Parity Calibration

Youngseog Chung, Aaron Rumack, Chirag Gupta

August 1, 2023

Carnegie Mellon University
MACHINE LEARNING DEPARTMENT
Auton Lab
DELPHI GROUP

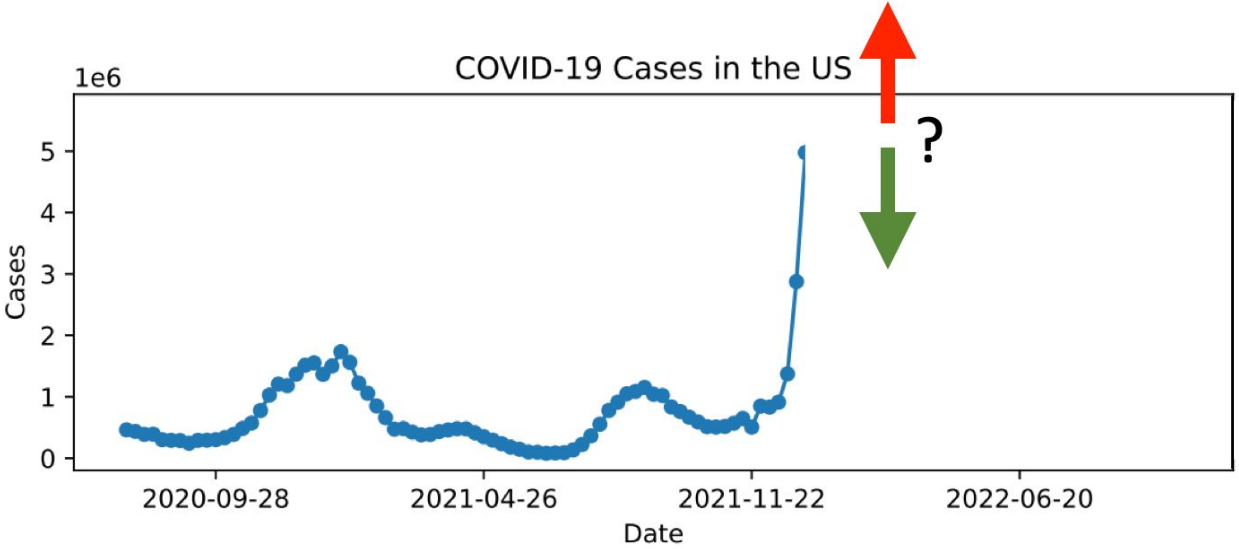
Youngseog Chung

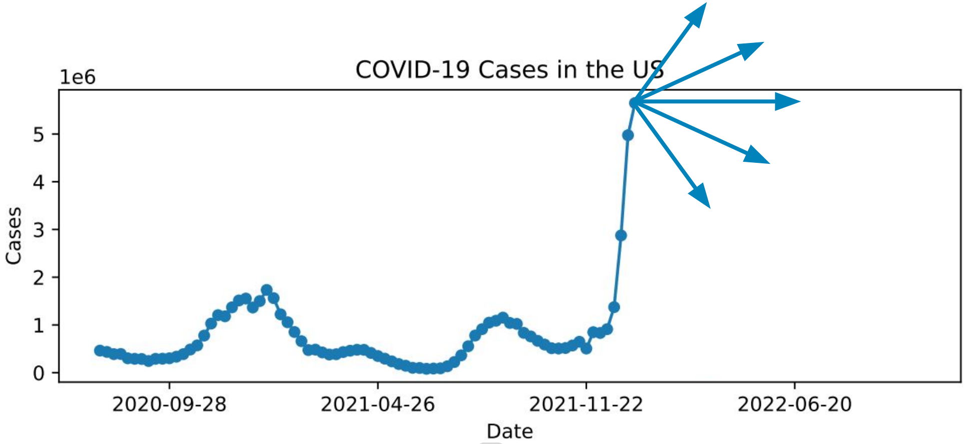
Aaron Rumack

Chirag Gupta

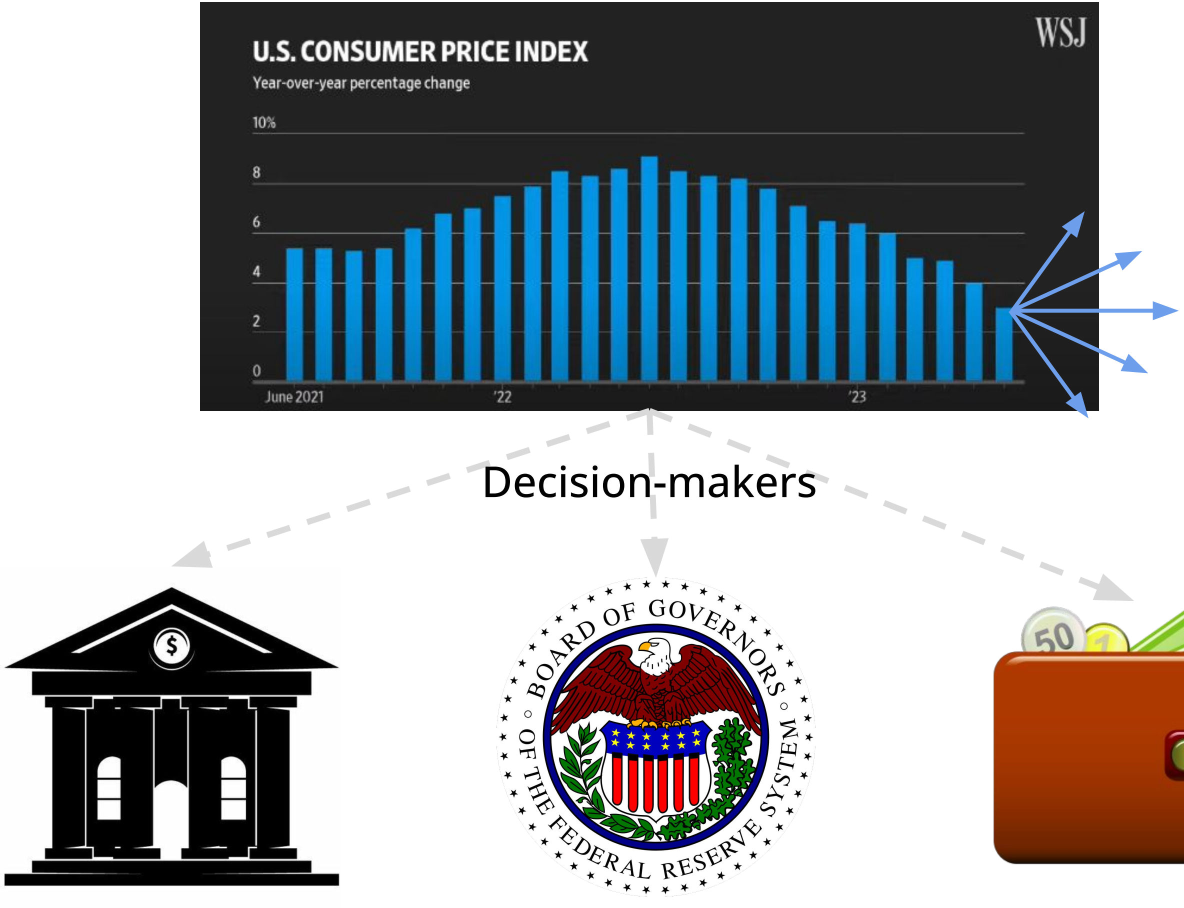
Carnegie Mellon University

MACHINE LEARNING DEPARTMENT


COVID-19 Cases in the US
1e6
5
4
3
2
1
0
Cases
2020-09-28
2021-04-26
2021-11-22
2022-06-20
Date
?



Decision-makers

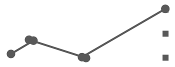
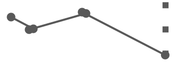

