# Mean, Median, Mode, More

Tilmann Gneiting

University of Washington

# Mean, Median, Mode, More

or

# Quantiles as Optimal Point Predictors

Tilmann Gneiting

University of Washington

# Optimal point prediction

## Point prediction problem

Consider the following **statistical decision problem**:

- you are supposed to predict a critical **real-valued** future **quantity of interest**, $Y$, with **verifying observation**, $y$
- after much dedicated and very hard work, your favorite statistical technique provides a conditional **predictive distribution**, in the form of a **density forecast**, $f$, or a **cumulative distribution function**, $F$
- only now you realize that you are required to issue a **point forecast**, $x$, rather than a **predictive distribution**

a common scenario, for reasons of **communication** (`probcast.washington.edu`, . . . ), **trading mechanisms** (energy markets, . . . ) or legal **reporting** requirements, among others

*Now, what are you going to do?*

## Loss function

you remember the vintage **decision theory** stuff you hated to study in graduate school, and pick the **optimal point predictor** or **Bayes rule**, which, as you recall, depends on the **loss function**

generally, the **loss** $L(x, y)$ is a function of both the **point predictor**, $x$, and the **verifying observation**, $y$

the **loss** may or may not depend on the **prediction error**, $x - y$, only, and may or may not be **symmetric** in $x$ and $y$

subsequent **assumptions** on the **loss function** $L(x, y)$ include

- (L0) $L(x, y) \geq 0$ with equality if $x = y$
- (L1) $L$ is continuous
- (L2a) the partial derivative $L_y(x, y)$ exists and is continuous
- (L2b) the partial derivative $L_x(x, y)$ exists if $x \neq y$

## Optimal point predictor

given a **loss function** $L$ and a **predictive distribution** $F$, the **optimal point predictor** or **Bayes rule** is the number

$$\hat{x} = \arg\min_x E_F \, L(x, Y)$$

*Any examples?*

## Optimal point predictor

given a **loss function** $L$ and a **predictive distribution** $F$, the **optimal point predictor** or **Bayes rule** is the number

$$\hat{x} = \arg\min_x E_F\, L(x, Y)$$

*Any examples?*

- if the **loss function** is **quadratic**

  $$L(x, y) = (x - y)^2$$

  the **optimal point predictor** is the **mean** of $F$

# Optimal point predictor

given a **loss function** $L$ and a **predictive distribution** $F$, the **optimal point predictor** or **Bayes rule** is the number

$$\hat{x} = \arg\min_x E_F\, L(x, Y)$$

*Any examples?*

- if the **loss function** is **quadratic**

  $$L(x, y) = (x - y)^2$$

  the **optimal point predictor** is the **mean** of $F$

- if the **loss function** is **linear**

  $$L(x, y) = |x - y|$$

  the **optimal point predictor** is any **median** of $F$

## Optimal point predictor

given a **loss function** $L$ and a **predictive distribution** $F$, the **optimal point predictor** or **Bayes rule** is the number

$$\hat{x} = \arg\min_x E_F \, L(x, Y)$$

*Any examples?*

- if the **loss function** is **quadratic**

$$L(x, y) = (x - y)^2$$

  the **optimal point predictor** is the **mean** of $F$

- if the **loss function** is **linear**

$$L(x, y) = |x - y|$$

  the **optimal point predictor** is any **median** of $F$

- if the **loss function** is **piecewise linear**

$$L(x, y) = \begin{cases} \alpha \, |x - y| & \text{if} \quad x \leq y \\ (1 - \alpha) \, |x - y| & \text{if} \quad x \geq y \end{cases}$$

  the **optimal point predictor** is any **$\alpha$-quantile** of $F$

## Some relevant decision theoretic questions

*Can the usual loss functions (quadratic, linear, piecewise linear) be considered realistic?*

