Assessing Point Forecast Accuracy by Stochastic Error Distance

Francis X. Diebold
University of Pennsylvania
and
Minchul Shin
University of Illinois

August 7, 2015

How Many Times Have you Ranked Forecasts' Accuracy by *RMSE*?

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e^2_{t/t-1}}$$

How Many Times Have you Ranked Forecasts' Accuracy by *RMSE*?

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_{t/t-1}^2}$$

What are you really *doing*?

What does it really *mean*?

What do you really *want*?

Does it matter whether you rank
using *RMSE* or other criteria like *MAE*?

Traditional Point-Forecast Accuracy Comparison: Emphasizes the Loss Function

Error: $e = y - \hat{y}$

Loss: $L(e)$, where $L(0) = 0$ and $L(e) \geq 0, \forall e$

The big three:

Absolute-error loss: $L(e) = abs(e)$

Squared-error loss: $L(e) = square(e)$

Check-error, or lin-lin, loss: $L(e) = check_\tau(e)$,

where

$$check_\tau(e) = \begin{cases} (1-\tau)|e|, & e < 0 \\ \tau|e|, & e \geq 0. \end{cases}$$

Accuracy comparison via expected loss: $E(L(e))$, e.g. $E(e^2)$

How to choose a loss function?

Does the choice matter for accuracy rankings?

This Paper's Point-Forecast Accuracy Comparison: Works Directly From First Principles

Compare:

$F(e)$ (c.d.f. of e)

vs.

$F^*(e)$ (c.d.f. of perfect forecast),

where

$$F^*(e) = \begin{cases} 0, & e < 0 \\ 1, & e \geq 0. \end{cases}$$

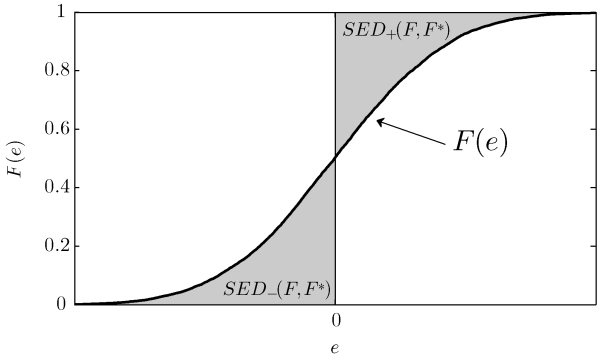
"Unit step function at zero"

Stochastic Error Distance (*SED*)

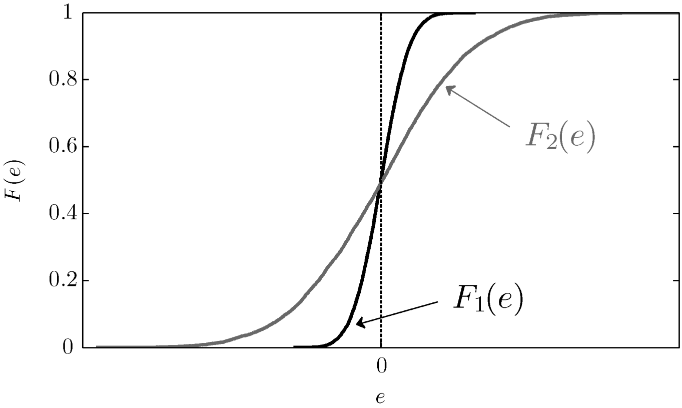
$$SED(F, F^*) = \int_{-\infty}^{\infty} |F(e) - F^*(e)| \, de$$

$$= \int_{-\infty}^{0} F(e) \, de + \int_{0}^{\infty} [1 - F(e)] \, de$$

$$= SED(F, F^*)_{-} + SED(F, F^*)_{+}$$



Example: Two Forecast Error Distributions



Under the *SED* criterion, we prefer F_1 to F_2.

SED and Expected Absolute Loss

$$SED(F, F^*) = \int_{-\infty}^{\infty} |F(e) - F^*(e)| \, de$$

Proposition (Equivalence of *SED* and Expected Absolute Loss):

If e is a forecast error with cumulative distribution function $F(e)$, such that $E(|e|) < \infty$, then SED equals expected absolute loss:

$$SED(F, F^*) = E(|e|).$$

SED accuracy evaluation is *MAE* accuracy evaluation!

Weighted Stochastic Error Distance (*WSED*)

$$WSED(F, F^*; \tau) = 2(1-\tau)SED(F, F^*)_- + 2\tau SED(F, F^*)_+,$$

where $\tau \in [0, 1]$.

