Comparing Possibly Misspecified Forecasts

Andrew J. Patton

Department of Economics
Duke University

October 2016

Asking for an expert forecast

- How should we ask experts for their (point, interval or density) forecasts?
- It is broadly accepted that the method used to **evaluate** experts should be tied to the quantity that forecasters were **asked to predict**:
  - If they are asked for their *conditional expectation*, we should evaluate experts in terms of their ability to match the *mean* of the target variable.
  - Similarly if they are asked for the median, Value-at-Risk, a prediction interval, a probability density forecast, etc.
  - Eg, the same number will likely not be a good forecast of the mean and VaR.
- Being clear about the quantity of interest to the forecast user provides structure (details below) on how we evaluate expert forecasts.

Some real-world complications

- This paper addresses whether the way we ask experts for their forecasts should also vary with the forecast "*environment*"?
  - **(a)** Are the best forecasting models possibly **misspecified**?
  - **(b)** Are competing forecasters using **non-nested** information sets?
  - **(c)** Are the forecasts subject to **estimation error**?
- This paper shows that the presence of *any* of the above features provides even greater structure on how we should ask for an expert forecast.
- What do economic surveys actually ask for?

What do surveys ask for?

- **Survey of Professional Forecasters**, run by the Philadelphia Fed:
  - *"What do you expect to be the annual average CPI inflation rate over the next 5 years?"* (Section 7 of the survey)
- Thomson Reuters/University of **Michigan Survey of Consumers**:
  - *"By about what percent do you expect prices to go (up/down) on the average, during the next 12 months?"* (Question A12b)
- **Livingston survey**:
  - *"What is your forecast of the annual rate of change in the CPI?"*

Outline

1. Introduction
2. Comparing mean forecasts using Bregman loss functions
   - Some surprising existing results
   - Results for the "ideal" forecasting environment
   - Results for economic forecasting environments
3. Results for quantile and distribution forecasts
4. Simulation-based examples
5. Application to survey forecasts of U.S. inflation
6. Conclusion and recommendations

Outline

1. Introduction
2. Comparing mean forecasts using Bregman loss functions
   - **Some surprising existing results**
   - Results for the “ideal” forecasting environment
   - Results for economic forecasting environments
3. Results for quantile and distribution forecasts
4. Simulation-based examples
5. Application to survey forecasts of U.S. inflation
6. Conclusion and recommendations

Mean forecasts and Bregman loss functions

- Minimizing expected **quadratic loss** leads to the conditional **mean**:

$$\hat{Y}^*_{t+h|t} \equiv \arg\min_{y \in \mathcal{Y}} \mathbb{E}\left[\left(Y_{t+h} - \hat{y}\right)^2 |\mathcal{F}_t\right] = \mathbb{E}\left[Y_{t+h}|\mathcal{F}_t\right]$$

- This is in fact true for an entire family of functions known as **Bregman loss** functions (Banerjee, et al. 2005 *IEEE;* Gneiting 2011 *JASA;* Bregman, 1967)

$$L\left(y, \hat{y}; \phi\right) = \phi\left(y\right) - \phi\left(\hat{y}\right) - \phi'\left(\hat{y}\right)\left(y - \hat{y}\right), \quad \phi \text{ convex}$$

- Bregman loss functions are said to be "**consistent**" for the mean, in that the mean minimizes the expected loss for any $L \in \mathcal{L}_{Bregman}$
  - No other quantity (median, mode, etc.) leads to lower expected loss
  - The mean is said to be "elicitable" using Bregman loss

Mean forecasts ⇔ Bregman loss functions

- Perhaps more surprising is that this class of loss functions is **sufficient *and* necessary** for the conditional mean to optimal:

$$\hat{Y}^*_{t+h|t} \equiv \arg\min_{y \in \mathcal{Y}} \mathbb{E}\left[L\left(Y_{t+h}, \hat{y}\right)|\mathcal{F}_t\right] = \mathbb{E}\left[Y_{t+h}|\mathcal{F}_t\right]$$
$$\Leftrightarrow \quad L \in \mathcal{L}_{Bregman}$$

- **Sufficiency** of Bregman loss is easy to verify: the first-order condition for the optimal forecast is

$$\text{FOC} \quad 0 = \phi''\left(\hat{Y}^*_{t+h|t}\right)\left(\mathbb{E}\left[\left(Y_{t+h}\right)|\mathcal{F}_t\right] - \hat{Y}^*_{t+h|t}\right)$$

- **Necessity** is a bit harder to establish and interpret
  - Patton (2011) uses a generalized Farkas lemma but other proofs are possible

A few interesting features of Bregman loss

- Bregman loss:

$$L(y,\hat{y};\phi) = \phi(y) - \phi(\hat{y}) - \phi'(\hat{y})(y-\hat{y}), \quad \phi \text{ convex}$$

- Savage (1971, *JASA*): The only Bregman loss function that is solely a function of the **difference** forecast error, y-$\hat{y}$, is the **quadratic** loss function.

- Patton (2011, *JoE*): The only Bregman loss function that is solely a function of the **ratio** forecast error, $y/\hat{y}$, is the **QLIKE** loss function:

$$L(y,\hat{y}) = \frac{y}{\hat{y}} - \log\frac{y}{\hat{y}} - 1$$

- Savage (1971, *JASA*): The only Bregman loss function that is **symmetric** is the **quadratic** loss function.

