Research Article

Forecasting Using Information and Entropy Based on Belief Functions

Woraphon Yamaka and Songsak Sriboonchitta

Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

Correspondence should be addressed to Woraphon Yamaka; woraphon.econ@gmail.com

Received 27 May 2020; Revised 11 August 2020; Accepted 13 August 2020; Published 1 September 2020

1. Introduction

Analysts are concerned with the uncertainty of the future. Reliable and timely information about future conditions is important for policy makers and planners to propose an appropriate plan and make a correct decision in the current situation. In addition, the assessment of future conditions is essential to reach reliable results and rational decisions. For example, in the economic research field, predicting the future of the economy has become a common practice (see, Mitchell [1] and Lahiri and Monokroussos [2]). In recent years, the need to forecast in economics has become even more important due to high fluctuation in the global economy. Whatever the approach used, a forecast cannot be trusted unless it is accompanied by some measure of uncertainty. Thus, in this context, it is highly desirable to have subjective probabilities or predicted intervals for the prediction [3]. In addition to the economic field, the prediction is also important in finance, medicine, engineering, and biology in order to provide an efficient solution for any current problems.

Since the introduction of the Dempster–Shafer theory of belief functions [4, 5], new formal frameworks for handling uncertainty have been proposed and have attracted increasing interest in various areas. In this approach, a piece of evidence can be represented by a belief function, which is mathematically equivalent to a random set [6]. Although the approach is conceptually simple and easy to apply in various problems and models, statistical methods based on belief functions have not gained much acceptance. Kanjanatarakul et al. [7] mentioned that this approach cannot be applied easily to the complex statistical models in many areas, such as machine learning or econometrics. Thus, they suggested Shafer’s approach to statistical inference using belief functions, based on the concept of likelihood. Shafer [5] extended Dempster’s seminal work and proposed a belief function that is constructed from the likelihood function. It was recently justified by Denoeux [8] from the following three basic principles: likelihood principle, compatibility with Bayesian inference, and the least commitment principle.

The likelihood-based belief function has gained an increasing interest in the last few years. Previous works using...
the likelihood-based belief function in forecasting models are found in many studies such as Kanjanatarakul et al. [3], Kanjanatarakul et al. [7], Thianpaen et al. [9], and Chakpitak et al. [10]. However, it is often very difficult to specify the appropriate likelihood distribution for such models. If the distribution assumption is not specified properly, it might lead to model misspecification and bring about a prediction bias. To overcome this problem, an approach based on information theory, the classical maximum entropy (ME) principle, which was introduced by Jaynes [11], is proposed in this study. The ME principle has opened the way for a whole new class of estimation methods. The estimation allows us to extract all the available information from the observed data with minimum assumptions and avoid specifying a likelihood function. In this study, we consider the Generalized Maximum Entropy (GME). We modify the likelihood-based belief functions by using an entropy function in place of the likelihood function. Therefore, we propose an approach in which a belief function Bel on $\mathcal{Q}$ is built from the entropy function. The advantage of this method is that the statistical inference will not assume any distribution in the plausibility $p_k$ which refers to the relative entropy. Hence, the method becomes more flexible to explain a wide range of variation in the observed data.

Although the likelihood-based belief function has been proved to be a reliable interval prediction method (Denoeux [8]) applications to forecasting have remained limited in some applications. The reason might be that the likelihood-based belief function requires the statistician to provide the likelihood distribution function, which is problematic when no likelihood knowledge, or only weak information, is available. The purpose of this paper is to introduce a new approach for forecasting future observation as well as providing a prediction interval. Different from the likelihood-based belief function, we replace the parametric likelihood function with the entropy function to derive the entropy-based belief function, and thus the model mis-specification is avoided since no distributional assumption is made.

