Sequential Quantile Prediction of Time Series.

Joined work with Gérard Biau.

Benoît Patra, benoit.patra@lokad.com

JDS BORDEAUX, MAY 2009.

Introduction.

Time series prediction.

- Time series prediction has a long history (`Yule, 1927`).
- Parametric approaches (Until 70's).
- Recently non parametric approach.

Quantile forecasting.

Given a stochastic process $Y_1, Y_2, \ldots$.

- Usually, estimate the conditional mean of Y_n given $Y_1, \ldots, Y_{n-1}$.
- Here: the conditional τth quantile of Y_n given $Y_1, \ldots, Y_{n-1}$.

Introduction.

Time series prediction.

- Time series prediction has a long history (`Yule, 1927`).
- Parametric approaches (Until 70's).
- Recently non parametric approach.

Quantile forecasting.

Given a stochastic process $Y_1, Y_2, \ldots$.

- Usually, estimate the conditional mean of Y_n given $Y_1, \ldots, Y_{n-1}$.
- Here: the conditional τth quantile of Y_n given $Y_1, \ldots, Y_{n-1}$.

Introduction.

What for?

- Understand conditional distributions.
- $\tau = 0.5$ robust forecasting.
- Build confidence interval.

Applications fields.

- Finance: CVAR. Also biology, medecine, telecoms...
- Here: call volumes (optimize staff in a call center).

Introduction.

What for?

- Understand conditional distributions.
- $\tau = 0.5$ robust forecasting.
- Build confidence interval.

Applications fields.

- Finance: CVAR. Also biology, medecine, telecoms...
- Here: call volumes (optimize staff in a call center).

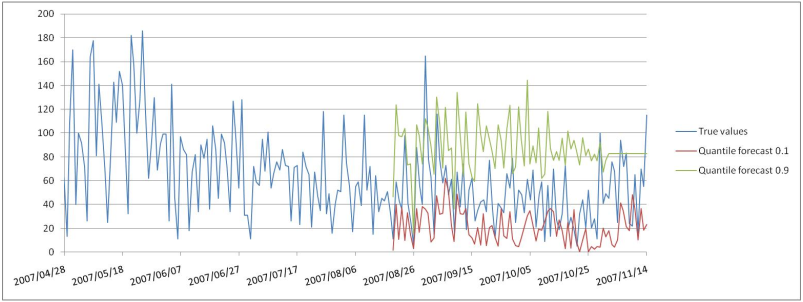


Figure: Quantile forecast with $\tau = 0.1, 0.9$.

Quantile Regression.

Conditional quantiles.

X multivariate random variable, Y real valued random variable,

$$q_\tau(X) \triangleq F^{\leftarrow}_{Y|X}(\tau) = \inf\{t \in \mathbb{R} : F_{Y|X}(t) \geq \tau\}.$$

$F_{Y|X}$ conditional cumulative distribution function.

Proposition (Koenker, 2005)

$$q_\tau(X) \in \operatorname*{argmin}_{q(.)\in\mathbb{R}} \mathbb{E}_{\mathbb{P}_{Y|X}} \left[\rho_\tau\left(Y - q(X)\right)\right].$$

Quantile Regression.

Conditional quantiles.

X multivariate random variable, Y real valued random variable,

$$q_\tau(X) \triangleq F^{\leftarrow}_{Y|X}(\tau) = \inf\{t \in \mathbb{R} : F_{Y|X}(t) \geq \tau\}.$$

$F_{Y|X}$ conditional cumulative distribution function.

Proposition (Koenker, 2005)

$$q_\tau(X) \in \underset{q(.)\in\mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}_{\mathbb{P}_{Y|X}} \left[\rho_\tau \left(Y - q(X)\right)\right].$$

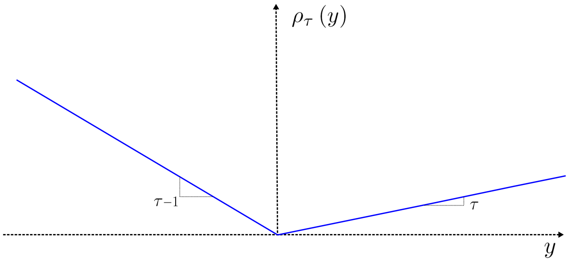


Figure: Pinball function graph.

Non parametric framework.

On line.

- Consider the sequential (= on-line) quantile prediction of time series.
- Including series that do not necessarily satisfy classical statistical assumptions for bounded, mixing or Markovian process.