WSED and Expected Lin-Lin Loss

Proposition (Equivalence of *WSED* and Expected Lin-Lin Loss):

If e is a forecast error with cumulative distribution function $F(e)$, such that $E(|e|) < \infty$, then WSED equals expected lin-lin loss:

$$\begin{aligned} WSED(F, F^*; \tau) &= 2(1-\tau)\int_{-\infty}^{0} F(e)\, de + 2\tau \int_{0}^{\infty} [1 - F(e)]\, de \\ &= 2E(L_\tau(e)), \end{aligned}$$

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

$$L_\tau(e) = \begin{cases} (1-\tau)|e|, & e < 0 \\ \tau|e|, & e \geq 0. \end{cases}$$

Generalized Weighted Stochastic Error Distance (*GWSED*)

$$GWSED(F, F^*; p, w) = \int |F(e) - F^*(e)|^p \, w(e) \, de,$$

where $p > 0$.

SED and *WSED* are nested special cases:

- $p = 1$ and $w(e) = 1 \; \forall \; e$ produces *SED*.
- $p = 1$ and
$$w(e) = \begin{cases} 2(1-\tau), & e < 0 \\ 2\tau, & e \geq 0 \end{cases}$$
produces *WSED*.
- Other choices of p and $w(e)$?

GWSED and Expected Loss: A Complete Characterization

$$GWSED(F, F^*; p, w) = \int |F(e) - F^*(e)|^p \, w(e) \, de$$

Proposition (Equiv. of $GWSED\left(F, F^*; 1, \left|\frac{dL(e)}{de}\right|\right)$ and $E(L(e))$):

Suppose that $L(e)$ is piecewise differentiable with $dL(e)/de > 0$ for $e > 0$ and $dL(e)/de < 0$ for $e < 0$, and suppose also that $F(e)$ and $L(e)$ satisfy $F(e)L(e) \to 0$ as $e \to -\infty$ and $(1 - F(e))L(e) \to 0$ as $e \to \infty$. Then:

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

Connections I: Cramér-von Mises Divergence

$GWSED(F, F^*; 2, f(e))$ is Cramér-von Mises divergence:

$$CVM(F^*, F) = \int |F^*(e) - F(e)|^2 f(e) de$$
$$= -F(0)(1 - F(0)) + \frac{1}{3}$$

$CVM(F^*, F)$ is minimized at $F(0) = \frac{1}{2}$.

That is, like $SED(F, F^*)$,
$CVM(F^*, F)$ is minimized by the conditional-median forecast.

Connections II: Kolmogorov-Smirnov Distance

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

$KS(F, F^*)$ is minimized at $F(0) = \frac{1}{2}$,
as is $CVM(F^*, F)$.

That is, like $SED(F, F^*)$,
$KS(F, F^*)$ is minimized by the conditional-median forecast.

Penn
UNIVERSITY of PENNSYLVANIA

Practical Implications

Switch from *RMSE* to *MAE* for forecast accuracy rankings.

– But is it really important to make the switch?

– That is, will rankings really change?

– In general, yes!

MSE vs. *MAE* Rankings

In general, *MSE* and *MAE* rankings differ.

Simplest Gaussian environment:

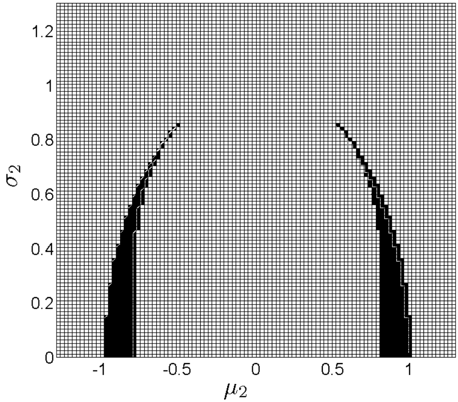
$$e \sim N\left(\mu, \sigma^2\right)$$

$$\implies \quad E(|e|) = \sigma\sqrt{2/\pi}\exp\left(-\frac{\mu^2}{2\sigma^2}\right) + \mu\left[1 - 2\Phi\left(-\frac{\mu}{\sigma}\right)\right]$$

Unbiased case ($\mu = 0$): $E(|e|) \propto \sigma$
MAE and *MSE* rankings must be identical

Biased case ($e_1 \sim N(0,1)$ and $e_2 \sim N(\mu_2, \sigma_2^2)$):
MAE and *MSE* rankings can diverge, even under normality.

MSE and *MAE* Divergence Regions, Gaussian Case



$$e_1 \sim N(0,1),\; e_2 \sim N(\mu_2, \sigma_2^2)$$

Conclusions

We have:

1. Approached forecast accuracy comparison from first principles. (*SED*.)
2. Arrived inescapably at *MAE* loss.
3. Clarified what it means to "select a loss function."
   (Select a $w(e)$ function in *GWSED*.)
4. Compared *SED* to *CVM* and *KS*.
   (Each is minimized by the conditional-median forecast.)
5. Shown that *MSE* forecast rankings do *not* match those of *SED*/*MAE* in general.