BGL Order Truncation and Model Selection in $|V_{cb}|$ Extraction

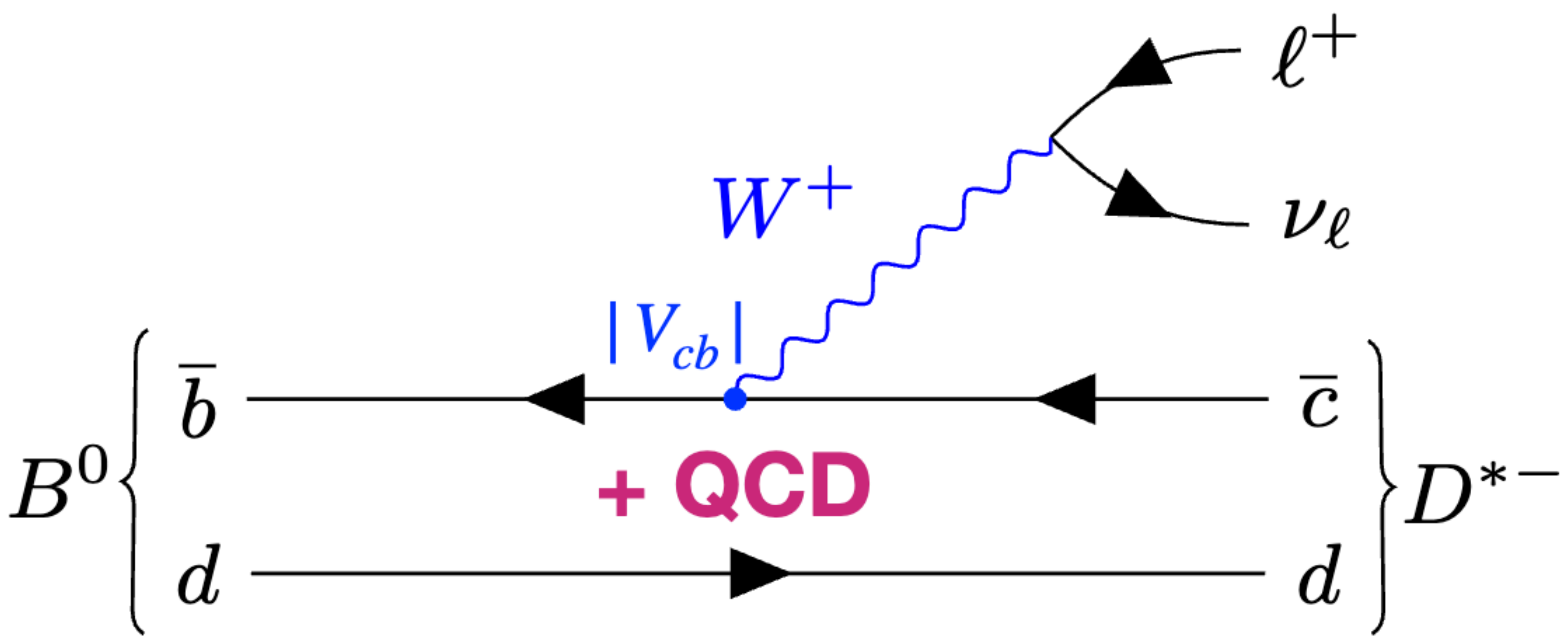
Eric Persson
Florian Bernlochner
Zoltan Ligeti
Markus Prim
Dean J. Robinson

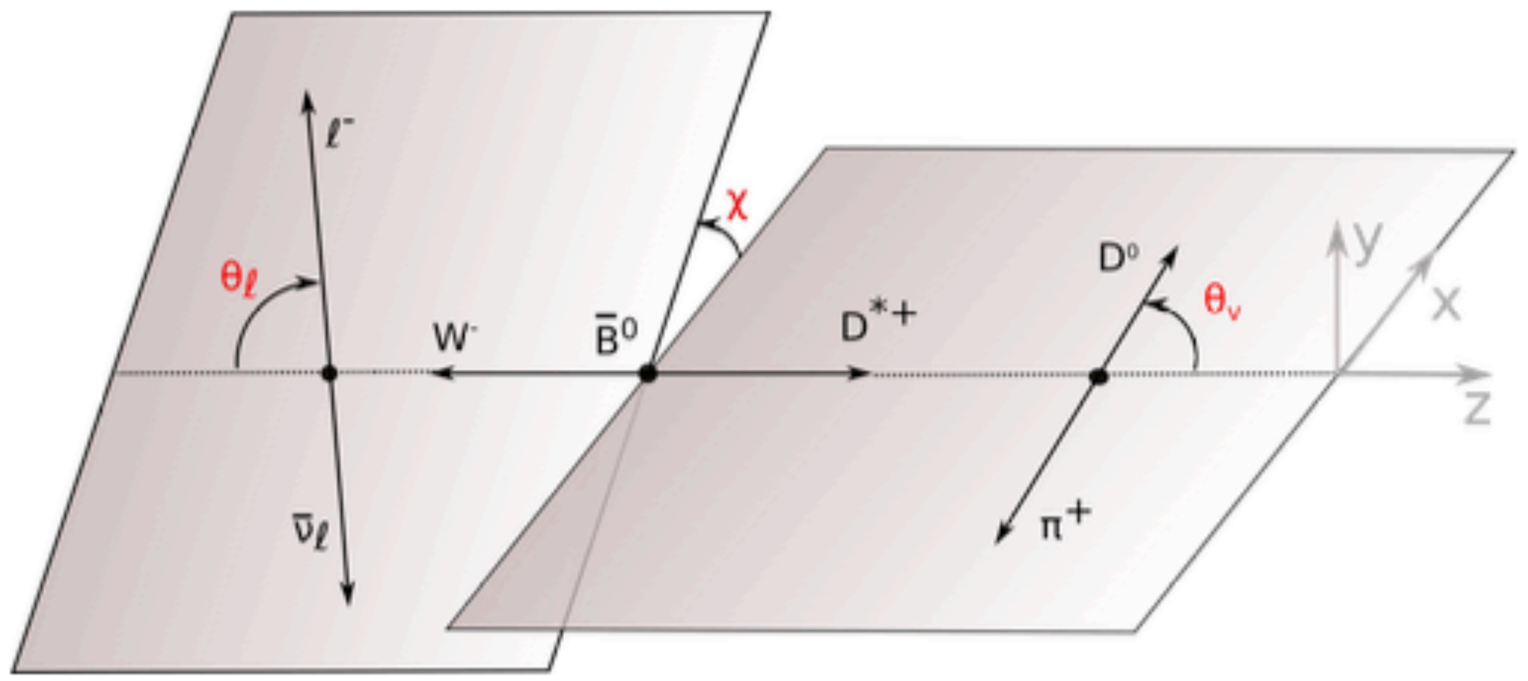
UNIVERSITÄT BONN

The $|V_{cb}|$ Puzzle

- **Discrepancy in $|V_{cb}|$ determinations:** There is a long-standing tension between inclusive and exclusive measurements of the CKM matrix element $|V_{cb}|$, with inclusive methods typically yielding higher values.

- **Focus on $B \to D^* l\nu$ channel:** Recent studies have concentrated on this exclusive decay channel due to its experimental accessibility and theoretical cleanliness.





BGL Parametrization and $|V_{cb}|$

- **Boyd-Grinstein-Lebed (BGL) form factor parametrization:** Expresses the form factors as a power series in a parameter z, incorporating unitarity constraints.

$$\begin{aligned}\langle D^*(\varepsilon,p')|\bar{c}\gamma^\mu b|\bar{B}(p)\rangle &= ig\epsilon^{\mu\nu\alpha\beta}\varepsilon^*_\nu p_\alpha p'_\beta,\\ \langle D^*(\varepsilon,p')|\bar{c}\gamma^\mu\gamma^5 b|\bar{B}(p)\rangle &= f\varepsilon^{*\mu} + (\varepsilon^*\cdot p)[a_+(p+p')^\mu + a_-(p-p')^\mu],\end{aligned}$$

$$g(z) = \frac{1}{P_g(z)\phi_g(z)}\sum_{n=0}^{N} a_n z^n, \qquad f(z) = \frac{1}{P_f(z)\phi_f(z)}\sum_{n=0}^{N} b_n z^n, \qquad \mathcal{F}_1(z) = \frac{1}{P_{\mathcal{F}_1}(z)\phi_{\mathcal{F}_1}(z)}\sum_{n=0}^{N} c_n z^n,$$

Combination of f and a_+

Conformal variable z:

$$z = \frac{\sqrt{w+1}-\sqrt{2a}}{\sqrt{w+1}+\sqrt{2a}}$$

QCD encoded in coefficients:

$$\{a_n, b_n, c_n\}$$

$c_0 =$ **constants** $\times\, b_0$

The Truncation Dilemma

- **Truncation order dilemma:** The choice of where to truncate the BGL expansion can impact the extracted $|V_{cb}|$ value:
  - **Truncate too soon:** Model dependence in extracted result for $|V_{cb}|$?
  - **Truncate too late:** Unnecessarily increase variance on $|V_{cb}|$?

