Assessing Point Forecast Accuracy by Stochastic Divergence from Zero

Francis X. Diebold and Minchul Shin
University of Pennsylvania

April 28, 2014
Motivation 1: Environment

- Two forecasters (forecasting models) for variable \( y_{t+h} \).

\[ f_{1,t+h} \text{ versus } f_{2,t+h} \]

where forecasts are made at \( t \).

- We want to compare point forecasts from two models.

- Forecast errors from two models (\( e_{i,t+h} = y_{t+h} - f_{i,t+h} \) for \( i = 1, 2 \)).

\[ e_{1,t+h} \text{ versus } e_{2,t+h} \]
Motivation 2: Decision theory based evaluation

- Forecasting evaluator acts as if he/she has some specific loss (or utility) function.

\[ E[L(e_1, t+h)] \quad \text{versus} \quad E[L(e_2, t+h)] \]

where \( L(0) = 0 \) and \( L(e) \geq 0 \) for all \( e \). Smaller the better.

- Most popular choice in practice is a quadratic loss

\[ E[e_1^2, t+h] \quad \text{versus} \quad E[e_2^2, t+h] \]
What we do

\[ E[L(e_1)] \] versus \[ E[L(e_2)] \]

- We study accuracy measures based on the “distributional divergence”

- We answer two questions in this paper
  1. Can we go beyond expectation?
  2. Can we escape from the loss function?
(Measure 1) Accuracy measure for $L(e)$

- Keep the loss function.

- We want to measure a **distributional distance** between c.d.f. of $L(e)$ and some reference distribution.

  - We want the distance to be non-negative. It should be zero when $e$ is a perfect forecast error.

  - Define the perfect forecast: $f_{t+h}^o = y_{t+h}$. Perfect forecast error is
    $$e_{t+h}^o = y_{t+h} - f_{t+h}^o = 0 \quad \forall \ t.$$  
    The best benchmark forecast. Nothing can dominate perfect forecast.

- c.d.f. of loss based on the perfect forecast is a unit step function
  $$F(L(e^o)) = \begin{cases} 
  0, & L(e^o) < 0 \\
  1, & L(e^o) \geq 0
  \end{cases}$$  
  Put all masses on 0.
(Measure 1) Accuracy measure for $L(e)$

The Stochastic Loss Divergence (SLD)

$$A(L(e)) = \int |F(L(e^0)) - F(L(e))| \, de$$

This is a divergence between
- the c.d.f. of the perfect forecast.
- the c.d.f. of $L(e)$
Example: Stochastic Loss Divergence (SLD)

- $F$ is the c.d.f. of $L(e)$.
- Grey area is the stochastic loss divergence (SLD).
- We prefer smaller $A(L(e))$. 
Example: Stochastic Loss Divergence (SLD)

Under the SLD criterion, we prefer $F_1$ to $F_2$. 
Result 1

Equivalence of SLD and Expected Loss

Let $L(e)$ be a forecast-error loss function satisfying $L(0) = 0$ and $L(e) \geq 0$, $\forall e$. If $E|L(e)| < \infty$, then

$$A(L(e)) = E(L(e)),$$  \hspace{1cm} (1)

where $F(L(e))$ is the cumulative distribution function of $L(e)$. That is, SLD equals expected loss for any loss function and error distribution.

We started from the distributional distance. Then we arrive to the expected loss.

- An evaluator who minimizes an expected loss also minimizes a distance between the c.d.f. of the loss of forecast error and of the perfect forecast.
(Measure 2) Accuracy measure for $e$

- Dispense with the loss function entirely.

- We want to measure a distributional distance between c.d.f. of $e$ and some reference distribution.
  - We want the distance to be non-negative. It should be zero when $e_t$ is a perfect forecast error.
  - Define the perfect forecast: $f^o_t = y_t$. Perfect forecast error is
    \[ e^o_t = y_t - f^o_t = 0 \quad \forall t. \]
    The best benchmark forecast error. Nothing can dominate perfect forecast.
  - c.d.f. of the perfect forecast error is a unit step function
    \[
    F(e^o) = \begin{cases} 
    0, & e^o < 0 \\ 
    1, & e^o \geq 0 
    \end{cases}
    \]
(Measure 2) Accuracy measure for e

The Stochastic Error Divergence (SED)

\[ A(e) = \int \left| F(e^o) - F(e) \right| de \]

This is a divergence between

- the c.d.f. of the perfect forecast
- the c.d.f. of e.
Example: Stochastic Error Divergence (SED)

c.d.f. of $e$. Under the SED criterion, we prefer smaller $A(e) = A_-(e) + A_+(e)$. 
Example: Stochastic Error Divergence (SED)

Two forecast error distributions. Under the (SED) criterion, we prefer $F_1$ to $F_2$. 
Equivalence of Stochastic Error Divergence (SED) and Expected Absolute Error Loss

For any forecast error $e$ with cumulative distribution function $F(e)$. If $E|e| < \infty$, then

$$A(e) = E(|e|).$$

(2)

That is, SED equals expected absolute loss for any error distribution.

