

# Ranking Forecasts by Stochastic Error Distance, Error Tolerance and Survival Information Risks

Omid M. Ardakani<sup>a</sup>, Nader Ebrahimi<sup>b</sup>, Ehsan S. Soofi<sup>c\*</sup>

<sup>a</sup>*Department of Economics, Armstrong State University, Savannah, Georgia, USA*

<sup>b</sup>*Division of Statistics, Northern Illinois University, DeKalb, Illinois, USA*

<sup>c</sup>*Sheldon B. Lubar School of Business and Center for Research on International Economics, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA*

## Abstract

The stochastic error distance (SED) representation of the mean absolute error and its weighted versions have been introduced recently. The SED facilitates establishing connections between ranking forecasts by the mean absolute error, the error variance, and error entropy. We introduce the representation of the mean residual absolute error (MRAE) function as a new weighted SED which includes a tolerance threshold for forecast error. Conditions for this measure to rank the forecasts equivalently with Shannon entropy are given. The global risk of this measure over all thresholds is the survival information risk (SIR) of the absolute error. The SED, MRAE, and SIR are illustrated for various error distributions and comparing empirical regression and times series forecast models.

Keywords: Convex order; dispersive order; entropy; forecast error; mean absolute error; mean residual function; survival function.

## 1 Introduction

Forecast distributions are usually evaluated according to the risk functions defined by expected values of various loss functions. The most commonly-used risk functions are mean squared error and the mean absolute error (MAE). Recently, Diebold and Shin (2014, 2015) introduced an interesting representation of the MAE in terms of the following  $L_1$  norm:

$$\begin{aligned} SED(F, F_0) &= \int_{-\infty}^{\infty} |F(e) - F_0(e)| de \\ &= E(|e|) \\ &= MAE(e), \end{aligned} \tag{1}$$

where  $SED(\cdot, \cdot)$  stands for “stochastic error distance”,  $F(e)$  is the probability distribution of the forecast error and

$$F_0(e) = \begin{cases} 0, & e < 0 \\ 1, & e \geq 0 \end{cases}$$

is the distribution of the ideal error-free forecast. Let  $e_k$  be the forecast error with distribution  $F_k, k = 1, 2$ . Clearly  $e_1$  is preferred to  $e_2$  when  $SED(F_1, F_0) \leq SED(F_2, F_0)$ , which will be denoted as  $e_1 \leq_{sed} e_2$ .

---

\*August 20, 2015

Diebold and Shin (2015) also defined weighted and more general versions of  $SED(F, F_0)$  in the following form:

$$SED_{p,w}(F, F_0) = \int_{-\infty}^{\infty} |F(e) - F_0(e)|^p w(e) de, \quad (2)$$

where  $p > 0$  and  $w(e)$  is a weight function. Diebold and Shin (2014) noted that the Cramér-von Mises divergence is  $SED_{2,f}(F, F_0)$ , where  $f$  is the probability density function (PDF) of  $F$ . They explored the connection between  $SED_{2,1}(F, F_0)$  and Cramér distance and noted a connection between SED and the Kolmogorov-Smirnov distance. They also noted that the Kullback-Leibler information divergence does not fit in the SED framework. Through an example, they illustrated that the mean squared error does not always fit in the SED framework.

This paper continues this line of research, via two objectives. Our first objective is to connect ranking forecasts by SED to the rankings by forecast variance and the Shannon entropy. This will be accomplished through a stronger variation ordering, known as the dispersive order and convex order. This exploration can be considered as continuation of Ebrahimi, Maasoumi, and Soofi (1999) who established conditions for the equivalence of entropy and variance orderings of probability distributions.

Our second objective is to offer the SED formulation of the mean residual absolute error (MRAE) function as a generalization of the MAE. This is accomplished by introducing a version of (2) which includes a forecast error tolerance threshold and provides a dynamic generalization of  $SED(F, F_0)$ . This formulation also identifies sufficient conditions for the forecast errors such that the weighted SED ranks forecasts similarly to the Shannon entropy of the forecasts. For a given tolerance threshold, the weighted  $SED(F, F_0)$  is a local risk function. Its global risk is found through averaging by the distribution of the threshold. This expedition leads us to a risk function which is an information measure, known as cumulative residual entropy (Rao et al. 2004), survival entropy (Zografos and Nadarajah 2005), and the entropy functional of the survival function (Asadi et al., 2014). The  $SED(F, F_0)$  framework reveals that this measure is an information measure, hence we call it the *survival information risk (SIR)*. Estimates of the SIR provide criteria for ranking forecast models. We compute an estimate based on the empirical survival function and illustrates its applications via ranking regression and some time series forecast models. These explorations are continuation of the works by several authors who have shown usefulness of reliability notions for economic problems; see Ebrahimi et al. (2014) and references therein.

Section 2 gives some sufficient conditions for the equivalence of ranking forecasts by SED and by entropy and variance. Section 3 shows the representation of MRAE a new weighted SED and gives some sufficient conditions for the equivalence of ranking forecasts by MRAE and the entropy. Section 4 presents the SIR as the global risk of MRAE. Section 5 illustrates empirical applications of SED, MRAE, and SIR to evaluation of regression and time series models.

## 2 SED, Variance, and Entropy

The error variance will be denoted by  $V(e) = E(e - E(e))^2$  and Shannon entropy of  $e$  with probability density function (PDF)  $f$  is defined by

$$H(e) = H(f) = - \int_{-\infty}^{\infty} f(e) \log f(e) de,$$

provided that the integral is finite. The orderings of forecast distributions by SED, variance, and entropy can be connected through two stronger dispersion orderings of random variables (distributions) defined as follows.

**Definition 1** Let  $e_k, k = 1, 2$  denote forecast errors with distributions  $F_k, k = 1, 2$ .

(a) The forecast error  $e_1$  is smaller than another forecast error  $e_2$  in dispersive order, denoted as  $e_1 \leq_{disp} e_2$ , if

$$F_1^{-1}(\beta) - F_1^{-1}(\alpha) \leq F_2^{-1}(\beta) - F_2^{-1}(\alpha), \text{ for all } 0 < \alpha < \beta < 1,$$

where  $F_k^{-1}(\alpha) = \sup\{e : F_k^{-1}(e) \leq \alpha\}$  is the  $\alpha$ th quantile.

