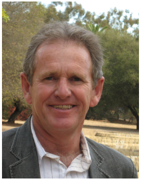
Smooth multi-period forecasting with application to prediction of COVID-19 cases

Elena Tuzhilina

Stanford University, Department of Statistics

elenatuz@stanford.edu



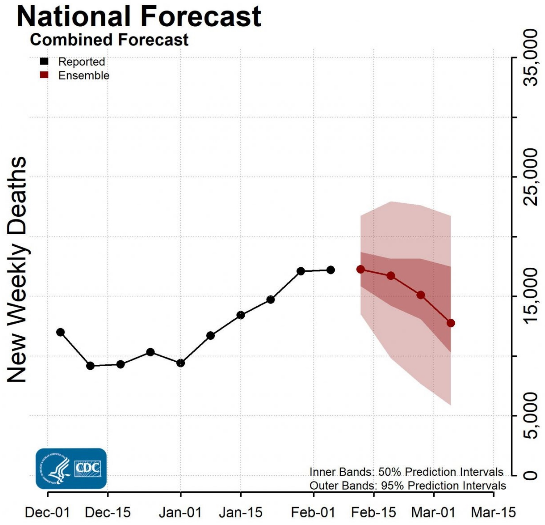
Trevor Hastie

Daniel McDonald

Kenneth Tay

Robert Tibshinari

Motivation



Goal: predict multiple ahead values of a signal.

Multi-period forecaster (MPF)

Notations:

- location i
- time stamp t
- response signal $Y_i(t)$
- feature signals $X_i(t) = \big(X_{i1}(t), \ldots, X_{ip}(t)\big)$
- ahead values $A = \{a_1, \ldots, a_q\} \in \mathbb{R}^q_{\geq 0}$
- lags for k-th feature signal $L_k = \{\ell_{k1}, \ldots, \ell_{km_k}\} \in \mathbb{R}^{m_k}_{>0}$
- set of all $L = \{L_1, \ldots, L_p\}$

Multi-period forecaster (MPF)

Goal: for location i and timestamp t predict the response variable for all the ahead values, i.e.

$$Y_i(t+A) = \big(Y_i(t+a_1), \ldots, Y_i(t+a_q)\big) \in \mathbb{R}^q,$$

using all the lagged features, i.e.

$$X_i(t-L) = \big(X_{i1}(t-L_1), \ldots, X_{ip}(t-L_p)\big) \in \mathbb{R}^m.$$

Here, $m = \sum_{k=1}^p m_k$ and, by analogy with the response,

$$X_{ik}(t-L_k) = \big(X_{ik}(t-\ell_{k1}), \ldots, X_{ik}(t-\ell_{km_k})\big) \in \mathbb{R}^{m_k}.$$

Baseline MPF

Baseline model: for each location i, timestamp t and ahead $a \in A$

$$Y_i(t+a) = \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t-\ell)\ b_{k\ell}(a) + \epsilon_i(t+a).$$

We collect the data for locations $i = 1, \ldots, n$ and timestamps $T = \{t_1, \ldots, t_N\}$ and form the **baseline MPF loss**

$$\sum_{i=1}^{n} \sum_{t \in T} \sum_{a \in A} \left(Y_i(t+a) - \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t-\ell)\ b_{k\ell}(a) \right)^2 .$$

Comment: the loss can be optimized separately for each a.

Baseline MPF

$$\underset{B \in \mathbb{R}^{m \times q}}{\text{minimize}} \|Y - X B^\mathsf{T}\|_F^2 \Longrightarrow \widehat{B}^\mathsf{T} = (X^\mathsf{T} X)^{-1} X^\mathsf{T} Y$$

Smooth MPF

Assume that $b_{k\ell}(a)$ is a *smooth function* of ahead values $\Longrightarrow$

$$b_{k\ell}(a) = \sum_{j=1}^{d} \theta_{jk\ell} h_j(a)$$

for some basis $h_1(a), \ldots, h_d(a)$ and coefficients $\theta_{jk\ell} \in \mathbb{R}$

Smooth MPF loss

$$\sum_{i=1}^{n} \sum_{t \in T} \sum_{a \in A} \left(Y_i(t+a) - \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t-\ell) \sum_{j=1}^{d} \theta_{jk\ell} h_j(a) \right)^2$$

