

Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
<http://economics.sas.upenn.edu/pier>

PIER Working Paper 14-038

“Assessing Point Forecast Accuracy by
Stochastic Error Distance”

by

Francis X. Diebold and Minchul Shin

<http://ssrn.com/abstract=2521551>

Assessing Point Forecast Accuracy by Stochastic Error Distance

Francis X. Diebold

Minchul Shin

University of Pennsylvania

University of Pennsylvania

November 2, 2014

Abstract: We propose point forecast accuracy measures based directly on distance of the forecast-error c.d.f. from the unit step function at 0 (“stochastic error distance,” or *SED*). We provide a precise characterization of the relationship between *SED* and standard predictive loss functions, showing that all such loss functions can be written as weighted *SED*’s. The leading case is absolute-error loss, in which the *SED* weights are unity, establishing its primacy. Among other things, this suggests shifting attention away from conditional-mean forecasts and toward conditional-median forecasts.

Acknowledgments: For helpful discussion we are especially grateful to Lorenzo Braccini, Laura Liu, Andrew Patton and Allan Timmermann. We also thank Ross Askanazi, Xu Cheng, Valentina Corradi, Roger Koenker, Mai Li, Oliver Linton, Essie Maasoumi, Norm Swanson, Mark Watson, Tiemen Woutersen, and seminar participants at University of Pennsylvania and the European University Institute. The usual disclaimer applies.

Key words: Forecast accuracy, forecast evaluation, absolute-error loss, quadratic loss, squared-error loss

JEL codes: C53

Contact: fdiebold@sas.upenn.edu

1 Introduction

One often wants to evaluate (that is, rank) competing point forecasts by accuracy. Invariably one proceeds by ranking by expected loss, $E(L(e))$, where e is forecast error and the loss function $L(e)$ satisfies $L(0) = 0$ and $L(e) \geq 0, \forall e$.¹ Typically, however, little thought is given to the loss function $L(e)$. Instead, Gauss' centuries-old quadratic loss, $L(e) = e^2$, remains routinely invoked, primarily for mathematical convenience.

Against this background, in this paper we take a different approach, based on the entire distribution of e . In particular, recognizing that any reasonable loss function must satisfy $L(0) = 0$, we study accuracy measures based directly on the distance between $F(e)$, the c.d.f. of e , and $F^*(e)$, the unit step function at 0,²

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

We compare $F(e)$ to $F^*(e)$, and we favor forecasts that minimize the integrated absolute distance between the two, or “stochastic error distance” (*SED*). This approach turns out to yield useful insights with important practical implications.

We proceed as follows. In sections 2 and 3 we introduce unweighted and weighted *SED*, respectively, and in each case we characterize the relationship between *SED* minimization and expected loss minimization. In section 4 we generalize *SED* in a way that facilitates relating it to Cramer-von-Mises divergence, among other divergence measures, and we provide a complete characterization of the relationship between generalized *SED* minimization and expected loss minimization. We conclude in section 5.

¹More general representations are possible, which recognize that the actual and forecasted values (y and \hat{y} , say) need not enter loss only through their difference, which is the forecast error, $e = y - \hat{y}$. See, for example, Patton (2014) and the references therein. We could instead rank by $E(L(y, \hat{y}))$, where the loss function $L(y, \hat{y})$ satisfies $L(y, y) = 0$ and $L(y, \hat{y}) \geq 0, \forall y, \hat{y}$. In the vast majority of the literature, however, the simple $L(e)$ form is used, and we shall follow suit here.

²In an abuse of notation, throughout this paper we use “ $F(\cdot)$ ” to denote any cumulative density function. The meaning will be clear from context.

2 Ranking Forecasts by Stochastic Error Distance

We propose simply using the distribution of e directly, ranking forecasts by stochastic distance of $F(e)$ from $F^*(\cdot)$, the unit step function at 0. That is, we rank forecasts by

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

where smaller is better.³ We call $SED(F, F^*)$ the *stochastic error distance*. We can split the $SED(F, F^*)$ integral at the origin, yielding

$$\begin{aligned} SED(F, F^*) &= SED_-(F, F^*) + SED_+(F, F^*) \\ &= \int_{-\infty}^0 F(e) de + \int_0^{\infty} (1 - F(e)) de. \end{aligned} \tag{1}$$

Hence $SED(F, F^*)$ has both “integrated c.d.f.” and “integrated survival function” components. In Figure 1a we show $SED(F, F^*)$ and its components, and in Figure 1b we provide an example of two error distributions such that one would prefer F_1 to F_2 under $SED(F, F^*)$.

