Semiparametric Estimation of Beyond Standard Model Physics in Parton Distribution Function Fits

Richard Zhu[1], Jaco ter Hoeve[2], Mikael Kuusela[1], Maria Ubiali[3], Larry Wasserman[1]

[1]Carnegie Mellon University, [2]University of Edinburgh, [3]University of Cambridge

Introduction

- **Parton Distribution Functions** (PDFs) are fundamental objects in high-energy physics that describe the distribution of fraction of momentum carried by quarks and gluons (known as partons) in proton collisions.
- PDFs are crucial for collider predictions and interpretation of LHC data.
- Standard PDF fits assume the **Standard Model (SM)**, potentially biasing inference for **Beyond Standard Model (BSM)** physics.
- We introduce a **semiparametric framework** that treats PDFs as infinite-dimensional nuisance functions.
- Inference focuses on a low-dimensional **BSM parameter**, reducing PDF-induced bias and allowing for uncertainty quantification.

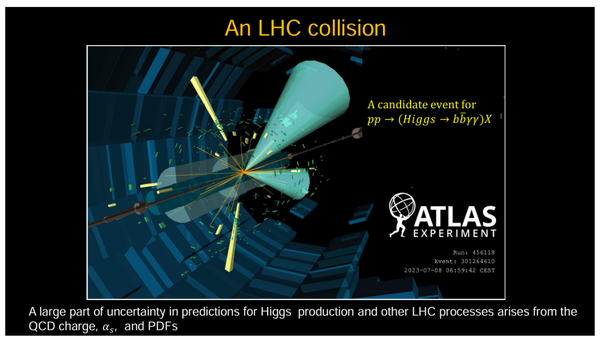


Figure adapted from Nadolsky, *Introduction and CTEQ view*, PHYSTAT Seminar: "Statistical Uncertainty quantification for Parton Density Distributions (PDFs)," CERN Indico, May 28, 2025.

Contact

huanbiaz@andrew.cmu.edu

Models

Effective Field Theory Modification of Standard Model Predictions

$$\mu(x, Q^2; c, f) = \nu(c, Q^2) T_{\mathrm{sm}}(x, Q^2; f),$$

with

$$\nu(c, Q^2) = 1 + k_{\mathrm{lin}}(Q^2)c + k_{\mathrm{quad}}(Q^2)c^2$$

- c: BSM parameter
- f: parton distribution function (PDF)
- T_{SM}: Standard Model predictions

Observed Data:

$$D \sim \mathcal{N}(\mu(c, f), \Sigma)$$

where

$$\mu(c, f) = \left[\mu(x_1, Q_1^2; c, f), \ldots, \mu(x_n, Q_n^2; c, f)\right]$$

Σ = Experimental Covariance Matrix

Methods

Goal: Inference on BSM Parameter c

1. **Initial estimation**
Fit PDF under SM assumption
(e.g. CTEQ-TEA, NNPDF, B-splines)
↓
Use fitted PDF to estimate c
↓
⇒ Initial estimates: $\hat{c}$ and $\hat{f}$.

2. Bias correction
One-step estimator

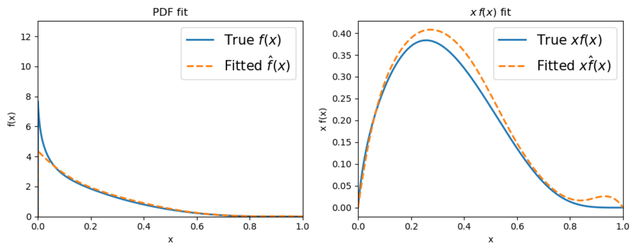
$$\tilde{c} = \hat{c} + \varphi(D, \hat{c}, \hat{f})$$

where φ is called the efficient influence function as a bias-correction term.
⇒ $\tilde{c}$ is asymptotically unbiased and normally distributed around c.