U.S. CONSUMER PRICE INDEX
Year-over-year percentage change
WSJ
10%
8
6
4
2
0
June 2021
'22
'23
Decision-makers
BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM
50

Parity Calibration

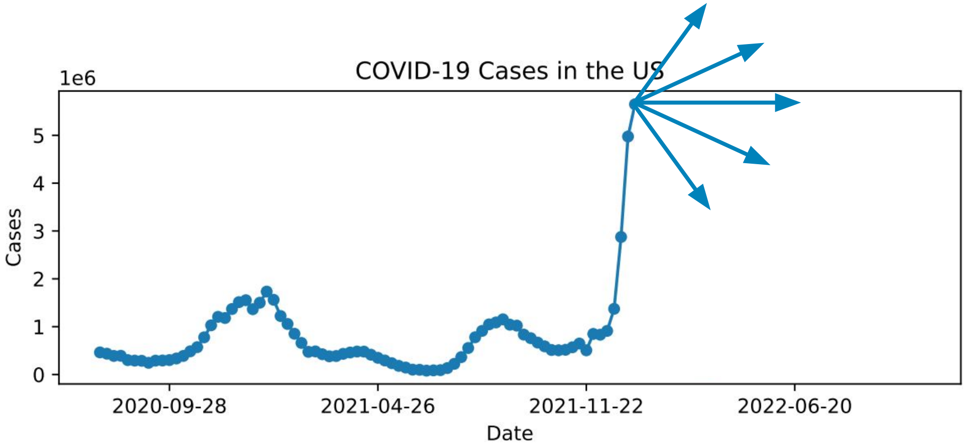
- Sequential observations of real valued targets $y_1, y_2, \cdots \in \mathbb{R}$
- At timestep $t \geq 2$, predict whether $y_t > y_{t-1}$ or $y_t \leq y_{t-1}$
- Define ***parity outcomes*** *as* $\tilde{y}_t := \mathbb{1}\{y_t \leq y_{t-1}\}$
- Problem: produce calibrated predictions $\hat{p}_t$ for the parity outcomes $\tilde{y}_t$