(b) The forecast errors  $e_1$  is smaller than another forecast error  $e_2$  in convex order, denoted as  $e_1 \leq_{cx} e_2$ , if  $E[\phi(e_1)] \leq E[\phi(e_2)]$  for all convex functions  $\phi : \mathfrak{R} \rightarrow \mathfrak{R}$ .

The dispersive order  $e_1 \leq_{disp} e_2$  implies that  $V(e_1) \leq V(e_2)$ ,  $H(e_1) \leq H(e_2)$ , and  $E|e_1 - E(e_1)| \leq E|e_2 - E(e_2)|$ ; (see Oja 1981). Thus, under dispersive order, MAE, variance, and entropy rank forecast equally biased forecasts similarly.

It is easy to verify convex order. Let  $SC(h)$  be the number of sign changes of function  $h$ . Then  $e_1 \leq_{cx} e_2$  and any of the following conditions hold:

- (a)  $SC(f_2 - f_1) = 2$  and the sign sequence is  $+, -, +$ .
- (b)  $SC(F_2 - F_1) = 1$  and the sign sequence is  $+, -$ .

See Shaked and Shanthikumar (2007) for details.

The following Proposition provides implications of the convex order for the equivalence ranking of forecasts by SED, variance, and entropy.

**Proposition 1** Let  $e_k$  be forecast error with distribution  $F_k, k = 1, 2$ .

(a) If  $e_1 \leq_{cx} e_2$ , then  $e_1 \leq_{sed} e_2$ .

(b) If  $e_1 \leq_{cx} e_2$ , then  $e_1 \leq_{sed} e_2 \iff V(e_1) \leq V(e_2)$ .

(c) If  $e_1 \leq_{cx} e_2$  and  $f_2$  is log-concave, then  $e_1 \leq_{sed} e_2 \iff H(e_1) \leq H(e_2)$ .

**Proof.** (a) The convex order,  $e_1 \leq_{cx} e_2$  is represented by the following two integral relationships:

$$\int_{-\infty}^b F_1(u)du \leq \int_{-\infty}^b F_2(u)du \quad \text{for all } b, \quad (3)$$

$$\int_b^{\infty} (1 - F_1(u))du \leq \int_b^{\infty} (1 - F_2(u))du \quad \text{for all } b; \quad (4)$$

(Shaked and Shanthikumar 2007). Summing up (3) and (4) with  $b = 0$  we obtain

$$\begin{aligned} SED(F_1, F_0) &= \int_{-\infty}^0 F_1(u)du + \int_0^{\infty} (1 - F_1(u))du \\ &\leq \int_{-\infty}^0 F_2(u)du + \int_0^{\infty} (1 - F_2(u))du = SED(F_2, F_0). \end{aligned}$$

(b) This result is obtained from part (a) and the known result  $e_1 \leq_{cx} e_2 \implies V(e_1) \leq V(e_2)$ .

(c) This result is obtained from part (a) and the following result. If  $e_1 \leq_{cx} e_2$  and the PDF of  $e_2$  is log-concave, then  $H(e_1) \leq H(e_2)$  (Yu 1988).

From parts (b) and (c) of Proposition 1 we have the following Corollary.

**Corollary 1** *If  $e_1 \leq_{cx} e_2$  and  $f_2$  is log-concave, then the MAE, variance, and entropy rank forecasts similarly.*

The inequality (3) is equivalent to an ordering referred to as the increasing concave order which with strict inequality at some  $b$  gives the second order stochastic dominance. The inequality (4) is equivalent to an ordering referred to as the increasing convex order. Each of these two orderings is weaker than and implied by the usual (first order) stochastic dominance. However, when  $E(X_1) = E(X_2)$ , then  $X_1 \leq_{cx} X_2$  if and only if one of the two relationships, (3) or (4), holds. Therefore, neither (3) nor (4) alone is sufficient for ranking the forecasts by  $SED(F, F_0)$ .

The dispersive order and/or convex order with log-concavity are sufficient, but not necessary for the similar rankings of the MAE, variance, and entropy. We illustrate this fact using two families of error distributions.

Both families considered include the normal model  $N(\mu, \sigma^2)$ . For unbiased (and equally biased forecasts), the forecast errors are dispersive ordered and convex ordered by the  $\sigma^2$ , and the PDF is log-concave. Thus the three measures rank the family similarly. For example for unbiased forecasts we have:

$$f(e) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}e^2}, \quad \text{Dispersive and convex ordered by } \sigma^2 \uparrow,$$

where  $\uparrow$  indicates the increasing order. The expressions for the three measures are well-known and order the family as follows:

$$MAE(e) = \sigma\sqrt{\frac{2}{\pi}} \uparrow, \quad V(e) = \sigma^2 \uparrow, \quad H(e) = \frac{1}{2} + \frac{1}{2}\log(2\pi\sigma^2) \uparrow.$$

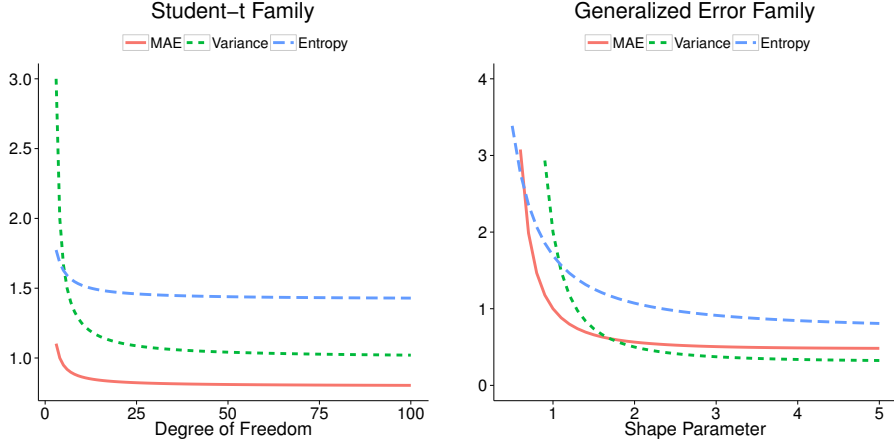


Figure 1: MAE, variance, and entropy of the Student- $t$  and Generalized Error Families.