*What type of loss function is realistic for a broad range of applied problems?*

## Let's think utility

- we are supposed to provide a **point forecast**, $x$, for a **future quantity** with **verifying observation**, $y$

  example: wind speed at a wind energy site

- **economic utility** can be measured in terms of a potentially nonlinear, **nondecreasing** function $g(y)$ of $y$

  example: the wind power generated at a wind farm is a nonlinear function $g(y)$, say

  $$g(y) = [\max(y, y_0)]^3 = \begin{cases} y^3 & \text{if} \quad 0 \leq y \leq y_0 \\ y_0^3 & \text{if} \quad y \geq y_0 \end{cases}$$

  of the wind speed, $y$

- in the case of an **underprediction**, $x \leq y$, the **loss** is **proportional to the difference** $g(y) - g(x)$

  example: energy surplus $g(y) - g(x)$ can only be sold at a loss

- in the case of an **overprediction**, $x \geq y$, the **loss** is **proportional to the difference** $g(x) - g(y)$, with a **potentially distinct** proportionality constant

  example: energy deficit $g(x) - g(y)$ needs to be aquired on the spot market

## A class of economically relevant loss functions

### Definition

A **loss function** $L(x,y)$ is **generalized piecewise linear (GPL) of order $\alpha$**, where $0 < \alpha < 1$, if there exists a nondecreasing function $g$ such that

$$L(x,y) = \begin{cases} \alpha\,(g(y) - g(x)) & \text{if} \quad x \leq y \\ (1-\alpha)\,(g(x) - g(y)) & \text{if} \quad x \geq y \end{cases}$$

### Some properties of note

- the **GPL family** is the class we have been talking about
- the **utility ratio** $\alpha/(1-\alpha)$ characterizes the possibly **distinct costs** of **under-** and **overpredicting**
- if $g$ is the identity function, we recover the **piecewise linear**, **linlin**, **hinge**, **tick** or **pinball** loss function

## Some decision theoretic questions

*Are the usual predictors, such as the mean or quantiles, optimal under realistic loss functions?*

## Some decision theoretic questions

*Are the usual predictors, such as the mean or quantiles, optimal under realistic loss functions?*

We are led to the following **decision theoretic problem:**

Give sufficient and/or necessary **conditions** on a **loss function** $L$ to be **consistent with** the **mean** (or any **quantile**), in the sense that the mean (or any quantile) is an optimal point predictor

## Some decision theoretic questions

*Are the usual predictors, such as the mean or quantiles, optimal under realistic loss functions?*

We are led to the following **decision theoretic problem:**

Give sufficient and/or necessary **conditions** on a **loss function** $L$ to be **consistent with** the **mean** (or any **quantile**), in the sense that the mean (or any quantile) is an optimal point predictor, whatever the predictive distribution

*Are there any loss functions other than the quadratic that are consistent with the mean?*

# Mean as optimal point predictor

## Definition

A **loss function** $L(x, y)$ is of the **Bregman type** if there exists a **convex** function $\phi$ with subgradient $\phi'$ such that

$$L(x, y) = \phi(y) - \phi(x) - \phi'(x)(y - x).$$

the only Bregman loss function that depends on the prediction error only is the quadratic loss, $L(x, y) = (x - y)^2$, that arises when $\phi(x) = x^2$

## Theorem (Savage 1971; Banerjee, Guo and Wang 2005)

If the loss function function is of the **Bregman type**, the **mean** is an **optimal point predictor**.

## Theorem (Savage 1971; Banerjee, Guo and Wang 2005)

If the loss function satisfies assumptions (L0), (L1) and (L2a) and the **mean** is an **optimal point predictor**, whatever the predictive distribution, the loss is of the **Bregman type**.

# Quantiles as optimal point predictors

## Definition

A **loss function** $L(x, y)$ is **generalized piecewise linear (GPL) of order** $\alpha$, where $0 < \alpha < 1$, if there exists a nondecreasing function $g$ such that

$$L(x, y) = \begin{cases} \alpha\,(g(y) - g(x)) & \text{if} \quad x \leq y \\ (1 - \alpha)\,(g(x) - g(y)) & \text{if} \quad x \geq y \end{cases}$$

## Theorem

If the loss function function is **GPL** of **order** $\alpha$, any $\alpha$**-quantile** is an **optimal point predictor**.