In this paper, we further explore this proposed approach by proving that, under weak distribution function information, the predictive entropy-based belief function converges to the true probability distribution of the not yet observed data. Several simulation studies are conducted to investigate the performance of our proposed method. Finally, we apply our proposed method to Autoregressive (AR) model in predicting the growth rate of Thailand’s GDP where data is scarce and involves high uncertainty. Besides, we further explore this method by showing that it can be used for predicting future observations based on the linear regression framework using the data from the operation of a plant for the oxidation of ammonia to nitric acid, measured on 21 consecutive days. We note that our efficient method allows for accurate decision-making based on reliable forecasts in various research fields.

The remainder of the paper is organized as follows. Section 2 provides the background on the linear entropy-based belief function concept and describes the estimation and prediction using the entropy-based belief function methods. Section 3 proposes a simulation study and real application study in various fields to evaluate the performance of our method through autoregressive and regression models. Section 4 summarizes the paper.

2. Methodology

This section briefly explains the necessary background notions related to the entropy approach and introduces the connection of entropy on the belief function. The Generalized Maximum Entropy approach and parameter estimation are first presented in Sections 2.1 and 2.2, respectively. The entropy-based belief function is then introduced in Section 2.3. Finally, the basic concept for predictive belief function is presented in Section 2.4.

2.1. Generalized Maximum Entropy Approach. In this study, we propose using a maximum entropy estimator to estimate our unknown parameter $\theta = \{\beta_1, \ldots, \beta_K\}$ in the econometric models, say AR and linear regression models. Before we discuss this estimator for regression and its statistical properties, we briefly summarize the entropy approach. The maximum entropy concept consists of inferring the probability distribution that maximizes information entropy given a set of various constraints. Let be a proper probability of variable $k$ on support $\mathcal{M}$ and the vector $z_k = z_{k1}, \ldots, z_{kM}$ be $M$-dimensional vector support of coefficients $k, k = 1, \ldots, K$. Following Shannon’s entropy (1948), the summation of the $K$ entropies is

$$H(p) = -\sum_{k=1}^{K} \sum_{m=1}^{M} p_{km} \log p_{km},$$  \hspace{1cm} (1)

where the probabilities $P$ are the unknown and unobservable probabilities in the entropy $\sum_{m=1}^{M} p_{km} = 1$. The entropy measures the uncertainty of distribution and reaches a maximum when $p_{k1} = p_{k2} = \cdots = p_{km} = 1/M$; in other words, if no constraints (data) are imposed, $H(P)$ reaches its maximum value for uniform distribution.

2.2. Parameter Estimation. To better understand the entropy estimation, we give a simple example. Let us consider the simple autoregressive AR model. Let $\{y_t\}, t = 1, \ldots, T$, be observed data which are expressed as a linear combination of past observations $y_{t-k}$. In the AR ($K$), we have

$$y_t = \sum_{k=1}^{K} \beta_k y_{t-k} + \varepsilon_t,$$  \hspace{1cm} (2)

where $K$ and $\beta_k$ are the autoregressive order and the autoregressive coefficient, respectively, and $\varepsilon_t$ is the independent and identically distributed random white noise that is not assumed to have any distribution and is assumed to be an uncorrelated random variable $y_{t-k}$.

To apply this concept to be an estimator of the regression model, we generalize the maximum entropy to the inverse problem in the regression framework. Rather than searching for the point estimates $\beta$, we can view these unknown parameters as expectations of random variables with $M$
support values for each estimated parameter value \((k)\), \(Z = [z_1, ..., z_K]\), where \(z_k = [z_{k1}, ..., z_{kM}]\) for all \(k = 1, ..., K\), where \(2 \leq M < \infty\), and \(z_{k1}\) and \(z_{kM}\) are the lower and upper bounds of \(\beta_k\). Thus, we can express parameter \(\beta_k\) as

\[
\beta_k = \sum_{m=1}^{M} p_{km} z_{km},
\]

where \(p_{km}\) are the \(M\)-dimensional estimated probability defined on the support vector \(z_{km}\). Next, similar to the above expression, \(\epsilon_i\) is also constructed as the mean value of some random variable \(\nu_i\). Here, each \(\epsilon_i\) is assumed to be a random vector with finite and discrete random variable with \(M\) support value, \(\nu_i = [\nu_{i1}, ..., \nu_{iM}]\). Let \(w_t\) be a \(M\) dimensioned proper probability weights defined on the set \(v_{tm}\) such that

\[
\epsilon_i = \sum_{m=1}^{M} w_{tm} v_{tm}.
\]