Classic bias-variance trade-off. Necessary to develop principled, rigorous procedure for model selection in the context of BGL parametrization.

Introduction to Model Selection

- **Model selection:** the task of choosing the best model from a set of candidate models based on data, balancing complexity and fit (principle of parsimony).

- **In the BGL context:** each possible truncation order of the BGL expansion represents a different model to be evaluated.

- **Connection to statistical literature:** Conceptualizing the choice of BGL order as a model selection problem lets us connect this specific issue in HEP to a much broader and more general statistical literature.

Components of Model Selection

- **Model evaluation metrics:** These are quantitative measures to assess model performance, such as:
  - SSE (Sum of Squared Errors) a.k.a. "Chi-squared"
  - AIC (Akaike Information Criterion)
  - BIC (Bayesian Information Criterion)

$$\mu : BGLOrder \to \mathbb{R}$$

- **Model selection decision rules:** The criteria used to choose between models based on their evaluation metrics, e.g., "select the model with the lowest model evaluation metric" or "select the more complex model only if it improves the evaluation metric by at least some value".

$$\delta_\mu : (BGLOrder, BGLOrder) \to BGLOrder$$

- **Model space search algorithms:** These are the procedures used to explore the space of possible models, such as forward selection, backward elimination, or exhaustive search.

$$f_{\delta,\mu} : \Omega_{BGL} \to BGLOrder$$

Different approaches on the market

	Evaluation Metric	**Selection Rule**	**Search Algorithm**
Bernlocher et al. (2019)	χ^2	Choose nested model if $\boldsymbol{\Delta}\chi^2 > 1$	Forward stepwise selection
Gambino, Jung, Schacht	χ^2 + unitarity penalty	Higher complexity until stable	Forward selection
Current paper	AIC	Lowest metric (w/ and w/o UT)	Exhaustive search

Akaike Information Criterion

$$AIC = 2k - 2log(\hat{L})$$

where k is the number of parameters and $\hat{L}$ the maximized value of the likelihood function for the model. It aims to find the model that minimizes information loss.

Advantages:

- Theoretically well-motivated
- Easy to implement
- Ubiquitous in other fields (e.g. time series analysis)
- Allows for straightforward comparison of non-nested models

Toy Study Design

- **Purpose:** To demonstrate the effectiveness of AIC in choosing BGL order. Compare result to the Nested Hypothesis Test (NHT) approach of Bernlochner et al. (2019).

- **Data generation:** Simulate $B \to D^* l\nu$ decay data assuming true underlying order of (3,3,3), with covariance matrix from HFLAV that reflects current world average precision.

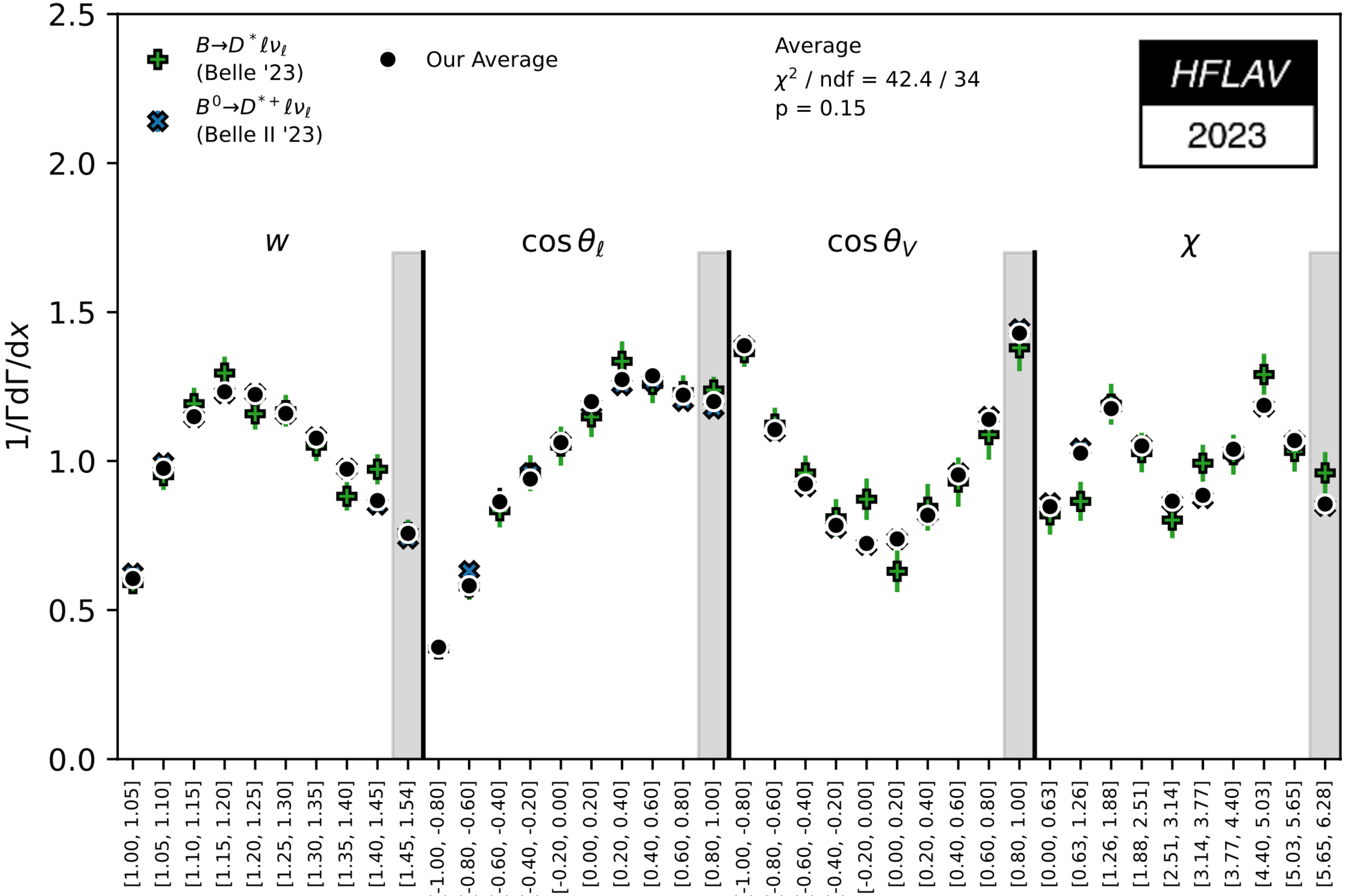
- **Model fitting:** Fit all permutations of BGL orders from (1, 1, 1) to (3, 3, 3) to the simulated data using standard least squares fit.

- **Comparison:** Apply AIC and NHT procedures to select the optimal BGL order. Show that our procedure produces unbiased estimates of $|V_{cb}|$ with correct coverage properties.