Illustrating the variety of Bregman loss functions

- Consider the following two parametric families of Bregman loss functions:

1. **Homogeneous Bregman:**

$$L(y,\hat{y};k) = |y|^k - |\hat{y}|^k - k\text{sgn}(\hat{y})|\hat{y}|^{k-1}(y-y), \; k>1$$

   - Nests quadratic loss at $k=2$

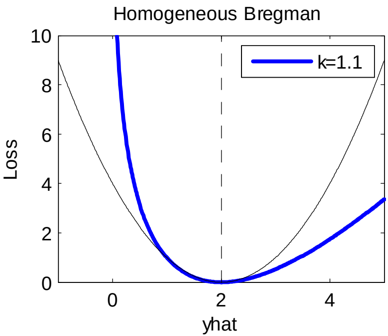
2. **Exponential Bregman:**

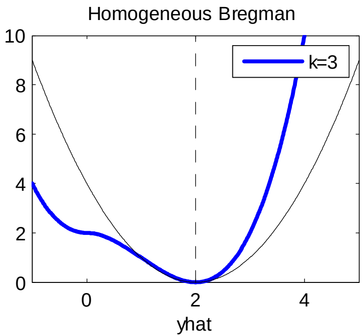
$$L(y,\hat{y};a) = \frac{2}{a^2}(\exp\{ay\} - \exp\{a\hat{y}\}) - \frac{2}{a}\exp\{a\hat{y}\}(y-\hat{y}), \quad a \neq 0$$

   - Nests quadratic loss as $a \longrightarrow 0$
   - Note the similarity to "linex" loss.

Homogeneous Bregman loss functions

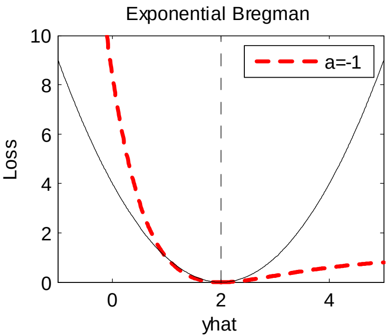
Bregman loss functions can be asymmetric in either direction, and be convex or not





Exponential Bregman loss functions

Bregman loss functions can be asymmetric in either direction



Evaluating possibly misspecified forecasts

- Gneiting (2011, *JASA*):

  "If point forecasts are to be issued and evaluated, it is essential that either the **scoring function** *be specified ex ante, or an elicitable* **target functional** *be named, such as the mean or a quantile of the predictive distribution, and scoring functions be used that are consistent for the target functional."*

Evaluating possibly misspecified forecasts

- Gneiting (2011, *JASA*):

  "If point forecasts are to be issued and evaluated, it is essential that either the **scoring function** *be specified ex ante, or an elicitable* **target functional** *be named, such as the mean or a quantile of the predictive distribution, and scoring functions be used that are consistent for the target functional."*

  $\Rightarrow$ If you ask for a forecast of the **mean**, you should evaluate using a **Bregman** loss function.

Evaluating possibly misspecified forecasts

- Gneiting (2011, *JASA*):

  "If point forecasts are to be issued and evaluated, it is essential that either the **scoring function** *be specified ex ante, or an elicitable* **target functional** *be named, such as the mean or a quantile of the predictive distribution, and scoring functions be used that are consistent for the target functional."*

  $\Rightarrow$ If you ask for a forecast of the **mean**, you should evaluate using a **Bregman** loss function.

- This is a sensible recommendation, but this paper shows that in practice, it is *not sufficient* to tell the survey respondents the target functional

  $\Rightarrow$ Respondents should be told the **specific loss function** that will be used to evaluate their forecasts.

Forecast evaluation using economic loss functions

Paper	**Target**	**Loss function**
Leitch & Tanner (1991, *AER*)	Interest rates	T-bill trading profits
West, Edison, Cho (1993, *JIE*)	FX volatility	Portfolio decisions
Fleming, Kirby, Ostdiek (2001, *JF*)	Asset volatility	Portfolio decisions
Christoffersen & Jacobs (2004, *JFE*)	Equity volatility	Option pricing errors

- Specific economic applications trace out particular loss functions, (and in turn imply a particular quantities as the target functionals)
- This paper is more related to cases where only the target functional is known, but arrives at similar conclusions

Outline

1. Introduction
2. Comparing mean forecasts using Bregman loss functions
   - Some surprising existing results
   - **Results for the "ideal" forecasting environment**
   - Results for economic forecasting environments
3. Results for quantile and distribution forecasts
4. Simulation-based examples
5. Application to survey forecasts of U.S. inflation
6. Conclusion and recommendations

Forecast comparison in ideal environments

- **Proposition:** Assume that

(i) the information sets of forecasters A and B are **nested**, so that $\mathcal{F}_t^B \subseteq \mathcal{F}_t^A\ \forall t$ or $\mathcal{F}_t^A \subseteq \mathcal{F}_t^B\ \forall t$, **and**

(ii) forecasts A and B are **optimal** under some Bregman loss function.

★ Then the ranking of these forecasts by MSE is sufficient for the ranking by **any** Bregman loss function. That is

$$MSE_A \lesseqqgtr MSE_B \Rightarrow \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^A\right)\right] \lesseqqgtr \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^B\right)\right]\ \forall L \in \mathcal{L}_{Bregman}$$

- **Proof** is based necessity of Bregman for mean forecast optimality and the law of iterated expectations (see paper for details).

- See also Holzmann and Eulert (2014, *AoAS*)

Model estimation in ideal environments

- **Proposition:** Assume that

(i) $\mathbb{E}\left[Y_t|\mathcal{F}_t\right] = m\left(X_t;\theta_0\right)$ for some $\theta_0 \in \Theta \subseteq \mathbb{R}^p$, $p < \infty$, and

(ii) $\partial m\left(X_t;\theta\right)/\partial\theta \neq 0$ $a.s.$ $\forall\ \theta \in \Theta$.

- Define

$$\theta^*_\phi \equiv \arg\min_{\theta\in\Theta}\ \mathbb{E}\left[L\left(Y_t, m\left(X_t;\theta\right);\phi\right)\right]$$

where L is a Bregman loss function characterized by ϕ.

★ Then $\theta^*_\phi = \theta_0\ \forall\phi$.

- Elliott, *et al.* (2016, REStat) has a related result for forecasts of binary variables.

Recap (or, "Why haven't I heard about Bregman before?")

- In the "ideal" forecasting environment:
  1. models are **correctly specified**,
  2. models are **free from parameter estimation error**, and
  3. competing forecasts are based on **nested information sets**.
- The variety of loss functions that are consistent for the mean may be **ignored** (in the limit):
  - ranking by MSE is the same as by any Bregman loss
  - estimation by OLS is the same as by any Bregman loss
- The *efficiency* of estimation and comparison by these loss functions may differ, but we do not focus on that here.