## Theorem

If a loss function satisfies assumptions (L0), (L1) and (L2b) and any $\alpha$**-quantile** is an **optimal point predictor**, whatever the predictive distribution, it is **GPL** of **order** $\alpha$.

# Simulation study: Forecasting a conditionally heteroscedastic process

## GARCH process

$$Y_{t+1} \sim \mathcal{N}(0, \sigma_{t+1}^2) \quad \text{where} \quad \sigma_{t+1}^2 = \alpha Y_t^2 + \beta \sigma_t^2 + \gamma,$$

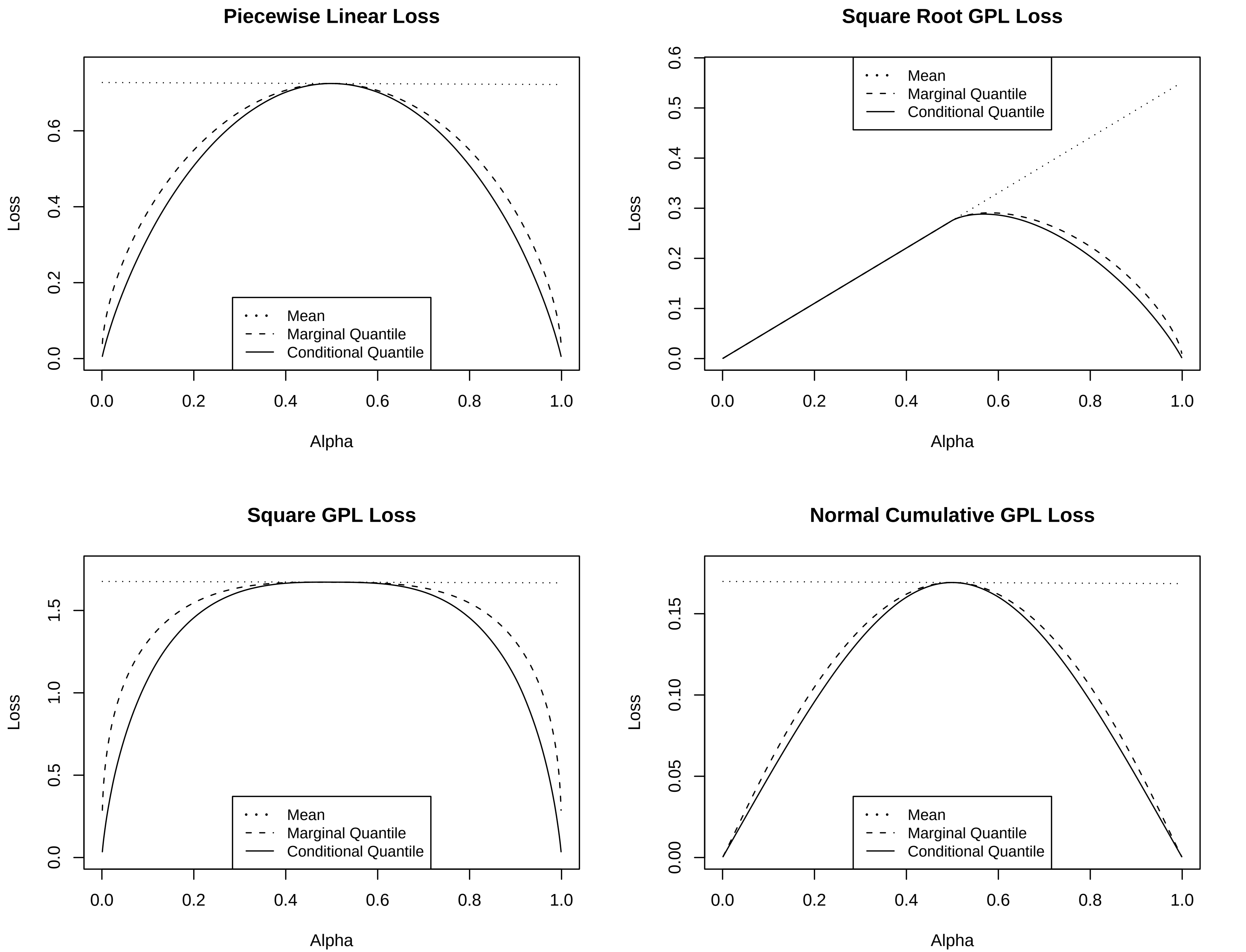
where $\alpha = 0.2$, $\beta = 0.75$ and $\gamma = 0.05$