2.1 The Relation Between $SED(F, F^*)$ and $E(L(e))$

We motivated $SED(F, F^*)$ as directly appealing and intuitive. It turns out, moreover, that $SED(F, F^*)$ is intimately connected to one, and only one, traditionally-invoked loss function, and it is not quadratic. We begin with a lemma and then proceed to the main result.

Lemma 2.1

(i) Let y be a positive random variable such that $E(|y|) < \infty$. Then

$$E(y) = \int_0^{\infty} [1 - F(y)] dy,$$

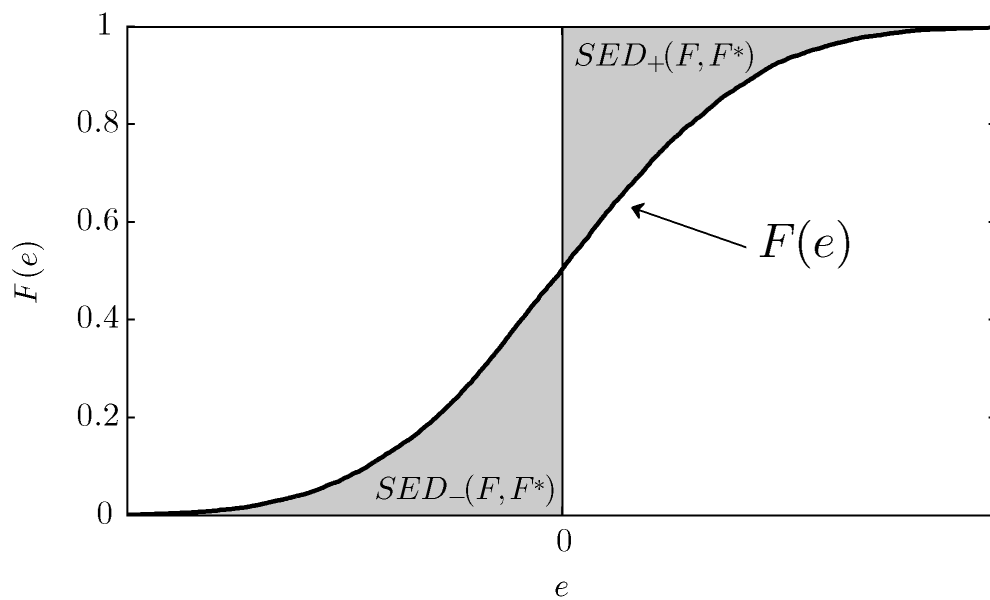
where $F(y)$ is the cumulative distribution function of y .