Outline

1. Introduction
2. Comparing mean forecasts using Bregman loss functions
   - Some surprising existing results
   - Results for the "ideal" forecasting environment
   - **Results for economic forecasting environments**
3. Results for quantile and distribution forecasts
4. Simulation-based examples
5. Application to survey forecasts of U.S. inflation
6. Conclusion and recommendations

Optimizing misspecified models

- If the model is misspecified, then the optimal model parameters will generally be **sensitive** to the choice of Bregman loss function used in estimation.

- Simple example:

$$\textbf{DGP} \qquad \begin{aligned} Y &= X^2 + \varepsilon, \quad \varepsilon \sim iid\ N(0,1) \\ X &\sim iid\ N\left(\mu, \sigma^2\right) \end{aligned}$$

$$\textbf{Model} \qquad Y = \alpha + \beta X + e$$

Optimizing misspecified models

- If the model is misspecified, then the optimal model parameters will generally be **sensitive** to the choice of Bregman loss function used in estimation.
- Simple example:

$$\begin{aligned} \textbf{DGP} \qquad Y &= X^2 + \varepsilon, \quad \varepsilon \sim iid\ N(0,1) \\ X &\sim iid\ N\left(\mu, \sigma^2\right) \\ \textbf{Model} \qquad Y &= \alpha + \beta X + e \end{aligned}$$

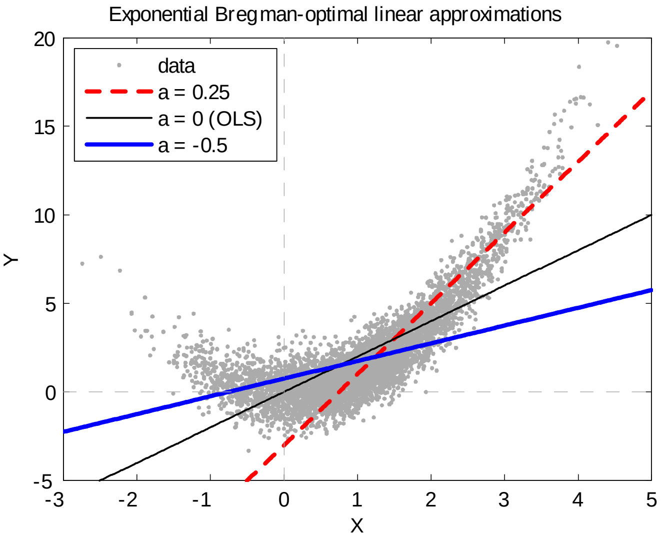
★ Then, across the class of Exponential Bregman loss functions with shape parameter a, we find:

$$\hat{\alpha}_a^* = \sigma^2 - \frac{\mu^2}{\left(1 - 2a\sigma^2\right)^2} \quad \text{and} \quad \hat{\beta}_a^* = \frac{2\mu}{1 - 2a\sigma^2} \quad \text{for } a \neq 1/\left(2\sigma^2\right)$$

- See the following figure for $\mu = \sigma^2 = 1$.

Optimal approximations

The optimal linear approximation varies with the choice of (consistent) loss function



Optimal approximations

Matching the estimation and the evaluation loss function can be helpful

Expected loss

		Estimation loss function		
		$a = 0.25$	$a = 0$	$a = -0.5$
Evaluation	$a = 0.25$	**1.21**	2.56	133.72
loss	$a = 0$	4.06	**2.99**	7.94
function	$a = -0.5$	18.42	13.22	**9.03**

★ If forecasters use misspecified models, they can improve their forecasts by calibrating their models using the forecast consumer's loss function.

- In population, this is always true
- In finite samples, there is a possible bias-variance trade-off:

Forecast comparison in not-so-ideal environments

- **Proposition:** If

(i) the information sets of forecasters A and B are **non-nested**, so that $\mathcal{F}_t^B \nsubseteq \mathcal{F}_t^A$ for some t or $\mathcal{F}_t^A \nsubseteq \mathcal{F}_t^B$ for some t, or

(ii) at least one of forecasts A and B are based on a **misspecified** model, or

(iii) at least one of the forecasts A and B contains **estimation error**

★ Then the ranking of these forecasts is, in general, *sensitive to the choice of Bregman loss function*.

- **Proof:** just requires an example – I present:
  1. Analytical results for three stylized examples
  2. Simulation-based results for more realistic examples

Comparing misspecified, error-ridden, nonnested models.

- Consider the following simple **DGP**:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} .... + \phi_5 Y_{t-5} + \varepsilon_t, \quad \varepsilon_t \sim iid \ N(0,1)$$

where $\phi_0 = 1$, and $[\phi_1, ..., \phi_5] = [0.8, 0.3, -0.5, 0.2, 0.1]$.

- I will consider a variety of forecasting models, and will compare them using the Exponential-Bregman **loss function**:

$$\begin{aligned} L(y, \hat{y}; a) &= \frac{2}{a^2}(\exp\{ay\} - \exp\{a\hat{y}\}) - \frac{2}{a}\exp\{a\hat{y}\}(y - \hat{y}), \quad a \neq 0 \\ &\to (y - \hat{y})^2 \quad \text{as } a \to 0 \end{aligned}$$

- The combination of Exp-Breg and Normality allows me to obtain analytical expressions for optimal forecasts, expected loss, etc.

