curve, $\sum_{l=0}^{N} \bar{c}_l (1 - x)^l$ with $0 \leq N < n$. The input PDFs satisfy $f(x = 1) = 0$. We therefore expect the $N = 0$ truncation to vanish if the Bézier interpolation is constructed properly, meaning that $c_n = \bar{c}_0 = 0$. Indeed, the $N = 0$ truncation is consistent with zero in all subfigures of Fig. 2.

Figure 2(a) shows $\mathcal{B}^{(n)}(x)$ for an input function given by a fourth-degree polynomial, $f(x) = 30x^2(1 - x)^2$. The red points denote $\{x_i, p_i = f(x_i)\}$ given by the exact input function. The interpolating polynomial is readily found as $\mathcal{B}^{(4)}(x)$ using Eq. (21), or equivalently as $\mathcal{B}^{(n)}(x)$ with $c_l = 0$ for $n, l > 4$. The monomial term with a linear $1 - x$ dependence, denoted as $N = 1$, vanishes in this simple case.

For a more elaborate input function assumed in Figs. 2(b), 2(c), and 2(d),

$$f(x) = 200x^2(1 - x)^2(1 - 2.9\sqrt{x(1 - x)} + 2.3x(1 - x)), \qquad (25)$$

we must resort to $\mathcal{B}^{(8)}(x)$ with $n = k = 8$ to obtain interpolation that goes through all points p_i according to Eq. (19). One can see from the figures that linear terms ($N = 1$) are present in the respective monomial expansions, even though the input function does not contain linear terms. The slopes of the $N = 1$ terms are different in the three panels, and the monomial expansion converges too slowly for $x < 0.8$. The coefficients of the linear and higher-l terms depend on the ranges and spacings of x_i, which are varied in Figs. 2(b), 2(c), and 2(d). The ambiguity in the $N = 1$ term is introduced by the correlation of the coefficient $\bar{c}_1$ with the coefficients $\bar{c}_l$ with high l. To pin down the linear term, we must restrict the fit to the highest portion of the x range, such as $x > 0.9$. It appears that finding the low-l monomial terms with good accuracy requires one to fit precise data in the interval that extends closely to the end point $x = 1$, where higher-twist effects are likely important.

With typical input PDF shapes and fewer than about ten input points, we find that interpolation by Bézier curves is numerically stable over the range $x_0 \leq x \leq x_k$: the Bézier
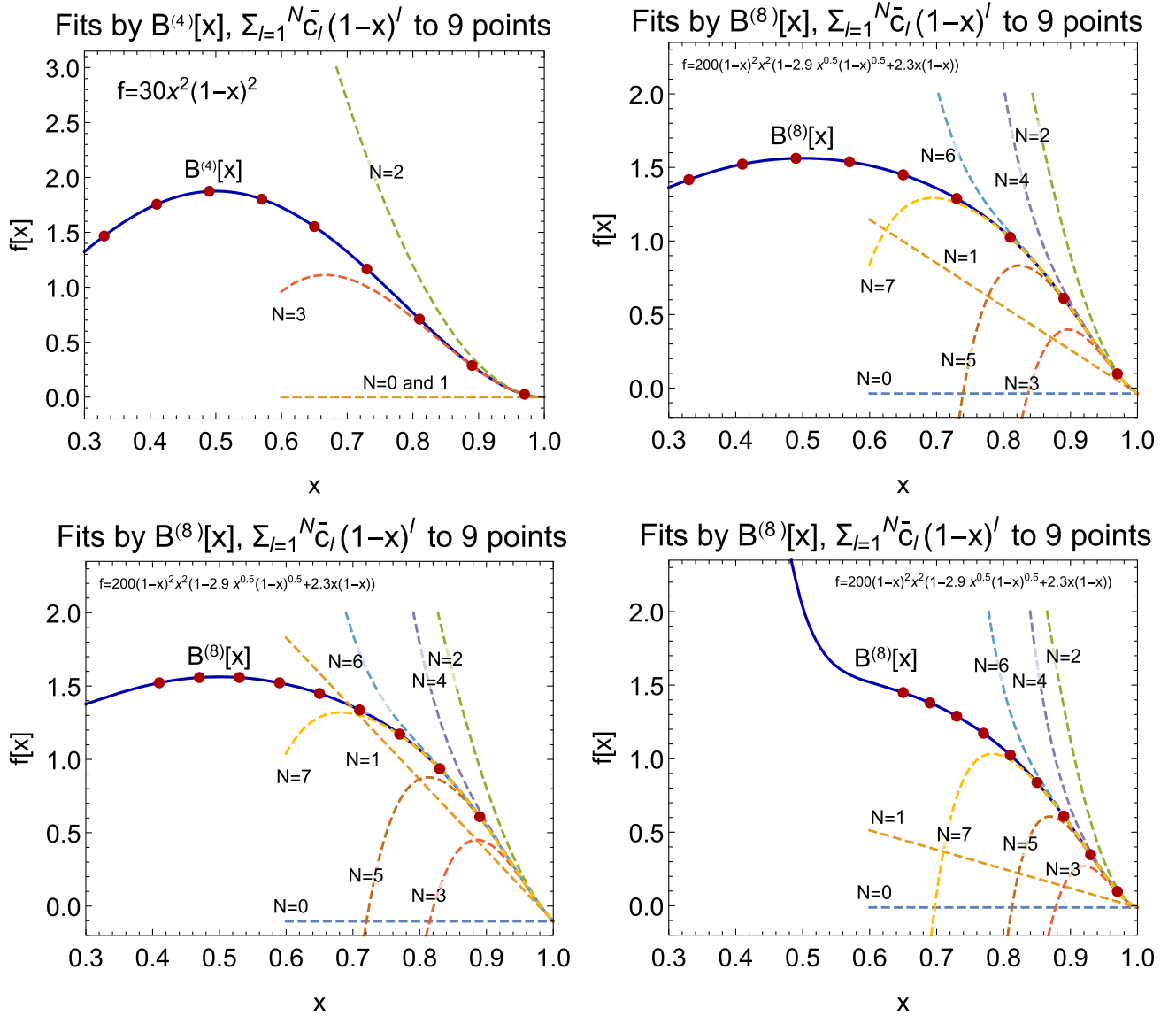
FIG. 2.   (a,b) Bézier and polynomial fits to nine discrete points sampled from the functions $f(x)$ specified in the figures. (c,d) Same as (b), for different ranges and spacings of x covered by the sampled points.

curve goes through all input points and may differ from the input function in high-order terms between the points. Outside of the x range covered by the data points, extrapolation can be unstable.

B. Effective large-x exponent

The example in Fig. 2 demonstrates that *mimicry* of the fitted functional forms impedes determination of the lowest powers in the monomial $(1 - x)$ expansion even in an idealized fit to a few "data" points without uncertainties. An interpolation or fit by a high-degree polynomial may render terms with low powers of $(1 - x)$ that are not present in the fitted function, and which depend on how the data are sampled. In QCD, there is no reason to expect that

coefficients $\bar{c}_l$ for high powers of $(1 - x)$ are suppressed in the PDFs. Statistical and systematic errors in the measurements also get in the way of the determination of the analytic $(1 - x)$ dependence.

The remainder of the article will follow a less pretentious path. Predictions of QCRs and various nonperturbative models suggest that, in the $x_{(B)} \rightarrow 1$ limit, the structure functions or PDFs, denoted collectively as $\mathcal{F}(x_{(B)}, Q^2)$, behave as

$$\mathcal{F}(x_{(B)}, Q^2) = (1 - x_{(B)})^{A_2} \times \Phi(1 - x_{(B)}), \qquad (26)$$

where $\Phi(1 - x_{(B)})$ is a slowly varied function. Based on this observation, it is natural to define

$$A_2^{\mathrm{eff}}[\mathcal{F}(x_{(B)}, Q^2)] \equiv \frac{\partial \ln \left(\mathcal{F}(x_{(B)}, Q^2)\right)}{\partial \ln \left(1 - x_{(B)}\right)}, \qquad (27)$$

with the expectation that $A_2^{\mathrm{eff}} \approx A_2$ when the logarithmic derivative of $\Phi(1 - x_{(B)})$ is small.[2] We will compare theoretical predictions for A_2 presented in Sec. II with the $A_2^{\mathrm{eff}}[\mathcal{F}(x_{(B)}, Q^2)]$ values obtained from a phenomenological PDF ensemble.

C. CT18 PDF ensemble and estimation of PDF uncertainty

In the present analysis, we focus on the results obtained with the CT18 global QCD analysis [23]. The CT18 PDFs are determined by fitting next-to-next-to-leading-order (NNLO) theoretical cross sections to 40 experimental datasets with a total of 3681 data points. The fitted scattering processes—DIS, production of lepton pairs, jets, and $t\bar{t}$ pairs—cover a large kinematic region that extends up to $x = 0.75$ in the typical momentum fraction and down to $Q = 2$ GeV in the factorization scale.