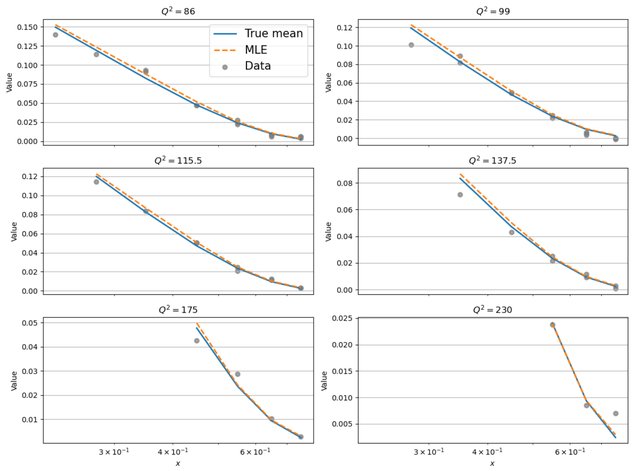
Results

T3 Dataset from DIS Experiments

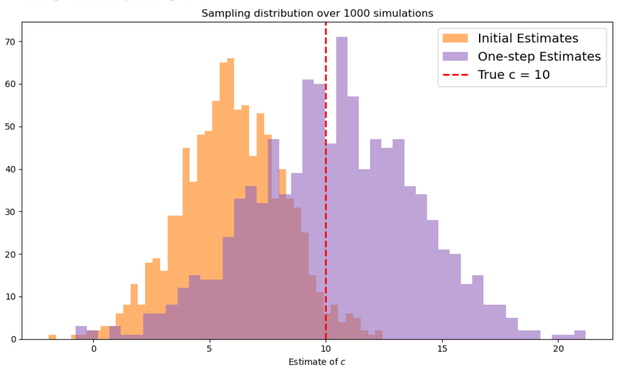
Initial SM-based PDF Fit



Data Fit Across Partial Q^2 values:

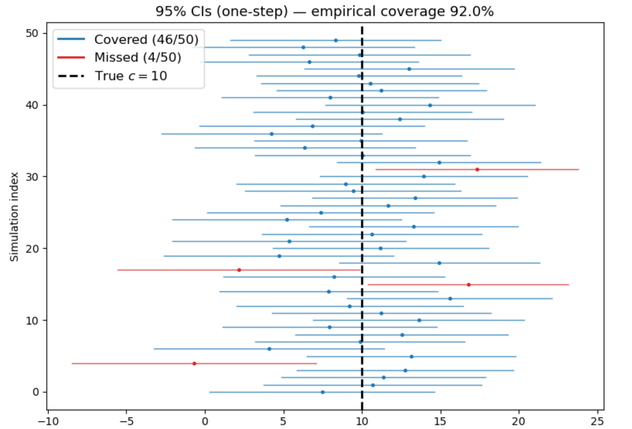


Bias Correction Improves BSM Estimation

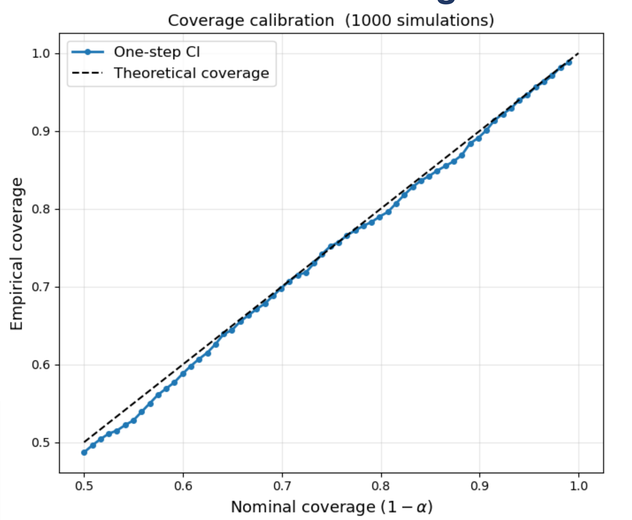


Truth: c = 10
Initial: mean = 5.97 bias = -4.03
One-step: mean = 10.35 bias = 0.35

One-step 95% Confidence Intervals



Near-nominal CI Coverage



Key Takeaways

- Treating PDFs as nuisance functions enables more robust estimation of BSM parameters.
- Semiparametric approach corrects first-order PDF-induced bias.
- Simulations show improved recovery of the true BSM parameter with correct frequentist coverage.