The ideal case

- First, consider the ideal case. Three competing models:

$$\begin{aligned} \text{AR}(1) \quad \hat{Y}_t &= \beta_0 + \beta_1 Y_{t-1} \\ \text{AR}(2) \quad \hat{Y}_t &= \gamma_0 + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} \\ \text{AR}(5) \quad \hat{Y}_t &= \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_5 Y_{t-5} \end{aligned}$$

  - ✓ These are all **correct**, given their limited information sets
  - ✓ I obtain the pop'n parameters for AR(1) and AR(2), so **no estimation error**
  - ✓ Their information sets are **nested**

The ideal case

- First, consider the ideal case. Three competing models:

$$\begin{array}{lrcl} \text{AR(1)} & \hat{Y}_t & = & \beta_0 + \beta_1 Y_{t-1} \\ \text{AR(2)} & \hat{Y}_t & = & \gamma_0 + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} \\ \text{AR(5)} & \hat{Y}_t & = & \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_5 Y_{t-5} \end{array}$$

  - ✓ These are all **correct**, given their limited information sets
  - ✓ I obtain the pop'n parameters for AR(1) and AR(2), so **no estimation error**
  - ✓ Their information sets are **nested**

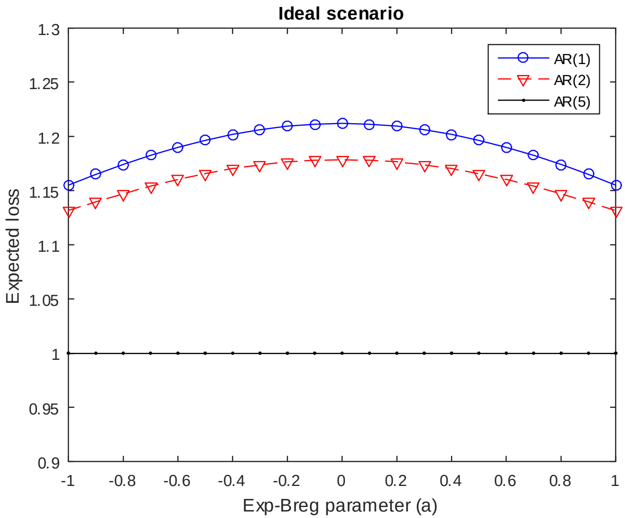
- The following figure shows

$$\frac{\mathbb{E}\left[L\left(Y_t, \hat{Y}_t^m; a\right)\right]}{\mathbb{E}\left[L\left(Y_t, \hat{Y}_t^*; a\right)\right]}, \quad m \in \{\text{AR(1), AR(2), AR(5)}\}$$

as a function of the Exp-Breg parameter a, where $\hat{Y}_t^*$ is the AR(5) forecast

The ideal case

Expected losses never cross; rankings are as expected



Misspecified models

- Recall the DGP:

$$Y_t = 1 + 0.8Y_{t-1} + 0.3Y_{t-2} - 0.5Y_{t-3} + 0.2Y_{t-4} + 0.1Y_{t-5} + \varepsilon_t$$

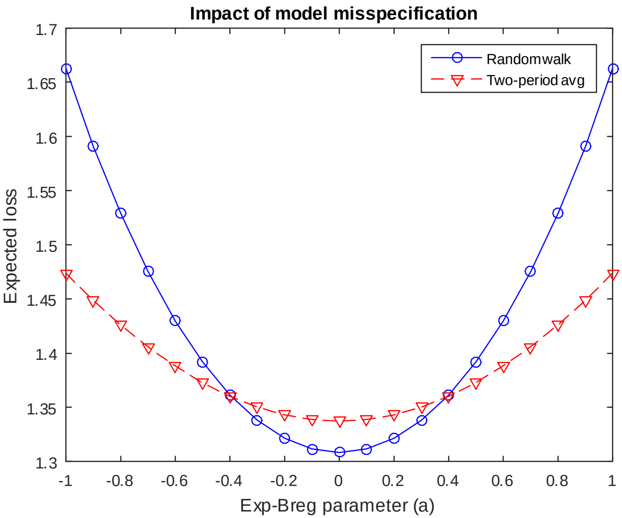
- Now consider two misspecified models:

$$\begin{array}{lrcl} \text{Random Walk} & \hat{Y}_t & = & Y_{t-1} \\ \text{Two-period Avg} & \hat{Y}_t & = & \frac{1}{2}\left(Y_{t-1} + Y_{t-2}\right) \end{array}$$

  - × These models are **misspecified**
  - ✓ But they contain **no estimation error**
  - ✓ Their information sets are **nested**

Comparing misspecified forecasting models

Expected losses cross: ranking depends on choice of Bregman loss function



Parameter estimation error

- Recall the DGP:

$$Y_t = 1 + 0.8Y_{t-1} + 0.3Y_{t-2} - 0.5Y_{t-3} + 0.2Y_{t-4} + 0.1Y_{t-5} + \varepsilon_t$$

- Now consider two models subject to estimation error:

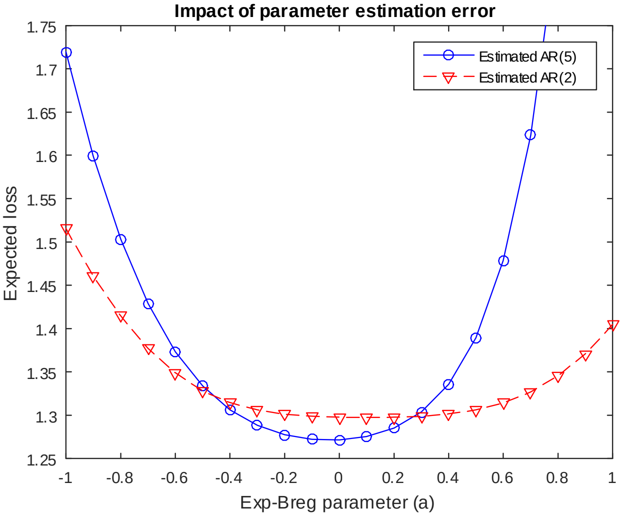
$$\begin{array}{llcl} \widehat{\text{AR}}(2) & \hat{Y}_t & = & \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t} Y_{t-1} + \hat{\gamma}_{2,t} Y_{t-2} \\ \widehat{\text{AR}}(5) & \hat{Y}_t & = & \hat{\phi}_{0,t} + \hat{\phi}_{1,t} Y_{t-1} \ldots + \hat{\phi}_{5,t} Y_{t-5} \end{array}$$

  - ✓ These models are **correctly specified**
  - × But they contain **estimation error** (rolling window OLS estimation, $n = 36$)
  - ✓ Their information sets are **nested**

- I use a simulation of 10 million OOS observations to approximate the expected loss.