at each of 1,000 replications, we draw a realization and predict one step ahead, based on the true model

three competing **point predictors**: mean, marginal $\alpha$-quantile and conditional $\alpha$-quantile

four classes of **loss functions**: GPL of order $\alpha \in (0, 1)$ where $g(x) = x$, $g(x) = \max(0, x)^{1/2}$, $g(x) = x^2$ and $g(x) = \Phi(x)$

*If we assess the three types of point forecasts by the average loss over 1,000 replications, what results does theory suggest?*


Piecewise Linear Loss
Square Root GPL Loss
Square GPL Loss
Normal Cumulative GPL Loss
Loss
Alpha
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.3
0.5
1.5
0.00
0.05
0.10
0.15
Mean
Marginal Quantile
Conditional Quantile

## Real world example:
## Forecasting wind speed and wind energy

**Stateline wind energy center** is a 300 megawatts, $300 million project located on the Washington/Oregon border

**2-hour** ahead **forecasts** of **hourly average wind speed** at **Stateline** in May — November 2003

## Regime-switching space-time (RST) method

merges **meteorological** and **statistical** expertize (Gneiting, Larson, Westrick, Genton and Aldrich 2006)

model formulation is parsimonious, yet **takes salient features of wind speed into account:** alternating atmospheric regimes, temporal and spatial autocorrelation, diurnal and seasonal non-stationarity, conditional heteroscedasticity, and non-Gaussianity

the RST **predictive distributions** are **truncated normal** with **predictive density**

$$f(y) = \left[ \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) \right] \Big/ \Phi\left(\frac{\mu}{\sigma}\right) \quad \text{for} \quad y \geq 0$$

and **quantile function**

$$q_\alpha = \mu + \sigma \, \Phi^{-1}(\alpha + (1-\alpha)\Phi(-\mu/\sigma))$$

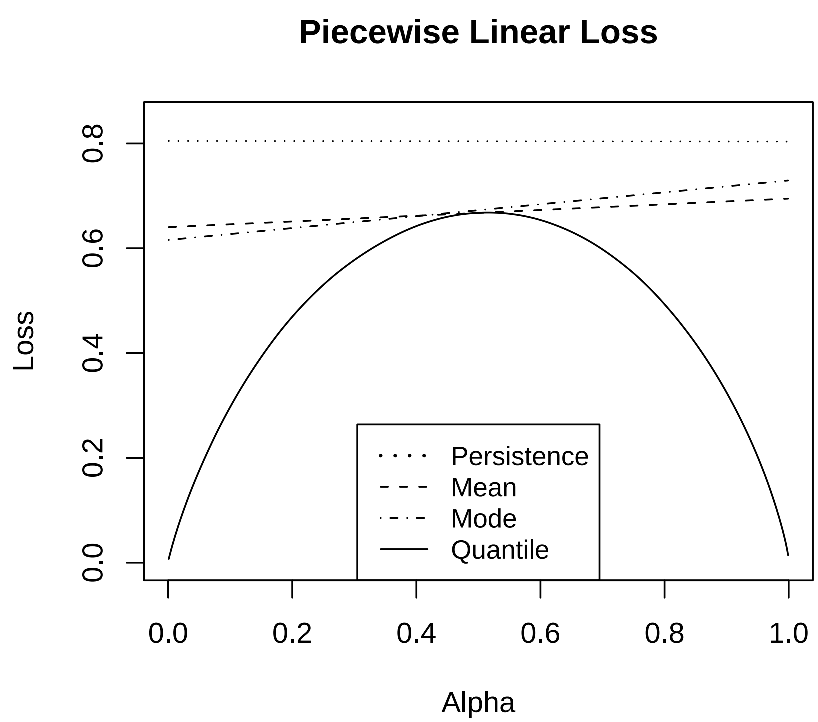
which respect the **non-negativity** of **wind speed**

