# Empirical Bayes Method for Short-Term Time Series Forecasting


Munther Dahleh [1 2 3], Yury Polyanskiy [1 2 3], Anzo Teh [1 2]

[1]EECS [2]LIDS [3]IDSS


## Prediction Task

Given observations of $n$ distributionally similar time series $X_{it}$, $1 \le i \le n$, $1 \le t \le k$, predict $X_{i(k+1)}$ for each $i = 1, \cdots, n$ that minimizes the following L1 loss:

$$\sum_{i=1}^{n} \mathbb{E}_{X_{i(k+1)}|X_{i1},\cdots,X_{ik}}[|\hat{X}_{i(k+1)} - X_{i(k+1)}|] \tag{1}$$

### Modeling of Data

We model each $X_{it} \sim \text{Lap}(\theta_i, 1)$, with $\theta_i \overset{\text{iid}}{\sim} \pi$.

$$p(X_{it}|\theta_i) = \frac{1}{2}\exp(-|X_{it} - \theta_i|) \qquad p_\pi(X_{it}) = \int_{\theta_i} p(X_{it}|\theta_i) d\pi(\theta_i)$$

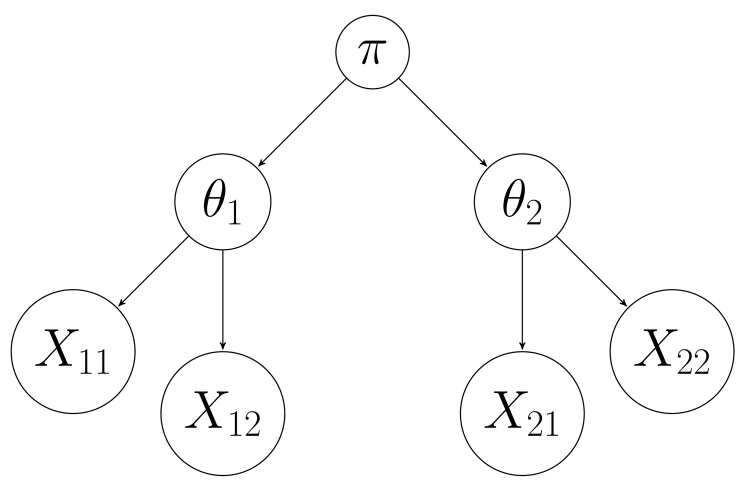


Figure 1. Graphical Model When $n = k = 2$

### Baseline

All $\theta_i$'s independent (no $\pi$) $\Rightarrow$ predict MLE of Laplace / sample median:

$$X_{i(k+1)} = \text{Median}(X_{i1}, \cdots, X_{ik})$$

## Related Work

### Parametric Methods

[1] (A2-CLPM), [2] (follow up on convergence), [5] (Gaussian-Laplace mixture).

- Restricted to limited number of components.

### Non-Parametric ML Estimators

General NPMLE: [3] (self-consistency property).

Gaussian mixture: [4] (statistical degree of NPMLE in $\Theta(\log n)$ for subgaussian $\pi$).

Laplace mixture: [6] (NPMLE is directly solvable for sparse observations).

## Non-Parametric Maximum Likelihood Estimator

We estimate $\pi$ by maximizing the following likelihood function:

$$\hat{\pi} := \arg\max_{\pi} = \sum_{i=1}^{n} \log \left( \int_{\theta_i} \prod_{t=1}^{k} p_\pi(X_{it}|\theta_i) d\pi(\theta_i) \right) \tag{2}$$

We establish the following theorem, extended from [6].

**Theorem.** Under mild conditions, $\hat{\pi}$ is supported only on the observations $X_{it}$'s.

- (2) reduced to a finite-dimensional problem;
- solve iteratively via convex optimization (e.g. the EM algorithm).