Goal.

- Show consistency results under a minimum of hypotheses.

Notation.

- $y_1^n = (y_1, \ldots, y_n)$ real sequence.
- $Y_1^n = (Y_1, \ldots, Y_n)$ random variables sequence.

Non parametric framework.

On line.

- Consider the sequential (= on-line) quantile prediction of time series.
- Including series that do not necessarily satisfy classical statistical assumptions for bounded, mixing or Markovian process.

Goal.

- Show consistency results under a minimum of hypotheses.

Notation.

- $y_1^n = (y_1, \ldots, y_n)$ real sequence.
- $Y_1^n = (Y_1, \ldots, Y_n)$ random variables sequence.

Non parametric framework.

On line.

- Consider the sequential (= on-line) quantile prediction of time series.
- Including series that do not necessarily satisfy classical statistical assumptions for bounded, mixing or Markovian process.

Goal.

- Show consistency results under a minimum of hypotheses.

Notation.

- $y_1^n = (y_1, \ldots, y_n)$ real sequence.
- $Y_1^n = (Y_1, \ldots, Y_n)$ random variables sequence.

Non parametric framework.

On line.

- Consider the sequential (= on-line) quantile prediction of time series.
- Including series that do not necessarily satisfy classical statistical assumptions for bounded, mixing or Markovian process.

Goal.

- Show consistency results under a minimum of hypotheses.

Notation.

- $y_1^n = (y_1, \ldots, y_n)$ real sequence.
- $Y_1^n = (Y_1, \ldots, Y_n)$ random variables sequence.

Non parametric framework.

Framework.

- Here, we observe a string realization y_1^{n-1} of a stationary and ergodic process $\{Y_n\}_{-\infty}^{\infty}\dots$
- $\dots$ and try to estimate $q_\tau(y_1^{n-1}) = F^{\leftarrow}_{Y_n|Y_1^{n-1}=y_1^{n-1}}(\tau)$, the conditional quantile at time n.

Strategy.

Sequence $g = \{g_n\}_{n=1}^{\infty}$ of τth quantile forecasting functions

$$g_n : \mathbb{R}^{n-1} \longrightarrow \mathbb{R}.$$

Quantile estimation at time n is $g_n(y_1^{n-1})$.

Non parametric framework.

Framework.

- Here, we observe a string realization y_1^{n-1} of a stationary and ergodic process $\{Y_n\}_{-\infty}^{\infty}\ldots$
- $\ldots$ and try to estimate $q_\tau(y_1^{n-1}) = F^{\leftarrow}_{Y_n|Y_1^{n-1}=y_1^{n-1}}(\tau)$, the conditional quantile at time n.

Strategy.

Sequence $g = \{g_n\}_{n=1}^{\infty}$ of τth quantile forecasting functions

$$g_n : \mathbb{R}^{n-1} \longrightarrow \mathbb{R}.$$

Quantile estimation at time n is $g_n(y_1^{n-1})$.

Errors.

Empirical measure criterion.

At time n the cumulative pinball error of the strategy g is

$$L_n(g) = \frac{1}{n}\sum_{t=1}^{n} \rho_\tau \left(y_t - g_t(y_1^{t-1}) \right).$$

A fundamental limit (Algoët, 1994).

For any stationary and ergodic process $\{Y_n\}_{n=-\infty}^{+\infty}$,

$$\liminf_{n\to\infty} L_n(g) \geq L^\star \quad \text{a.s.},$$

where

$$L^\star = \mathbb{E}\left[\min_{q(.)} \mathbb{E}_{\mathbb{P}_{Y_0|Y_{-\infty}^{-1}}}\left[\rho_\tau\left(Y_0 - q(Y_{-\infty}^{-1})\right)\right]\right].$$

Errors.

Empirical measure criterion.

At time n the cumulative pinball error of the strategy g is

$$L_n(g) = \frac{1}{n}\sum_{t=1}^{n} \rho_\tau \left(y_t - g_t(y_1^{t-1}) \right).$$

A fundamental limit (Algoët, 1994).

For any stationary and ergodic process $\{Y_n\}_{n=-\infty}^{+\infty}$,

$$\liminf_{n\to\infty} L_n(g) \geq L^\star \quad \text{a.s.},$$

where

$$L^\star = \mathbb{E}\left[\min_{q(.)} \mathbb{E}_{\mathbb{P}_{Y_0|Y_{-\infty}^{-1}}} \left[\rho_\tau \left(Y_0 - q(Y_{-\infty}^{-1}) \right)\right]\right].$$

A NN based aggregation scheme.