$$\frac{\sum_{t=2}^{T} \tilde{y}_t \mathbb{1}\{\hat{p}_t = p\}}{\sum_{t=2}^{T} \mathbb{1}\{\hat{p}_t = p\}} \to p, \forall p \in [0,1]$$

- E.g. whenever I predicted 30%, the current observation *decreased* relative to the previous observation 30% of the time

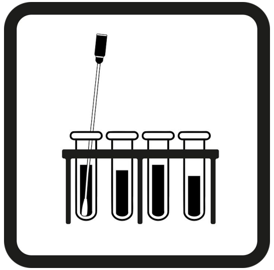
Target Timeseries	Parity Outcome
	$\tilde{y}_t = 0$
	$\tilde{y}_t = 1$

Timestep = t



Domain expert

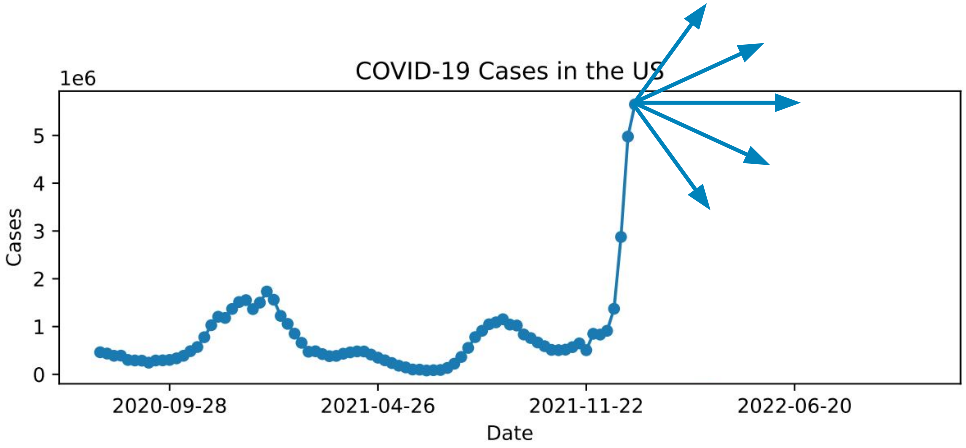
COVID-19 ForecastHub