Using the reparameterized unknowns \(\beta_k\) and \(\epsilon_i\), we can rewrite equation (2) as

\[
y_t = \sum_{k=1}^{K} y_{t-k} \sum_{m=1}^{M} p_{km} z_{km} + \sum_{m=1}^{M} w_{tm} v_{tm},
\]

where the vector support \(z_{km}\) and \(v_{tm}\) are convex set that is symmetric around zero. Then, we can construct our Generalized Maximum Entropy (GME) estimator as

\[
H(p, w) = \arg \max_{p, w} \{H(p) + H(w)\}
\]

\[
\equiv -\sum_{k=1}^{K} \sum_{m=1}^{M} p_{km} \log p_{km} - \sum_{m=1}^{M} w_{tm} \log w_{tm},
\]

subject to

\[
y_t = \sum_{k=1}^{K} y_{t-k} \sum_{m=1}^{M} p_{km} z_{km} + \sum_{m=1}^{M} w_{tm} v_{tm},
\]

\[
\sum_{m=1}^{M} p_{km} = 1,
\]

\[
\sum_{m=1}^{M} w_{tm} = 1,
\]

where \(p_{km}\) and \(w_{tm}\) are on the interval \([0, 1]\).

This optimization problem can be solved using the Lagrangian method which takes the following form:

\[
L = H(p, w) + \sum_{t=K+1}^{T} \lambda_{1t} \left( y_t - \sum_{k=1}^{K} y_{t-k} \sum_{m=1}^{M} p_{km} z_{km} - \sum_{m=1}^{M} w_{tm} v_{tm} \right)
+ \sum_{k=1}^{K} \lambda_{2k} \left( 1 - \sum_{m=1}^{M} p_{km} \right) + \sum_{t=K+1}^{T} \lambda_{3t} \left( 1 - \sum_{m=1}^{M} w_{tm} \right),
\]

where the unknown multipliers \(\lambda_1, \lambda_2, \lambda_3\) are \(T \times 1, K \times 1\), and \(T \times 1\) vectors of Lagrangian multipliers. Thus, the resulting first-order conditions are

\[
\frac{\partial L}{\partial p_{km}} = -\log p_{km} - 1 - \sum_{t=K+1}^{T} \left( y_t - k \lambda_{1t} z_{km} - \lambda_{2k} \right) = 0,
\]

\[
\frac{\partial L}{\partial w_{tm}} = -\log w_{tm} - 1 - \left( \lambda_{1t} v_{tm} \right) = 0,
\]

\[
\frac{\partial L}{\lambda_{1t}} = \left( y_t - k \sum_{k=1}^{K} y_{t-k} \sum_{m=1}^{M} p_{km} z_{km} - \sum_{m=1}^{M} w_{tm} v_{tm} \right) = 0,
\]

\[
\frac{\partial L}{\lambda_{2k}} = 1 - \sum_{m=1}^{M} p_{km} = 0,
\]

\[
\frac{\partial L}{\lambda_{3t}} = 1 - \sum_{m=1}^{M} w_{tm} = 0.
\]

Thus, we have

\[
p_{km} = \exp \left( -1 - \lambda_{2k} - \sum_{t=K+1}^{T} \left( y_t - k \lambda_{1t} z_{km} \right) \right) = 0,
\]

\[
w_{tm} = \exp \left( -1 - \lambda_{3t} - \lambda_{1t} v_{tm} \right) = 0.
\]