Comparing forecasting models subject to estimation error

Expected losses cross: ranking depends on choice of Bregman loss function



Nonnested information sets I

- Unfortunately the Gaussian AR(p) + Exponential Bregman design does not allow me to show the problems that arise with nonnested information sets
  - Can show that nonnested info sets cause no problems in this particular case
- Instead, consider the following simple DGP:

$$
\begin{aligned}
Y_t &= \underbrace{X_t \mu_L + (1 - X_t)\,\mu_H}_{\text{small low or small high}} + \underbrace{W_t \mu_C + (1 - W_t)\,\mu_M}_{\text{crisis or not}} + Z_t \\
\text{where } X_t &\sim iid\ \ Bernoulli\,(p) \quad , \quad W_t \sim iid\ \ Bernoulli\,(q) \\
Z_t &\sim iid\ \ N\,(0, 1)
\end{aligned}
$$

  - Forecaster X has access to a Bernoulli signal that is regular ($p = 0.5$) but not very strong ($\mu_L = -1$, $\mu_H = +1$).
  - Forecaster W has access to a Bernoulli signal that is irregular ($q = 0.05$) but valuable when it arrives ($\mu_C = -5$, $\mu_M = 0$).

Nonnested information sets II

- Recall the DGP:

$$Y_t = \underbrace{X_t \mu_L + (1 - X_t)\,\mu_H}_{\text{small low or small high}} + \underbrace{W_t \mu_C + (1 - W_t)\,\mu_M}_{\text{crisis or not}} + Z_t$$

$$\text{where } \begin{aligned} X_t &\sim iid\ \ Bernoulli\,(p) \\ W_t &\sim iid\ \ Bernoulli\,(q) \\ Z_t &\sim iid\ \ N\,(0,1) \end{aligned}$$

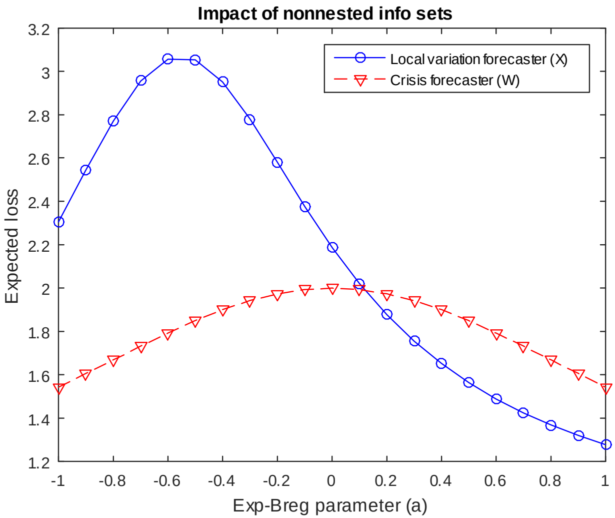
- Forecasts:

$$\hat{Y}_t^x = q\mu_C + (1-q)\,\mu_M + \mu_H + (\mu_L - \mu_H)\,X_t$$
$$\hat{Y}_t^w = p\mu_L + (1-p)\,\mu_H + \mu_M + (\mu_C - \mu_M)\,W_t$$

✓ These models are **correctly specified** (given their information sets)
✓ They contain no **estimation error**
× But their information sets are **nonnested**

Comparing forecasting models with nonnested info sets

Expected losses cross: ranking depends on choice of Bregman loss function



Discussion of the results

- **"Ideal environment" result**: if information sets are nested and models are perfect, then survey respondents only need to be told the statistical functional of interest (the mean, in this case).
  - Ranking can be done using *any* Bregman loss function (eg, MSE)
  - Estimated model parameters have same probability limit
  - No gains from using multiple Bregman loss functions
- **"Real-world environment" result**: in the presence of realistic deviations from the "ideal" environment, survey respondents should be told the specific (consistent) loss function
  - Rankings of forecasts can switch depending on choice of Bregman loss function
  - Optimal approximating model can vary with choice of Bregman loss function
  - Averaging across multiple Bregman loss functions may cloud results

Outline

1. Introduction
2. Comparing mean forecasts using Bregman loss functions
   - Some surprising existing results
   - Results for the "ideal" forecasting environment
   - Results for economic forecasting environments
3. **Results for quantile and distribution forecasts**
4. Simulation-based examples
5. Application to survey forecasts of U.S. inflation
6. Conclusion and recommendations

Extensions: Quantile forecasts

- All of the above results (positive and negative) generalize to **quantile** forecasts, using the fact that

$$\alpha = \mathbb{E}\left[\mathbf{1}\left\{Y_{t+h} \leq \hat{Y}^*_{t+h|t}\right\}\Big|\mathcal{F}_t\right] \Leftrightarrow L \in \mathcal{L}^{\alpha}_{GPL}$$

where $\mathcal{L}^{\alpha}_{GPL}$ is the class of **generalized piecewise linear (GPL)** loss functions, given by

$$L(y, \hat{y}; \alpha) = (\mathbf{1}\{y \leq \hat{y}\} - \alpha)(g(\hat{y}) - g(y)), \quad g \text{ increasing}$$

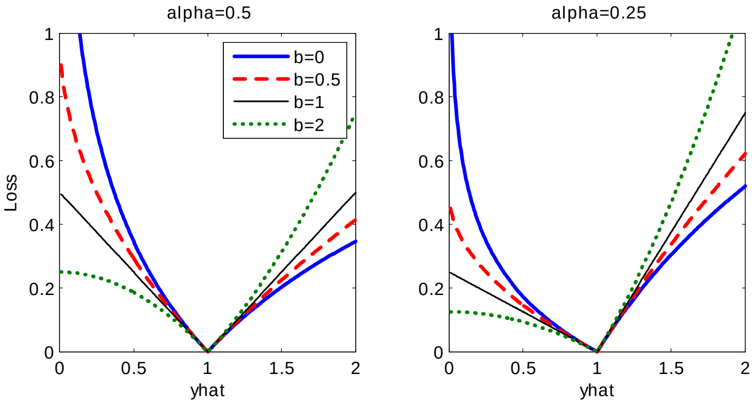
See Gneiting (2011, *IJF*).