The CT18 functional form is given by

$$f_{a/A}(x, Q_0^2) = x^{A_{1,a}} (1 - x)^{A_{2,a}} \times \Phi_a(x). \qquad (28)$$

Functions $\Phi_a(x)$ are parametrized by Bézier curves $\mathcal{B}^{(n)}(y)$ with $n = 4$–5 and $y \equiv \sqrt{x}$ or a similar scaling function; see Appendix C in Ref. [23]. At the initial scale of the fit, $Q_0 = 1.3$ GeV, the functional forms provide the initial condition for DGLAP equations that predict PDFs at an arbitrary $Q > Q_0$. Each interpolating polynomial $\Phi_a(x)$ reduces to a nonzero constant at $x \to 0$ or 1, so that the exponents $A_{1,a}$ and $A_{2,a}$ control the behavior of the individual PDFs when approaching these limits. While in principle all $A_{2,a}$ parameters can be determined from the data, in practice, not all their combinations result in non-negative PDFs or physically acceptable values of flavor-dependent observables. The CT18 fit imposes a requirement $A_{2,u_v} = A_{2,d_v}$ to guarantee a finite value of $d(x, Q^2)/u(x, Q^2)$ at $x \to 1$ or, equivalently, a nontrivial asymptotic value of $F_2^p(x, Q^2)/F_2^n(x, Q^2)$. We thus expect

$$\lim_{x \to 1} \frac{A_{2,d_v}^{\mathrm{eff}}(x)}{A_{2,u_v}^{\mathrm{eff}}(x)} = \frac{A_{2,d_v}}{A_{2,u_v}} = 1, \qquad (29)$$

where the limit is reached only at very high x values that are outside of the x region covered by the experimental measurements.

On the other hand, at x values below 0.8, where a sufficient amount of experimental data exists, the polynomials $\Phi_a(x)$ with $a = u$ or d are flexible enough to allow a variety of functional behaviors of the u and d PDFs.

---

[2]A similar definition was introduced in Ref. [24].

Therefore, at $x < 0.8$, $A_{2,d_v}^{\mathrm{eff}}(x)$ can be quite different from $A_{2,u_v}^{\mathrm{eff}}(x)$.

The CT18 ensemble consists of the central PDF set and 58 error PDF sets that can be used to estimate the PDF uncertainty in A_2^{eff} according to the master formulas of Refs. [52,53]. A variety of sources contribute to this PDF uncertainty, including experimental, theoretical, parametrization, and methodological uncertainties [21]. Among these, the uncertainty introduced by the choice of the PDF functional forms has been examined in the CT18 study by estimated the spread of the PDFs in 250 candidate fits with alternative functional forms or methodological settings, as explained in Sec. III. C. 3 of Ref. [23]. The tolerance on the nominal Hessian error sets has been selected so as to cover, on average, the spread of the best-fit values in the alternative fits. Thus, the CT18 Hessian uncertainty covers the spread of results with the alternative parametrization choices, with the exception of the extrapolated x regions, $x > 0.7$, where some of the explored best fits fall outside of the nominal CT18 error band. We therefore quote the uncertainty on A_2^{eff} as the envelope constructed from the CT18 Hessian uncertainty at the 68% probability level and the extreme variations of A_2^{eff} obtained with the extended set of 363 alternative functional forms.

D. Effective exponent for a DIS structure function

In deep inelastic scattering, a structure function $F(x_B, Q^2)$ of the proton can be predicted in terms of phenomenological PDFs $f_{a/p}(x, \mu^2)$ as in the QCD factorization formula in Eq. (10). Figure 3 shows the effective $(1 - x_B)$-power $A^{\mathrm{eff}}[F_2(x_B, Q^2)]$ computed according to Eq. (27) with the CT18 NNLO set [23] for three values of Q^2. These are leading-twist predictions for $F_2(x_B, Q^2)$ that reflect a combination of constraints from 40 diverse experiments, and which may not be directly comparable to the actual DIS data because of the limitations discussed in Sec. II B. At $Q = 2$ GeV, the interval $x_B \gtrsim 0.8$ corresponds to DIS in the resonant region, see Sec. II B, where the smooth behavior of $F_2(x_B, Q^2)$ that we predict is modified by complex local features that do not obey the QCRs. We indicate the x_B values lying in the resonant DIS region, approximately corresponding to $W^2 < 2m_p^2$, by using the semitransparent fill for a part of the error band for $Q = 2$ GeV.

Each error band in Fig. 3 consists of an inner (darker) part, indicating the 68% probability level Hessian uncertainty, and the outer (lighter) part, representing the envelope formed by the Hessian uncertainty and the exponents (27) for the alternative functional forms, as explained in Sec. III C. We see that the magnitudes of the Hessian and envelope uncertainties are similar for the region where data are available, i.e., $x < 0.75$. On the other hand, in the region in which the PDFs are unconstrained by the data, the error reflecting the choice of parametrization dominates.
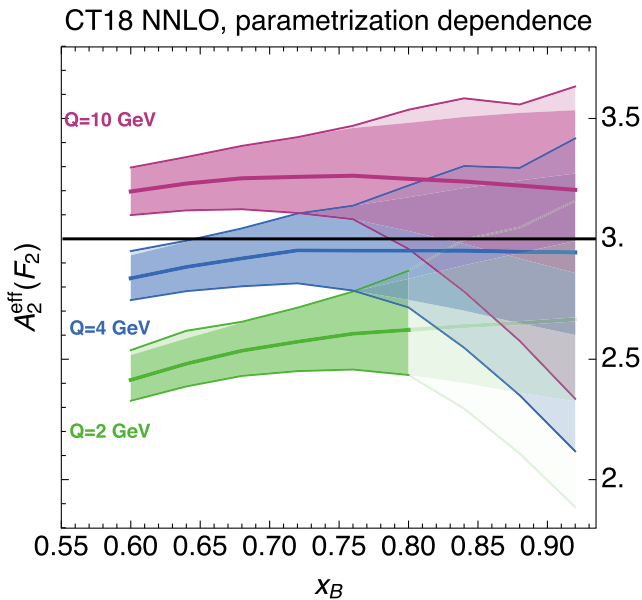
FIG. 3. The effective exponent A_2^{eff} for the structure function $F_2(x_B, Q^2)$ as a function of x_B and for the Q values of 2,4, and 10 GeV represented in green, blue, and magenta, respectively. The central curve of each error band represents the CT18 NNLO central value, the dark shaded band is the asymmetric Hessian error [52,53] at the 68% probability level. The extreme curves correspond to the envelope of the Hessian and parametrization uncertainties estimated as in Sec. III C. The transparent part of the $Q = 2$ GeV band corresponds to the region with $W^2 < 2m_p^2$, approximately corresponding to the resonance region in DIS. The reference prediction from the QCRs is shown by a line at $A_2^{\mathrm{eff}}(F_2) = 3$.

The envelope indicated by the outer error band represents a more conservative estimate of the uncertainty at each Q^2.

We now compare the $A^{\mathrm{eff}}[F_2(x_B, Q^2)]$ from the CT18 fit to the prediction $A_2 = 3$ from the QCRs for $F_2(x_B, Q^2)$ in a proton given by Eq. (2). The A_2^{eff} values for $Q = 2$ GeV in Fig. 3 clearly fall below the QCR prediction. On the other hand, starting from a scale of about 4 GeV, the expected exponent of 3 is attained at the highest x within the error bands.

The A_2^{eff} values in Fig. 3 substantially depend on x_B and Q^2. The true limit of the QCRs would imply weak dependence on either x_B or Q^2, indicating that neither multiparticle final Fock states nor anomalous dimensions are important. In accordance with the discussion in Sec. II, the QCRs would be realized when the Feynman diagrams in Fig. 1 dominate.

As we alluded to in Sec. II B, we do not expect the conditions for the QCRs to be fully met in a global PDF fit. Various simplifications in the fitted processes limit the anticipated accuracy, for example, due to the neglect of process-dependent power-suppressed terms. Nevertheless, Fig. 3 shows that the CT18 values for $A^{\mathrm{eff}}[F_2(x_B, Q^2)]$ are consistent with the QCR prediction of 3 within about one

unit. Keep in mind that the bound-state wave function for the hadronic target predicts parton distributions in a free hadron. On the other hand, the nucleons—and *a fortiori* the pions—probed in the global fits are not truly free at some level; their partons feel the other hadrons present in the scattering event before or after the hard scattering. At the leading power, the effect of the soft QCD background field created by spectator hadrons on the propagation of partons entering the scattering is given by the Wilson line in the $\overline{MS}$ PDF in Eq. (11). Additional radiation of this kind and power-suppressed terms may modify the x dependence compared to the QCR prediction.