## 2.1 Student- $t$ Family

Consider the Student- $t$  error model  $t(\mu, \sigma, \nu)$ , where  $\nu$  is the degrees of freedom. For a given  $\nu$ , the errors are ordered by the scale parameter as in the case of normal errors. For given  $\mu$  and  $\sigma$ , the errors are dispersive and convex ordered by the degrees of freedom. For example, for  $t(0, 1, \nu)$  we have:

$$f(e) = C_\nu \left(1 + \frac{e^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad \text{Dispersive and convex ordered by } \nu \downarrow,$$

where  $C_\nu = \frac{\Gamma(\nu/2 + 1/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)}$  is the normalizing factor. The PDF is not log-concave for all  $\nu$ , so parts (a) and (b) of Proposition 1 are applicable, but part (c) is not applicable to the entire family. However, due to the dispersive order, the three measures order the family similarly. These measures order the family as follows: Therefore,

$$\begin{aligned} MAE(e) &= \frac{2\nu C_\nu}{\nu-1}, \nu > 1 \downarrow, \quad V(e) = \frac{\nu}{\nu-2}, \nu > 2 \downarrow, \\ H(e) &= -\log C_\nu + \frac{\nu+1}{2} \left[ \psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) \right] \downarrow, \end{aligned}$$

where  $\psi(\cdot)$  is digamma function. The expressions for the variance and entropy of the  $t$  distribution are well-known. The expression for the MAE is the mean of folded- $t$  given by Psarakis and Panaretos (1990). Note that for  $\nu = 1$  the SED is not applicable and for  $\nu = 2$ , the variance is not defined, however the SED is applicable.

## 2.2 Generalized Error Family

This example illustrates that the dispersive and convex orders are sufficient, but not necessary for SED, variance, and entropy order forecasts similarly. The Generalized Error family

$GE(\mu, \sigma, \beta)$ , also known as the exponential power distribution family, and the PDF in the following form:

$$f(e) = \frac{\beta}{2\sigma\Gamma(1/\beta)} e^{-|(e-\mu)/\sigma|^\beta}, \beta > 0.$$

This distribution is the maximum entropy error model subject to the moment condition  $E|(e - \mu)/\sigma|^\beta$ . Specific cases include the Laplace distribution for  $\beta = 1$  and  $N(\mu, \sigma^2/2)$  for  $\beta = 2$ . For a given  $\beta$ , the errors are ordered by the scale parameter as in the case of normal errors. For the unbiased forecasts with equal scale parameters of the errors, the family is convex ordered by  $\beta \downarrow$  when,  $\beta \leq 2$  and when  $\beta_1 < 2, \beta_2 \geq 2$ . The PDF is log-concave for  $\beta \geq 1$ . It can also be shown that the family is not dispersive ordered by  $\beta$ . Yet for given  $\mu$  and  $\sigma$ , and all  $\beta > 0$ , the MAE, variance, and entropy order forecasts similarly. The expressions for the variance and entropy are known. The expression for the MAE can be obtained by noting that the distribution of  $|e|$  is a generalized gamma with shape parameters  $1/\beta$  and  $\beta$  and scale parameter  $\sigma$ . For example, for  $GE(0, 1, \beta)$  we have:

$$\begin{aligned} MAE(e) &= \frac{\Gamma(2/\beta)}{\Gamma(1/\beta)}, \beta > 0 \downarrow, \quad V(e) = \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}, \beta > 0 \downarrow, \\ H(e) &= \frac{1}{\beta} + \log \frac{2\Gamma(1/\beta)}{\beta}, \beta > 0 \downarrow. \end{aligned}$$

Figure 1 shows the plots of the three measures for the  $t$  family (left panel) and for the GE family (right panel). Because  $\nu$  and  $\beta$  determine thickness of the tails, the plots show similar patterns for the two families.

### 3 SED With Tolerance Threshold

Consider the weighted SED (2) with  $p = 1$  and

$$w_\tau(e) = \begin{cases} 0, & |e| \leq \tau \\ \frac{1}{P(|e| > \tau)}, & |e| > \tau, \end{cases} \quad (5)$$

where  $\tau$  is the error tolerance threshold. The forecast errors whose magnitudes are less than  $\tau$  are negligible.

Using (5) in (2) gives the following weighted SED:

$$\begin{aligned} WSED_\tau(F, F_0) &= \int_{-\infty}^{\infty} |F(e) - F_0(e)| w_\tau(e) de \\ &= \frac{1}{P(|e| > \tau)} \left[ \int_{-\infty}^{-\tau} |F(e) - F_0(e)| de + \int_{\tau}^{\infty} |F(e) - F_0(e)| de \right] \\ &= \frac{1}{P(|e| > \tau)} \left[ \int_{-\infty}^{-\tau} F(e) de + \int_{\tau}^{\infty} S(e) de \right] \\ &= \frac{1}{P(|e| > \tau)} E(e - \tau) \mathbf{1}(|e| > \tau) \\ &= MRAE(\tau), \end{aligned}$$

where  $\mathbf{1}(A)$  is the indicator function of set  $A$  and  $MRAE(\tau)$  is the mean residual function of  $|e|$  with the following representations:

$$MRAE(\tau) = E(|e| - \tau \mid |e| > \tau) \quad (6)$$

$$= \frac{\int_{\tau}^{\infty} S_{|e|}(|e|)d|e|}{S_{|e|}(\tau)} \quad (7)$$

$$= \frac{E(|e|) - \int_0^{\tau} S_{|e|}(|e|)d|e|}{S_{|e|}(\tau)}, \quad (8)$$

where  $S_{|e|}$  denotes the survival function. See Poynor (2010) for an exposition of the mean residual function.