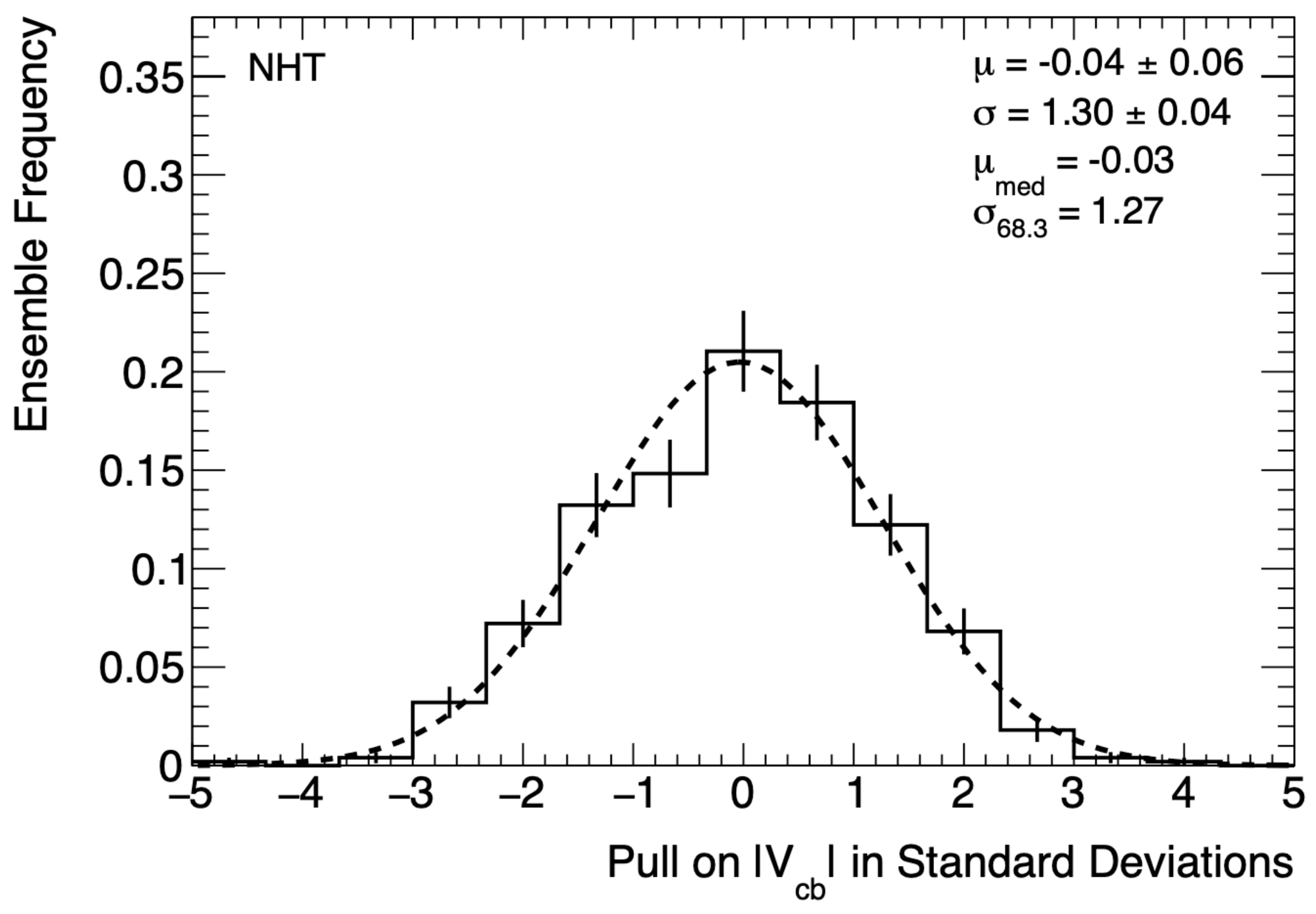
$|V_{cb}|$ pulls

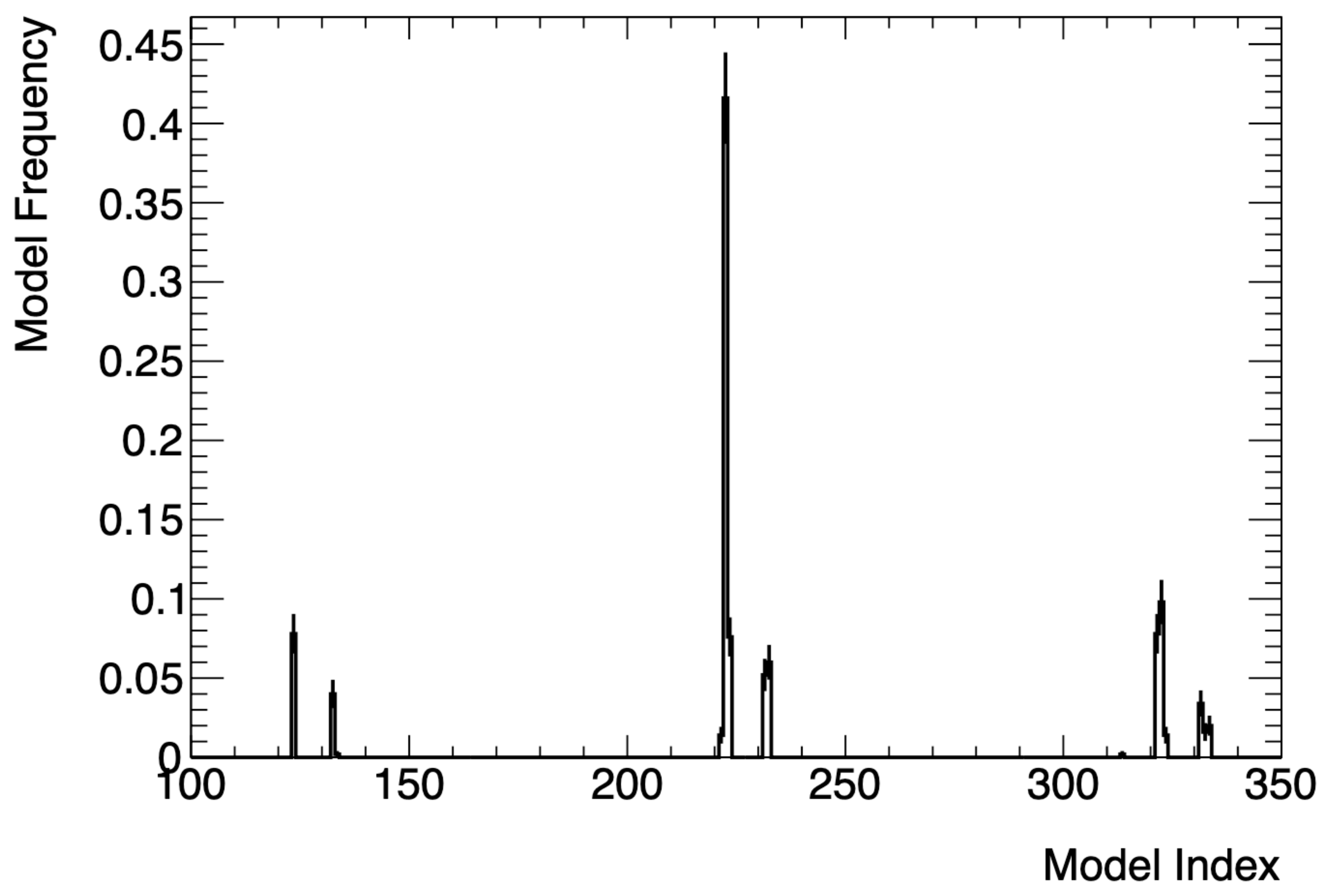
$$\text{pull}_i = \frac{V_{cb}^{i} - V_{cb}^{TRUE}}{\sigma_i}$$