- Class of GPL loss functions is also flexible: consider homogeneous GPL loss:

$$g(x) = \text{sgn}(x)|x|^b, \quad b > 0$$

Homogeneous GPL loss functions

Many loss functions are consistent for a given quantile



Extensions: Distribution forecasts

- All of the above results (positive and negative) generalize to **distribution** forecasts, using the fact that all **proper scoring rules** must be of the form

$$L(F, y) = \Psi(F) + \Psi^*(F, y) - \int \Psi^*(F, y)\, dF(y)$$

where Ψ is convex and Ψ^* is a subtangent of Ψ at F, see Gneiting and Raftery (2007, *JASA*).

- The **KLIC is proper**:

$$KLIC(F, y) = -\log F'(y)$$

- So is the class of "weighted continuous ranked probability scores":

$$wCRPS(F, y) = \int_{-\infty}^{\infty} \omega(z) \left(F(z) - \mathbf{1}\{y \leq z\}\right)^2 dz$$

and $\omega : \mathbb{R} \rightarrow \mathbb{R}_+$.

Most general case

- If a given target functional is elicitable (ie, there exists a loss function that is consistent for it), then:

1. Under nested information sets & correctly specified models, the "ideal environment" result holds (Holzmann and Eulert, 2014, *AoAS*)

- If in addition the class of consistent loss functions has more than one element, then:

2. In the presence of misspecified models, nonnested info sets, or estimation error, it is possible to construct cases where the **ranking is sensitive** to the chosen loss function
    - This paper: mean, quantile and distribution forecasts
    - Merkle and Steyvers (2013, *DA*): probability forecasts of binary variables
    - Elliott, Ghanem, Krüger (2016, REStat): binary forecasts of binary variables
    - *Conjecture*: possible to find examples for *any* case covered by (2)

Outline

Some realistic examples

- In this section I will illustrate the sensitivity of the rankings of forecasts to the choice of **consistent** loss function using some realistic examples.
- The objective here is to show that for *empirically-relevant* simulation designs, the choice of consistent loss function can lead to important differences in rankings of forecasts, and in estimated parameters
  - i.e., it does *not* take crazy DGPs to see this sensitivity in practice
- I will consider:
  1. A mean forecasting example (Bregman loss)
  2. A quantile forecasting example (GPL loss)
  3. A distribution forecasting example (Proper scoring rule)

Mean forecasting

- **DGP** is a persistent AR(5):

$$\begin{aligned} Y_t &= Y_{t-1} - 0.02Y_{t-2} - 0.02Y_{t-3} - 0.01Y_{t-4} - 0.01Y_{t-5} + \varepsilon_t \\ \varepsilon_t &\sim N(0,1) \end{aligned}$$

- **Models**:

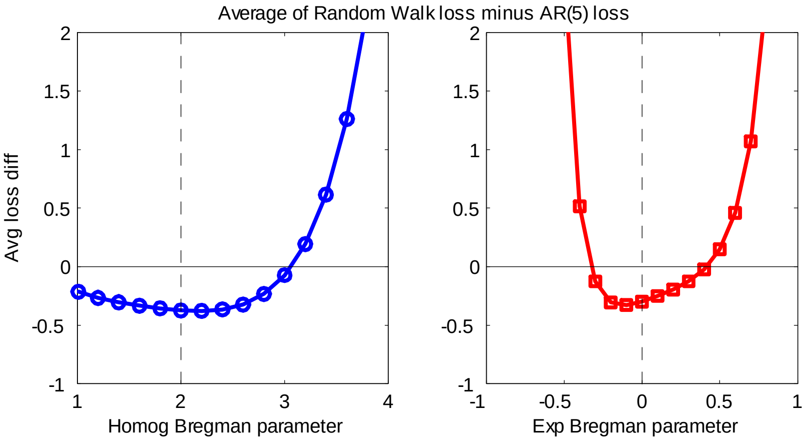
$$\begin{aligned} \hat{Y}_t^A &= Y_{t-1} \\ \hat{Y}_t^B &= \hat{\phi}_{0,t} + \hat{\phi}_{1,t}Y_{t-1} + \hat{\phi}_{2,t}Y_{t-2} + \hat{\phi}_{3,t}Y_{t-3} + \hat{\phi}_{4,t}Y_{t-4} + \hat{\phi}_{5,t}Y_{t-5} \end{aligned}$$

where $\left[\hat{\phi}_{0,t}, ..., \hat{\phi}_{5,t}\right]$ based on rolling window of 100 observations

- Simulate for 10,000 observations, and report the differences in average loss for homogeneous and exponential Bregman loss functions in the following figure.

Random Walk loss minus AR(5) loss

Random walk wins for most Homog loss fns, and for Exp loss 'near' the quadratic



Quantile forecasting

- **DGP** is a GARCH(1,1):

$$\begin{aligned} Y_t &= \mu_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid\ N(0,1) \\ \text{where } \mu_t &= 0.03 + 0.05 Y_{t-1} \\ \sigma_t^2 &= 0.05 + 0.9\sigma_{t-1}^2 + 0.05\sigma_{t-1}^2 \varepsilon_{t-1}^2 \end{aligned}$$

- **Models** are based on non-nested information sets:

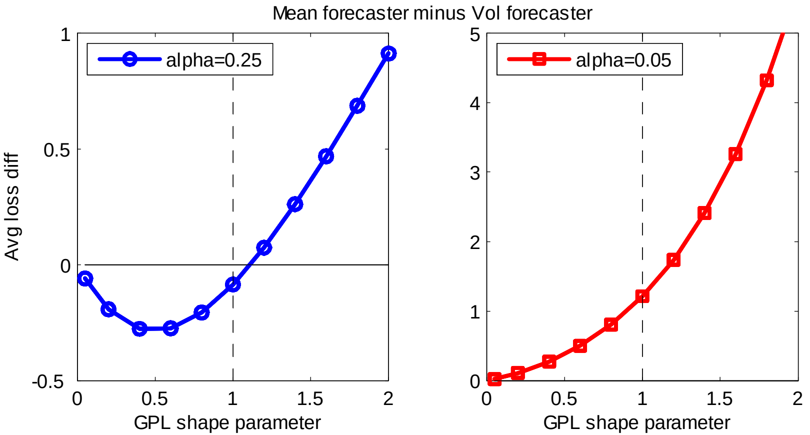
$$\begin{aligned} \hat{Y}_t^A &= \mu_t + \bar{\sigma}\Phi^{-1}(\alpha) \\ \hat{Y}_t^B &= \bar{\mu} + \sigma_t \Phi^{-1}(\alpha) \end{aligned}$$

where $\bar{\mu} = E[Y_t]$ and $\bar{\sigma}^2 = V[Y_t]$.

- Consider two quantiles, $\alpha = 0.05$ and $\alpha = 0.25$.
- Compare forecasts using the homogeneous GPL loss function and report results based on a simulation of 10,000 observations.

"Mean" forecaster loss minus "Vol" forecaster loss

Vol forecaster wins for all loss when alpha=0.05; Results mixed for alpha=0.25



Distribution forecasting

- **DGP** is a GARCH(1,1) with skew t shocks

$$\begin{aligned} Y_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid\ Skew\ t\,(0,1,6,-0.25) \\ \text{where} \quad \sigma_t^2 &= 0.05 + 0.9\sigma_{t-1}^2 + 0.05\sigma_{t-1}^2 \varepsilon_{t-1}^2 \end{aligned}$$

- **Forecasts**:

$$\begin{aligned} \hat{F}_{A,t}(x) &= \Phi\left(x/\hat{\sigma}_t\right), \quad \text{where} \quad \hat{\sigma}_t^2 = 1/100 \textstyle\sum_{j=1}^{100} Y_{t-j}^2 \\ \hat{F}_{B,t}(x) &= 1/100 \textstyle\sum_{j=1}^{100} \mathbf{1}\left\{Y_{t-j} \leq x\right\} \end{aligned}$$

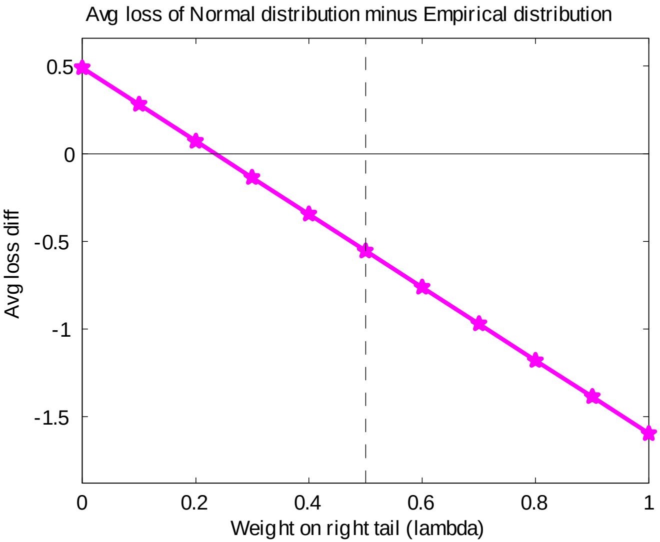
- Use the *wCRPS* scoring rule with weights based on a standard Normal CDF:

$$\omega(z;\lambda) \equiv \lambda\Phi(z) + (1-\lambda)(1-\Phi(z)), \quad \lambda \in [0,1]$$

  - $\lambda = 1 \Rightarrow$ more weight on the right tail than the left tail. (Reverse for $\lambda = 0$.)
  - $\lambda = 0.5 \Rightarrow$ weighting is flat and weights both tails equally.

Normal dist'n loss minus Empirical dist'n loss

Empirical distribution wins only when weight is particularly high on left tail



Outline

1. Introduction
2. Comparing mean forecasts using Bregman loss functions
   - Some surprising existing results
   - Results for the "ideal" forecasting environment
   - Results for economic forecasting environments
3. Results for quantile and distribution forecasts
4. Simulation-based examples
5. **Application to survey forecasts of U.S. inflation**
6. Conclusion and recommendations

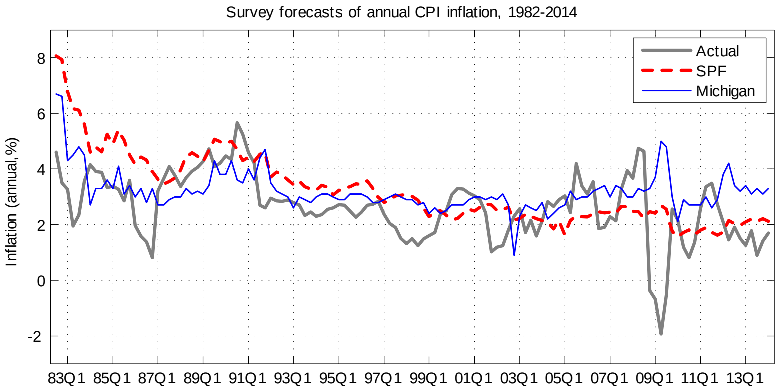
Evaluating survey forecasts of US inflation

- To illustrate these ideas in practice, I consider two forecast comparisons, both for U.S. inflation over 1982Q3 to 2014Q1 (127 observations):

1. Compare consensus forecasts from **Survey of Professional Forecasters** and **Michigan Survey of Consumers**
    - Likely to have non-nested information sets
    - *Very* likely to be based on models with estimation error or misspecification
    - Forecast horizon is one year (to match Michigan), updated quarterly (to match SPF)
2. Compare **individual forecasters** from the Survey of Professional Forecasters
    - Want to avoid criticism for using a "consensus" forecast
    - Forecast horizon is one quarter, updated quarterly