E. Effective exponents for PDFs

We will now investigate effective exponents A_2^{eff} for individual PDFs. In the CT18 global fit, *universal* PDFs enter theoretical cross sections for the fitted processes as shown in Eqs. (10) and (13) for DIS and DY. The hadronic cross sections are evaluated up to NNLO in α_s. Equation (9) states the QCR predictions for the A_2 exponents for PDFs of various flavors. These predictions can be compared with A_2^{eff} computed for the respective phenomenological PDFs.

First, we plot, in the left subfigure of Fig. 4, the nominal parameters A_{2,u_v} and A_{2,d_v} in the parametrizations for u and d valence quarks, introduced as in Eq. (28). As summarized in Sec. III C, the CT18 fit assumes these parameters to be the same. The figure shows the red line that corresponds to the best-fit CT18 value for $A_{2,u_v} = A_{2,d_v}$ and its 68% C.L. Hessian uncertainty.

In the same subfigure, the green ellipse shows the 68% C.L. region for the *effective* exponents A_2^{eff} computed using the CT18 Hessian error PDF set at $Q = 1.3$ GeV and $x = 0.875$ according to Eq. (27). Finally, blue scattered points in the left subfigure are for the A_2^{eff} combinations obtained with 363 alternative parametrizations of CT18 PDFs. The lines indicate the QCR prediction of 3 for each exponent.

We see in the left Fig. 4 that, at this high value of x and the initial scale Q_0, the nominal parameters A_{2,u_v} and A_{2,d_v} are consistent with the QCR predictions within the PDF uncertainty. While the nominal A_{2,u_v} and A_{2,d_v} are set to be equal in the initial functional forms, the effective coefficients for u_v and d_v at $x = 0.875$ turn out to be slightly different. The distribution of the scatter points is narrower for the up valence than for the down valence, implying that the PDF for d_v is less constrained by the data at large x.

Figure 5 shows how A_2^{eff} for u_v and d_v obtained with the alternative parametrizations change when Q takes the values of 1.3, 4, and 10 GeV, and x varies between 0.675 and 0.875. The left column, corresponding to $Q = Q_0 = 1.3$ GeV, illustrates the x dependence of A_2^{eff} at the initial scale of DGLAP evolution. The scattered clusters for u_v become narrower toward smaller x values— going down from the top to the bottom row. Going from the
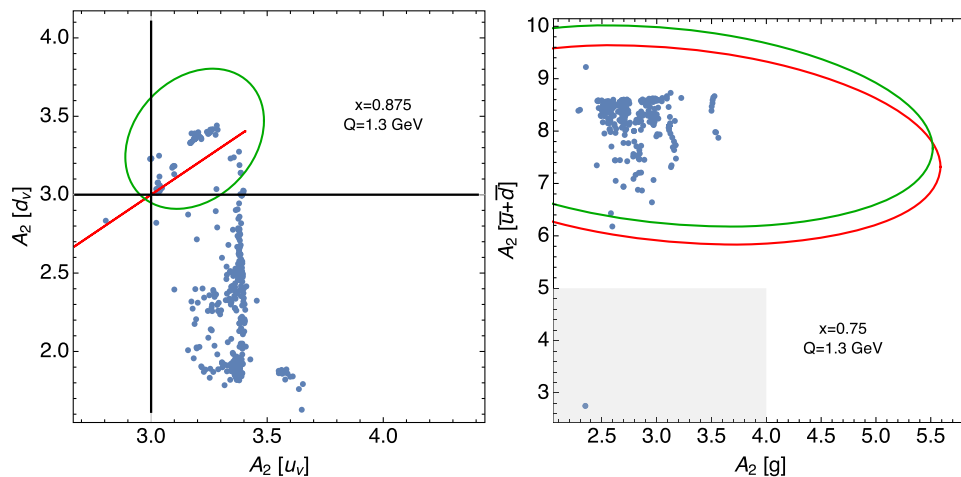
FIG. 4. The effective exponent A_2^{eff} for 363 alternative parametrizations of CT18 NNLO (blue points), compared to the CT18 NNLO Hessian error ellipse at 68% C.L. for $A_2^{\text{eff}}[\text{CT18NNLO}]$ (green ellipse) and the tabulated A_2 parameters of the CT18NNLO error ensemble at 68% C.L. (red line or red ellipse). Both plots are shown at $Q_0 = 1.3$ GeV. The left panel is for the PDF flavors u_v vs d_v at $x = 0.875$. The right panel is for g vs $\bar{u} + \bar{d}$ at $x = 0.75$. The lines in the left subfigure show the expected values. The gray rectangle in the right subfigure shows the region forbidden by the QCRs; see the text.

left to the right in each row, we observe the effect of DGLAP evolution when increasing Q. The shapes of the point distributions are largely preserved when increasing Q, while the distributions shift toward higher A_2^{eff} as a whole. The effective $A_2^{\text{eff}}(u_v(x, Q))$ is larger than the expected value of 3 for all the scales considered. On the other hand, $A_2^{\text{eff}}(d_v(x, Q))$ is as low as 2 for some parametric forms. Overall, the figure demonstrates non-negligible dependence of A_2^{eff} on x (possibly caused by more Fock states contributing to scattering at smaller x) and on Q (reflecting the anomalous dimensions for the A_2 exponents).

Similar plots for g and $\bar{u} + \bar{d}$ PDFs are shown in the right Fig. 4 for the A_2^{eff} values at $Q_0 = 1.3$ GeV and $x = 0.75$ and in Fig. 6 for the x and Q dependence. These PDFs quickly vanish at very large x; thus, we limit the respective comparisons to the region $x \leq 0.75$ to avoid numerical issues.

Since the nominal A_2 parameters are not constrained to be the same for g and $\bar{u} + \bar{d}$, the 68% C.L. Hessian uncertainty region in the right Fig. 4 is given by an ellipse and not by a line. The Hessian uncertainty region (red ellipse) on A_2^{eff} values for these flavors agrees well with the respective nominal A_2 values (green ellipse), as well as with the A_2^{eff} obtained with alternative parametrizations.

According to Fig. 6, the A_2^{eff} values for g and $\bar{u} + \bar{d}$ depend weakly on the x value, taken to be $x = 0.625, 0.7,$ and 0.75. The Q dependence is significant for the gluon A_2^{eff} and much weaker for $\bar{u} + \bar{d}$. At $Q_0 = 1.3$ GeV, neither A_2^{eff} is truly compatible with the respective QCR predictions; the gluon A_2^{eff} of 2–4 is systematically lower than the QCR prediction of 4, while the sea quark A_2^{eff} of 6–9 tends to be higher than 5.

The Q dependence due to the singlet DGLAP evolution is very pronounced for both types of PDFs. The gluon effective exponent grows above 4 starting from about 4 GeV, now in compliance with the QCRs. The A_2^{eff} for the sea combination $\bar{u} + \bar{d}$ has a large uncertainty at Q_0 but quickly correlates with the gluon once the DGLAP evolution is turned on.

We can compare the effective exponents for CT18 NNLO PDFs with those computed for the MMHT14 and NNPDF3.0 sets. We find reasonable agreement between the effective exponents obtained based on these three PDF ensembles. We also broadly agree with the observations presented in Ref. [24] for a low Q scale.[3]

We learn several things from these comparisons. The phenomenological effective exponents and QCR predictions better agree for the valence u and d quarks. The agreement is not so good for the gluon and especially the sea-quark PDFs.

The assumptions justifying the quark counting rules hold when the scale dependence is small. In reality, Figs. 5 and 6 demonstrate pronounced scale dependence, suggesting that higher-order QCD radiation, producing multiparton final states, is not entirely negligible in the fitted processes.