DELPHI GROUP

Forecasts

Decision-maker

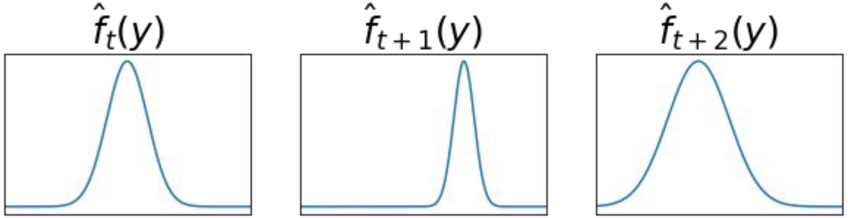


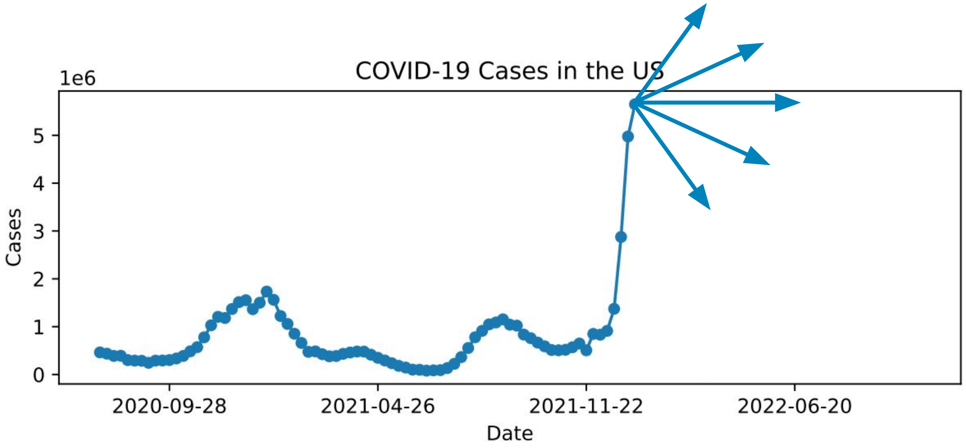
Domain expert

COVID-19 ForecastHub

DELPHI GROUP

Predictive pdf's

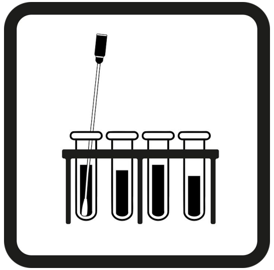




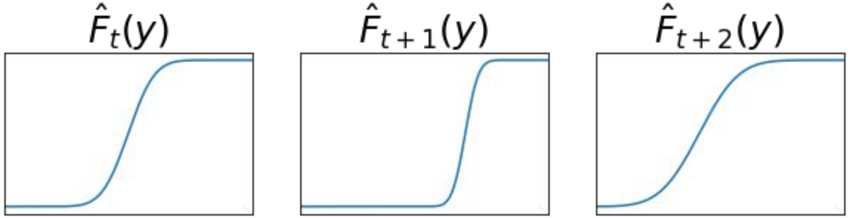
Domain expert

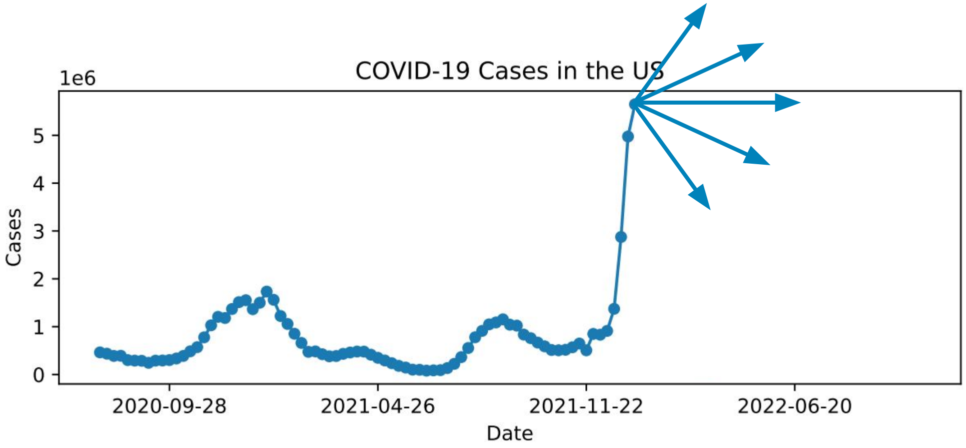
Predictive cdf's

COVID-19 ForecastHub



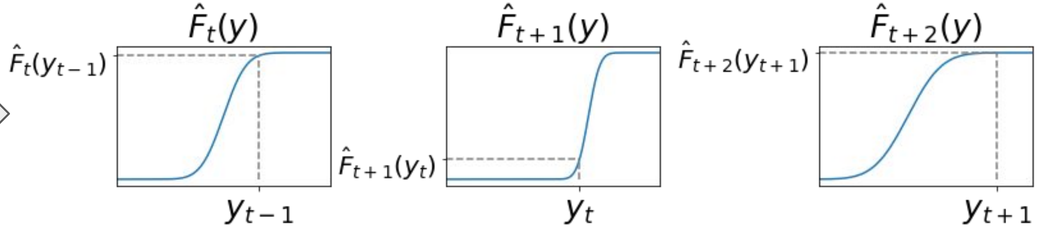
DELPHI GROUP

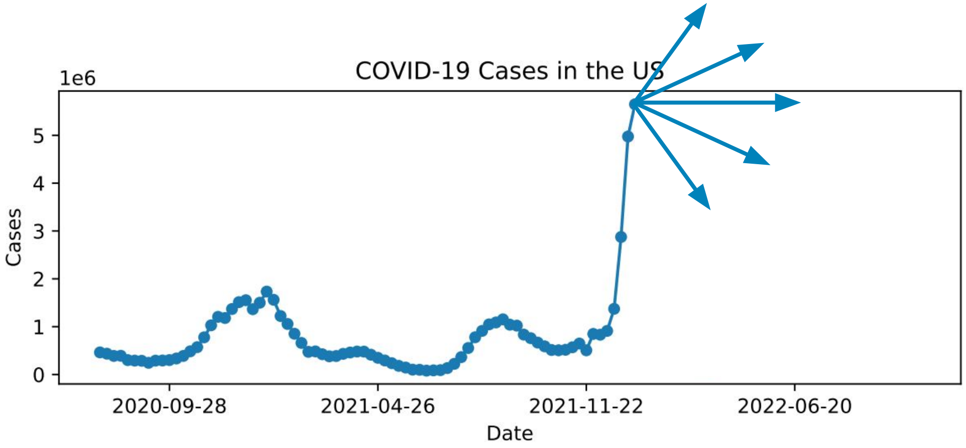