On line learning.

Scheme inspired from prediction of individual sequences.
Reference : Cesa-Bianchi, N. and Lugosi, G. Prediction, Learning, and Games, Cambridge University Press, New York, 2006.

Previous works.

- Sequential prediction of bounded time series. Györfi, Lugosi, 2001.
- ... portfolio optimization. Györfi, Lugosi, 2006.
- ... prediction of time series (fourth moment). Biau, Beakley, Györfi, Ottucsàk, 2009.

A NN based aggregation scheme.

On line learning.

Scheme inspired from prediction of individual sequences.
Reference : Cesa-Bianchi, N. and Lugosi, G. Prediction, Learning, and Games, Cambridge University Press, New York, 2006.

Previous works.

- Sequential prediction of bounded time series. Györfi, Lugosi, 2001.
- ... portfolio optimization. Györfi, Lugosi, 2006.
- ... prediction of time series (fourth moment). Biau, Beakley, Györfi, Ottucsàk, 2009.

A NN based aggregation scheme.

On line learning.

Scheme inspired from prediction of individual sequences.
Reference : Cesa-Bianchi, N. and Lugosi, G. Prediction, Learning, and Games, Cambridge University Press, New York, 2006.

Previous works.

- Sequential prediction of bounded time series. Györfi, Lugosi, 2001.
- ... portfolio optimization. Györfi, Lugosi, 2006.
- ... prediction of time series (fourth moment). Biau, Beakley, Györfi, Ottucsàk, 2009.

Nearest neighbors strategy.

Elementary predictors.

- Define infinite array of experts $h_n^{(k,\ell)}$: $k, \ell = 1, 2, \dots$

Each expert has a job.

- At time n, expert $h_n^{(k,\ell)}$ searches for the $\bar{\ell}$ nearest neighbors of length k.

Nearest neighbors strategy.

Elementary predictors.

- Define infinite array of experts $h_n^{(k,\ell)}$: $k, \ell = 1, 2, \ldots$

Each expert has a job.

- At time n, expert $h_n^{(k,\ell)}$ searches for the $\bar{\ell}$ nearest neighbors of length k.

1,34	1,78	2,56	1,88	0,57	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31

2,27	2,89	2,12	1,78	2,67	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	0,99

0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73

?

Figure: Work of fundamental expert with $k = 3$ and $\bar{\ell} = 4$.

1,34	1,78	2,56	1,88	0,57	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31

2,27	2,89	2,12	1,78	2,67	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	0,99

0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**

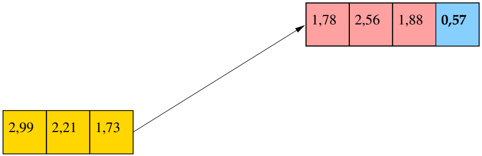
2,99	2,21	1,73	**?**

1,34	1,78	2,56	1,88	**0,57**	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31

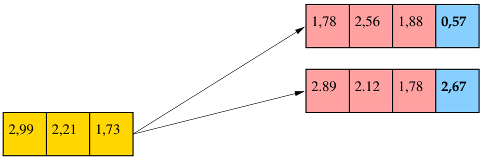
2,27	2,89	2,12	1,78	2,67	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	0,99

0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**

2,99	2,21	1,73	**?**



1,34	1,78	2,56	1,88	**0,57**	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31
2,27	2,89	2,12	1,78	**2,67**	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	0,99
0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**
																2,99	2,21	1,73	**?**

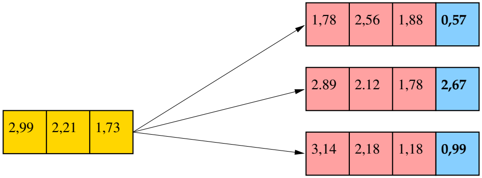


1,34	1,78	2,56	1,88	**0,57**	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31

2,27	2,89	2,12	1,78	**2,67**	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	**0,99**

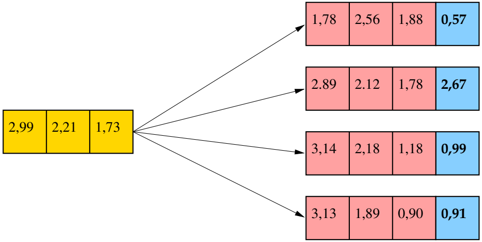
0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**