Focusing on DIS for a moment, the numerical effect of the multiparton final states at various Q^2 and W^2 is twofold. When Q^2 is increased at a fixed x value, QCD radiation, evaluated in the logarithmic approximation by DGLAP equations, increases the effective power A_2^{eff} for valence quarks and gluons—see Ref. [24] and references therein. The growth of the corresponding A_2^{eff} in our figures is

---

[3] Let us remark that the NNPDF3.0 results must be averaged, as mentioned in the cited reference.
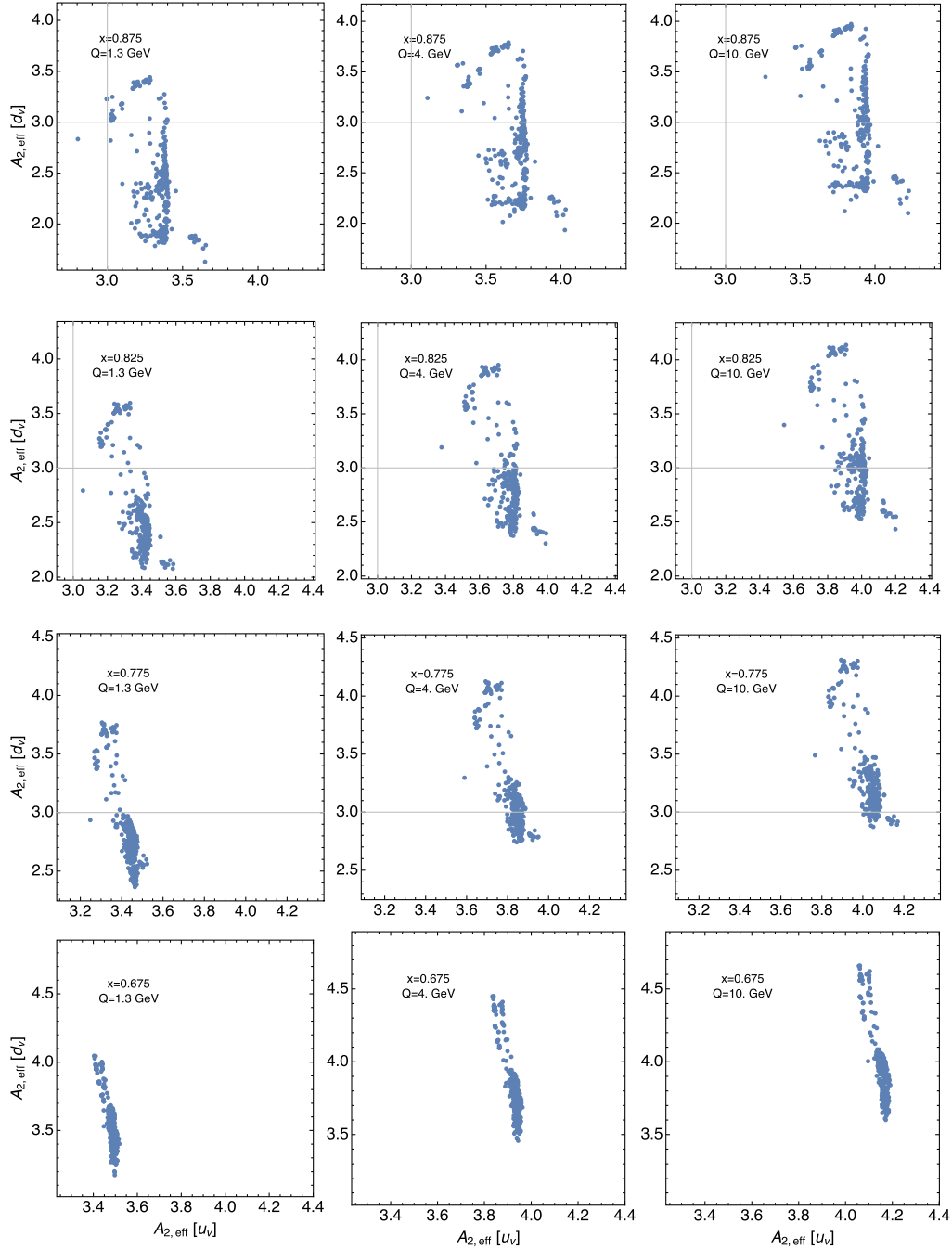
FIG. 5.   The effective exponent A_2^{eff} for the alternative parametrizations for the PDF flavors u_v vs d_v. The x values, from the upper to lower row, are 0.875, 0.825, 0.775, 0.675. The Q values, from left to right, are 1.3, 4, and 10 GeV. The lines $A_2^{\text{eff}}(u_v, d_v) = 3$ are shown for reference.

consistent with this expectation, while the Q^2 dependence for antiquark A_2^{eff} is less clear.

When W^2 is increased (and x_B is decreased) at a fixed Q^2, new final states may be produced and introduce terms with high powers of $1 - x$ in the PDF expressions. As discussed in Sec. III A, such terms can *increase* or *reduce* the effective leading power A_2^{eff}, as compared to the QCR prediction, and introduce dependence of A_2^{eff} on $x_{(B)}$.

F. Process dependence of effective exponents

In Secs. III D and III E, we presented the effective exponents A_2^{eff} and their uncertainties for leading-power DIS structure functions and PDFs determined based on the totality of 40 experiments fitted in the CT18 global analysis. We will now examine the agreement of individual experimental datasets in their preferences for the large-x behavior of PDFs quantified by A_2^{eff}. Toward this goal,
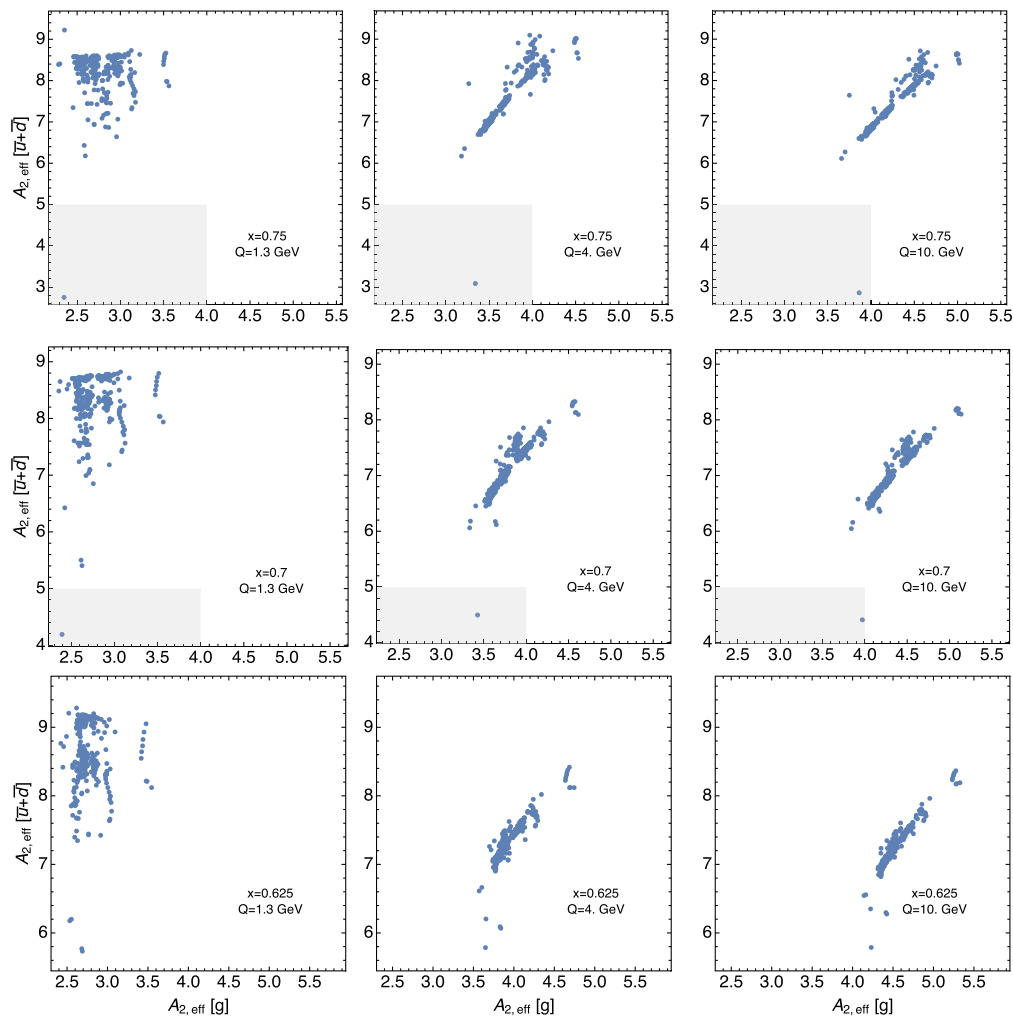
FIG. 6. The effective exponent A_2^{eff} for the PDF flavors g vs $\bar{u} + \bar{d}$. The x values, from the upper to lower row, are 0.75, 0.7, 0.625. The Q values, from left to right, 1.3, 4, and 10 GeV. The rectangle shows the region $A_2^{\text{eff}}(g) < 4$ and $A_2^{\text{eff}}(\bar{u} + \bar{d}) < 5$ that is disfavored by the QCRs.

we will employ a statistical indicator called the L_2 sensitivity [54], following its applications in the CT18 analysis [23] to examine agreement between the experimental datasets. The L_2 sensitivity is constructed from the values A_2^{eff} and goodness-of-fit function χ_E^2 for each fitted experiment E, computed for each Hessian eigenvector set of the CT18 NNLO ensemble. See the relevant equations in Ref. [54].