Domain expert

Predictive cdf's



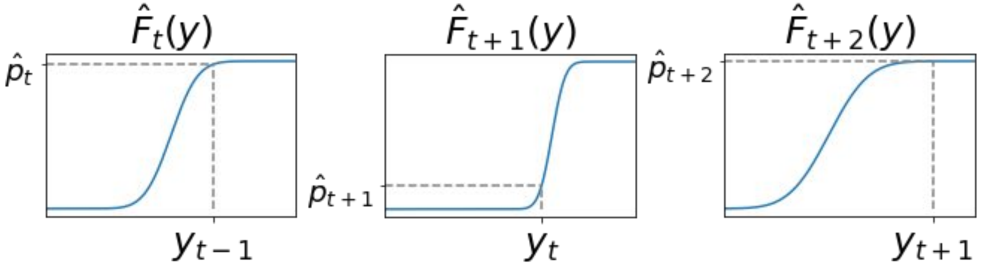


Domain expert

COVID-19 ForecastHub

DELPHI GROUP

Predictive cdf's



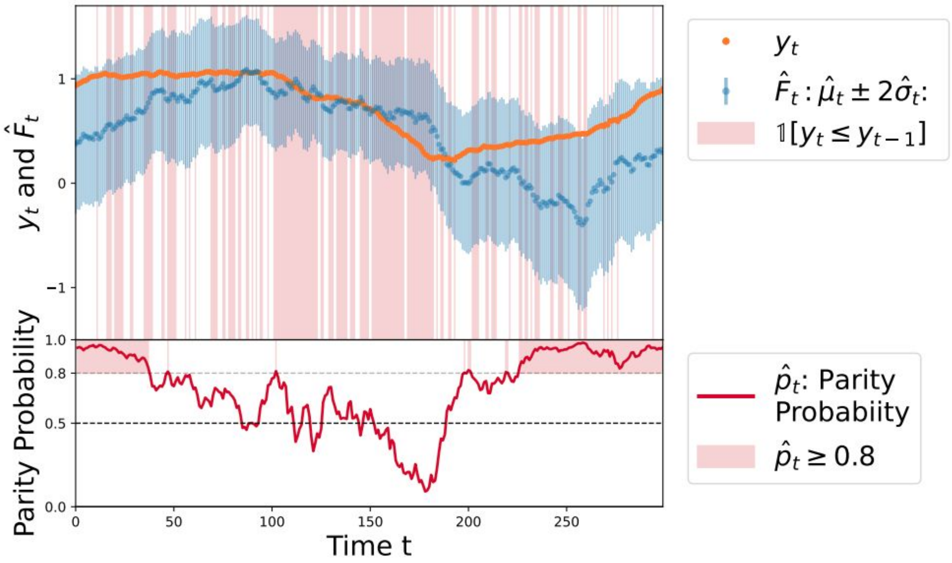
"Well-calibrated" $\hat{F}_t$ $\xrightarrow{?}$ "Well-calibrated" $\hat{p}_t$

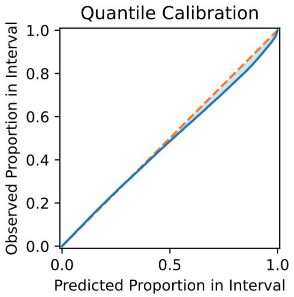
Calibration in regression does not imply parity calibration

....for many common and well-known notions of calibration in regression:

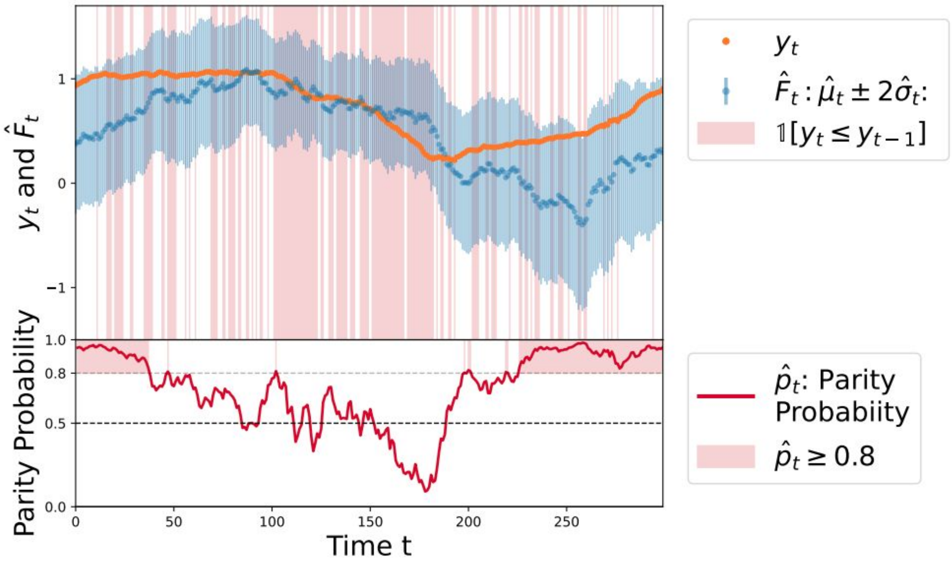
- Probabilistic calibration (or quantile calibration) [Gneiting et al., 2007]
- Distribution calibration [Song et al., 2019]
- Threshold calibration [Sahoo et al., 2021]

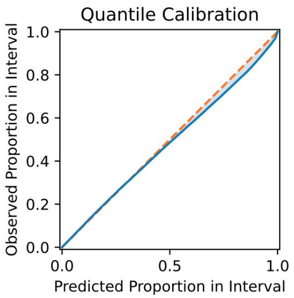
Calibration in regression does not imply parity calibration

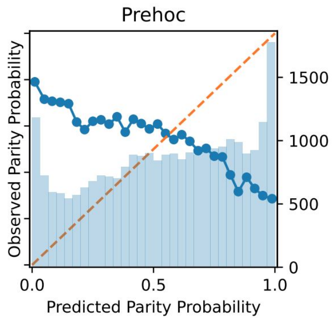




Calibration in regression does not imply parity calibration







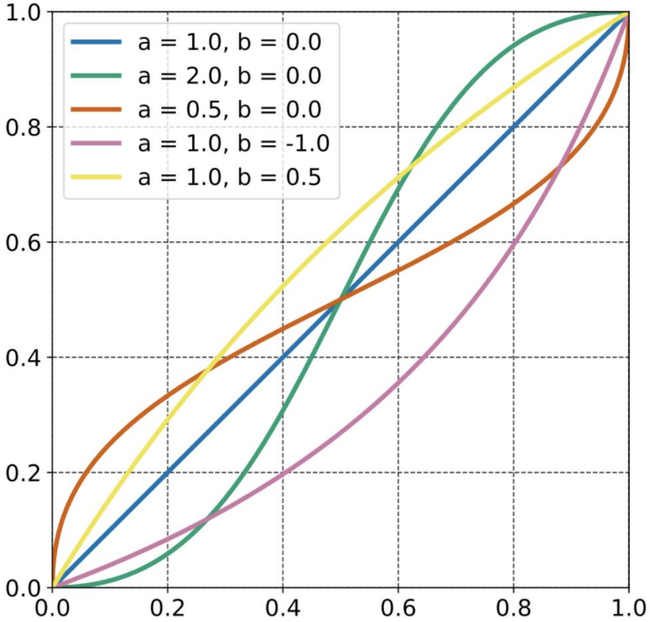
Parity Calibration via Post-hoc Calibration

- Online learn post-hoc mapping $m_t : [0, 1] \to [0, 1]$ for adaptivity to distribution drifts/shifts
- Prehoc forecast: $\hat{p}_t = \hat{F}_t(y_{t-1})$
- Post-hoc calibrated forecast: $m_t(\hat{p}_t)$

Platt Scaling

$$m^{a,b} = \text{sigmoid}(a * \text{logit}(\cdot) + b)$$

- $\text{sigmoid}(z) = \frac{1}{1+e^{-z}}$
- $\text{logit}(z) = \log \frac{z}{1-z}$



Parameters (a, b) are learned by minimizing the log loss:

$$l(p, y) = -y \log(p) - (1 - y) \log(1 - p)$$

where $p = m^{a,b}(\hat{p})$

Online versions of Platt scaling

- Windowed Platt scaling:
  - Re-fit (a_t, b_t) with a recent history of the target timeseries