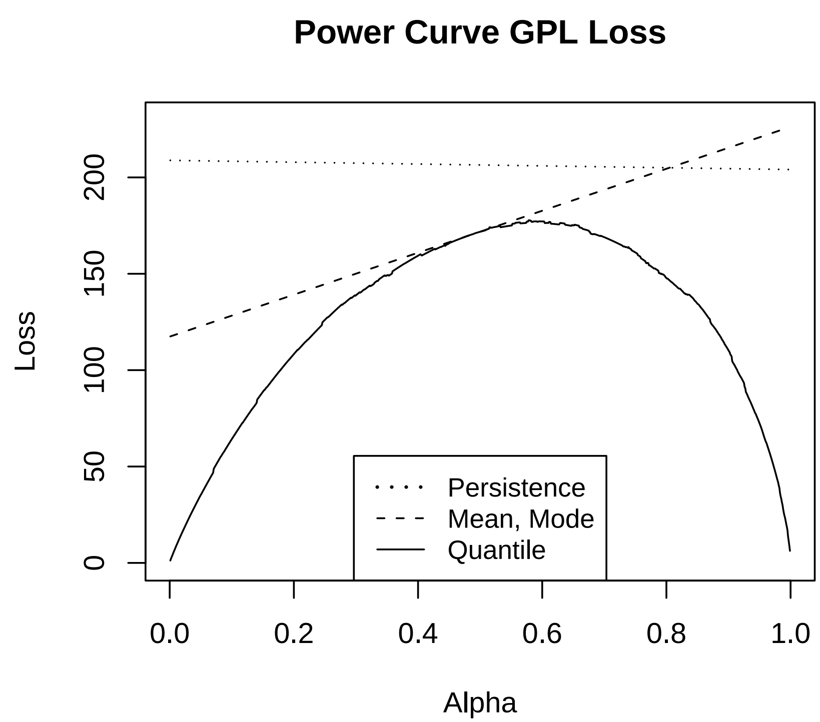
## Point prediction

three competing **point predictors**: **persistence** forecast, **mean** and **$\alpha$-quantile** of the **truncated normal** predictive distribution

two classes of **loss functions**: GPL of order $\alpha \in (0, 1)$ where $g(x) = x$ and $g(x) = [\max(x, 20)]^3$

the latter is the economically relevant **power curve loss** function with a saturation at 20 $\mathrm{m} \cdot \mathrm{s}^{-1}$