Figure 7 plots the L_2 sensitivity of several fitted experiments to the effective A_2^{eff} for the proton structure function, $F_2^p(x_B, Q^2)$, evaluated at the x_B values shown on the horizontal axis, for $Q = 1.3$ (top), 4 (lower left), and 10 GeV (lower right subfigure). The L_2 sensitivity is approximately equal to the variation in χ_E^2 for experiment E when $A_2^{\text{eff}}[F_2^p(x_B, Q^2)]$ is increased by the 68% C.L. Hessian uncertainty above its best-fit value at the specified x_B. In other words, we increase $A_2^{\text{eff}}[F_2^p(x_B, Q^2)]$ to the upper boundary of the dark error band for the respective Q in Fig. 3 and ask how χ_E^2 changes under this variation.

The curves in Fig. 7 are for several experiments in the CT18 NNLO fit that show the highest sensitivity to $A_2^{\text{eff}}[F_2^p(x_B, Q^2)]$ at $x_B > 0.6$. At $Q = 1.3$ GeV, these are BCDMS ep and ed DIS cross sections [55,56], the NMC ratio of ep and ed DIS cross sections [57], CDHSW F_2 and F_3 measurements for charged-current DIS on a heavy nucleus [58], E866/NuSea pp Drell-Yan cross sections [59], and CMS jet production cross sections at 7 and 8 TeV [60,61]. Although the coverage by these data in the CT18 fit ends roughly at $x \approx 0.75$, they predict A_2^{eff} at larger x values through extrapolation.

While there is a reasonable agreement between the experiments on the value of $A_2^{\text{eff}}[F_2^p(x_B, Q^2)]$, the uncertainty bands on $A_2^{\text{eff}}[F_2^p(x_B, Q^2)]$ in Fig. 3 emerge as a compromise between the opposite pulls of the contributing experiments. At $Q = 1.3$ GeV and $x_B > 0.85$ in the left Fig. 7, we observe that the positive variation of $A_2^{\text{eff}}[F_2^p(x_B, Q^2)]$ leads to a decrease of χ_E^2 for the BCDMS ep DIS and E866/NuSeA
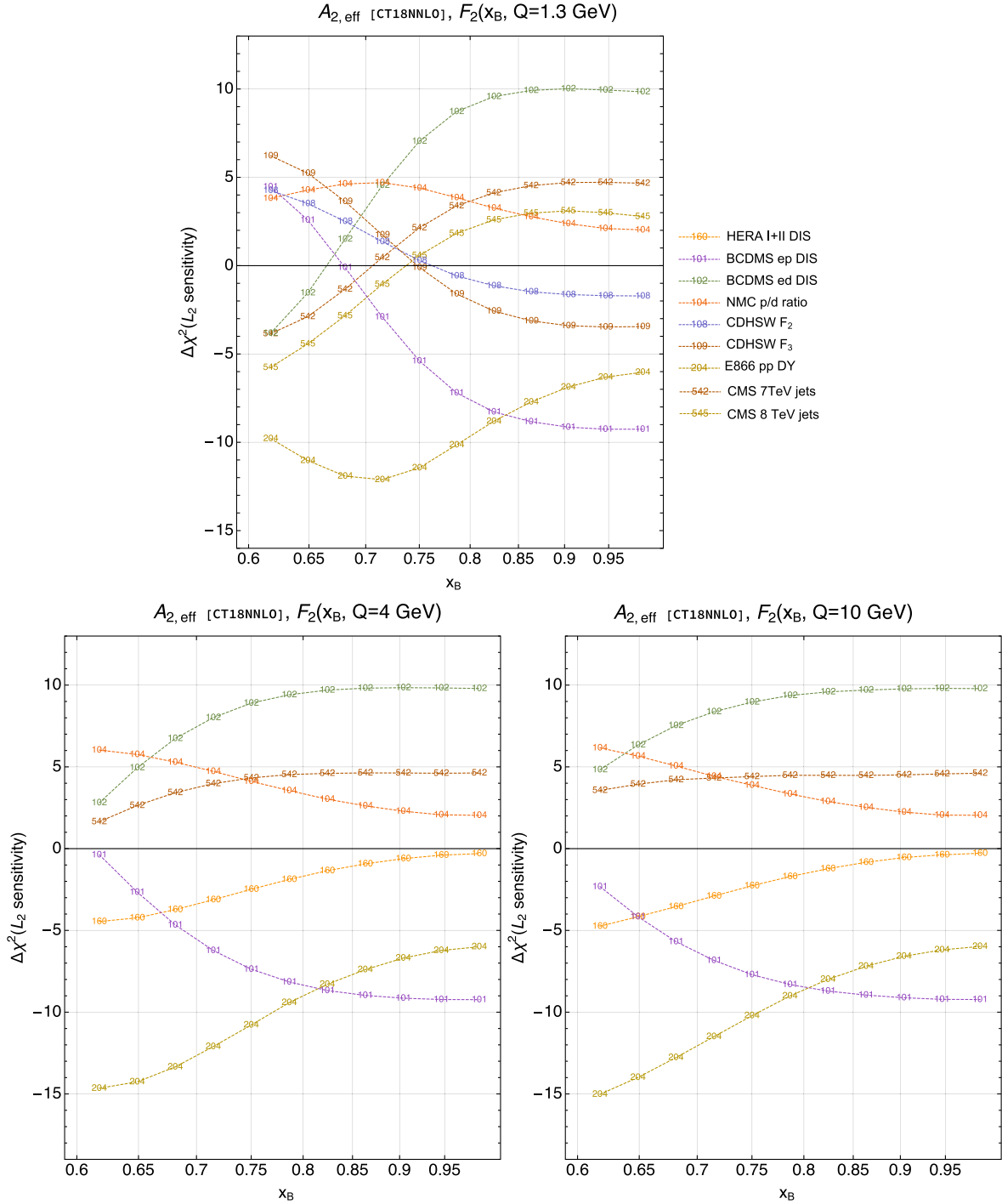
FIG. 7. L_2 sensitivity to the effective exponent A_2^{eff} for the proton structure function $F_2(x_B, Q^2)$ evaluated with the CT18NNLO PDF ensemble, as a function of x_B for the Q values of 1.3, 4, and 10 GeV.

pp Drell-Yan cross sections by up to 10 units, while at the same time, it increases χ_E^2 for the BCDMS ed DIS cross section by a comparable amount. At $x_B = 0.6$–0.7, the E866/NuSeA data, together with BCDMS ed DIS and CMS jet production datasets, prefer a larger-than-nominal $A_2^{\text{eff}}[F_2^p(x_B, Q^2)]$, while they are opposed by downward pulls on $A_2^{\text{eff}}[F_2^p(x_B, Q^2)]$ from BCDMS ep DIS, CDHSW

charged-current DIS, NMC ep/ed ratio, and other measurements.

At $Q = 4$ and 10 GeV, the CDHSW datasets play a less prominent role, while the combined HERA I + II DIS dataset [62] imposes some constraints.

Next, we turn to the sensitivities to A_2^{eff} of u_v and d_v distributions at $Q = 1.3$ GeV in Fig. 8. The pattern of
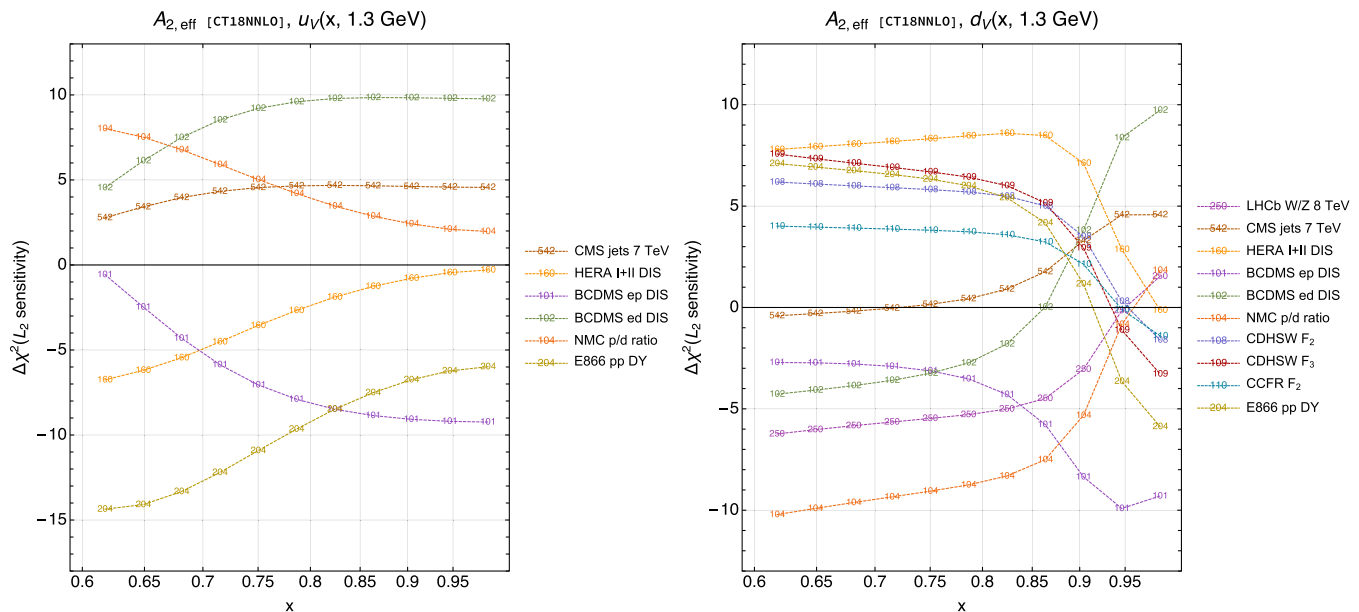
FIG. 8. L_2 sensitivity to the effective exponent A_2^{eff} for $u_v(x, Q^2)$ (left) and $d_v(x, Q^2)$ (right) vs x at $Q = 1.3$ GeV.