- Online Platt Scaling (OPS) [Gupta & Ramdas, 2023]
  - Re-frame online learning of Platt scaling model as *online logistic regression* with 1-dimensional pseudo-features $\text{logit}(\hat{p})$

$$p_t^{\text{OPS}} = \text{sigmoid}(a_t * \text{logit}(\hat{p}_t) + b_t)$$

Learn a_t, b_t that minimize $\sum_{t=1}^{T} l(p_t^{\text{OPS}}, y_t)$

l is logistic regression loss
$l(p, y) = -y\log(p) - (1-y)\log(1-p)$

Or, minimize regret

$$R_T = \sum_{t=1}^{T} l(p_t^{\text{OPS}}, y_t) - \min_{(a,b)\in\mathcal{B}} \sum_{t=1}^{T} l(m^{a,b}(f(\mathbf{x}_t)), y_t)$$

Empirical Case Studies: COVID-19 forecasting

COVID-19 ForecastHub

Website: www.covid19forecasthub.org

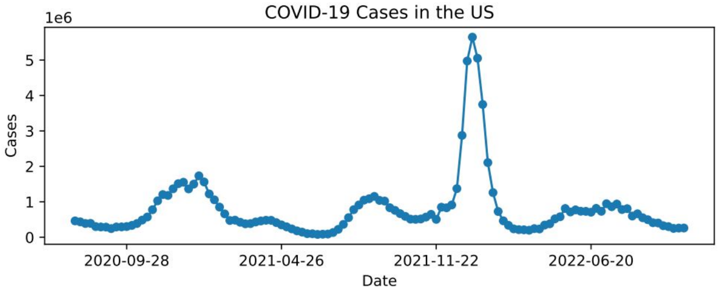
The COVID-19 Forecast Hub

This site maintains the authoritative record for real-time forecasts of COVID-19 hospitalizations in the US, as well as archival forecasts for COVID-19 cases and deaths, created by dozens of leading infectious disease modeling teams from around the globe, in coordination with the US CDC.

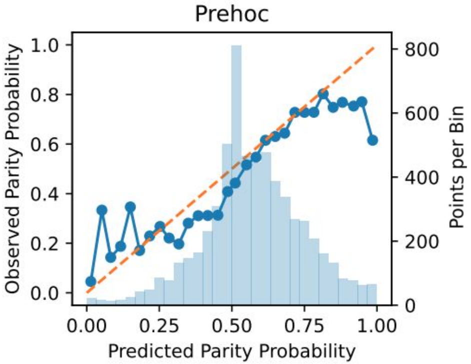
Up-to-date forecasts

Every week dozens of modeling teams from across the globe submit forecasts of the trajectory of the COVID-19 pandemic in the US to our forecast data repository. In collaboration with the US CDC, we take these data and build a single ensemble forecast which is later analyzed by, and communicated to the general public by CDC.

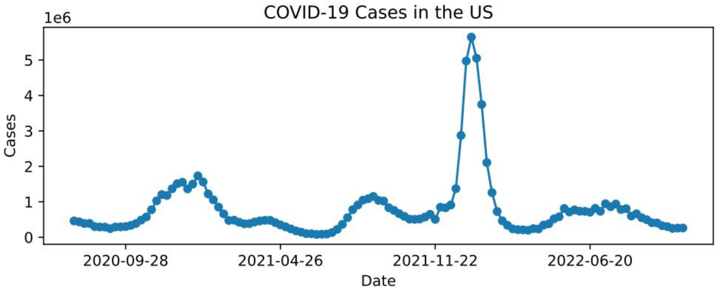
Empirical Case Studies: COVID-19 forecasting