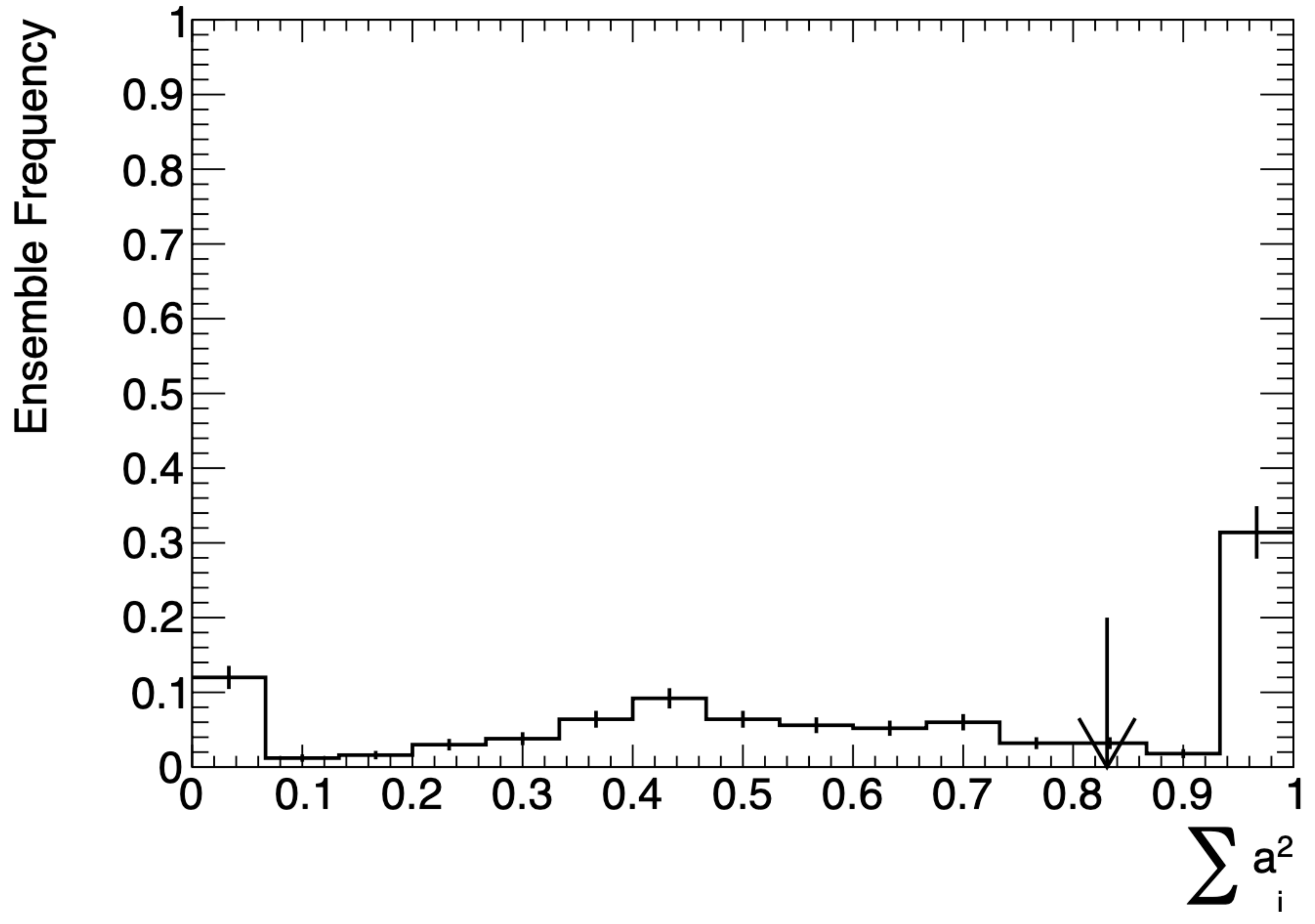
Our D* averaged spectrum

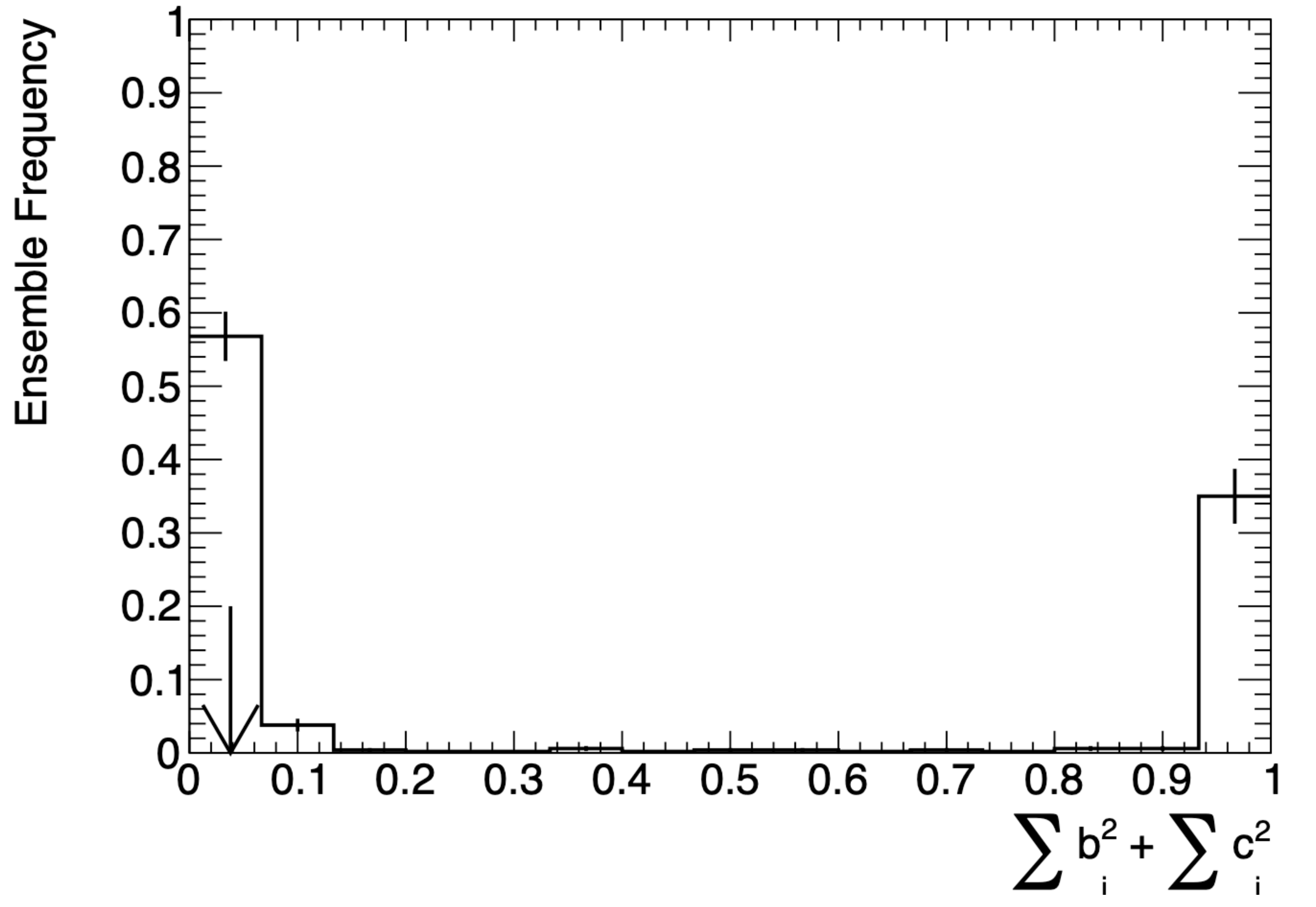


NHT (without unitarity constraints)