(ii) Let y be a negative random variable such that $E(|y|) < \infty$.⁴ Then

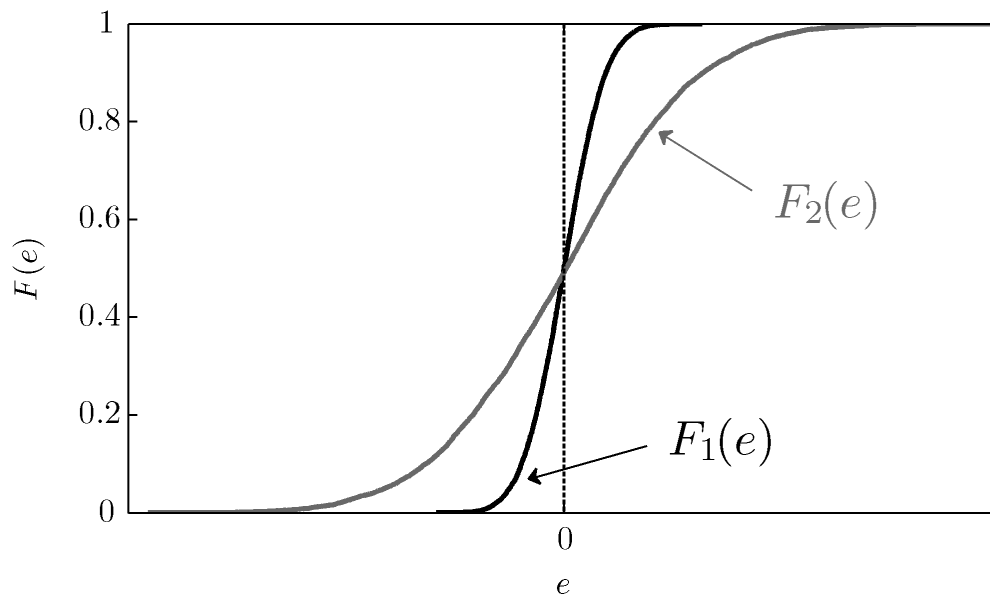
$$E(y) = - \int_{-\infty}^0 F(y) dy,$$

³Note that in the symmetric case $SED(F, F^*) = 2 \int_{-\infty}^0 F(e) de$.

⁴In yet another abuse of notation, we use “ y ” throughout to denote either a generic random variable or its realization.



(a) c.d.f. of e . Under the $SED(F, F^*)$ criterion, we prefer smaller $SED(F, F^*) = SED_-(F, F^*) + SED_+(F, F^*)$.



(b) Two forecast error distributions. Under the $SED(F, F^*)$ criterion, we prefer $F_1(e)$ to $F_2(e)$.

Figure 1: Stochastic Error Distance ($SED(F, F^*)$)

where $F(y)$ is the cumulative distribution function of y .

Proof We prove (i). Integrating by parts, we have

$$\begin{aligned}\int_0^c yf(y)dy &= -y[1 - F(y)]\Big|_0^c + \int_0^c [1 - F(y)]dy \\ &= -c(1 - F(c)) + \int_0^c [1 - F(y)]dy\end{aligned}$$

where $c > 0$. The first term goes to zero as $c \rightarrow \infty$, because

$$\begin{aligned}c(1 - F(c)) &= cP(Y > c) \\ &= c \int_c^\infty dP(y) \\ &= \int_c^\infty c dP(y) \\ &\leq \int_c^\infty y dP(y) \quad (\text{replacing } c \text{ with } y) \\ &= \int_0^\infty y dP(y) - \int_0^c y dP(y).\end{aligned}$$

But this converges to zero as $c \rightarrow \infty$, because

$$\int_0^\infty y dP(y) \leq \int_{-\infty}^\infty |y| dP(y) < \infty. \quad \blacksquare$$

The proof of (ii) proceeds identically, so we omit it. To the best of our knowledge, (i) has not appeared in the forecasting literature. It does appear, however, in the hazard and survival modeling literature, in whose jargon “expected lifetime equals the integrated survival function.”

We now arrive at a key result.

Proposition 2.2 (*Equivalence of SED and Expected Absolute Error Loss*)

For any forecast error e , with cumulative distribution function $F(e)$ such that $E(|e|) < \infty$, we have

$$SED(F, F^*) = \int_{-\infty}^0 F(e) de + \int_0^\infty [1 - F(e)] de = E(|e|). \quad (2)$$

That is, $SED(F, F^*)$ equals expected absolute loss for any error distribution.

Proof ⁵

$$\begin{aligned}
SED(F, F^*) &= SED_-(F, F^*) + SED_+(F, F^*) \\
&= \int_{-\infty}^0 F(e) de + \int_0^{\infty} (1 - F(e)) de \\
&= - \int_{-\infty}^0 ef(e) de + \int_0^{\infty} ef(e) de \quad (\text{by Lemma 2.1 (i) for } SED_- \text{ and (ii) for } SED_+) \\
&= \int_0^{\infty} ef(-e) de + \int_0^{\infty} ef(e) de \\
&= \int_0^{\infty} e(f(-e) + f(e)) de \\
&= \int_{-\infty}^{\infty} |e|f(e) de \\
&= E(|e|). \quad \blacksquare
\end{aligned}$$

Hence if one is comfortable with $SED(F, F^*)$ and wants to use it to evaluate forecast accuracy, then one must also be comfortable with expected absolute-error loss and want to use it to evaluate forecast accuracy. The two criteria are *identical*.