sensitivities to A_2^{eff} of u_v in the left panel is visually similar to that for $F_2^p(x_B, Q^2)$ at $Q = 4$ GeV in the second Fig. 7. Once again, competing pulls of the BCDMS ep cross sections and E866/NuSea pp Drell-Yan cross sections against the BCDMS ed DIS cross sections stand out at $x > 0.8$. At $x = 0.6$, the E866/NuSea data and to some extent the HERA DIS data prefer higher A_2^{eff} than the BCDMS ed and NMC p/d DIS measurements.

The pattern on the pulls on A_2^{eff} is more elaborate for d_v in the right panel of Fig. 8. The d_v is less constrained at large x than u_v because the relevant data in the global fit are dominated by neutral-current DIS measurements that are four times more sensitive to up-type quark PDFs than to down-type ones. In Figs. 4 and 5, we observed a moderate PDF uncertainty on A_2^{eff} for u_v and a much larger uncertainty on d_v, especially at $x > 0.75$ where the parametrization uncertainty dominates.

Since the published CT18 parametrization sets $A_{2,d_v} = A_{2,u_v}$, in the region $x > 0.9$ with no data, A_2^{eff} for d_v essentially follows that for u_v. Namely, its value reflects a tradeoff between the opposing pulls of the BCDMS ep and ed DIS datasets. At $x < 0.9$, we see a different pattern, whereby a fairly strong upward pull on A_2^{eff} by the NMC p/d ratio, complemented by LHCb W/Z production at 8 TeV [63] and both BCDMS ep and ed DIS, is opposed by the combined HERA DIS, E866/NuSea pp, as well as by CDHSW and CCFR [64] inclusive charged-current DIS data on heavy nuclei.

In these comparisons, we see some differences between the preferences of scattering experiments on the proton versus deuteron and heavy-nuclei scattering. While these differences do not rise to clear disagreements, they nevertheless suggest importance of the treatment of nuclear effects in future PDF fits.

IV. IMPLICATIONS FOR LOW-ENERGY DYNAMICS

We can now address the question of whether phenomenological PDFs reflect manifestations of low-energy dynamics, based on the discussion of the physical meaning of PDFs in nonperturbative approaches and phenomenological fits in Sec. II B, the mathematical arguments of Sec. III A, and numerical results in Secs. III D–III F. We will focus on two points relevant for further analyses: the relation between the nonperturbative approaches and phenomenological PDFs and also studies of PDFs in the pion.

A. Relation to nonperturbative approaches

It seems appropriate to assume that the quark counting rules are realized when the QCD coupling in semihard scattering is reasonably small, so that the final Fock states are dominated by contributions with a small number of perturbative partons interacting through nearly perturbative interactions. The reality is more complex: radiation in the PQCD regime results in logarithmic evolution of the partonic probability from large x at low Q^2 toward smaller x values at higher Q^2, which in turn introduces Q^2 dependence of the $(1 - x)$ exponents quantified by their anomalous dimensions [24,31,32]. The ideal window in $\{x, Q^2\}$ for QCR studies overlaps with the resonance region in DIS, from which the structure functions can still be extracted, but the global fits of PDFs in this domain must include target-mass corrections—a kinematic effect dependent on a specific scattering process—and other

higher-twist terms [14]. Threshold resummation [65,66] and related nonperturbative effects, such as the modified running of the QCD coupling constant [67], are important at the highest x.

Reconciliation of the nonperturbative and phenomenological definitions of PDFs runs into important differences between the degrees of freedom adopted in various theoretical approaches. In the PQCD collinear factorization framework exemplified by Eq. (10) for deep inelastic scattering, universal PDFs correspond to the long-distance part of the hadronic cross section that is perturbatively expanded as a series of the small QCD coupling and power-suppressed (twist) terms. Here, both expansions are made possible by the presence of a scale $Q > 1$ GeV in the hard cross section. Phenomenological PDFs are defined in an $\overline{MS}$ or another factorization scheme introduced to separate long- and short-distance radiative contributions.

Nonperturbative predictions for the hadron structure do not have an inherent large energy scale that sets the small expansion parameters. They describe the internal structure at a low hadronic scale $\mu_0 < 1$ GeV and must be matched to the factorized PQCD predictions at an intermediate scale $Q_0 > \mu_0$. The bridge between these two scales, namely the one-to-one connection between the low-scale dynamic degrees of freedom to the PQCD quarks and gluons, remains an unsolved problem, with hints available in Dyson-Schwinger approaches and lattice QCD, e.g., in Refs. [6,68–70].

In the absence of such a clear connection at present, one might resort to a model of the spin and flavor dependence of the whole operator matrix element in the $\overline{MS}$ definition (11) of $f_{a/p}(x, Q^2)$. Results obtained with an $SU(6)$-symmetric wave function or an $SU(6)$-broken [71], quark-diquark configuration [72–75], to name a few, reveal useful hints about the x dependence of PDFs at large x. We relied on these considerations when equating A_2, the $(1 - x)$ exponents of the valence up and down quarks, in the PDF parametrizations adopted in the CT18 analysis.

A similar spin-flavor consideration for the pion, discussed in the next subsection, seems less relevant at mild energies due to the pion's pseudo-Nambu–Goldstone origin.

B. Pion case

In Refs. [22,29–31], the counting rules were also formulated for the structure function of the pion, predicting a $(1 - x)^2$ falloff near the threshold. This behavior is often predicted based on the expression of the pion PDF, or distribution amplitude, in terms of the long-distance pion wave function ϕ and semihard scattering contribution dominated by the $q\bar{q}$ state, as reviewed in Sec. II A 2 and Fig. 1.

The pion structure provides a fascinating window on QCD dynamics. Kinematics of the target meson in neutral-current DIS (Sec. II B) takes a new meaning in light of the

pion mass generation, a key emerging feature for pion-related observables. Chiral symmetry and its breaking govern the pion structure at low to mid energies. The nonperturbative quark-quark interaction cannot be replaced by a hard-gluon exchange at energies at which manifestations of chiral symmetry are substantial compared to PQCD interactions. This point is also highlighted for a related case of hard exclusive processes in Ref. [76]. The latter are best understood by comparing a fully nonperturbative approach for predicting the pion electromagnetic form factor to a large-Q^2 description in terms of distribution amplitudes and a hard-scattering part [9,77]. The addition of nonperturbative effects to the hard-gluon exchange in Ref. [76] improves the description of the form factor at low/moderate Q^2 and provides a better transition to the asymptotic "perturbative" behavior [77] associated with the QCRs for Q^2 up to a least 10 GeV^2. The role played by the large-x distribution amplitude (or the structure function) in the behavior of the pion electromagnetic form factor at large Q^2 has been emphasized numerous times; see, e.g., Refs. [78–80]. In other words, for the pion, the concepts of weak coupling and a loosely bound initial state assumed in the QCR picture cannot be approached without also considering the long-distance effects induced by chiral symmetry.

Present and future experiments—at JLab, EIC, or AMBER/COMPASS++—aim to unveil the pion structure in DIS and Drell-Yan pair production. The questions examined throughout this manuscript apply to the fits of pion PDFs. Given the simpler valence structure of the pion and the pion's low mass, we anticipate considerable simplifications with respect to the case of the proton. The main experimental constraints on the pion PDFs at large x for now come from the E615 Drell-Yan pair production in pion-nucleus scattering [81]. In this process, a large momentum fraction x_1 for the pion corresponds to a small x_2 for a nucleus, except in the true threshold limit when $s \approx Q^2$, where no measurements currently exist. In such kinematic regime, when the nuclear beam remnant creates high hadronic multiplicities in the final state, one must carefully revisit the justifications for PQCD factorization for the Drell-Yan process presented at the end of Sec. II B. Nuclear shadowing in the initial state and interactions with the nuclear remnant in the final state may elevate the power-suppressed contributions as compared to the nucleon scattering. A concern about having a genuinely free pion target arises in prompt photon production in pion-proton scattering, as well as in leading neutron electroproduction. On the positive side, the finely binned E615 data points extend to $x_1 = 0.99$, very close to the end point. Due to the smallness of the pion mass, target-mass corrections for pion DIS are almost negligible [78].