$MRAE(\tau)$  is the dynamic generalization of the MAE,  $MRAE(0) = E(|e|)$ , hence  $WSED_{\tau}(F, F_0)$  is the dynamic generalization of  $SED(F, F_0)$ ; clearly,  $WSED_0(F, F_0) = SED(F, F_0)$ . For a given  $\tau \geq 0$ ,  $WSED_{\tau}(F, F_0)$  is the risk of the loss function  $L(e) = |e| - \tau \mid |e| > \tau$  according to which forecasts with error magnitudes below threshold  $\tau$  are not penalized.

The mean residual functions of distributions can be monotone (decreasing or increasing) or non-monotone. For representation of  $WSED_{\tau}(F, F_0)$ , the monotone model of a bounded PDF is meaningful. We give two examples.

**Example 1** Consider the normal error model  $f_e = N(0, \sigma^2)$ . The distribution of absolute error  $|e|$  is half-normal (folded-normal) at zero and the mean residual function is given by

$$MRAE(\tau) = \frac{\sigma \sqrt{\frac{2}{\pi}} e^{-\frac{\tau^2}{2\sigma^2}}}{1 - 2(\Phi(\tau/\sigma) - .5)} - \tau, \quad \tau \geq 0,$$

where  $\Phi(\cdot)$  is the standard normal CDF. Figure 2 shows plots of  $MRAE(\tau)$  and corresponding plots of PDF and CDF for two normal error models with  $\sigma = 1, 2$ . The CDF plot also includes the error distribution of the perfect forecast  $F_0$ . The total area between  $F_0$  and  $F$  gives the  $SED(F : F_0) = E(|e|)$ , which graphically display that  $N(0, 1)$  is preferred to  $N(0, 4)$ . For the normal models  $MRAE(\tau)$  is decreasing, which implies that the more tolerance is allowed, the lower will be the loss. Hence the MAE is the maximum loss,  $WSED_{\tau}(F, F_0) \leq SED(F, F_0)$  for all  $\tau \geq 0$ .

The mean residual function uniquely determines the distribution and provides a measure for ranking random variables. That is, for the nonnegative random variable  $|e|$ ,

$$S_{|e|}(x) = \frac{MRAE(0)}{MRAE(\tau)} \exp \left\{ - \int_0^x \frac{1}{MRAE(\tau)} d|e| \right\}.$$

The following example illustrates a well-known case.

**Example 2** Let  $MRAE(\tau) = A\tau + B$ ,  $A > -1, B > 0$ . Oakes and Dasu (1990) showed that the survival function corresponding to the linear MR is

$$S(|e|) = \left( \frac{B}{A|e| + B} \right)^{\frac{1}{A} + 1}, \quad |e| \geq 0. \quad (9)$$

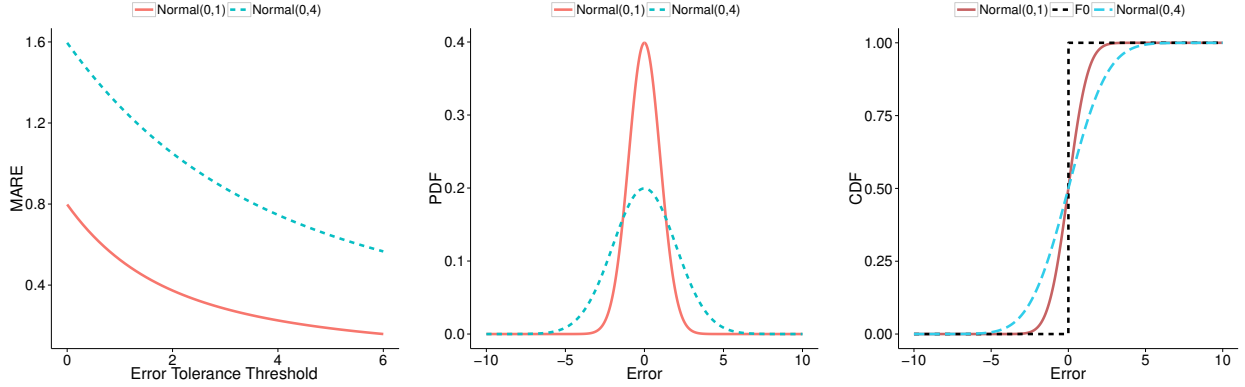


Figure 2: Plots of the mean residual absolute error and corresponding plots of PDF and CDF for two normal error models.

This is the survival function of the Generalized Pareto (GP) distribution, which includes three distributions.

- (a) For  $A = B > 0$ , (9) is the survival function of Pareto Type II distribution with PDF usually parameterized in terms of tail index as follows:

$$f_{\alpha}(|e|) = \frac{\alpha}{(1 + |e|)^{\alpha}}, \quad |e| \geq 0, \quad \alpha = \frac{1}{A} + 1 > 1.$$

The mean residual function is

$$MRAE(\tau) = \frac{\tau + 1}{\alpha - 1}, \quad \alpha > 1.$$

The left panel of Figure 3 shows the  $WSED_{\tau}(F, F_0)$  for  $A = 1, 2$  ( $\alpha = 2, 1.5$ ). The other two panels of Figure 3 show the PDF (middle panel) of the corresponding Double Pareto error distributions for  $\alpha = 1.5, 2$  and their distribution functions along with  $F_0$  (right panel). For the Pareto models  $MRAE(\tau)$  is increasing, which implies that the more tolerance is allowed, the higher will be the loss. Hence the MAE is the minimum loss,  $WSED_{\tau}(F, F_0) \geq SED(F, F_0)$  for all  $\tau \geq 0$ . We also should note that the variance is defined when  $\alpha > 2$ , so the errors having these distributions cannot be compared by the variance. However, the entropy is defined for these distributions and orders them similarly.

- (b) For  $A = 0$ ,  $MRAE(\tau) = B$  and (9) is the survival function of the exponential distribution and the error distribution is Laplace.
- (c) For  $-1 < A < 0$ , (9) is the survival function of a Beta distribution with unbounded PDF, which is not a suitable model for  $|e|$ .