3 Weighted Stochastic Error Distance

In other circumstances, however, one may feel that the basic idea behind $SED(F, F^*)$ is appropriate, but that divergence of $F(\cdot)$ from $F^*(\cdot)$ on one side of the origin is more harmful than on the other. This leads to the idea of a *weighted SED* ($WSED$) criterion, given by a *weighted* sum of $SED_-(F, F^*)$ and $SED_+(F, F^*)$.

3.1 A Natural Generalization

In particular, let,

$$\begin{aligned}
WSED(F, F^*; \tau) &= 2(1 - \tau)SED(F, F^*)_- + 2\tau SED(F, F^*)_+ \\
&= 2(1 - \tau) \int_{-\infty}^0 F(e) de + 2\tau \int_0^{\infty} (1 - F(e)) de,
\end{aligned}$$

where $\tau \in (0, 1)$.⁶ The following result is immediate.

⁵We provide an alternative proof of Proposition 2.2 in Appendix A.

⁶Note that when $\tau = 0.5$, $WSED(F, F^*; \tau)$ is just $SED(F, F^*)$.

Proposition 3.1 (*Equivalence of WSED and Expected Lin-Lin Loss*)

For any forecast error e , with cumulative distribution function $F(e)$ such that $E(|e|) < \infty$, we have

$$WSED(F, F^*; \tau) = 2(1 - \tau) \int_{-\infty}^0 F(e) de + 2\tau \int_0^{\infty} [1 - F(e)] de = 2E(L_\tau(e)), \quad (3)$$

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

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

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

Proof We have

$$\begin{aligned} WSED(F, F^*; \tau) &= 2(1 - \tau) \int_{-\infty}^0 F(e) de + 2\tau \int_0^{\infty} (1 - F(e)) de \\ &= 2(1 - \tau) \int_{-\infty}^0 (-e)f_e(e) de + 2\tau \int_0^{\infty} ef_e(e) de \quad (\text{by Lemma 2.1}) \\ &= 2(1 - \tau) \int |e|1\{e \leq 0\}f_e(e) de + 2\tau \int |e|1\{e > 0\}f_e(e) de \\ &= 2 \int [(1 - \tau)|e|1\{e \leq 0\} + \tau|e|1\{e > 0\}]f_e(e) de \\ &= 2E(L_\tau(e)). \quad \blacksquare \end{aligned}$$

The loss function $L_\tau(e)$ appears in the forecasting literature as a convenient and simple potentially asymmetric loss function.⁷ It is often called “lin-lin” loss (i.e., linear on each side of the origin), and sometimes also called “check function” loss, again in reference to its shape. Importantly, it is the loss function underlying quantile regression; see Koenker (2005).

Remark 3.2 (*WSED and optimal prediction under asymmetric loss, I*).

Because $WSED(F, F^*; \tau)$ is proportional to expected lin-lin loss as established by Proposition 3.1, we are led inescapably to the insight that point forecast accuracy evaluation by $WSED(F, F^*; \tau)$ is actually point forecast accuracy evaluation by expected lin-lin loss. The

⁷See Christoffersen and Diebold (1997).

primacy of lin-lin loss in the $WSED(F, F^*; \tau)$ case, and the primacy of absolute error loss in the leading special case of $WSED(F, F^*; \tau)$ ($SED(F, F^*)$), emerges clearly.

Remark 3.3 (*WSED and optimal prediction under asymmetric loss, II*).

Patton and Timmermann (2007) suggest a different and fascinating route that also leads directly and exclusively to lin-lin loss. Building on the work of Christoffersen and Diebold (1997) on optimal prediction under asymmetric loss, they show that if loss $L(e)$ is homogeneous and the target variable y has no conditional moment dependence beyond the conditional variance, then the L -optimal forecast is always a conditional quantile of y . Hence under their conditions $WSED(F, F^*; \tau)$ loss is the only asymmetric loss function of relevance.

Our results and those of Patton and Timmermann are highly complementary but very different, not only in the perspective from which they are derived, but also in the results themselves. If, for example, y displays conditional moment dynamics beyond second-order, then the L -optimal forecast is generally *not* a conditional quantile (and characterizations in such cases remain elusive), whereas the $WSED(F, F^*; \tau)$ -optimal forecast is *always* a conditional quantile.

Remark 3.4 (*WSED as an estimation/combination criterion*).

