Assessing Point Forecast Accuracy by Stochastic Error Distance

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

June 29, 2017

Point Forecast Accuracy Comparison

Realization and forecast:

$$y, \ \hat{y}$$

Error:

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

Loss:

$$L(e)$$

$$L(0) = 0 \text{ and } L(e) \geq 0, \ \forall e$$

e.g. $L(e) = e^2$ (squared-error loss), $L(e) = |e|$ (absolute-error loss)

Accuracy comparison via expected loss:

$$E(L(e))$$

e.g., mean-squared error (MSE), mean-absolute error (MAE)

"The Big Three"

Squared-Error Loss: $L(e) = e^2$

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

"Check"-error loss:

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

(Of course $L_{.5}(e) = |e|$)

Dealing with Loss Function Choice

Avoidance
(perhaps via stochastic dominance)

– But that's special and rare...

Introspection in specific cases
(perhaps via firm-level engineering considerations)

– But that's special and rare...

Introspection in general

– But that seems really hard...

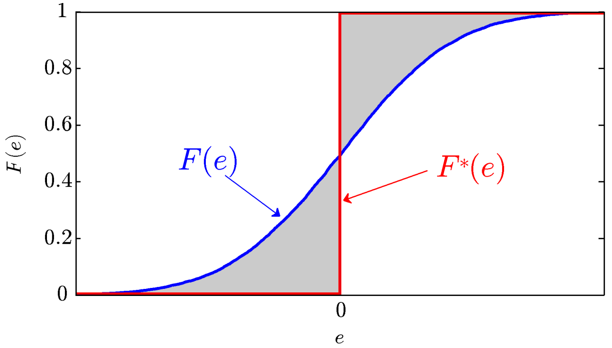
Introspection: A Loss Function Based on First Principles

Compare:

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

vs.

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

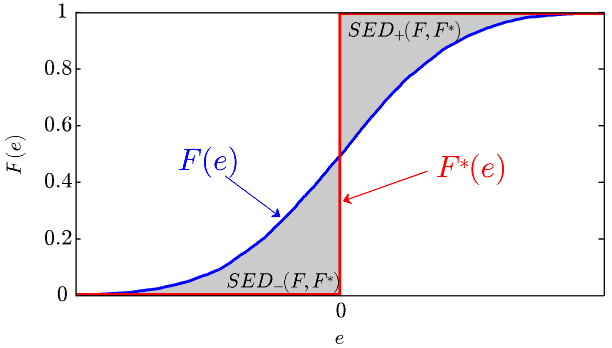


Stochastic Error Distance (*SED*)

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

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

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



SED and Expected Absolute Loss

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

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*) and Expected Check-Error Loss

$$WSED(F, F^*; \tau) = (1-\tau) SED(F, F^*)_- + \tau SED(F, F^*)_+, \ \tau \in [0, 1]$$

Weighted Stochastic Error Distance (*WSED*) and Expected Check-Error Loss

$$WSED(F, F^*; \tau) = (1-\tau)SED(F, F^*)_- + \tau SED(F, F^*)_+, \ \tau \in [0, 1]$$

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

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

$$WSED(F, F^*; \tau) = E(L_\tau(e)),$$

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

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

“The Big Three”, Redux

Squared-error loss

Absolute-error loss

Check-error loss

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 $w(e)$?

Penn
University of Pennsylvania

GWSED and Expected Loss: A Complete Characterization

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

GWSED and Expected Loss: A Complete Characterization

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

Proposition

(Equivalence 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:

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

Cramér Distance

$GWSED(F, F^*; 2, 1)$ is Cramér distance:
(Mallows, Monge-Kantorovich, earth-movers, ...)

$$C(F,\ F^*) = \int_{-\infty}^{\infty} \left[F(e) - F^*(e)\right]^2 de$$

$$= SED(F, F^*) - \int_{-\infty}^{\infty} F(e)(1 - F(e))\, de.$$

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.

Kolmogorov-Smirnov Distance

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

$KS(F, F^*)$ 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.

"The Big Three", Redux, Redux

Squared-error loss

Absolute-error loss

Check-error loss

Practical Conclusions/Implications

– Use *MAE* for forecast accuracy rankings.

– Recognize that selection of a loss function is selection of a *GSED* weighting function.

– Use the *GSED* loss representation to make new progress on old questions.

(e.g., When will *MSE* and *MAE* accuracy rankings match?)

Addendum: Ranking Forecasts Under MSE vs. MAE

General Gaussian environment $\left(e \sim N\left(\mu, \sigma^2\right)\right)$:

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

So *MAE* and *MSE* rankings can diverge, *even under normality*.

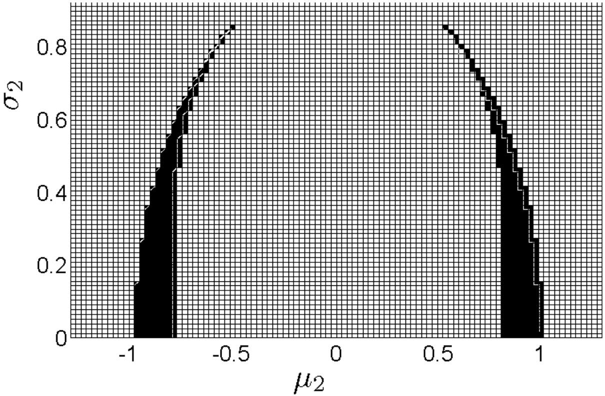
– Very little is known, even under normality.

Unbiased Gaussian environment $\left(e \sim N\left(0, \sigma^2\right)\right)$:

$$E(|e|) \propto \sigma$$

So *MAE* and *MSE* rankings must be identical.

MSE and *MAE* Divergence Regions (Black)



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