In general,  $MRAE(\tau)$  is not available in a closed form. However, like the mean residual order of random variables,  $WSED_{\tau}(F, F_0)$  provides a stochastic distance for comparison of forecasts, denoted as  $e_1 \leq_{mrae} e_2$ . The results available for the mean residual order are applicable to ranking forecasts by  $SED(F, F_0)$ . The mean residual order can be easily verified by the following stronger orderings.



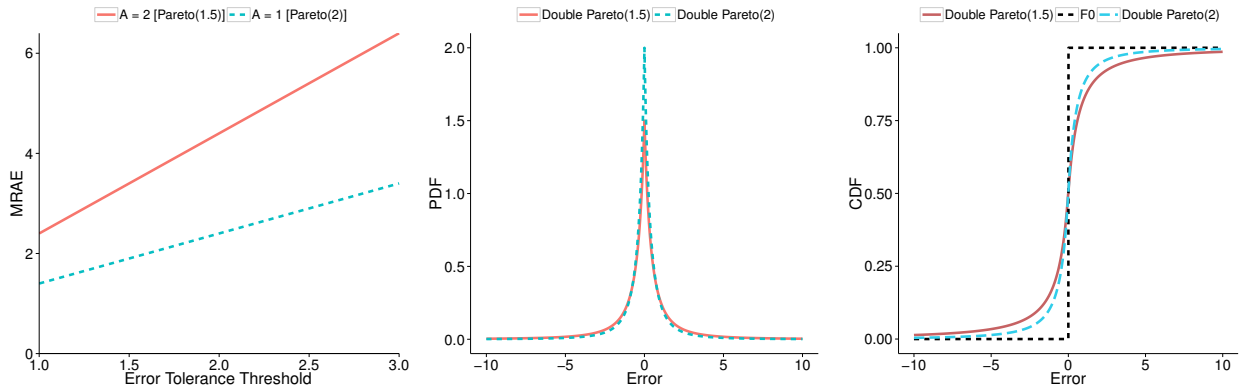


Figure 3: Linear mean residual function for absolute error and corresponding plots of PDF and CDF for two Double Pareto error distributions.

**Definition 2** Consider two nonnegative random variables  $X_k, k = 1, 2$  with distributions PDF  $f_k$  and hazard functions  $\lambda_k = \frac{f_k(x)}{S_k(x)}$ .

- (a)  $X_1$  is smaller than  $X_2$  in likelihood ratio order, denoted as  $X_1 \leq_{lr} X_2$ , if  $\frac{f_1(x)}{f_2(x)}$  is decreasing in  $x$  over the union of the supports of  $F_1$  and  $F_2$ .
- (b)  $X_1$  is smaller than  $X_2$  in hazard order, denoted as  $X_1 \leq_{hr} X_2$ , if  $\lambda_1 \geq \lambda_2$ .

The likelihood ratio order implies the hazard order and hazard order implies the mean residual order; see Shaked and Shanthikumar (2007) for details. The hazard rate order and a monotone PDF condition implies ordering by the Shannon entropy (Asadi et al. 2004). Thus we have the following result.

**Proposition 2** Let  $e_k, k = 1, 2$  be two forecast errors. If  $|e_1| \leq_{hr} |e_2|$  and  $f(|e_2|)$  is decreasing, then  $e_1 \leq_{mrae} e_2 \iff H(e_1) \leq H(e_2)$ .

The assumption of decreasing density for  $|e|$  is quite reasonable for the unbiased forecasts. For example, Proposition 2 applies when the distribution of  $|e|$  is half-normal, exponential, and Pareto with tail index larger than one.

## 4 Survival Information Risk

As a risk function,  $WSED_\tau(F, F_0)$  is conditional on the threshold. That is, for each  $\tau \geq 0$ ,  $WSED_\tau(F, F_0)$  is a local measure. The global risk of  $WSED_\tau(F, F_0)$  over  $\tau \geq 0$  gives the following measure:

$$\begin{aligned}
 E[WSED_\tau(F, F_0)] &= \int_0^\infty WSED_\tau(F, F_0) f_{|e|}(\tau) d\tau \\
 &= E[MRAE(\tau)] \\
 &= SIR(|e|),
 \end{aligned} \tag{10}$$

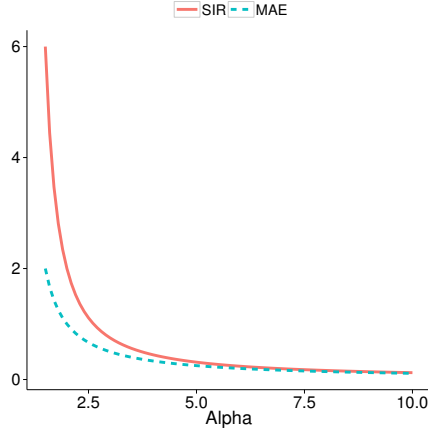


Figure 4: Survival information risk and MAE of Double Pareto error distribution.

where

$$SIR(|e|) = h(S_{|e|}) = - \int_0^\infty S_{|e|}(\tau) \log S_{|e|}(\tau) d\tau \geq 0. \quad (11)$$

The inequality becomes equality if and only if  $F(e) = F_0(e)$  almost everywhere. Proof of the nonnegativeness of  $h(S_{|e|})$  is given by Rao et al. (2004) and proof of (10) is given by Asadi and Zohrevand (2007).

The measure (11) is a well-known measure. Rao et al. (2004) introduced (11) as “a new measure of information” and as “an alternative measure of uncertainty in a random variable” to the Shannon entropy. They called it the cumulative residual entropy and illustrated that this measure is useful for applications in image processing. Zografos and Nadarajah (2005) studied extensions of (11) in terms of analogs of some generalizations of Shannon entropy which were referred to as the Survival Exponential Entropies. Asadi et al. (2014) referred to (11) as the entropy functional of the survival function because of its functional similarity to (??) but with an important conceptual distinction between the two measures. The global maximum of Shannon entropy in the discrete and continuous cases is the uniform (rectangular) PDF. This makes  $H(f)$  a measure of uncertainty in terms of the lack of concentration of  $f$ . But  $h(S)$  is a nonnegative concave function of  $S$  and its global maximum is attained by a rectangular survival function, a characterization of a degenerate distribution. However,  $h(S) = h(1 - F)$  is a nonnegative convex function of  $F$  and the degenerate distribution  $F_0$  is its global minimum. Thus, contrary to its functional similarity to the Shannon entropy which is a measure of divergence of  $F$  from the least concentrated distribution,  $h(S)$  is a measure of divergence of  $F$  from the most concentrated distribution,  $F_0$ . Hence,  $h(S)$  is an information measure, hereafter referred to as the *survival information risk*.