$WSED(F, F^*; \tau)$, which of course includes $SED(F, F^*)$ as a special case, can be used as a forecast model estimation criterion. By Proposition 3.1, this amounts to estimation using quantile regression, with the relevant quantile governed by τ . When $\tau = 1/2$, the quantile regression estimator collapses to the least absolute deviations (LAD) estimator. Similarly, because the forecast combination problem is a regression problem (Granger and Ramanathan (1984)), forecast combination under $WSED(F, F^*; \tau)$ simply amounts to estimation of the combining equation using quantile regression, with the relevant quantile governed by τ .

4 Generalized Weighted Stochastic Error Distance

As always let $F(e)$ be the forecast error c.d.f., and let $F^*(e)$ be the unit step function at zero. Now consider the following generalized weighted stochastic error distance ($GWSED$) measure:

$$GWSED(F, F^*; p, w) = \int |F(e) - F^*(e)|^p w(e) de, \quad (4)$$

where $p > 0$. All of our stochastic error distance measures are of this form. When $p = 1$ and $w(x) = 1 \forall x$, we have $SED(F, F^*)$, and when $p = 1$ and

$$w(x) = \begin{cases} 2(1 - \tau), & x < 0 \\ 2\tau, & x \geq 0, \end{cases}$$

we have $WSED(F, F^*; \tau)$. The $GWSED(F, F^*; p, w)$ representation facilitates comparisons of $WSED(F, F^*; \tau)$ to other possibilities that emerge for alternative choices of p and/or $w(\cdot)$.

4.1 Connections Between $GWSED(F, F^*; p, w)$ and Other Distance and Divergence Measures

Several connections emerge.

4.1.1 Cramér Distance

When $p = 2$ and $w(x) = 1$, $GWSED(F, F^*; p, w)$ is Cramér distance, also known as Mallows distance, or Monge-Kantorovich distance, or earth-movers distance; see Levina and Bickel (2001). Closely related, moreover, are the “energy distance” used in higher dimensions (e.g., Székely and Rizzo (2013)) and the “continuous ranked probability score” of Gneiting and Raftery (2007).⁸

We can decompose Cramér distance as

$$\begin{aligned} \int_{-\infty}^{\infty} [F(e) - F^*(e)]^2 de &= \int [F(e)(1 - F^*(e)) + (1 - F(e))F^*(e) \\ &\quad - F(e)(1 - F(e)) - F^*(e)(1 - F^*(e))] de \\ &= \int_{-\infty}^0 F(e)de + \int_0^{\infty} [1 - F(e)] de - \int_{-\infty}^{\infty} F(e)(1 - F(e)) de \\ &= SED(F, F^*) - \int_{-\infty}^{\infty} F(e)(1 - F(e)) de, \end{aligned} \tag{5}$$

where e and e' are random variables independently and identically distributed with distribution function $F(\cdot)$. Equation (5) is particularly interesting insofar as it shows that Cramér distance is closely related to $SED(F, F^*)$, yet not exactly equal to it, due to the adjustment

⁸The continuous ranked probability score, however, is not used to rank point forecasts, but rather to assess density forecasts; see Gneiting and Raftery (2007).

term, $\int F(e)(1 - F(e)) de$. One can show that

$$\int F(e)(1 - F(e)) de = \frac{1}{2}E(|e - e'|),$$

where e' is a stochastic copy of e , revealing that the adjustment term, like the leading term, is a measure of forecast error variability.

4.1.2 Cramér-von Mises Divergence

When $p = 2$ and $w(e) = f(e)$, the density corresponding to $F(e)$, $GWSED(F, F^*; p, w)$ is Cramér-von Mises divergence,

$$CVM(F^*, F) = \int |F^*(e) - F(e)|^2 f(e) de. \quad (6)$$

Note that the weighting function $w(e)$ in Cramer-von Mises divergence $CVM(F^*, F)$ is distribution-specific, $w(e) = f(e)$.

