# Assessing Point Forecast Accuracy by Stochastic Error Distance

Francis X. Diebold and Minchul Shin
University of Pennsylvania

November 15, 2014

# Point Forecast Accuracy Comparison

Traditional:

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

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

Expected loss: $E(L(e))$, e.g. $E(e^2)$

This Paper:

We assess by comparing $F(e)$, the c.d.f. of $e$, and $F^*(e)$, where

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

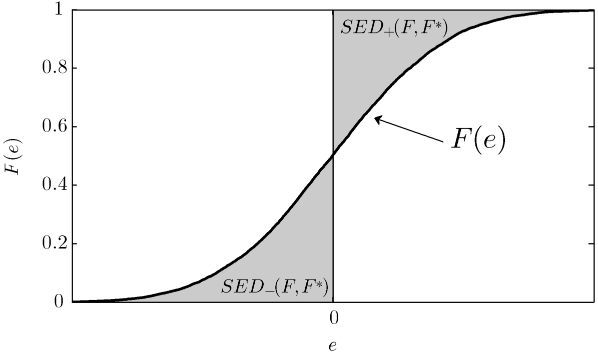
Penn
University of Pennsylvania

# Stochastic Error Distance (*SED*)

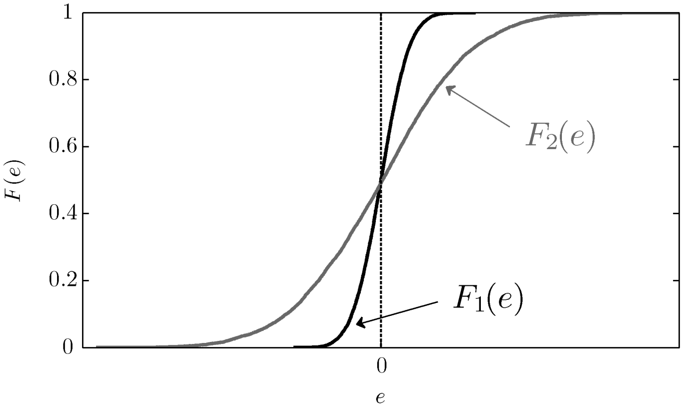
$$SED(F, F^*) = \int_{-\infty}^{\infty} |F(e) - F^*(e)|\, de$$

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

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



# 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:*

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

# 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 $w(e)$?

Penn
University of Pennsylvania

# *GWSED* and Expected Loss: A Complete Characterization

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

**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: Cramér-von Mises Divergence

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

$$\begin{aligned} CVM(F^*, F) &= \int |F^*(e) - F(e)|^2 f(e) de \\ &= -F(0)(1 - F(0)) + \frac{1}{3} \\ &\geq \frac{1}{12}, \end{aligned}$$

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

Hence, like $SED(F, F^*)$, $CVM(F^*, F)$ ranks forecasts according to expected absolute error.

# Connections: Kolmogorov-Smirnov Distance

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

where the lower bound is achieved at $F(0) = \frac{1}{2}$ as in the $CVM(F^*, F)$ case.

Hence, like $SED(F, F^*)$ and $CVM(F^*, F)$, $KS(F, F^*)$ ranks forecasts according to expected absolute error.

# Concluding Remarks

We have:

# Concluding Remarks

We have:

1. Stayed within the $E(L)$ framework (there is no escaping). ✓

# Concluding Remarks

We have:

1. Stayed within the $E(L)$ framework (there is no escaping). ✓
2. Established (surprising) primacy of absolute and lin-lin loss. ✓

Penn
University of Pennsylvania

# Concluding Remarks

We have:

1. Stayed within the $E(L)$ framework (there is no escaping). ✓
2. Established (surprising) primacy of absolute and lin-lin loss. ✓
3. Clarified what it means to "select a loss function." ✓

# Concluding Remarks

We have:

1. Stayed within the $E(L)$ framework (there is no escaping). ✓
2. Established (surprising) primacy of absolute and lin-lin loss. ✓
3. Clarified what it means to "select a loss function." ✓
4. Built bridges to *CVM* divergence, *KS* distance, etc. ✓

# Concluding Remarks

We have:

1. Stayed within the $E(L)$ framework (there is no escaping). ✓
2. Established (surprising) primacy of absolute and lin-lin loss. ✓
3. Clarified what it means to "select a loss function." ✓
4. Built bridges to *CVM* divergence, *KS* distance, etc. ✓
5. Raised important related questions. ✓

# Question: When Do MAE and MSE Rankings Diverge?

Simplest Gaussian environment:

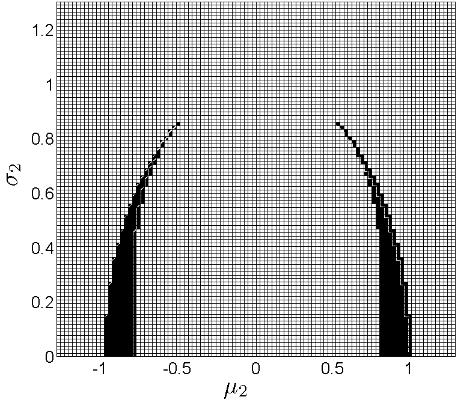
$$e \sim N\left(\mu, \sigma^2\right)$$

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

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




$e_1 \sim N(0, 1)$, $e_2 \sim N(\mu_2, \sigma_2^2)$
We show divergence regions in black.