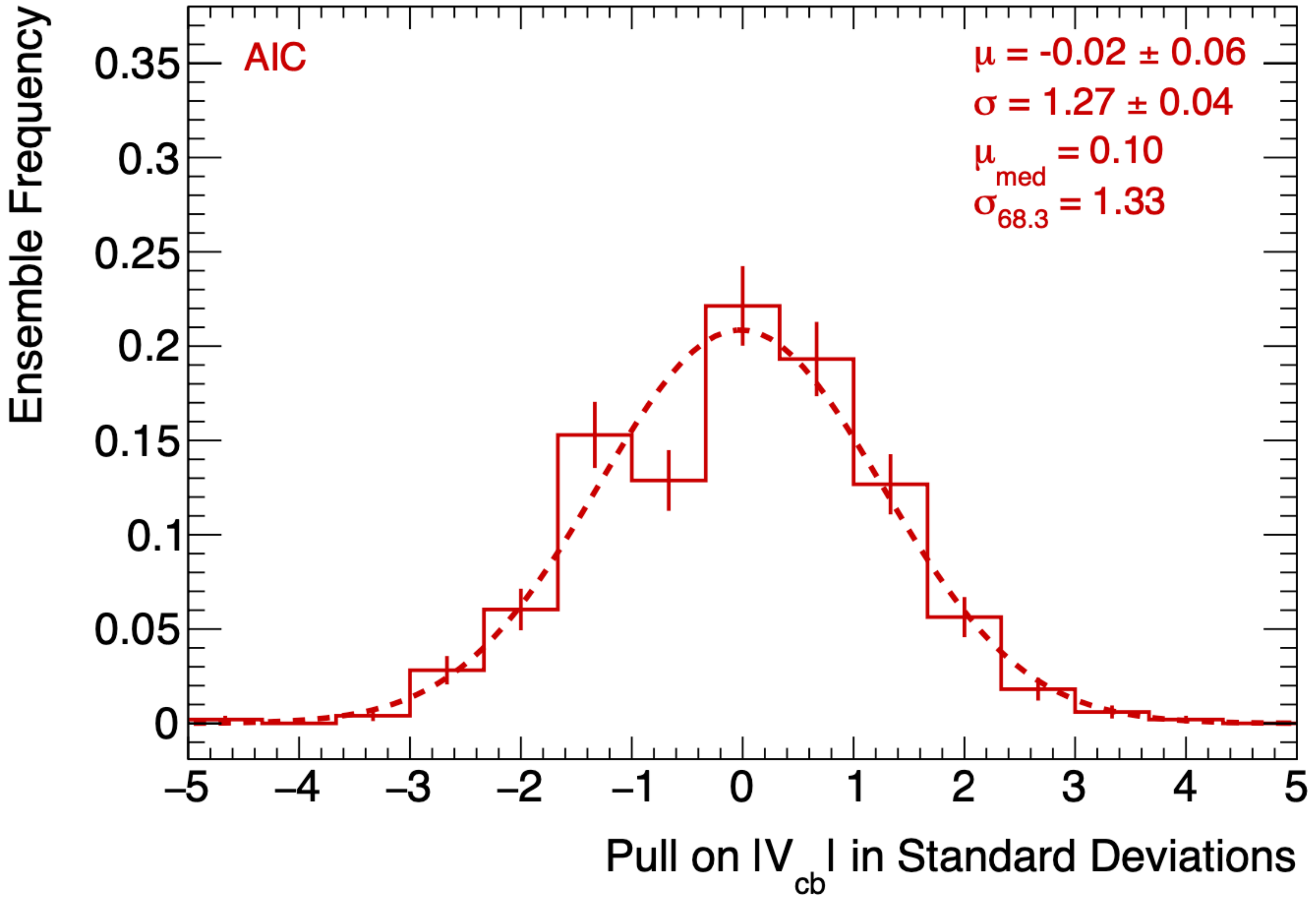


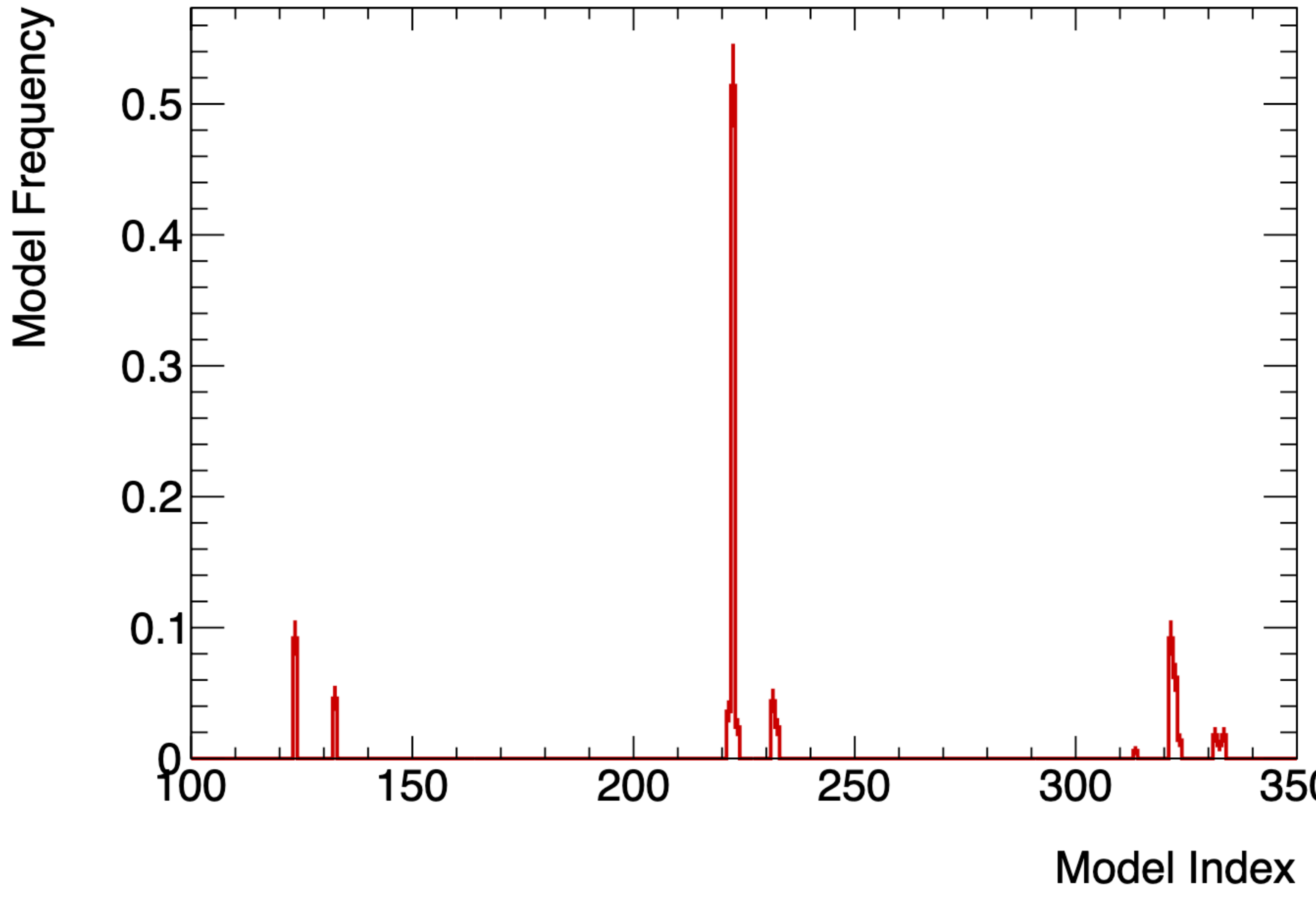