We can decompose Cramer-von-Mises divergence as

$$\begin{aligned} CVM(F^*, F) &= \int |F^*(e) - F(e)|^2 f(e) de \\ &= \int [F(e)(1 - F^*(e)) + (1 - F(e))F^*(e) \\ &\quad - F(e)(1 - F(e)) - F^*(e)(1 - F^*(e))] f(e) de \\ &= \int_{R_-} F(e)f(e)de + \int_{R_+} (1 - F(e))f(e)de - \int_R F(e)(1 - F(e))f(e) de \\ &= \int_0^{F(0)} p dp + \int_{F(0)}^1 (1 - p) dp - \int_0^1 p(1 - p) dp \quad (\text{by change of variable, } e = F^{-1}(p)) \\ &= F(0)^2 - F(0) + \frac{1}{3} \\ &\geq \frac{1}{12}. \end{aligned}$$

Note that $CVM(F^*, F)$ achieves its lower bound of $1/12$ if and only if $F(0) = 1/2$, which implies that, like $SED(F, F^*)$, $CVM(F^*, F)$ ranks forecasts according to expected absolute error.

Remark 4.1 (*CVM, Kolmogorov-Smirnov distance, and expected absolute error*).

Kolmogorov-Smirnov distance is

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

Like $CVM(F^*, F)$, $KS(F, F^*)$ achieves its lower bound at $F(0) = 1/2$. Hence, like $SED(F, F^*)$ and CVM , $KS(F, F^*)$ ranks forecasts according to expected absolute error.

Remark 4.2 (*Directional properties of CVM*).

Although $CVM(F^*, F)$ is well-defined, $CVM(F, F^*)$ is not, because

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

where $f^*(e)$ is Dirac's delta.

Remark 4.3 (*Comparative directional properties of CVM and Kullback-Leibler divergence*).⁹

The Kullback-Leibler divergence $KL(F^*, F)$ between $F^*(e)$ and $F(e)$ is

$$KL(F^*, F) = \int \log \left(\frac{f^*(e)}{f(e)} \right) f^*(e) de,$$

where $f^*(x)$ and $f(x)$ are densities associated with distributions F^* and F . Unlike $CVM(F^*, F)$, $KL(F^*, F)$ does not fit in our $GWSED(F, F^*; p, w)$ framework as it is ill-defined in *both* directions.

4.2 A Complete Characterization

Equivalence of $GWSED(F, F^*; p, w)$ 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 proposition.

Proposition 4.4 (*Equivalence of GWSED minimization and $E(L(e))$ minimization*)

⁹There are of course many other distance/divergence measures, exploration of which is beyond the scope of this paper. On Hellinger distance, for example, see Maasoumi (1993).

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 $F(e)L(e) \rightarrow 0$ as $e \rightarrow -\infty$ and $(1 - F(e))L(e) \rightarrow 0$ as $e \rightarrow \infty$. Then

$$\int_{-\infty}^{\infty} |F(e) - F^*(e)| \left| \frac{dL(e)}{de} \right| de = E(L(e)). \quad (7)$$

That is, minimization of $GWSED(F, F^*; p, w)$ when $p = 1$ and $w(e) = |dL(e)/de|$ corresponds to minimization of expected loss $E(L(e))$.

Proof

$$\begin{aligned} \int_{-\infty}^{\infty} |F(e) - F^*(e)| \left| \frac{dL(e)}{de} \right| de &= - \int_{-\infty}^0 F(e) \frac{dL(e)}{de} de + \int_0^{\infty} (1 - F(e)) \frac{dL(e)}{de} de \\ &= \int_{-\infty}^0 f(e)L(e) de + \int_0^{\infty} f(e)L(e) de \\ &= \int_{-\infty}^{\infty} f(e)L(e) de \\ &= E[L(e)], \end{aligned}$$

where we obtain the second line by integrating by parts and exploiting the the assumptions on $L(e)$ and $F(e)$. In particular,

$$\int_{-\infty}^0 F(e) \frac{dL(e)}{de} de = F(e)L(e) \Big|_{-\infty}^0 - \int_{-\infty}^0 f(e)L(e) de,$$

by integration by parts, but the first term is zero because $F(e)L(e) \rightarrow 0$ as $e \rightarrow -\infty$, and similarly,

$$\int_0^{\infty} (1 - F(e)) \frac{dL(e)}{de} de = (1 - F(e))L(e) \Big|_0^{\infty} + \int_0^{\infty} f(e)L(e) de,$$

again by integration by parts, and again the first term is zero because $(1 - F(e))L(e) \rightarrow 0$ as $e \rightarrow \infty$. ■

Remark 4.5 (*GWSED weightings other than those corresponding to WSED and SED*). Note that the $E(L(e))$ minimizers that match various $GWSED(F, F^*; p, w)$ minimizers generally correspond to non-standard and intractable loss functions $L(e)$ in all cases but the ones we have emphasized, namely $WSED(F, F^*; \tau)$ and its leading case $SED(F, F^*)$.