Smooth MPF

Assume that $b_{k\ell}(a)$ is a *smooth function* of ahead values $\Longrightarrow$

$$b_{k\ell}(a) = \sum_{j=1}^{d} \theta_{jk\ell} h_j(a)$$

for some basis $h_1(a), \ldots, h_d(a)$ and coefficients $\theta_{jk\ell} \in \mathbb{R}$

Smooth MPF loss

$$\sum_{i=1}^{n} \sum_{t \in T} \sum_{a \in A} \left(Y_i(t+a) - \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t-\ell) \sum_{j=1}^{d} \theta_{jk\ell} h_j(a) \right)^2$$

Comment: unlike the baseline MPF, the loss cannot be optimized separately for each a.

Smooth MPF

$$\underset{\Theta \in \mathbb{R}^{d \times m}}{\text{minimize}} \| Y - X\Theta^{\mathsf{T}} H^{\mathsf{T}} \|_F^2 \Longrightarrow \widehat{\Theta}^{\mathsf{T}} = (X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}} Y H$$

Weighted smooth MPF

Assume that for each location and timestamp the set of aheads vary
$\Longrightarrow$ we replace A by $A_i(t) \subseteq A$

$$\sum_{i=1}^{n} \sum_{t \in T} \sum_{a \in A_i(t)} \left(Y_i(t+a) - \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t-\ell) \sum_{j=1}^{d} \theta_{jk\ell} h_j(a) \right)^2$$

Example: some counties did not submit the data.

Weighted smooth MPF

Assume that for each location and timestamp the set of aheads vary $\implies$ we replace A by $A_i(t) \subseteq A$

Weighted smooth MPF loss

$$\sum_{i=1}^{n} \sum_{t \in T} \sum_{a \in A} W_i(t+a) \left(Y_i(t+a) - \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t-\ell) \sum_{j=1}^{d} \theta_{jk\ell} h_j(a) \right)^2$$

where $W_i(t+a) = \begin{cases} 1 & \text{if } a \in A_i(t), \\ 0 & \text{otherwise.} \end{cases}$

Weighted smooth MPF

Assume that for each location and timestamp the set of aheads vary
$\Longrightarrow$ we replace A by $A_i(t) \subseteq A$

Weighted smooth MPF loss

$$\sum_{i=1}^{n} \sum_{t \in T} \sum_{a \in A} W_i(t+a) \left(Y_i(t+a) - \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t-\ell) \sum_{j=1}^{d} \theta_{jk\ell} h_j(a) \right)^2$$

Comment: because of the weights pre-multiplication trick does not work anymore.

Weighted smooth MPF

$$\underset{\Theta \in \mathbb{R}^{d \times m}}{\text{minimize}} \| W \circ (Y - X\Theta^{\mathsf{T}} H^{\mathsf{T}}) \|_F^2 \Longrightarrow \underset{\theta \in \mathbb{R}^{dm}}{\text{minimize}} \| w \circ (y - \widetilde{X}\theta) \|_2^2$$

COVIDCast example

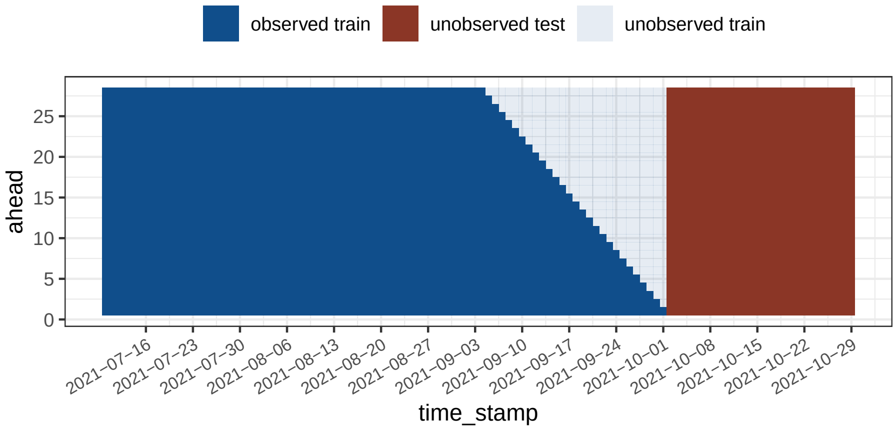
Signals:

- `confirmed_7dav_incidence_prop` – the daily number of new confirmed COVID-19 cases (per 100,000 people);
- `smoothed_cli` – the estimated percentage of people with COVID-like illness;
- `smoothed_hh_cmnty_cli` – the estimated percentage of people reporting illness in their local community.