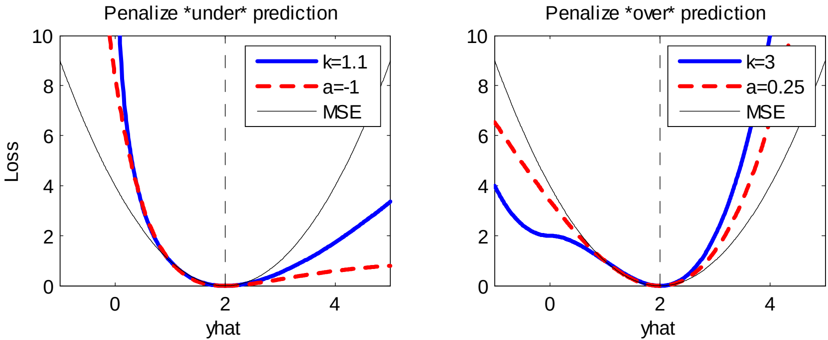
- "Actual" inflation figures taken from 2014Q2 vintage of real-time CPI data from the Philadelphia Fed

Survey forecasts of annual CPI inflation, 1982-2014



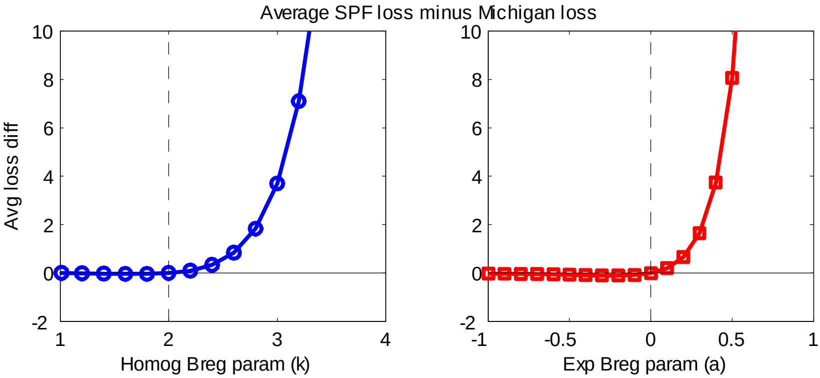
Recall: Homog and Exp Bregman loss functions

Penalty is high for under prediction if k<2 or a<0; High for over prediction if k>2 or a>0.



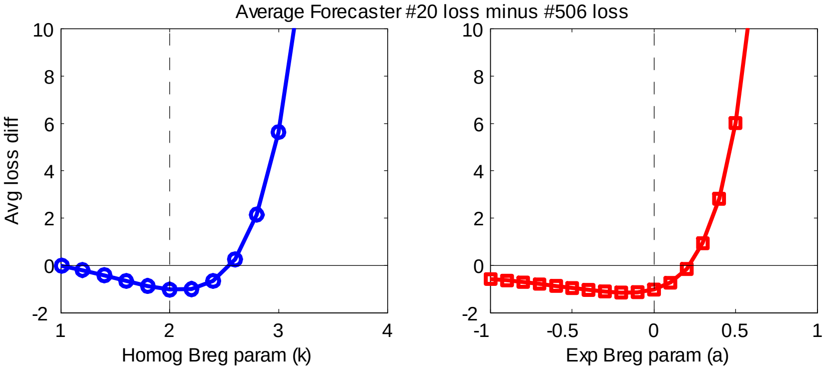
Average SPF loss minus Michigan loss

SPF beats Michigan for k<2 and a<0, loses for the rest



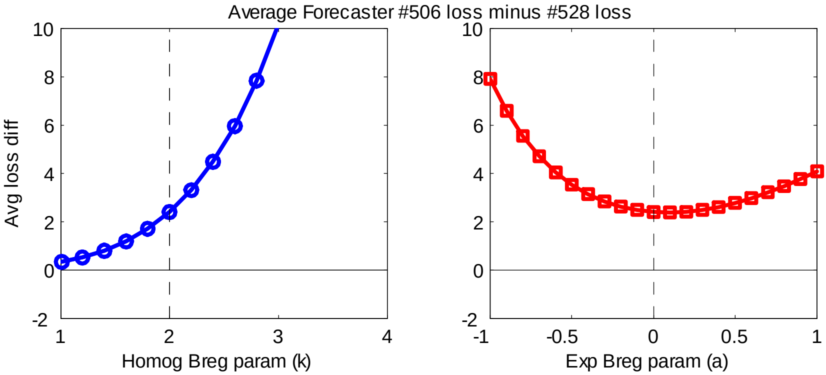
Average Forecaster 20 loss minus Forecaster 506 loss

#20 beats #506 for k<2.5 and a<0.25, loses for the rest



Average Forecaster 506 loss minus Forecaster 528 loss

#528 beats #506 for all loss functions considered



Summary and conclusion

- Clearly, forecasters will do better if they are told which statistical functional of the target variable's predictive distribution is of interest to the consumer
  - Eg: mean, median, Value-at-Risk, density
  - There is no "one" point forecast for a variable
- Forecasters should then be compared *only* using loss functions consistent for that functional
  - If mean, use Bregman loss
  - If α-quantile, use GPL^{α} loss
  - If distribution, use a proper scoring rule
  - ★ It is no surprise if forecast rankings change across loss functions that are consistent for *different* functionals

Key recommendations for forecast survey design

- In the ideal-world case, all forecasters' information sets are nested *and* all models are correct and free from estimation error, then:
  - Ranking by one consistent loss fn is sufficient for *any* consistent loss fn
  - Forecasters need only be told the functional of interest to the user
- But in the presence of real-world features like (any of):
  - model misspecification
  - non-nested information sets
  - estimation error

★ Forecasters should be told the **specific** loss function to be used for evaluation

  - They can then optimally tailor their model towards the user's loss function
  - It is *not* sufficient to ask forecasters for their expectation (or median, or even distribution) of the target variable.