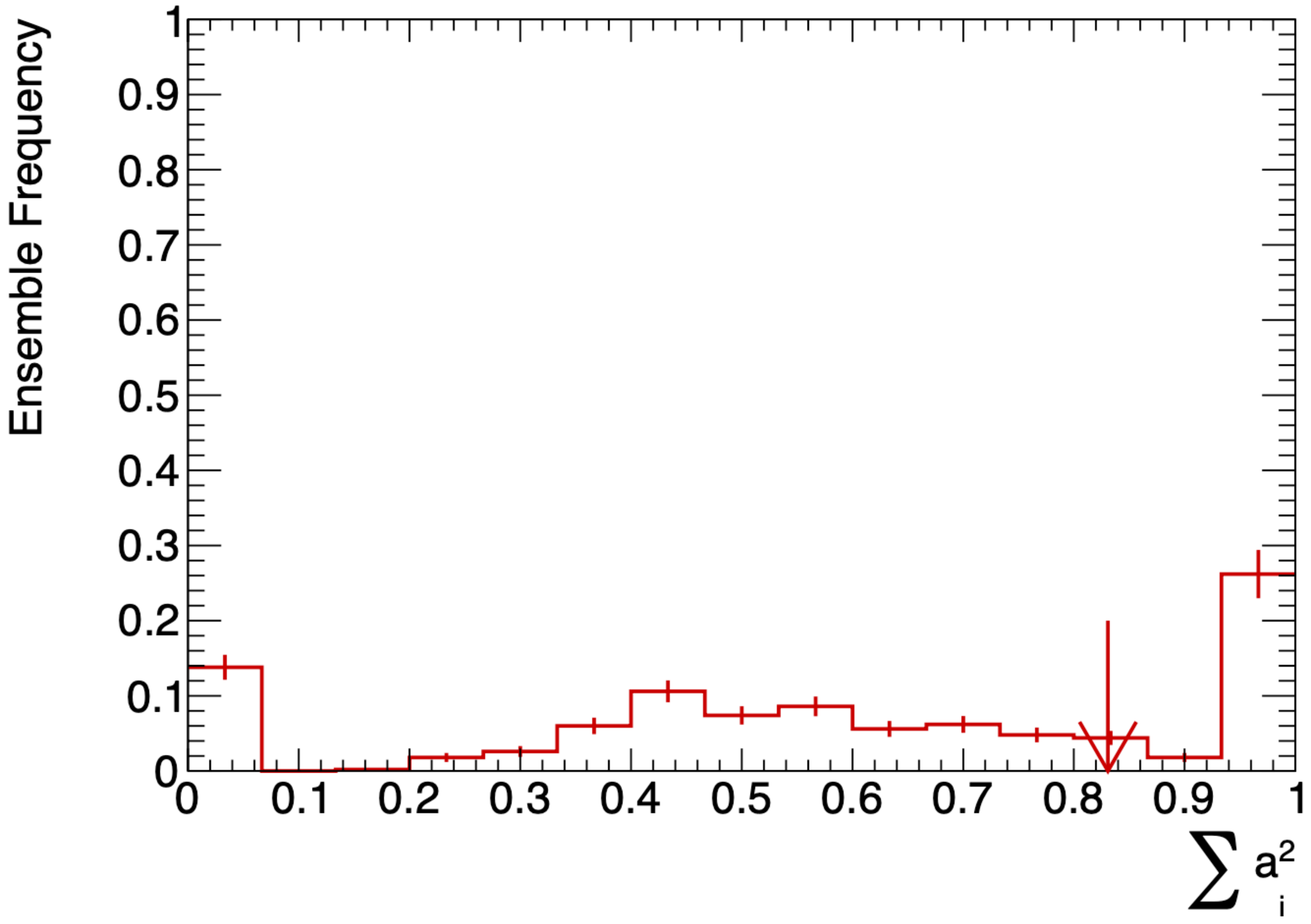


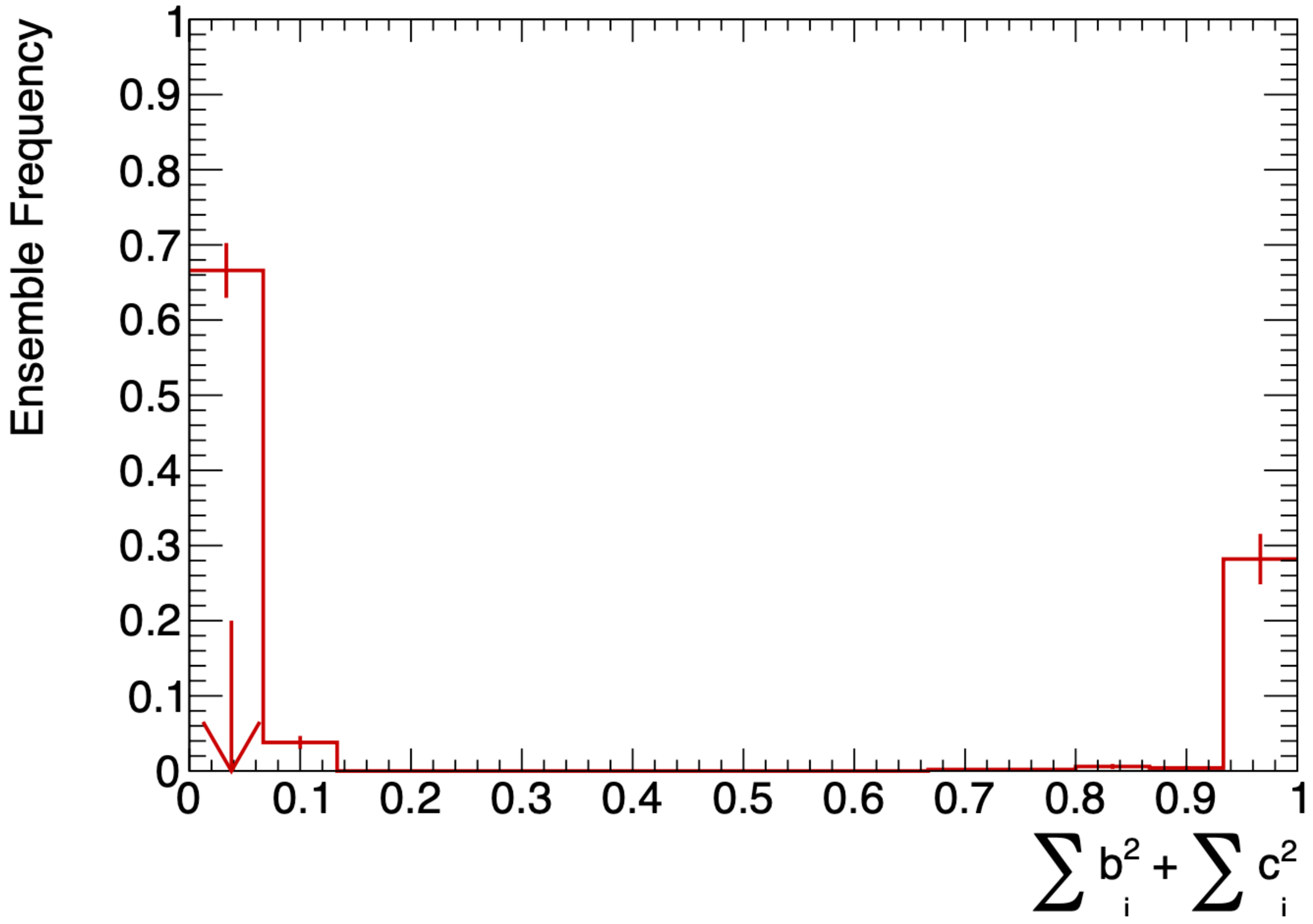


AIC (without unitarity constraints)