The following Example illustrates  $SIR(|e|)$ .

**Example 3** Consider the case of Example 2. The global risk of  $WSED_\tau(F, F_0)$  for the Pareto model is given by

$$SIR(|e|) = \alpha E^2(|e|) = \frac{\alpha}{(\alpha - 1)^2}, \quad \alpha = \frac{1}{A} + 1 > 1.$$

Figure 4 shows the plots of  $SIR(|e|)$  and the MAE of the Pareto model for  $|e|$ . It is seen that both measures are decreasing functions of the tail index parameter and  $SIR(|e|)$  dominates the MAE.

## 5 Empirical SED, MRAE, and SIR

Sample versions of SED, MRAE, and SIR provide criteria for evaluating empirical forecast models. Consider a sample of forecast errors  $e_1, \dots, e_n$  which are distributed according to  $F$  with a finite mean. There is no other restriction on  $F$ . This requirement is also important for estimating  $F$  in the representation (1) of  $E(|e|)$  by the sample version

$$MAE_n = \frac{1}{n} \sum_{i=1}^n |e_i|.$$

The the sample version of MRAE is defined by the empirical estimate of (7):

$$MRAE_n(\tau) = \frac{\int_{\tau}^{\infty} \hat{S}(u) du}{\hat{S}(|e|)} \cdot \mathbf{1}(\hat{S}(|e|) > 0),$$

where

$$\hat{S}(|e|) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(|e_i| > |e|)$$

is the empirical survival function. The the sample version of SIR is given by using the empirical survival function (11):

$$\begin{aligned} SIR_n = h(S_n) &= - \int_0^{\infty} \hat{S}(|e|) \log \hat{S}(|e|) d|e| \\ &= - \sum_{i=1}^n \int_{|e_{(i-1)}|}^{|e_{(i)}|} \hat{S}(|e|) \log \hat{S}(|e|) d|e| \\ &= - \sum_{i=1}^n (|e_{(i)}| - |e_{(i-1)}|) \left(1 - \frac{i}{n}\right) \log \left(1 - \frac{i}{n}\right), \end{aligned}$$

where  $0 = |e_{(0)}| < |e_{(1)}| < \dots < |e_{(n)}|$  are the ordered absolute errors. The empirical estimator of MRAE is uniformly strong consistent (Yang 1978, Lemma 2) and  $SIR_n$  is almost surely consistent (Rao et al., 2004, Theorem 9).

Next we illustrate applications of  $MAE_n$ ,  $MRAE_n$  and  $SIR_n$ .

### 5.1 Regression Forecasting

This example uses a subset of variables chosen from the Stock Liquidity data described in Frees (1996, p. 263). The variables chosen for the purpose of illustration are as follows. The trading volume for a three month period in millions shares ( $Y$ ) to be predicted by the price ( $X_1$ ), the number of shares outstanding at the end of the three month period in millions ( $X_2$ ), and the market value in billion dollars ( $X_3$ ).

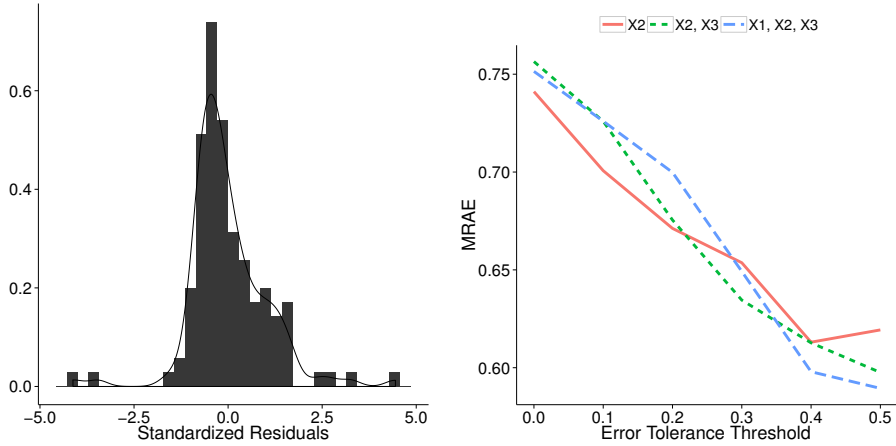


Figure 5:  $MRAE_n$  of three regression models for financial data.

We compare the top three subsets of predictors selected by AIC or BIC. The forecast errors are found by leave-one-out cross-validation forecast  $\tilde{y}_{(i)}$  of the deleted observation  $y_i$ . The raw residuals  $e_i = y_i - \tilde{y}_{(i)}$ ,  $i = 1, \dots, n$  have different variances, so we use Studentized residuals given by

$$se_i = \frac{e_i}{s(e_i)}.$$

Table 1 list the subsets of predictors. The left panel of Figure 5 shows histogram of  $se_i$  for the full model  $X_1, X_2, X_3$  (the histograms for the other two models were similar, thus are not shown). The right panel of Figure 5 shows the  $MRAE_n$  plots of the three regression models with  $\tau \leq .5$  (the scale of thresholds is standard deviation). These plots are produced based on between  $n = 123$  observations at  $\tau = 0$  (full sample) and at least 63 observations at  $\tau = .5$ . The plots indicate decreasing patterns, where none of the models uniformly dominates another model. For  $\tau > .12$ , the best model according to the  $MAE_n$  ( $X_1, X_2$ ) loses to the second best model ( $X_2$ ) and for  $\tau > .25$ , the best model becomes the worse.

Table 1 shows  $MAE_n$  and  $SIR_n$  for the three models. The numbers in the parentheses indicate the top three choices according to each criteria. In this example,  $MAE_n$  and  $SIR_n$  rank the models differently. For all three models  $MAE_n < SIR_n$ .