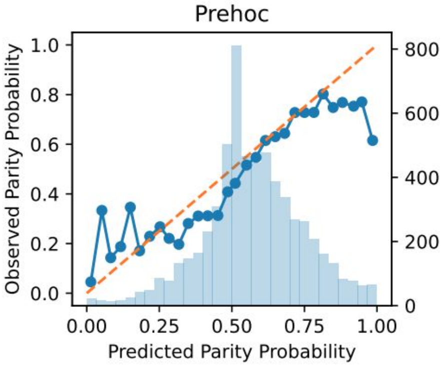
COVID-19
ForecastHub

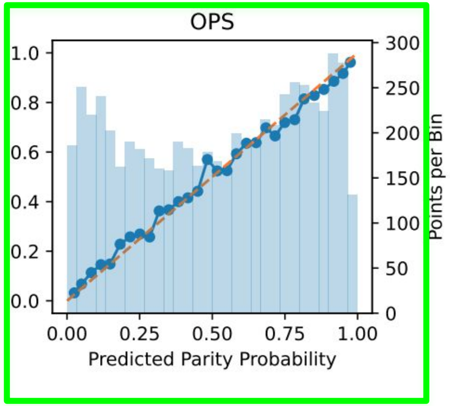


Empirical Case Studies: COVID-19 forecasting

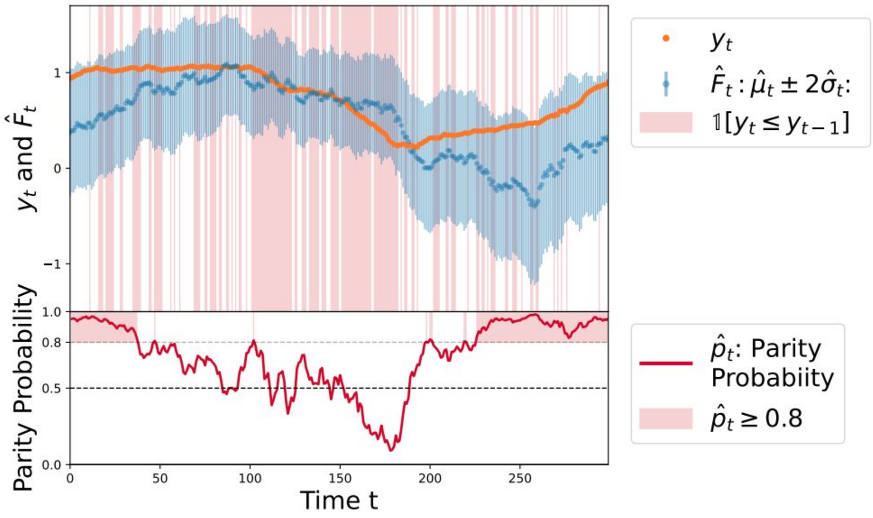


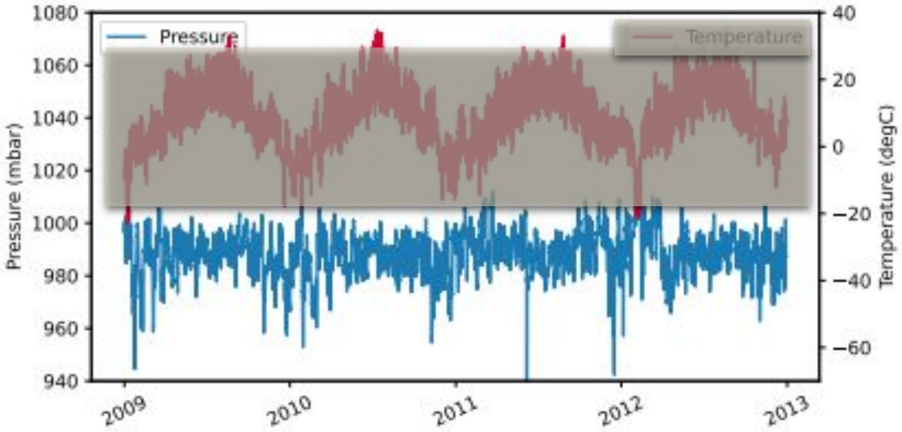
COVID-19
ForecastHub



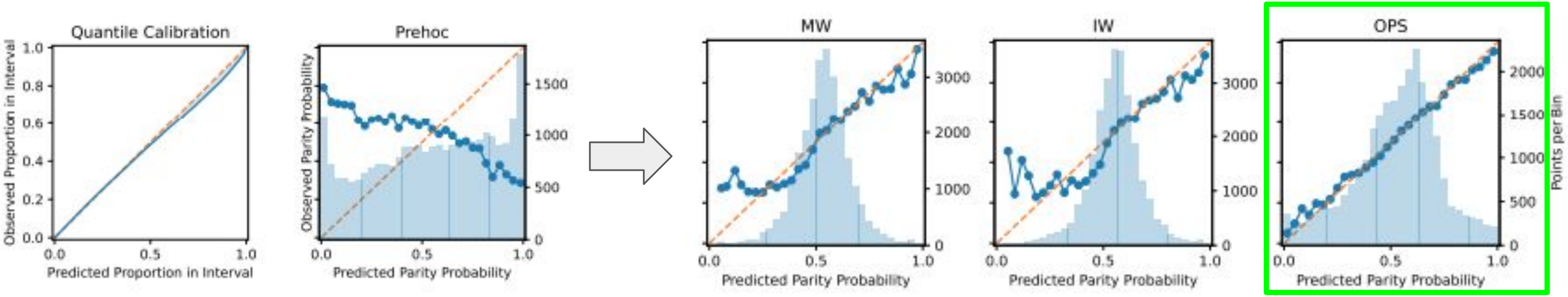


Empirical Case Studies: Weather forecasting





Empirical Case Studies: Weather forecasting



Thank you