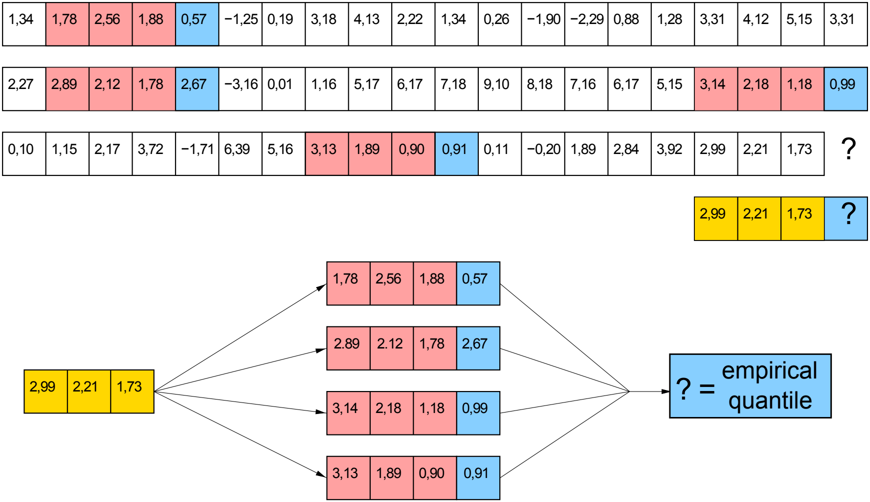
2,99	2,21	1,73	**?**



1,34	1,78	2,56	1,88	**0,57**	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31
2,27	2,89	2,12	1,78	**2,67**	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	**0,99**
0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	**0,91**	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**

2,99	2,21	1,73	**?**




1,34 1,78 2,56 1,88 0,57 −1,25 0,19 3,18 4,13 2,22 1,34 0,26 −1,90 −2,29 0,88 1,28 3,31 4,12 5,15 3,31
2,27 2,89 2,12 1,78 2,67 −3,16 0,01 1,16 5,17 6,17 7,18 9,10 8,18 7,16 6,17 5,15 3,14 2,18 1,18 0,99
0,10 1,15 2,17 3,72 −1,71 6,39 5,16 3,13 1,89 0,90 0,91 0,11 −0,20 1,89 2,84 3,92 2,99 2,21 1,73 ?
2,99 2,21 1,73 ?
1,78 2,56 1,88 0,57
2.89 2.12 1,78 2,67
2,99 2,21 1,73
3,14 2,18 1,18 0,99
3,13 1,89 0,90 0,91
? = empirical quantile

Prediction and Aggregation.

Prediction of one expert.

$$h_n^{(k,\ell)}(y_1^{n-1}) = \operatorname*{argmin}_{q\in\mathbb{R}} \sum_{\{t\in J_n^{(k,\ell)}\}} \rho_\tau(y_t - q).$$

[Can be easily computed by sorting the sample.]

Aggregated prediction of all experts.

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Where do the $p_{k,\ell,n}$ come from?

Exponentially weight the experts based on their past performance.

Prediction and Aggregation.

Prediction of one expert.

$$h_n^{(k,\ell)}(y_1^{n-1}) = \operatorname*{argmin}_{q\in\mathbb{R}} \sum_{\{t\in J_n^{(k,\ell)}\}} \rho_\tau(y_t - q).$$

[Can be easily computed by sorting the sample.]

Aggregated prediction of all experts.

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Where do the $p_{k,\ell,n}$ come from?

Exponentially weight the experts based on their past performance.

Prediction and Aggregation.

Prediction of one expert.

$$h_n^{(k,\ell)}(y_1^{n-1}) = \underset{q \in \mathbb{R}}{\operatorname{argmin}} \sum_{\{t \in J_n^{(k,\ell)}\}} \rho_\tau(y_t - q).$$

[Can be easily computed by sorting the sample.]

Aggregated prediction of all experts.

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Where do the $p_{k,\ell,n}$ come from?

Exponentially weight the experts based on their past performance.

Aggregation continued. . .

Definitions.

- Let $\{q_{k,\ell}\}$ be a probability distribution over all pairs (k, ℓ) of positive integers such that $q_{k,\ell} > 0$ for all (k, ℓ).
- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h_n^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction.

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Aggregation continued. . .

Definitions.