Modern global analyses for the pion PDFs [26,27,82] now apply advanced theoretical frameworks, e.g., threshold resummation [65,83]. Depending on the theoretical

framework, the recent analyses find that the pion data are compatible with a nominal $(1-x)$ or $(1-x)^2$ behavior of the PDF parametrization at Q_0. When the large-x resummation is included for the DY data, the authors [65,83] find a fast falloff of the valence pion PDFs in $(1-x)$, consistent with $A_2 = 2$. While threshold resummation must come into play at large x, it is not a sufficient condition for testing that the extracted PDFs fulfill the QCR predictions. None of these analyses addresses the functional mimicry of high-degree polynomial fits discussed in Sec. III A.

As we emphasized in that section, without knowing the exact functional form of the PDFs, one must include the end-point region, ideally $x > 0.9$, to pin down the low powers of $(1-x)$ in the monomial expansion. Physical uncertainties grow in the end-point region. For example, if threshold resummation is necessary, one must account for its uncertainties due to the choice of factorization scales, matching on the fixed-order prediction, and power-suppressed terms.

On the other hand, a fit that is restricted to smaller values of x introduces a spurious correlation between the coefficients with low and high powers of $(1-x)$. This correlation depends on the fitted data sample and strongly modifies the lowest-power monomial terms, as illustrated in Figs. 2(b)–2(d).

An alternative approach in Sec. III B computes the effective exponent A_2^{eff} that can be compared against theoretical predictions without reconstructing the analytic form of the PDFs. The value of A_2^{eff} depends on the range of x. Its *global* trend, reflecting slow variations over x, must be distinguished from the *local* one, existing in a small neighborhood of the examined data bins. For complex PDF parametrizations, like the ones used by NNPDF, the effective exponent may have large local variations, due to mimicry, and require averaging over PDF replicas and/or a range of x in order to determine its $x \to 1$ limit. For smooth PDF parametrizations, like the CT18 or JAM ones, averaging is not necessary; the A_2^{eff} values computed based on such PDFs follow smooth trends and can be determined for comparisons against the QCRs at the momentum fractions of about 0.8, as has been done in Fig. 3.

Analogous considerations apply to PDFs predicted by low-energy models. Ideally, these models should provide uncertainty bands for the cross sections or PDFs that can be verified (or falsified) by the experimental data *over the full x range*. Such uncertainties are difficult to estimate faithfully. In the absence of the uncertainty bands, one must focus on the aspects of the low-energy predictions that are preserved in hard scattering. For example, a low-energy dynamic effect like the broadening of the parton distributions due to the emergence of dynamical mass [2,84–86] may favor a particular $(1-x)$ falloff power, yet consistency of this power with the experimental data in some kinematic region is not sufficient for validating such a prediction as the only viable one. The local trend quantified by the empirical A_2^{eff} values may differ from the global trend in complex nonperturbative models. A comparison against the QCRs requires experimental access to the end point $x = 1$, where additional dynamical effects are most pronounced.

Sections III D and III E demonstrate that A_2^{eff} depends on the factorization scale Q. The pion valence PDFs obey the same nonsinglet evolution equation as the proton ones; thus, A_2^{eff} for the pion PDFs may change by 0.5–1 units within the typical Q range, like in the proton case. The evolution of a PDF that fulfills the QCRs at a low, prefactorization scale may either increase or decrease A_2^{eff} at a higher Q, reflecting the functional mimicry.

V. CONCLUSIONS

Complementary to the constraints based on first principles, the quark counting rules offer predictions for DIS structure functions and PDFs near the elastic threshold. In this picture, when a hadron target is almost unperturbed, and the strong coupling constant is small, the structure functions in a nucleon exhibit a $(1-x)^p$ falloff in the limit $x \to 1$, reflecting exchanges of semihard gluons between the quarks in the incoming bound state, as discussed in Sec. II A.

We examined two possible strategies for testing the quark counting rules with experimental data. From a purely mathematical point of view, the concept of polynomial mimicry, demonstrated in Sec. III A by employing the Bézier curve technique, reveals a limitation in reconstructing the exact functional forms of PDFs from discrete data, whether based on an interpolation or a fit. It is not possible to uniquely determine the powers of a $(1-x)$ monomial expansion except very closely to the end point. Associated uncertainties can be large.

As an alternative, Sec. III B showed that it is possible to define an effective exponent to examine the large-x behavior of any functional form for the PDFs. The leading-power structure function $F_2^p(x, Q^2)$ reconstructed within the CT18 NNLO global analysis [23] agrees, within error bands at moderate scales Q, with the predicted power law. However, a non-negligible shift in the effective $(1-x)$-exponent with increasing Q^2 is observed in Fig. 3, as expected from DGLAP evolution. Moreover, the resonance region in DIS forbids a reliable analysis at large-x values and photon virtuality of a few GeVs.

This has led us to our bottom line: exploration of the relevance of quark counting rules for high-energy processes must address various factors arising from both theory and statistics. We have investigated the pertinent issues in the case of PDFs for the nucleon, for which the situation is best understood. These issues include differences between the quark counting rules for hadronic observables, such as $F_2(x, Q^2)$, as opposed to $\overline{MS}$ PDFs; the universality of the PDFs and the role of power-suppressed and soft corrections; the phenomenological PDF uncertainty reflecting the

scarcity of the data at large x as well as the choice of the PDF parametrization. The latter point has been developed by studying $N = 363$ replicas of the CT18NNLO analysis; see Sec. III C.

Global analyses are based on experimental data from several processes, such as DIS and DY. When considering the effective power laws at the PDF level in Sec. III E, we have found that not all flavors behave on the same footing. The valence up distribution is consistent with the running exponent of 3, the valence down and gluon distributions on average tend to have lower-than-expected exponents, and the sea-quark exponent is too high; see Fig. 4. The $\bar{u} + \bar{d}$ exponent only slightly decreases with Q^2. In the same spirit, the preferred effective exponents depend at some level on the fitted experimental process, as highlighted in Sec. III F.

In conclusion, we emphasize that the quark counting rules emerge in the limit of weak coupling in processes with little underlying hadronic activity. Violations of these conditions in at least some high-energy processes put in question the universality of the rules, which is especially relevant for uses in global analyses. On the other hand, there may be experimental measurements that favor the QCRs, such as a subclass of DIS events with low final-state hadronic multiplicities. The other possibility is offered by pion scattering discussed in Sec. IV B, which is less affected by collateral factors present in the nucleon or nuclear cases.

An experimental observation of the x dependence predicted by a nonperturbative calculation constitutes an insufficient, but nonredundant, condition for validating the calculation. Functional forms of fitted PDFs are unnecessary but sufficient for describing the data. Only by measuring the structure functions/PDFs near the end point $x = 1$, one may reveal evidence of the primordial power law, as we have demonstrated in Sec. III A based on the comparison of the monomial and Bézier expansions. Reconciliation of the predictive phenomenological fits and the interpretative nonperturbative approaches within uncertainties will require more data at large x as well as efforts to address theoretical issues that hinder our understanding of cross sections at the end point $x = 1$.