Model:

- i represents a U.S. county;
- four weeks of aheads $A = \{0, 1, \ldots, 27\}$;
- four weeks of lags $L = \{1, 2, \ldots, 28\}$;

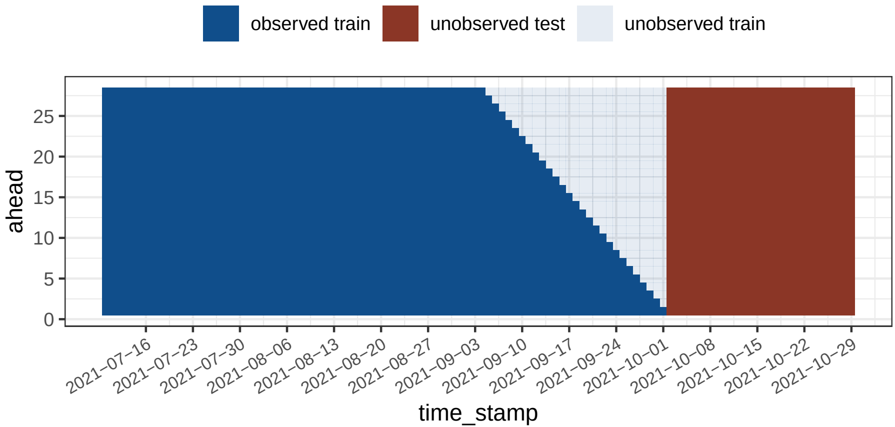
COVIDCast example



Scenario 1: timestamps $T_{train} = \{$10-Jul-2021, ... 4-Sep-2021$\}$
$T_{test} = \{$2-Oct-2021, ..., 29-Oct-2021$\}$

Methods: baseline MPF and smooth MPF.

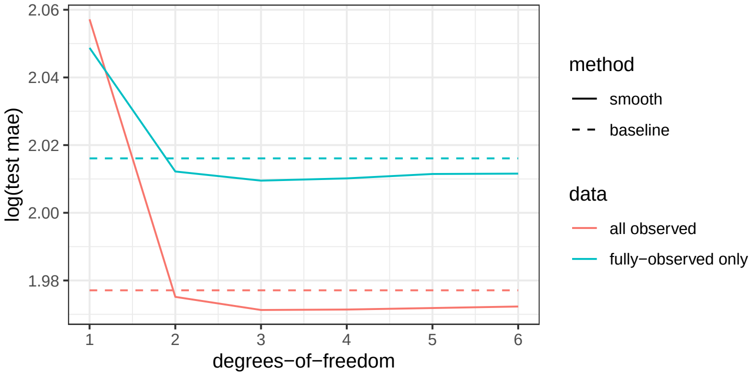
COVIDCast example



Scenario 2: timestamps $T_{train} = \{\text{10-Jul-2021}, \ldots, \text{1-Oct-2021}\}$
$T_{test} = \{\text{2-Oct-2021}, \ldots, \text{29-Oct-2021}\}$

Methods: weighted baseline MPF and weighted smooth MPF.

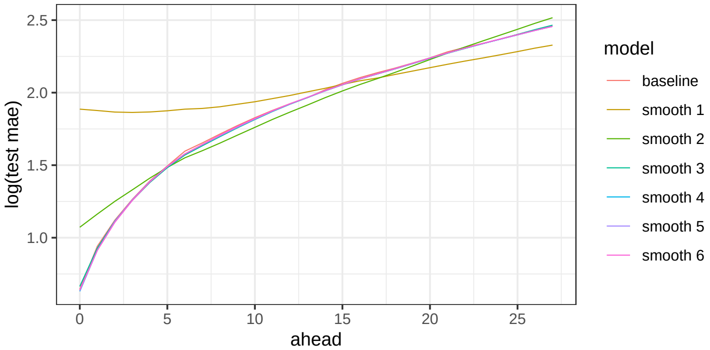
COVIDCast example: point prediction



Interpretation:

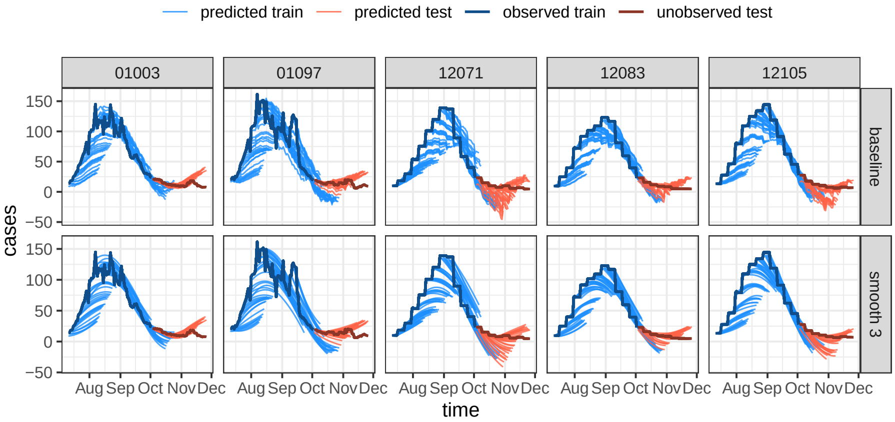
- superior performance of the smooth model;
- weights allow us to include more recent information thereby improving the performance;

COVIDCast example: point prediction



Interpretation: forecasting is more challenging for times which are further in the future.

COVIDCast example: point prediction



Interpretation: the baseline MPF fit demonstrates irregular behavior which is moderated by smoothing.

Quantile multi-period forecaster (QMPF)

For a quantile $\tau \in [0, 1]$ consider **the pinball loss function**

$$\rho_\tau(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}) & \text{if } y \geq \hat{y}, \\ (1 - \tau)(\hat{y} - y) & \text{otherwise.} \end{cases}$$

Quantile MPF loss

$$\sum_{i=1}^{n} \sum_{t \in T} \sum_{a \in A_i(t)} \rho_\tau \Bigg(Y_i(t + a), \; \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t - \ell) \; b_{k\ell}(a) \Bigg)$$

Quantile multi-period forecaster (QMPF)

For a quantile $\tau \in [0, 1]$ consider **the pinball loss function**

$$\rho_\tau(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}) & \text{if } y \geq \hat{y}, \\ (1 - \tau)(\hat{y} - y) & \text{otherwise.} \end{cases}$$

Smooth QMPF loss

$$\sum_{i=1}^{n} \sum_{t \in T} \sum_{a \in A_i(t)} \rho_\tau \left(Y_i(t + a),\ \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t - \ell) \sum_{j=1}^{d} \theta_{jk\ell} h_j(a) \right)$$

Quantile multi-period forecaster (QMPF)

For a quantile $\tau \in [0, 1]$ consider **the pinball loss function**

$$\rho_\tau(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}) & \text{if } y \geq \hat{y}, \\ (1 - \tau)(\hat{y} - y) & \text{otherwise.} \end{cases}$$

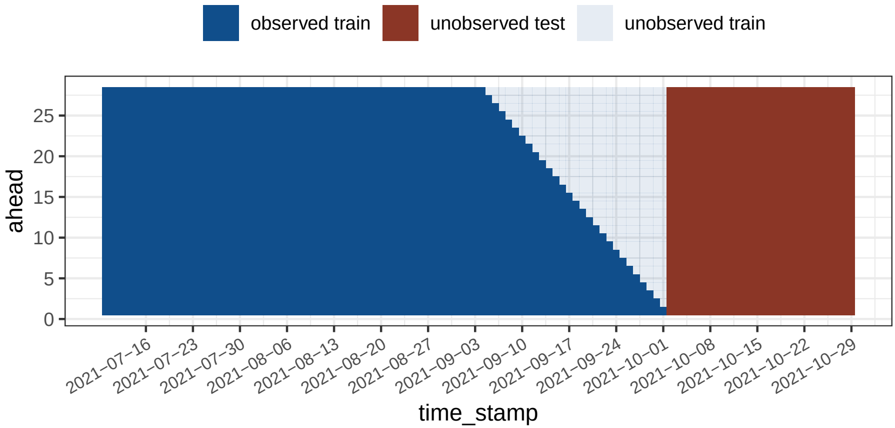
Weighted smooth QMPF loss

$$\sum_{i=1}^{n} \sum_{t \in T} \sum_{a \in A} W_i(t + a) \cdot \rho_\tau\left(Y_i(t + a),\ \sum_{k=1}^{p} \sum_{\ell \in L_k} X_{ik}(t - \ell) \sum_{j=1}^{d} \theta_{jk\ell} h_j(a) \right)$$