We start from the distributional distance. Then we arrive to the expected absolute loss function.
Extension: Weighted Stochastic Error Divergence (WSED)

Try the weighted average:

\[ \tilde{A}(e) = \tau A_{-}(e) + (1 - \tau) A_{+}(e) \]

for \( \tau \in [0, 1] \).
Weighted stochastic error divergence

Equivalence of Weighted Stochastic Error Divergence (WSED) and Expected Linlin Error Loss

For any forecast error $e$ with cumulative distribution function $F(e)$. If $E|e| < \infty$, then

$$\tilde{A}(e) = 2 \left[ \tau A_-(e) + (1 - \tau)A_+(e) \right] = 2E(L_\tau(e)), \quad (3)$$

where $L_\tau(e)$ is the loss function

$$L_\tau(e) = \begin{cases} 
(1 - \tau)|e|, & e \leq 0 \\
\tau|e|, & e > 0
\end{cases}$$

for $\tau \in (0, 1)$.

- The loss function $L_\tau(e)$ appears in the forecasting literature as a convenient and simple potentially asymmetric loss function. Christoffersen and Diebold (1996, 1997) call it “lin-lin” loss (i.e., linear on each side of the origin).

- It is also sometimes called “check function” loss (again in reference to its shape). Importantly, it is the loss function underlying quantile regression; see Koenker (2005).
Now consider the following generalized divergence measure:

\[
D(F, F^*; p, w) = \int |F^*(e) - F(e)|^p w(e)de.
\]

**SED and weighted SED are special cases:**

- When \( p = 1 \) and \( w(e) = 1, \forall e \), it is the SED.
- When \( p = 1 \) and \( w(e) = \begin{cases} 2(1 - \tau), & e < 0 \\ 2\tau, & e \geq 0 \end{cases} \), it is the weighted SED.
- Other choices of \( p \) and \( w(e) \)?
When $p = 2$ and $w(x) = 1$, $D$ is the **energy distance** (see Szekely and Rizzo, 2005 and Gneiting and Raftery, 2007)

$$E(F, F^*) = \int |F^*(e) - F(e)|^2 \, de.$$

We can decompose the energy distance as

$$E(F, F^*) = E|e| - E|e - e'|$$

where $e$ and $e'$ are independent copies of a random variable with distribution function $F$. 
When $p = 2$ and $w(e) = f(e)$, where $f(e)$ is the density corresponding to $F(e)$, $D$ is Cramer-von-Mises divergence,

$$CVM(F, F^*) = \int |F^*(e) - F(e)|^2 f(e) \, de$$

$$= -F(0)(1 - F(0)) + \frac{1}{3}$$

$$\geq \frac{1}{12},$$

where equality holds if and only if $F(0) = \frac{1}{2}$.

This means that this divergence ranks the median unbiased forecast the best.
Another popular statistical distance measure is the Kolmogorov-Smirnov statistic,

$$KS(F, F^*) = \sup_{e} \left| F^*(e) - F(e) \right| = \max(F(0), 1 - F(0))$$

where the lower bound is achieved at $F(0) = \frac{1}{2}$ as in the Cramer-von-Mises divergence case.

This means that this divergence ranks the median unbiased forecast the best.
Discussion: Generalizations and connections

- Equivalence of $D(F^*, F)$ minimization and $E(L(e))$ minimization can actually be obtained for a wide class of loss functions $L(e)$. In particular, we have the following result.

**Result (Complete characterization)**

Under some assumptions on $L(\cdot)$ and $F(\cdot)$,

$$\int_{-\infty}^{\infty} |F^*(e) - F(e)| \left| \frac{dL(e)}{de} \right| de = E(L(e)).$$

That is, minimization of stochastic error divergence $D(F^*, F; p, w)$ when $p = 1$ and $w(e) = |dL(e)/de|$ corresponds to minimization of expected loss.
Conclusion

We answered two questions using accuracy measures based on distributional divergence measures.

1. Can we go beyond expectation?
   - We introduce a measure: stochastic loss divergence (SLD).
   - It is the same as the expected loss.

2. Can we go without the loss function?
   - We introduce a measure: stochastic error divergence (SED).
   - It is the same as the absolute loss.
   - Other divergence measures are also related to the absolute loss.

- A graphical interpretation of the expected losses. Help forecast evaluators understand their choice of the loss function.