[1] R. Jaffe, Phys. Rev. D **11**, 1953 (1975).
[2] R. Davidson and E. R. Arriola, Phys. Lett. B **348**, 163 (1995).
[3] H.-W. Lin *et al.*, Prog. Part. Nucl. Phys. **100**, 107 (2018); arXiv:1711.07916.
[4] M. Constantinou *et al.*, arXiv:2006.08636.
[5] K. Kumericki, S. Liuti, and H. Moutarde, Eur. Phys. J. A **52**, 157 (2016).
[6] C. D. Roberts and A. G. Williams, Prog. Part. Nucl. Phys. **33**, 477 (1994).
[7] S. J. Brodsky and G. R. Farrar, Phys. Rev. Lett. **31**, 1153 (1973).
[8] S. J. Brodsky and G. R. Farrar, Phys. Rev. D **11**, 1309 (1975).
[9] G. Lepage and S. J. Brodsky, Phys. Rev. D **22**, 2157 (1980).
[10] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C.-P. Yuan, Phys. Rev. D **93**, 033006 (2016).
[11] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, Eur. Phys. J. C **75**, 204 (2015).
[12] R. D. Ball *et al.* (NNPDF Collaboration), Eur. Phys. J. C **77**, 663 (2017).
[13] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, Phys. Rev. D **96**, 014011 (2017).
[14] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato, Phys. Rev. D **93**, 114017 (2016).
[15] L. A. Harland-Lang, A. D. Martin, R. Nathvani, and R. S. Thorne, Eur. Phys. J. C **79**, 811 (2019).
[16] V. Bertone, S. Carrazza, N. P. Hartland, and J. Rojo (NNPDF Collaboration), SciPost Phys. **5**, 008 (2018).
[17] A. V. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, J. High Energy Phys. 12 (**2017**) 046.
[18] M. Stratmann, Z. Phys. C **60**, 763 (1993).
[19] M. Traini, A. Mair, A. Zambarda, and V. Vento, Nucl. Phys. **A614**, 472 (1997).
[20] J. Gao, J. High Energy Phys. 02 (2018) 026.
[21] K. Kovařík, P. M. Nadolsky, and D. E. Soper, Rev. Mod. Phys. **92**, 045003 (2020).
[22] G. R. Farrar and D. R. Jackson, Phys. Rev. Lett. **35**, 1416 (1975).
[23] T.-J. Hou *et al.*, Phys. Rev. D **103**, 014013 (2021).
[24] R. D. Ball, E. R. Nocera, and J. Rojo, Eur. Phys. J. C **76**, 383 (2016).
[25] A. C. Aguilar *et al.*, Eur. Phys. J. A **55**, 190 (2019).
[26] P. Barry, N. Sato, W. Melnitchouk, and C.-R. Ji, Phys. Rev. Lett. **121**, 152001 (2018).
[27] I. Novikov *et al.*, Phys. Rev. D **102**, 014040 (2020).
[28] R. J. Holt and C. D. Roberts, Rev. Mod. Phys. **82**, 2991 (2010).

[29] Z. Ezawa, Nuovo Cimento A **23**, 271 (1974).
[30] E. L. Berger and S. J. Brodsky, Phys. Rev. Lett. **42**, 940 (1979).
[31] D. E. Soper, Phys. Rev. D **15**, 1141 (1977).
[32] M. L. Goldberger, D. E. Soper, and A. H. Guth, Phys. Rev. D **14**, 1117 (1976).
[33] S. Drell and T.-M. Yan, Phys. Rev. Lett. **24**, 181 (1970).
[34] G. B. West, Phys. Rev. Lett. **24**, 1206 (1970).
[35] M. Diehl, Phys. Rep. **388**, 41 (2003).
[36] M. Diemoz, F. Ferroni, E. Longo, and G. Martinelli, Z. Phys. C **39**, 21 (1988).
[37] V. N. Gribov and L. N. Lipatov, Yad. Fiz. **15**, 781 (1972) [Sov. J. Nucl. Phys. **15**, 438 (1972)].
[38] L. N. Lipatov, Yad. Fiz. **20**, 181 (1974) [Sov. J. Nucl. Phys. **20**, 94 (1975)].
[39] G. Altarelli and G. Parisi, Nucl. Phys. **B126**, 298 (1977).
[40] Y. L. Dokshitzer, Zh. Eksp. Teor. Fiz. **73**, 1216 (1977) [Sov. Phys. JETP **46**, 641 (1977)].
[41] G. Parisi and R. Petronzio, Phys. Lett. B **62**, 331 (1976).
[42] D. Diakonov, V. Petrov, P. Pobylitsa, M. V. Polyakov, and C. Weiss, Nucl. Phys. **B480**, 341 (1996).
[43] M. Hecht, C. D. Roberts, and S. Schmidt, Phys. Rev. C **63**, 025213 (2001).
[44] E. D. Bloom and F. J. Gilman, Phys. Rev. Lett. **25**, 1140 (1970).
[45] C. Armstrong, R. Ent, C. Keppel, S. Liuti, G. Niculescu, and I. Niculescu, Phys. Rev. D **63**, 094008 (2001).
[46] S. Liuti, R. Ent, C. Keppel, and I. Niculescu, Phys. Rev. Lett. **89**, 162001 (2002).
[47] J. Collins, Foundations of Perturbative QCD (Cambridge University Press, Cambridge, England, 2013).
[48] J. C. Collins, Phys. Rev. D **58**, 094002 (1998).
[49] J. C. Collins, Phys. Lett. B **536**, 43 (2002).
[50] G. Farin, Curves and Surfaces for CAGD: A Practical Guide (Morgan Kaufmann, Burlington, MA, 2001).
[51] M. Kamermans (Pomax), A primer on Bézier curves, https://pomax.github.io/bezierinfo/.
[52] P. M. Nadolsky and Z. Sullivan, eConf **C010630**, P510 (2001).
[53] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, Phys. Rev. D **82**, 074024 (2010).
[54] T. Hobbs, B.-T. Wang, P. M. Nadolsky, and F. I. Olness, Phys. Rev. D **100**, 094040 (2019).
[55] A. C. Benvenuti et al. (BCDMS Collaboration), Phys. Lett. B **237**, 592 (1990).
[56] A. C. Benvenuti et al. (BCDMS Collaboration), Phys. Lett. B **223**, 485 (1989).
[57] M. Arneodo et al. (New Muon Collaboration), Nucl. Phys. **B483**, 3 (1997).
[58] J. P. Berge et al., Z. Phys. C **49**, 187 (1991).
[59] J. C. Webb et al. (NuSea Collaboration), arXiv:hep-ex/0302019.
[60] S. Chatrchyan et al. (CMS Collaboration), Phys. Rev. D **90**, 072006 (2014).
[61] V. Khachatryan et al. (CMS Collaboration), J. High Energy Phys. 03 (2017) 156.
[62] H. Abramowicz et al. (H1 and ZEUS Collaborations), Eur. Phys. J. C **75**, 580 (2015).
[63] R. Aaij et al. (LHCb Collaboration), J. High Energy Phys. 01 (2016) 155.
[64] U.-K. Yang et al. (CCFR/NuTeV Collaboration), Phys. Rev. Lett. **86**, 2742 (2001).
[65] M. Aicher, A. Schafer, and W. Vogelsang, Phys. Rev. Lett. **105**, 252003 (2010).
[66] M. Bonvini, S. Marzani, J. Rojo, L. Rottoli, M. Ubiali, R. D. Ball, V. Bertone, S. Carrazza, and N. P. Hartland, J. High Energy Phys. 09 (2015) 191.
[67] A. Courtoy and S. Liuti, Phys. Lett. B **726**, 320 (2013).
[68] A. Aguilar, D. Binosi, and J. Papavassiliou, Phys. Rev. D **78**, 025010 (2008).
[69] D. Binosi, L. Chang, J. Papavassiliou, and C. D. Roberts, Phys. Lett. B **742**, 183 (2015).
[70] I. Bogolubsky, E. Ilgenfritz, M. Muller-Preussker, and A. Sternbeck, Phys. Lett. B **676**, 69 (2009).
[71] F. Close, Phys. Lett. B **43**, 422 (1973).
[72] R. D. Carlitz, Phys. Lett. B **58**, 345 (1975).
[73] R. D. Carlitz and J. Kaur, Phys. Rev. Lett. **38**, 673 (1977); **38**, 1102(E) (1977).
[74] J. Kaur, Nucl. Phys. **B128**, 219 (1977).
[75] R. Jakob, P. Mulders, and J. Rodrigues, Nucl. Phys. **A626**, 937 (1997).
[76] E. Shuryak and I. Zahed, Phys. Rev. D **103**, 054028 (2021)..
[77] G. Lepage and S. J. Brodsky, Phys. Lett. **87B**, 359 (1979).
[78] W. Melnitchouk, Eur. Phys. J. A **17**, 223 (2003).
[79] L. Chang, I. Cloët, C. Roberts, S. Schmidt, and P. Tandy, Phys. Rev. Lett. **111**, 141802 (2013).
[80] M. Chen, M. Ding, L. Chang, and C. D. Roberts, Phys. Rev. D **98**, 091505 (2018).
[81] J. Conway et al., Phys. Rev. D **39**, 92 (1989).
[82] C. Bourrely, F. Buccella, and J.-C. Peng, Phys. Lett. B **813**, 136021 (2021).
[83] P. Barry, N. Sato, W. Melnitchouk, and C.-R. Ji, JAM pion PDF analysis including resummation, Workshop on Perceiving the Emergence of Hadronic Mass through AMBER@CERN, 2020, https://tinyurl.com/BarryAmber2020.
[84] V. Petrov, M. V. Polyakov, R. Ruskov, C. Weiss, and K. Goeke, Phys. Rev. D **59**, 114018 (1999).
[85] K. D. Bednar, I. C. Cloët, and P. C. Tandy, Phys. Rev. Lett. **124**, 042002 (2020).
[86] M. Ding, K. Raya, D. Binosi, L. Chang, C. D. Roberts, and S. M. Schmidt, Phys. Rev. D **101**, 054014 (2020).