Remark 4.6 (*The GWSED weighting that produces quadratic loss*).

The weighting function in (7) that produces expected squared-error loss ($E(L(e)) = E(e^2)$) is immediately

$$\left| \frac{dL(e)}{de} \right| = |2e|.$$

It is not obvious why one would generally want to adopt such a weighting, with $e = e^*$ weighted e^* times more than $e = 1$, for any reason other than mathematical convenience.

Remark 4.7 (*Relationship between GWSED and Elliott et al. (2005) loss*).

$GWSED(F, F^*; p, w)$ (4) resembles the Elliott et al. (2005) (ETK) loss function,

$$L_{ETK}(e; p, \alpha) = |e|^p (\alpha + (1 - 2\alpha)I(e < 0)).$$

However, it differs fundamentally in that $GWSED(F, F^*; p, w)$ is based on *distributional* distance, $|F - F^*|$, whereas ETK loss is based on the usual *forecast error* distance, $(y - \hat{y})$. Ultimately, ETK loss is a special case of $GWSED(F, F^*; p, w)$, corresponding to a particular choice of exponent p and weighting function $w(e)$, as per Proposition 4.4, as are *all* $L(e)$ loss functions that satisfy the regularity conditions of the proposition.

5 Conclusions and Directions for Future Research

Starting from first principles, we have proposed and explored several “stochastic error distance” (SED) measures of point forecast accuracy, based directly on the distance between the forecast-error c.d.f. and the unit step function at 0. SED -type criteria sharply focus attention on a *particular* loss function, absolute loss (and its lin-lin generalization), as opposed to the ubiquitous quadratic loss, or anything else. Our results elevate the status of absolute and lin-lin loss for both point forecast evaluation and for estimation.

Several interesting directions for future research are apparent. One direction concerns multivariate extensions, in which case it’s not clear how to define the unit step function at zero, $F^*(e)$. Consider, for example, the bivariate case, in which the forecast error is $e = (e_1, e_2)'$. It seems clear that we want $F^*(e) = 0$ when both errors are negative and $F^*(e) = 1$ when both are positive, but it’s not clear what to do when the signs diverge.

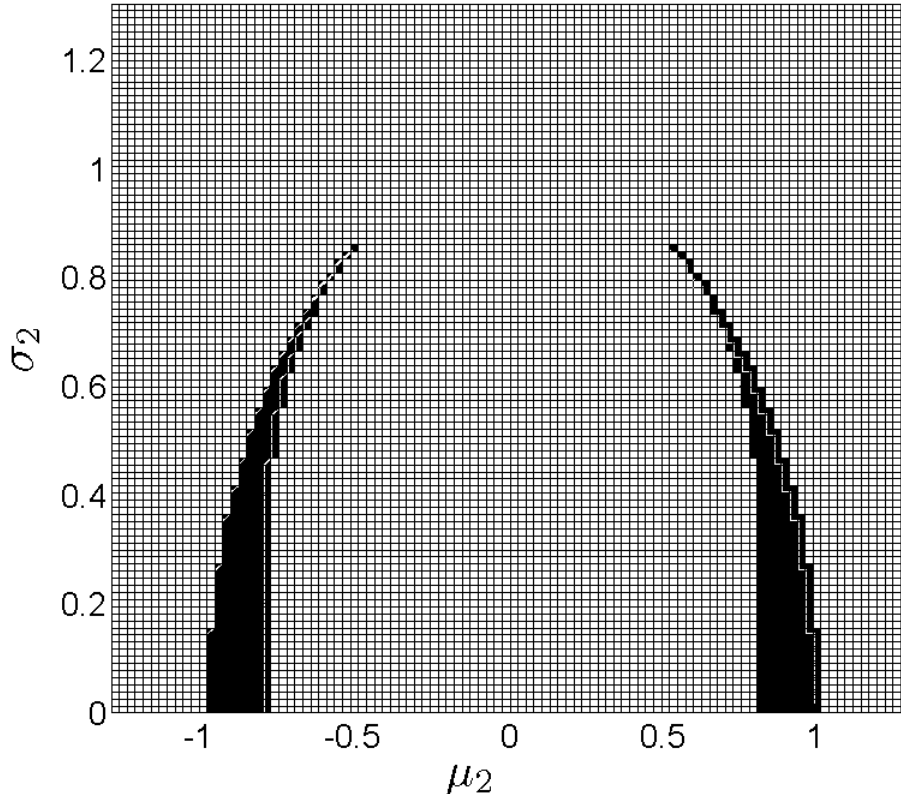


Figure 2: Absolute-Error Loss vs. Squared-Error Loss, $e_1 \sim N(0, 1)$, $e_2 \sim N(\mu_2, \sigma_2^2)$. We show the disagreement region in black.