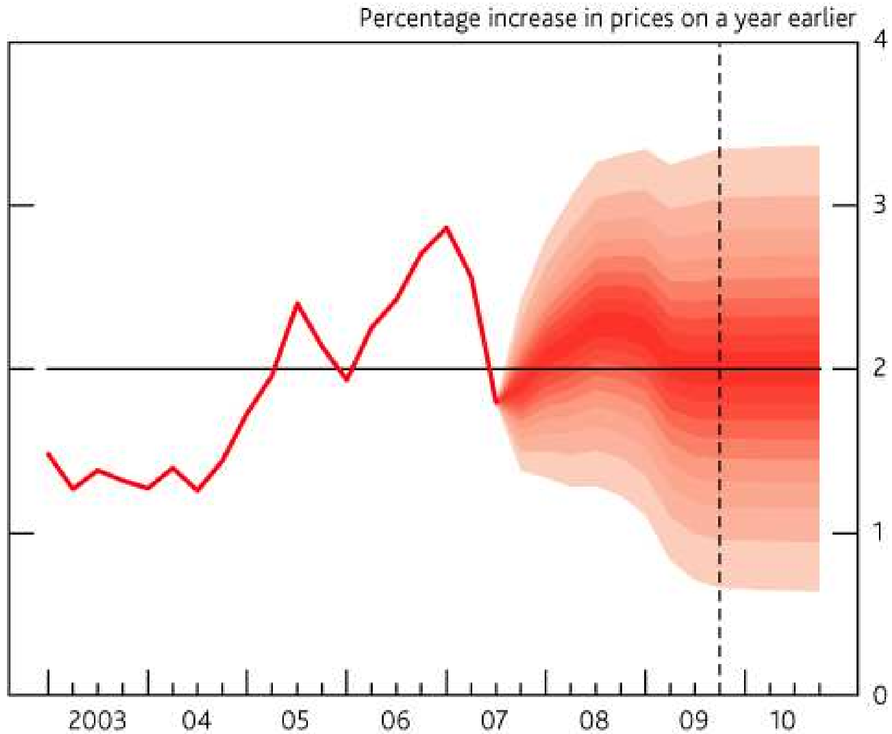


## Real world example:

## Bank of England inflation projections

the **Bank of England** has issued probabilistic forecasts of United Kingdom **inflation rates** every quarter since February 1996




using **fan charts** to visualize the **predictive distribution** for **percentage increase** in prices **on a year earlier**

## Bank of England density forecasts

the Bank of England **predictive distributions** for inflation rate are **two-piece normal** with **predictive density**

$$f(y) = \begin{cases} \left(\frac{\pi}{2}\right)^{-1/2} (\sigma_1 + \sigma_2)^{-1} \exp\left(-\frac{(y-\mu)^2}{2\sigma_1^2}\right) & \text{if} \quad y \le \mu \\ \left(\frac{\pi}{2}\right)^{-1/2} (\sigma_1 + \sigma_2)^{-1} \exp\left(-\frac{(y-\mu)^2}{2\sigma_2^2}\right) & \text{if} \quad y \ge \mu \end{cases}$$

and **quantile function**

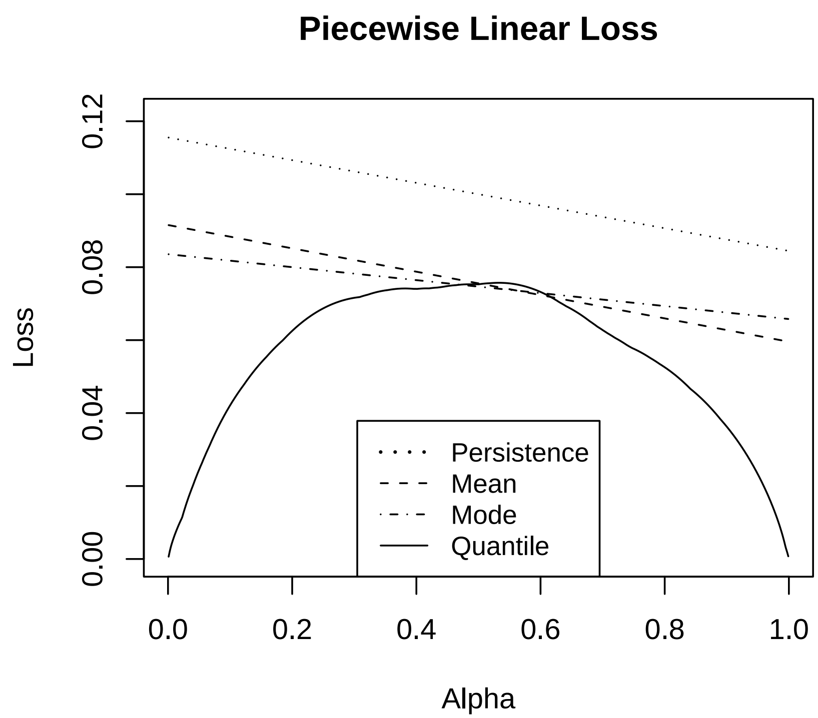
$$q_\alpha = \begin{cases} \mu + \sigma_1 \Phi^{-1}\left(\frac{\sigma_1+\sigma_2}{2\sigma_1}\alpha\right) & \text{if} \quad \alpha \le \frac{\sigma_1}{\sigma_1+\sigma_2} \\ \mu + \sigma_2 \Phi^{-1}\left(\left(\alpha - \frac{\sigma_1-\sigma_2}{\sigma_1+\sigma_2}\right)\frac{\sigma_1+\sigma_2}{2\sigma_2}\right) & \text{if} \quad \alpha > \frac{\sigma_1}{\sigma_1+\sigma_2} \end{cases}$$