The following example illustrates  $MAE_n$ ,  $MRAE_n$  and  $SIR_n$  of time series forecasts.

Table 1:  $SIR_n$ ,  $MAE_n$  forecasting loss of regression models selected by AIC and BIC.

Subset	Models selected by		Forecast assessment	
	AIC	BIC	$MAE_n$	$SIR_n$
$X_2$	470.8 (3)	476.4 (1)	.751 (1)	.652 (3)
$X_2, X_3$	468.0 (1)	476.5 (2)	.789 (3)	.625 (2)
$X_1, X_2, X_3$	468.6 (2)	479.8 (3)	.756 (2)	.624 (1)

## 5.2 Time Series Forecasting

We illustrate applications to three univariate time series.

### 5.2.1 U.S. Inflation

The left column of Figure 6 pertains to the U.S. monthly inflation from January 1947 to December 2014 at an annual rate calculated by  $y_t = 1200 \ln(P_t/P_{t-1})$ , where  $P$  is the price index. The series consists of 815 observations. The upper-left panel shows the sequence plot of the data. We considered two models for the entire series: ARIMA(5,1,5) identified as the optimal model by an R program which first tests for the unit root and then selects the model using AIC (Hyndman and Khandakar 2008), and the random walk with drift drawn from the literature. We then used data from January 1947 to December 1999 for estimation and produced rolling sample one-step ahead forecasting for January 2000 to December 2014. This procedure provided 635 observations for estimation and 180 forecasts. The histogram of the standardized forecast errors of the ARIMA(5,1,5) is shown Figure 6. The lower left panel of Figure 6 shows the plots of  $MRAE_n$  for the two models for  $\tau \leq .5$ . These plots show that neither of the two models uniformly dominates the other. For  $\tau < .4$ , the ARIMA model dominates the random walk and  $\tau > .4$  the dominance is reversed.

### 5.2.2 Chemical Process Concentration

The middle column of Figure 6 pertains to a chemical process concentration labeled by Box, Jenkins, and Reinsel (1994) as Series A. The series consists of 197 observations. We considered three models for the entire series: ARIMA(1,1,1) identified as the optimal model by an R program as described above, and two models ARIMA(0,1,1) and ARIMA(1,0,1) used by Box, Jenkins, and Reinsel (1994). We then used the first 80 observations for estimation and produced rolling sample 117 one-step ahead forecast. The histogram of the standardized forecast errors of the ARIMA(0,1,1) is shown Figure 6. (The histograms for the other two models were similar, hence not shown here). The lower middle panel of Figure 6 shows the plots of  $MRAE_n$  for the three models for  $\tau \leq .5$ . The ARIMA(1,0,1) uniformly dominates the ARIMA(1,1,1) in this range of tolerance thresholds. However, the plot of ARIMA(0,1,1) crosses the plots for the other two models.

### 5.2.3 Chemical Process Temperature

The left column of Figure 6 pertains to a chemical process temperature labeled by Box, Jenkins, and Reinsel (1994) as Series C. The series consists of 225 observations. We considered three models for the entire series: ARIMA(2,0,0) identified as the optimal model by an R program as described above, and two models ARIMA(1,1,0) and ARIMA(0,2,2) used by Box, Jenkins, and Reinsel (1994). Note that ARIMA(2,0,0) is the AIC optimal among ARIMA(p,0,q) because the unit root test concluded stationarity. We then used the first 80 observations for estimation and produced rolling sample 145 one-step ahead forecast. The histogram of the standardized forecast errors of the ARIMA(0,2,2) is shown Figure 6. (The histograms for the other two models were similar, hence not shown here). The lower left panel of Figure 6 shows the plots of  $MRAE_n$  for the three models for  $\tau \leq .5$ . The plots for