## Empirical Bayes Predictor

Take partial derivatives with respect to (1) $\Rightarrow$ L1 loss minimized when $\hat{X}_{i(k+1)}$ satisfy:

$$\mathbb{P}_{X_{i(k+1)}|X_{i1},\cdots,X_{ik}}[X_{i(k+1)} < \hat{X}_{i(k+1)}] = \mathbb{P}_{X_{i(k+1)}|X_{i1},\cdots,X_{ik}}[X_{i(k+1)} > \hat{X}_{i(k+1)}] \tag{3}$$

I.e. $\hat{X}_{i(k+1)}$ is the posterior median given observations $X_{i1}, \cdots, X_{ik}$ based on estimated prior $\hat{\pi}$.

Posterior density of $X_{i(k+1)}$:

$$\begin{aligned} p(X_{i(k+1)}|X_{i1}, \cdots, X_{ik}) &= \frac{\int_{\theta_i} p(X_{i(k+1)}|\theta_i) \prod_{t=1}^{k} p(X_{it}|\theta_i) d\hat{\pi}(\theta_i)}{\int_{\theta_i} \prod_{t=1}^{k} p(X_{it}|\theta_i) d\hat{\pi}(\theta_i)} \\ &= \frac{1}{2} \cdot \frac{\int_{\theta_i} \exp(-\sum_{t=1}^{k+1} |X_{it} - \theta_i|) d\hat{\pi}(\theta_i)}{\int_{\theta_i} \exp(-\sum_{t=1}^{k} |X_{it} - \theta_i|) d\hat{\pi}(\theta_i)} \end{aligned}$$

Hence (3) can also be expressed as

$$\begin{aligned} &\int_{-\infty}^{\hat{X}_{i(k+1)}} \int_{\theta_i} \exp\left(-\sum_{t=1}^{k+1} |X_{it} - \theta_i|\right) d\hat{\pi}(\theta_i) dX_{i(k+1)} \\ = &\int_{\hat{X}_{i(k+1)}}^{\infty} \int_{\theta_i} \exp\left(-\sum_{t=1}^{k+1} |X_{it} - \theta_i|\right) d\hat{\pi}(\theta_i) dX_{i(k+1)} \end{aligned}$$

## Experimental Setup and Results

### Currency Exchange Rates Against Euro

1. Data range: 1999 - 2022 (daily).
2. Number of currency, $n$: $\simeq 32$ for each time window.
3. Window size: $k = 5$.
4. Prediction interest: log returns.

### S&P 500 Stock Returns

1. Data range: 2017 - 2019 (daily).
2. Number of stocks, $n$: $\simeq 500$ for each time window.
3. Window size: $k = 5$.
4. Prediction interest: log returns (using adjusted closing for each day).

### Evaluation Procedure

1. Partition dataset into sliding windows, length $k$ each;
2. Get MAE of running sample median vs NPMLE on each time window;
3. Report 95% confidence interval of MAEs across all time windows.

### Results

| Datasets | Median | NPMLE + EB |
|---|---|---|
| Exchange Rates | $(4.309 \pm 0.059) \times 10^{-3}$ | $\mathbf{(3.927 \pm 0.057) \times 10^{-3}}$ |
| S&P 500 | $(1.231 \pm 0.045) \times 10^{-2}$ | $\mathbf{(1.122 \pm 0.046) \times 10^{-2}}$ |

Table 1. The 95% Confidence Interval of MAE Achieved Across All Time Windows

## Next Steps / Ongoing

1. Incorporate other related information for each time series to improve performance.
2. Establish statistical bounds on error bounds of Laplace mixture.