Weighted smooth quantile MPF

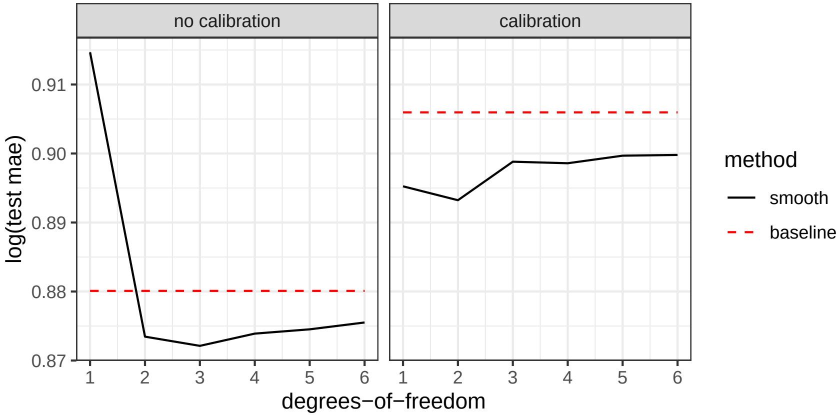
$$\underset{\theta \in \mathbb{R}^{dm}}{\text{minimize}} \sum_{i=1}^{Nnq} w_i \cdot \rho_\tau\big(y_i,\ \widetilde{X}_i^\mathsf{T}\theta\big).$$

Calibration



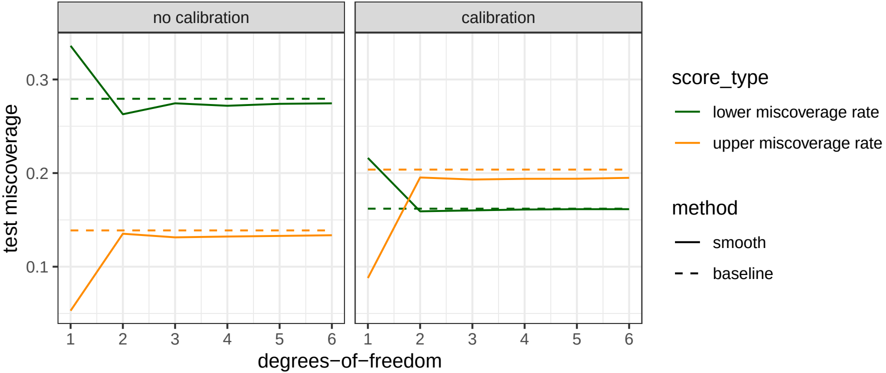
$$T_{\text{train}}^{\text{fit}} = \{\text{10-Jul-2021}, \ldots, \text{3-Sep-2021}\}$$
$$T_{\text{train}}^{\text{cal}} = \{\text{4-Sep-2021}, \ldots, \text{1-Oct-2021}\}$$
$$T_{test} = \{\text{2-Oct-2021}, \ldots, \text{29-Oct-2021}\}$$

COVIDCast example: interval prediction



Interpretation: smoothing improves the performance.

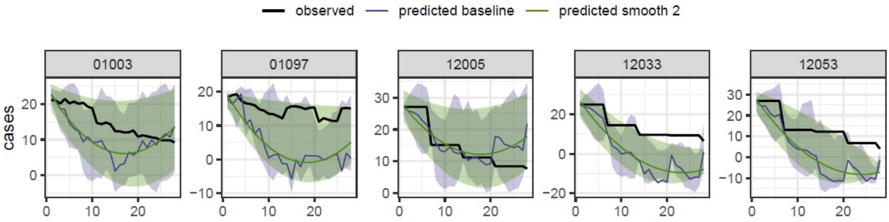
COVIDCast example: interval prediction



Interpretation:

- calibration improves coverage;
- improved performance of the smooth model relative to the baseline.

COVIDCast example: interval prediction



Interpretation: SQMPF results in a smooth trajectory and smooth bounds of prediction interval.

Thank you for your attention!