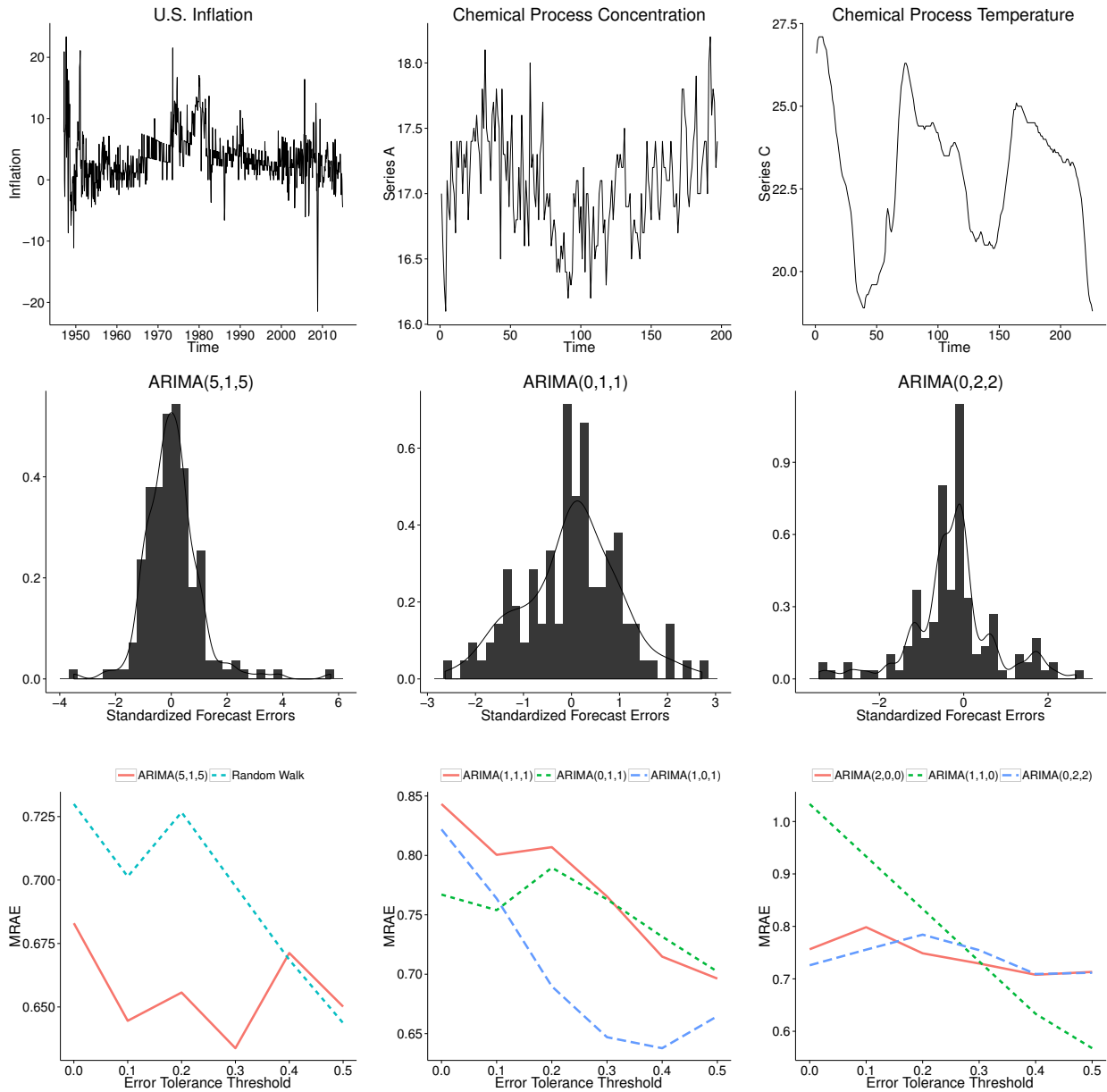


Figure 6: Sequence plots of U.S. Inflation rate and two chemical processes measurements, sequence plots and histograms of the forecast errors of the optimal models, and MRAE plots of the optimal models and random walk.

ARIMA(2,0,0) and ARIMA(0,2,2) are similar and flat in this range of tolerance thresholds. However, the plot of ARIMA(1,1,0) is linealy decreasing and crosses the plots for the other two models.

Table 2 summarizes the forecasting results in terms  $MAE_n$  and  $SIR_n$ . We note that these measures rank the models differently. The ranks given by  $MAE_n$  are similar to the ranks given by AIC, whereas, the ranks given by  $SIR_n$  are quite different, illustrating that tolerance for small forecast errors makes a difference.

Table 2: Forecast error measures of models for three time series

	AIC	$MAE_n$	$SIR_n$
U.S. Inflation			
ARIMA(5,1,5)	4039.09(1)	0.683(1)	0.708(2)
Random Walk	4509.06(2)	0.730(2)	0.675(1)
Chemical Process Concentration			
ARIMA(1,1,1)	108.74(1)	0.743(1)	0.647(3)
ARIMA(0,1,1)	111.02(3)	0.767(2)	0.615(2)
ARIMA(1,0,1)	109.49(2)	0.822(3)	0.568(1)
Chemical Process Temperature			
ARIMA(2,0,0)	-256.94(2)	0.705(2)	0.739(3)
ARIMA(1,1,0)	-259.34(1)	0.703(1)	0.732(2)
ARIMA(0,2,2)	-240.80(3)	0.726(3)	0.721(1)

## References

- Asadi, M. and Zohrevand, Y. (2007). On the dynamic cumulative residual entropy. *Journal of Statistical Planning and Inference*, 137, 1931-1941.
- Asadi, M., Ebrahimi, N., Hamedani, G. G, and Soofi, E. S. (2004). Maximum dynamic entropy models. *Journal of Applied Probability*, 41, 379-390.
- Asadi, M., Ebrahimi, N., Soofi, E. S., Zarezadeh, S. (2014). New maximum entropy methods for modeling lifetime distribution, *Naval Research Logistics*, 61, 427-434.
- Box, G.E.P., Jenkins, G.M, and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd Ed. Prentice-Hall, Englewood Cliffs, NJ.
- Diebold, F.X. and Shin, M. (2014). Assessing point forecast accuracy by stochastic error distance. Manuscript, University of Pennsylvania.
- Diebold, F.X. and Shin, M. (2015). Assessing point forecast accuracy by stochastic loss distance. *Economics Letters*, 130, 37-38.
- Ebrahimi, N., Maasoumi, E., and Soofi, E. S. (1999). Ordering univariate distributions by entropy and variance. *Journal of Econometrics*, 90, 317-336.
- Ebrahimi, N., Jalali, N.Y., Soofi, E.S., and Soyer, R. (2014). Importance of components for a system. *Econometric Reviews*, 33, 395-420.
- Frees, E.W. (1996). *Data Analysis Using Regression Models: The Business Perspective*. Prentice-Hall, Englewood Cliffs, NJ.
- Hyndman, R.J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27, <http://www.jstatsoft.org/>.
- Oakes, D. and Dasu, T. (1990). A note on mean residual life. *Biometrika*, 77, 409-410.

- Oja, H. (1981). On location, scale, skewness and Kurtosis of univariate distributions. *Scandinavian Journal of Statistics*, 8, 154-168.
- Psarakis, S. and Panaretos, J. (1990). The folded-t distribution. *Communications in Statistics A, Theory and Methods*, 19, 2717-2734.
- Poynor, V.A. (2010). *Bayesian Inference for Mean Residual Life Functions in Survival Analysis*. M.S. Thesis, University of California Santa Cruz  
<https://users.soe.ucsc.edu/~vpoynor/MastersRevised.pdf>.
- Rao, M. Chen, Y., Vemuri, B.C., and Wang, F. (2004). Cumulative residual entropy: A new measure of information, *IEEE Transactions on Information Theory*, IT 50, 1220-1228.
- Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic orders*. Springer, Netherlands.
- Yang, G.L. (1978). Estimation of a biometric function. *Annals of Statistics*, 6, 112-116.
- Yu, Y. (2008). On an equality of Karlin and Rinott concerning weighted sums of i.i.d. random variables. *Advances in Applied Probability*, 40, 1223-1226.
- Zografos, K. and Nadarajah, S. (2005). Survival exponential entropies, *IEEE Transactions on Information Theory*, IT 51, 1239-1246.