- Let $\{q_{k,\ell}\}$ be a probability distribution over all pairs (k,ℓ) of positive integers such that $q_{k,\ell} > 0$ for all (k,ℓ).
- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h_n^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction.

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Aggregation continued. . .

Definitions.

- Let $\{q_{k,\ell}\}$ be a probability distribution over all pairs (k,ℓ) of positive integers such that $q_{k,\ell} > 0$ for all (k,ℓ).
- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h_n^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction.

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Aggregation continued. . .

Definitions.

- Let $\{q_{k,\ell}\}$ be a probability distribution over all pairs (k, ℓ) of positive integers such that $q_{k,\ell} > 0$ for all (k, ℓ).
- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h_n^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction.

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Theoretical Results.

Theorem

- *Let $\mathcal{C}$ be the class of all jointly stationary and ergodic processes $\{Y_n)\}_{-\infty}^{\infty}$ such that $\mathbb{E}\{Y_0^2\} < \infty$ and $F_{Y_0|Y_{-\infty}^{-1}}$ is a.s. increasing.*
- *Then the nearest neighbor quantile forecasting strategy is universally consistent with respect to the class $\mathcal{C}$, that is, for all process $Y \in \mathcal{C}$*

$$\lim_{n\to\infty} L_n(g) = L^{\star} \quad \text{almost surely.}$$

Theoretical Results.

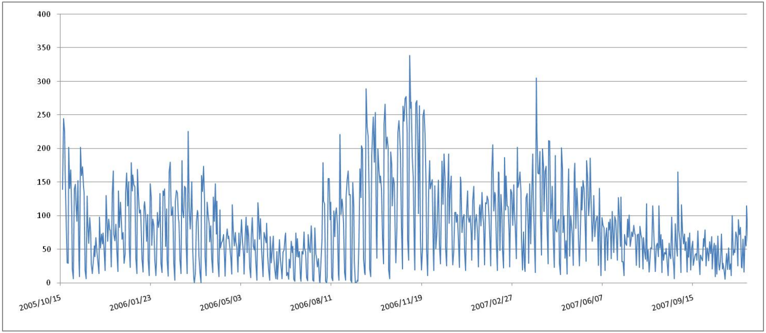
Theorem

- *Let $\mathcal{C}$ be the class of all jointly stationary and ergodic processes $\{Y_n)\}_{-\infty}^{\infty}$ such that $\mathbb{E}\{Y_0^2\} < \infty$ and $F_{Y_0|Y_{-\infty}^{-1}}$ is a.s. increasing.*
- *Then the nearest neighbor quantile forecasting strategy is universally consistent with respect to the class $\mathcal{C}$, that is, for all process $Y \in \mathcal{C}$*

$$\lim_{n\to\infty} L_n(g) = L^\star \quad \textit{almost surely.}$$

Call center data set.

- Daily call volumes entering a call center.
- Long series between 382 and 826 time values. 21 series.



Future outcome predictions results.

$\tau = 0.5$ median base forecaster : robustness.

Model Name	Avg Abs Error	Avg Sqr Error	Mape (%)
$AR(7)$	65.80	9738	31.6
$QAR(8)_{0.5}$	57.8	9594	24.9
DayOfTheWeekMean	53.95	7099	22.8
HoltWinters	49.84	6025	**21.5**
$\underline{QuantileExpertMixture_{0.5}}$	**48.1**	**5731**	21.6
MeanExpertMixture	52.37	6536	22.3
MA	179	62448	0.52

Figure: Forecasting future outcomes.

Quantile forecasting.

Model Name	PinBall Loss (0.1)	Ramp Loss
$QuantileExpertMixture_{0.1}$	13.71	0.80
$QAR(7)_{0.1}$	**13.22**	**0.88**

Figure: $\tau = 0.1$

Model Name	PinBall Loss (0.9)	Ramp Loss
$QuantileExpertMixture_{0.9}$	**12.27**	**0.07**
$QAR(7)_{0.9}$	19.31	**0.07**

Figure: $\tau = 0.9$

Quantile forecasting.

Model Name	PinBall Loss (0.1)	Ramp Loss
$QuantileExpertMixture_{0.1}$	13.71	0.80
$QAR(7)_{0.1}$	**13.22**	**0.88**

Figure: $\tau = 0.1$

Model Name	PinBall Loss (0.9)	Ramp Loss
$QuantileExpertMixture_{0.9}$	**12.27**	**0.07**
$QAR(7)_{0.9}$	19.31	**0.07**

Figure: $\tau = 0.9$

Questions?