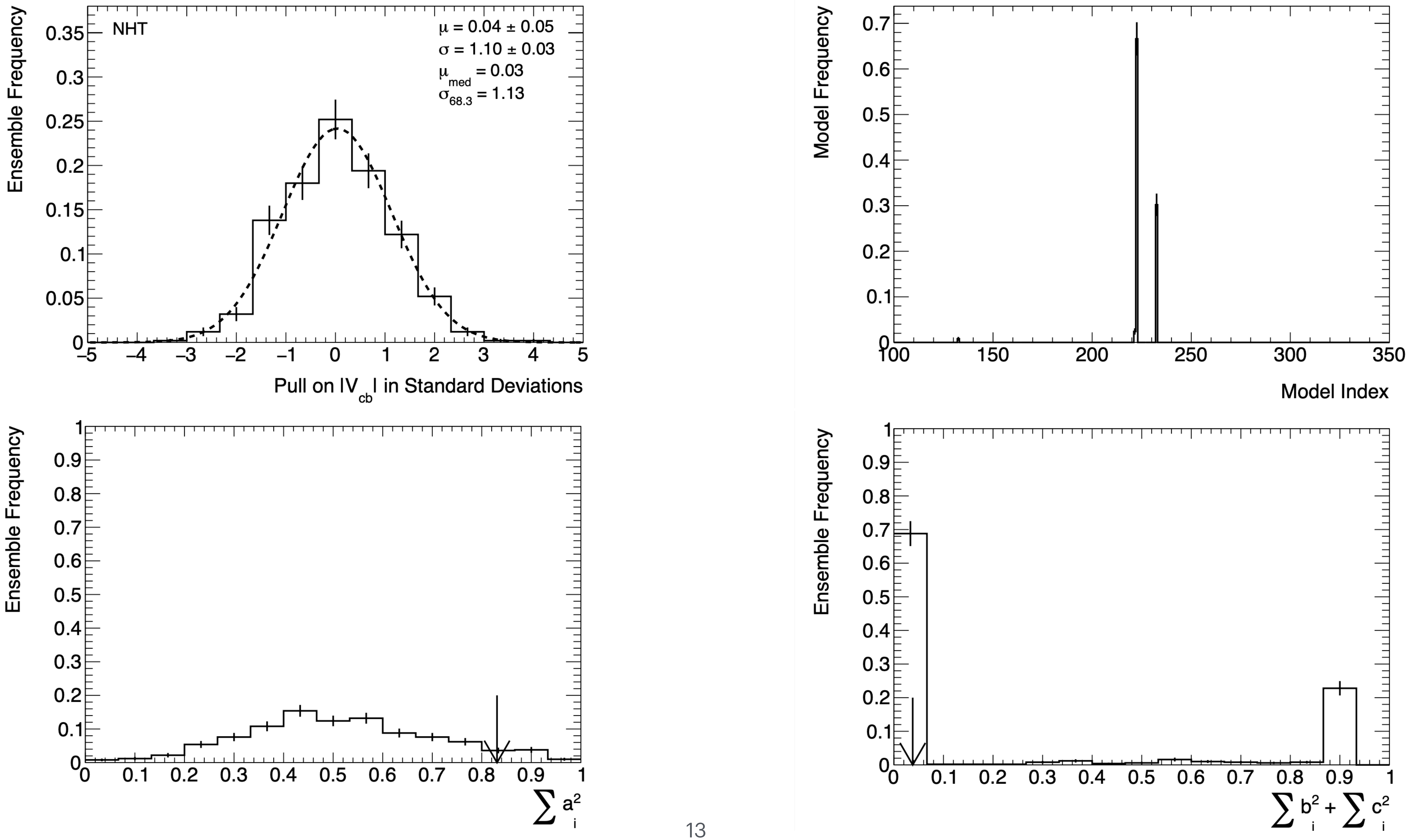




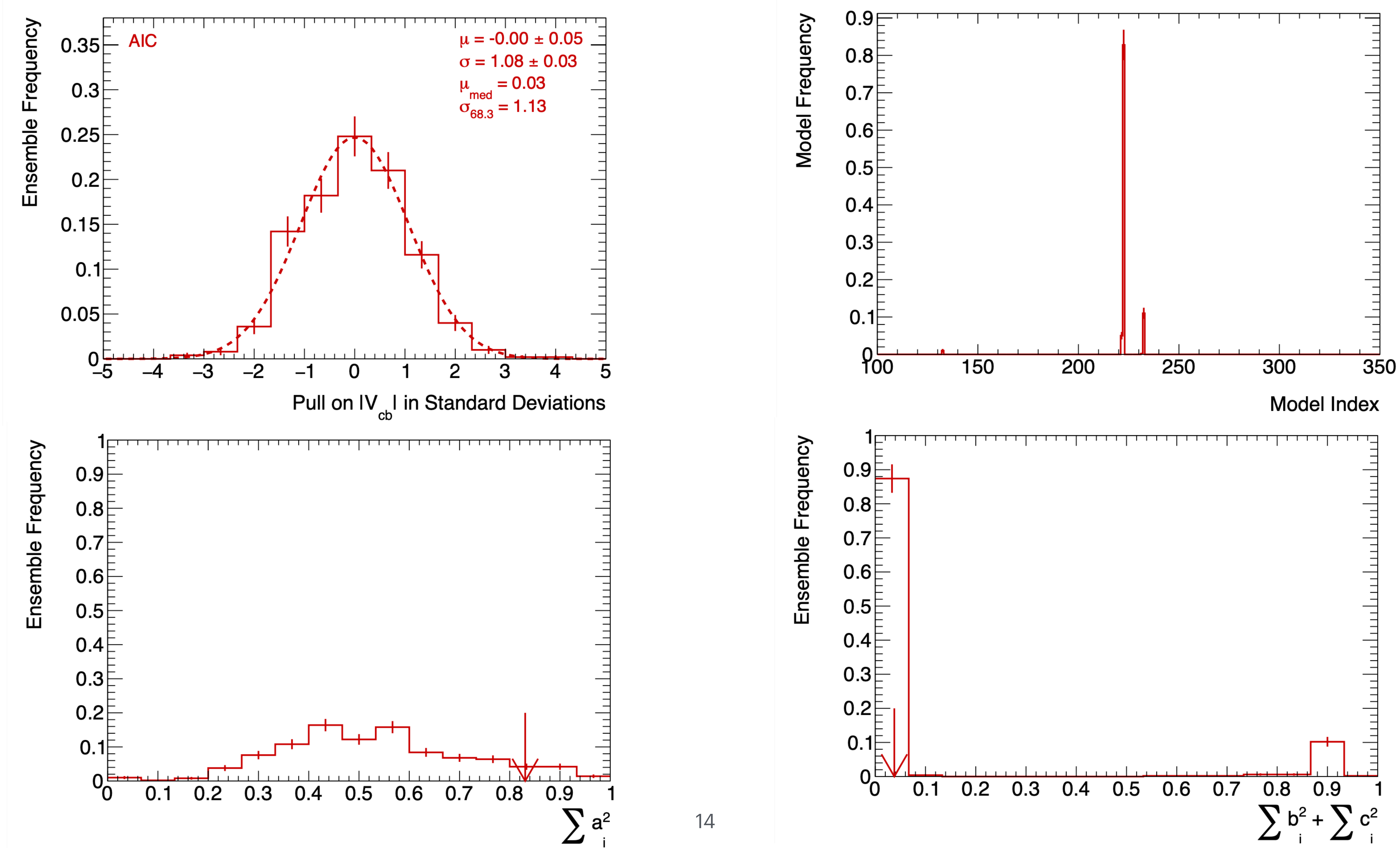




NHT (with unitarity constraints)



AIC (with unitarity constraints)



Beyond Single Model Selection

- **Limitations of single model selection:** Choosing a single "best" model ignores model uncertainty and can lead to overconfident inferences. Unnecessarily dichotomous.
- **Model averaging approaches:** These methods consider multiple models, weighing their contributions based on their relative support from the data.
- **Accounting for model uncertainty:** By considering multiple models, we can more accurately reflect our uncertainty about the true underlying process.

Global AIC (gAIC)

An approach that weighs multiple models based on their AIC scores, rather than selecting a single best model:

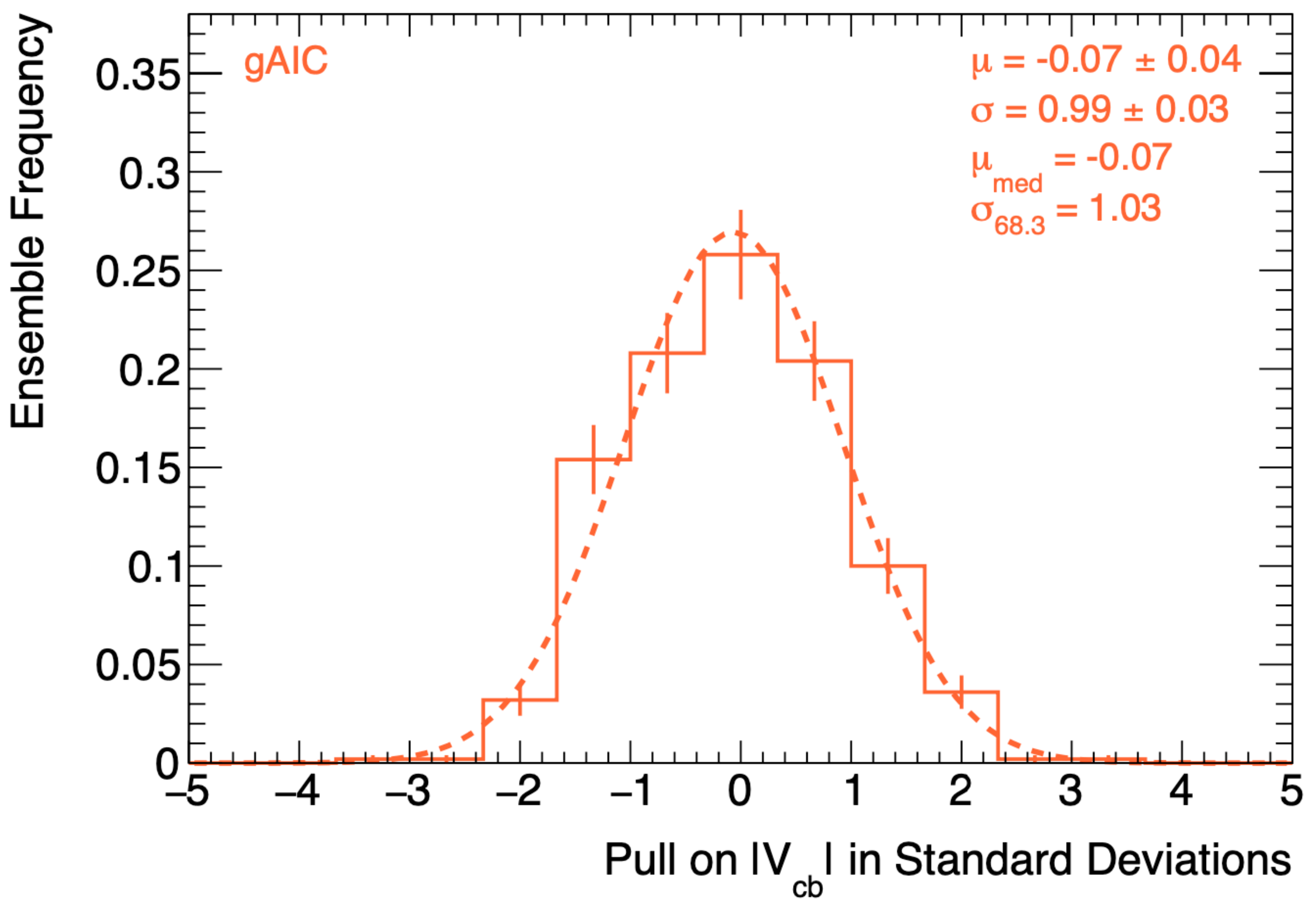
$$w_i = \exp(-\frac{1}{2}\Delta_i) / \sum_j \exp(-\frac{1}{2}\Delta_j) \qquad \Delta_i = \text{AIC}_i - \text{AIC}_{\min} \qquad |V_{cb}| = \sum_i w_i |V_{cb}|_i$$

where ΔAIC is the difference between a model's AIC and the minimum AIC in the set.