## Experimental Plots (MAE)

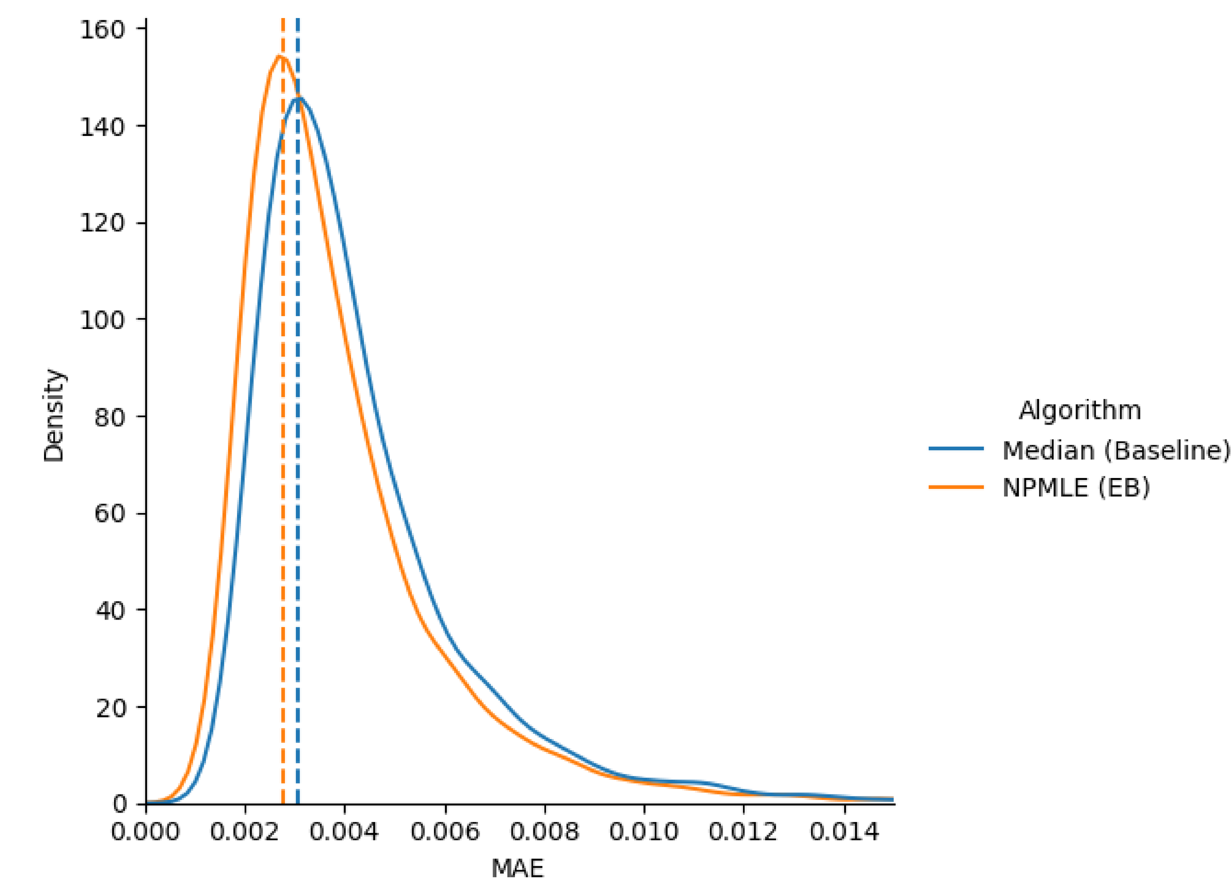


Figure 2. MAE Distribution on Exchange Rates Dataset

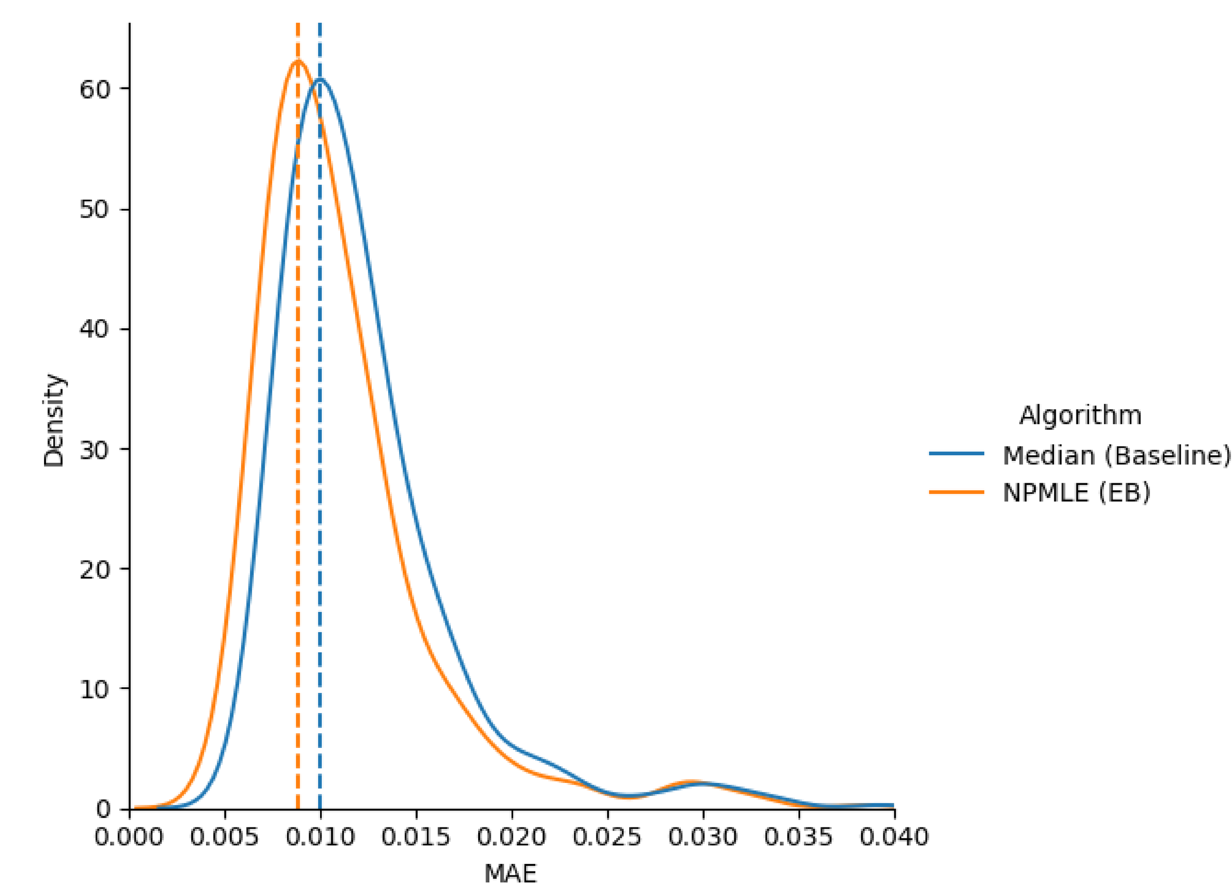


Figure 3. MAE Distribution on S&P 500 Dataset

## References


[1] Zakiah I. Kalantan and Faten Alrewely.
A 2-Component Laplace Mixture Model: Properties and Parametric Estimations.
*Mathematics and Statistics*, 7(4A):9–16, September 2019.
Publisher: Horizon Research Publishing.

[2] Zakiah I. Kalantan, Faten Alrewely, and Kamarul Ariffin Mansor.
Study on the Convergence Behavior of Expectation Maximization Algorithm for Laplace Mixture Model.
3:35–45, September 2019.

[3] Nan Laird.
Nonparametric Maximum Likelihood Estimation of a Mixing Distribution.
*Journal of the American Statistical Association*, 73(364):805–811, 1978.
Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[4] Yury Polyanskiy and Yihong Wu.
Self-regularizing Property of Nonparametric Maximum Likelihood Estimator in Mixture Models.
*arXiv:2008.08244 [math, stat]*, September 2020.
arXiv: 2008.08244.

[5] A. J. Scallan.
Maximum Likelihood Estimation for a Normal/Laplace Mixture Distribution.
*Journal of the Royal Statistical Society. Series D (The Statistician)*, 41(2):227–231, 1992.
Publisher: [Royal Statistical Society, Wiley].

[6] Gilbert G. Walter and Julius R. Blum.
A Simple Solution to a Nonparametric Maximum Likelihood Estimation Problem.
*The Annals of Statistics*, 12(1):372–379, March 1984.
Publisher: Institute of Mathematical Statistics.





anzoteh@mit.edu