Another direction concerns determining conditions under which squared-error and absolute-error loss disagree. That is, our results argue for the routine use of absolute-error loss and its lin-lin generalization, but the importance of the difference between ranking forecasts by absolute-error loss vs. other loss functions, and in particular, absolute-error loss vs. squared-error loss, remains to be explored.¹⁰

In certain cases the answer is clear. If, for example, forecast errors are Gaussian, $e \sim N(\mu, \sigma^2)$, then $|e|$ follows the folded normal distribution with mean

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

Hence for unbiased forecasts ($\mu = 0$) we have $E(|e|) \propto \sigma$, so that absolute and quadratic loss rankings must be identical. But even in the restrictive Gaussian case the rankings can

¹⁰Patton (2014) performs some related, but nevertheless different, explorations.

diverge if one (or both) of the forecasts are biased. Consider, for example, two forecast errors, $e_1 \sim N(0, 1)$ and $e_2 \sim N(\mu_2, \sigma_2^2)$, with $\mu_2 \in [-1.3, 1.3]$ and $\sigma_2 \in (0, 1.3]$. By simulation we identify situations where absolute-error and squared-error rankings diverge, which we show in Figure 2. The regions are not large, but they are certainly not negligible. We conjecture, moreover, that divergences may be much more common in *non*-Gaussian situations involving asymmetry and/or fat tails.

Appendices

A Alternative Proof of Proposition 2.2

Here we supply a different and shorter, if less instructive, proof.

Proposition

$$E(|e|) = \int_0^\infty [1 - F(e)] de = SED(F, F^*).$$

Proof

$$\begin{aligned} SED(F, F^*) &= - \int_0^c F^{-1}(p) dp + \int_c^1 F^{-1}(p) dp \quad (\text{where } c = F(0)) \\ &= \int_{-\infty}^0 -ef(e) de + \int_0^\infty ef(e) de \quad (\text{change of variables with } p = F(e)) \\ &= \int_{-\infty}^\infty |e|f(e) de \\ &= E(|e|). \quad \blacksquare \end{aligned}$$

References

- Christoffersen, P.F. and F.X. Diebold (1997), “Optimal Prediction Under Asymmetric Loss,” *Econometric Theory*, 13, 808–817.
- Elliott, G., I. Komunjer, and A. Timmermann (2005), “Estimation and Testing of Forecast Rationality under Flexible Loss,” *Review of Economic Studies*, 72, 1107–1125.
- Gneiting, T. and A. E. Raftery (2007), “Strictly Proper Scoring Rules, Prediction, and Estimation,” *Journal of the American Statistical Association*, 102, 359–378.
- Granger, C.W.J. and R. Ramanathan (1984), “Improved Methods of Forecasting,” *Journal of Forecasting*, 3, 197–204.
- Koenker, R. (2005), *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press, 2005.
- Levina, E. and P. Bickel (2001), “The Earth Mover’s Distance is the Mallows Distance: Some Insights from Statistics,” *Proceedings Eighth IEEE International Conference on Computer Vision*, 251–256.
- Maasoumi, E. (1993), “A Compendium to Information Theory in Economics and Econometrics,” *Econometric Reviews*, 12, 137–181.
- Patton, A. J. (2014), “Comparing Possibly Misspecified Forecasts,” *Working paper*.
- Patton, A.J. and A. Timmermann (2007), “Testing Forecast Optimality Under Unknown Loss,” *Journal of the American Statistical Association*, 102, 1172–1184.
- Székeley, G.J. and M.L. Rizzo (2013), “Energy Statistics: A Class of Statistics Based on Distances,” *Journal of Statistical Planning and Inference*, 143, 1249–1272.