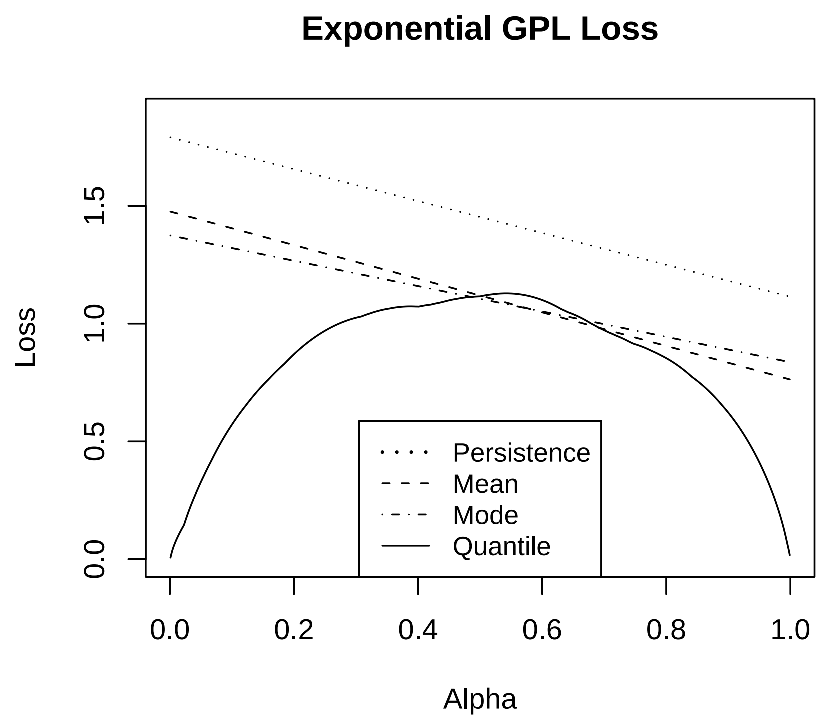
we consider a **prediction horizon** of **one quarter** ahead

# Point prediction

three competing **point predictors**: **persistence** forecast, **mean** and **$\alpha$-quantile** of the **two-piece normal** predictive distribution

two classes of **loss functions**: GPL of order $\alpha \in (0,1)$ where $g(x) = x$ and $g(x) = \exp(x)$



# Discussion

in a rapidly growing range of applied problems, **probabilistic forecasts** in the form of **predictive distributions** are the preferred forecast format

if we must issue **point forecasts**, they need to be tailored to the **loss structure** at hand

the class of loss structures which are such that the **mean** is an **optimal** point predictor, whatever the predictive distribution, is that of the **Bregman loss functions** (Savage 1971)

the class of loss structures which are such that any **$\alpha$-quantile** is an **optimal** point predictor, whatever the predictive distribution, is that of the **generalized piecewise linear (GPL)** loss functions **of order $\alpha$**