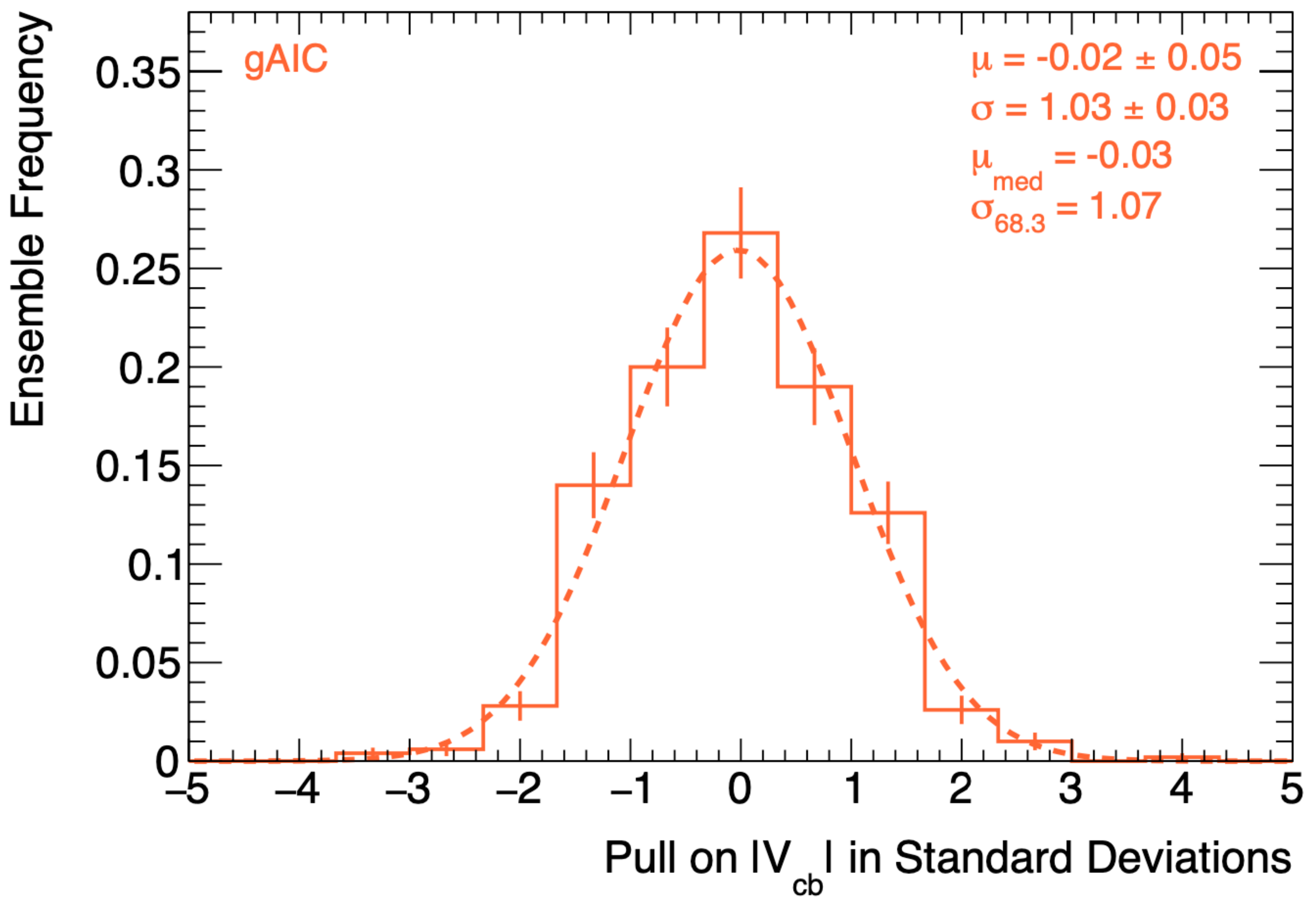
- **Advantages:** gAIC provides a more nuanced view of model performance, captures model selection uncertainty, and can lead to more robust predictions and parameter estimates.
- **Model uncertainty:** Accounts for the fact that multiple models may be plausible given the data.
- **Comprehensive view:** Offers a more nuanced understanding of the model space than single model selection.

gAIC

Without unitarity constraints



With unitarity constraints



Outlook

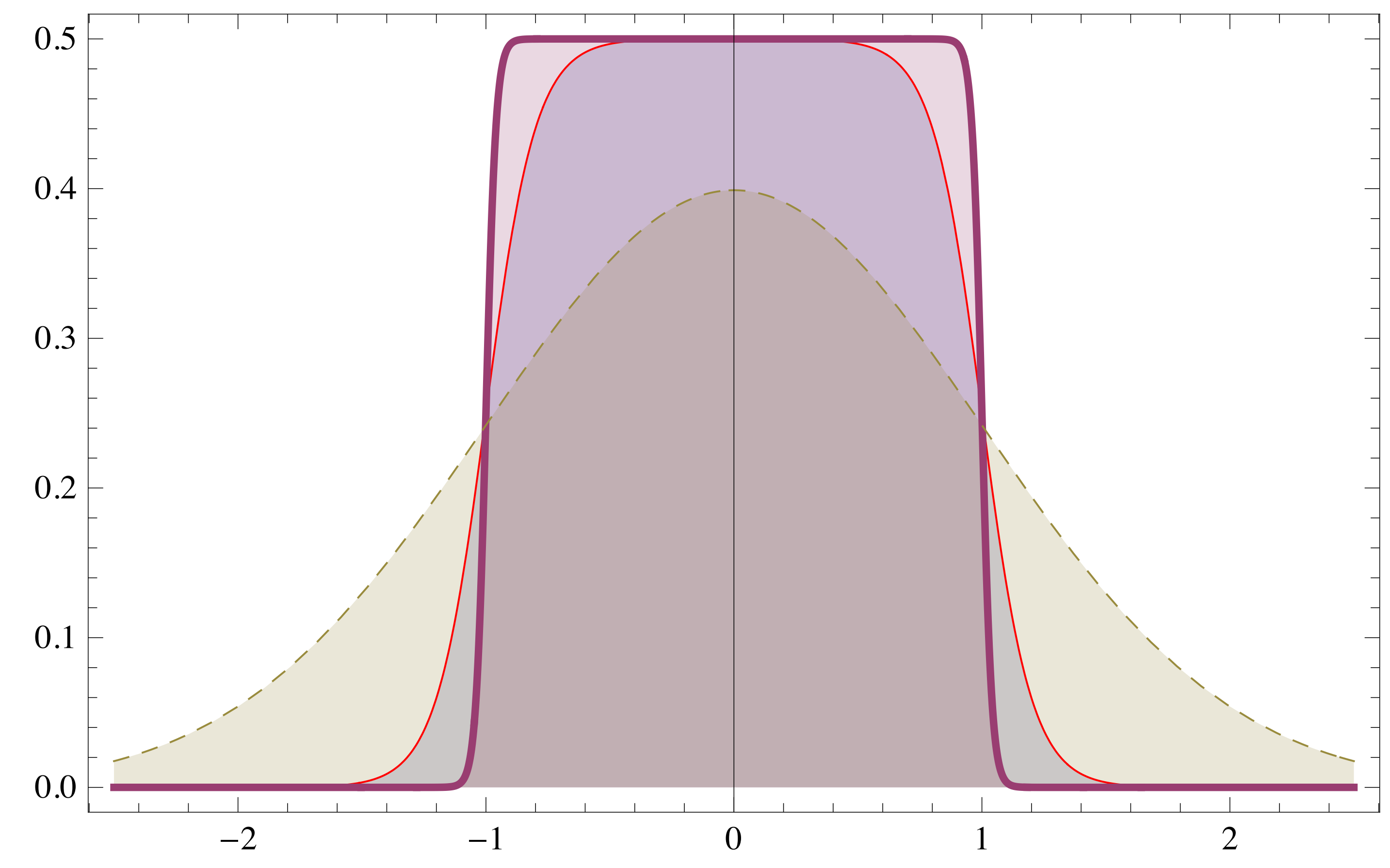
- **Finalize current studies:** Complete ongoing analyses and perform robustness checks.
- **Alternative metrics:** Explore other model evaluation criteria like adjusted R^2 or Mallows's C_p to show that AIC also outperforms these.
- **Incorporate external constraints:** Investigate the impact of including lattice QCD constraints in the model selection process.

Imposing unitarity

- To impose unitarity, we include a penalty into the function of the form $\chi^2 \to \chi^2 - 2 \sum_{a_i, \{b_i, c_i\}} \log \mathrm{DFD}$
- I.e. for each BGL coefficient we check if unitarity is violated, e.g. via

$$\sum_i a_i^2 / |V_{cb}|^2 \leq 1$$

$$\sum_i b_i^2 / |V_{cb}|^2 + \sum_i c_i^2 / |V_{cb}|^2 \leq 1$$



$DFD(x, w) = 1 / \left(2 \left(1 + e^{w(x-1)}\right) \left(1 + e^{-w(x+1)}\right)\right)$

The double Fermi Dirac function (DFD) provides an approximate top-hat function and penalizes the χ^2 only if a boundary is hit. We use $w = 50$ for the transition.