Since the constraint requires that \(\sum_{m=1}^{M} p_{km} = 1\) and \(\sum_{m=1}^{M} w_{tm} = 1\), we sum both \(p_{km}\) and \(w_{tm}\) over \(m\), and we have

\[
\sum_{m=1}^{M} p_{km} = \exp \left( -1 - \lambda_{2k} \right) \sum_{m=1}^{M} \exp \left( - \sum_{t=K+1}^{T} \left( y_t - k \lambda_{1t} z_{km} \right) \right) = 1,
\]

\[
\sum_{m=1}^{M} w_{tm} = \exp \left( -1 - \lambda_{3t} \right) \sum_{m=1}^{M} \exp \left( - \lambda_{1t} v_{tm} \right) = 1,
\]

where \(\lambda\) is the estimated multiplier. To estimate the Lagrangian multipliers, we can substitute the candidate probabilities, \(\tilde{p}_{km}\) and \(\tilde{w}_{tm}\), into the original Lagrangian (equation 9) Then, the estimated \(\lambda\) can be obtained by differentiating the concentrated Lagrangian (equation 9) (conditional on the candidates \(p_{km}\) and \(w_{tm}\) with respect to \(\lambda\) and set it to be equal be zero. Then, substitute these estimated parameters in equations (13)-(14). Note that \(\exp(-1-\lambda_{2k})\) and \((-1-\lambda_{3t})\) are constants for a given parameter, so the optimal probability id15equations (11) and (12) can be rewritten as

\[
\tilde{p}_{km} = \frac{\exp \left( -z_{km} \sum_{t=K+1}^{T} y_t - k \lambda_{1t} \right)}{\sum_{m=1}^{M} \exp \left( -z_{km} \sum_{t=K+1}^{T} y_t - k \lambda_{1t} \right)},
\]

\[
\tilde{w}_{tm} = \frac{\exp \left( -\lambda_{1t} v_{tm} \right)}{\sum_{m=1}^{M} \exp \left( -\lambda_{1t} v_{tm} \right)}.
\]

Summing up the above equations, we maximize the joint-entropy objective equation (6) subject to the regression
equation (7), by adding restrictions equation (8). The solution to this maximization problem is unique by forming the Lagrangian and solving for the first-order conditions to obtain the optimal solutions $\hat{p}_{km}$, $\hat{w}_{tm}$, and $\hat{\lambda}_1$ (estimated values). Then, these estimated probabilities are used to derive the point estimates for the regression coefficients and error term, see equations (3) and (4).

**Example 1.** To better understand the entropy estimation, we give a simple example. Let consider the first-order linear autoregressive AR(1) model as defined by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t \sim N(0, 1). \quad (16)$$

The true values for $\beta_0$ and $\beta_1$ are set to be 1 and $-1$, respectively. The error is drawn randomly from a normal distribution with mean 0 and variance 1. We generate equation (16) using sample size $T = 10$. The support spaces of $\beta_0$ and $\beta_1$ are defined as $[-5, 0, 5]$ while $\epsilon_t$ is defined as $[-3, 0, 3]$. The simulated data are shown in Figure 1. And the results are reported in Tables 1 and 2.

To estimate equation (16), we need to construct our GME problem as

$$H(p, w) = -\sum_{k=1}^{3} \sum_{m=1}^{3} p_{km} \log p_{km} - \sum_{t=1}^{10} \sum_{m=1}^{3} w_{tm} \log w_{tm},$$

$$y_t = \sum_{m=1}^{3} p_{1m} z_{1m} + \sum_{m=1}^{3} p_{2m} z_{2m} y_{t-1} + \sum_{m=1}^{3} w_{tm} v_{tm},$$

$t = 1, ..., 10,$

$$\sum_{m=1}^{3} p_{km} = 1,$$

$$\sum_{m=1}^{3} w_{tm} = 1,$$ \quad (17)

where $p_{km}$ and $w_{tm}$ are the parameter estimates in this problem, see equations (3)-(4).

2.3. The Entropy-Based Belief Function. Following the three basic principles of the likelihood-based belief function, consisting the likelihood principle, compatibility with Bayesian inference, and the least commitment principle, which are justified by Denoeux [8], we modify these approaches by using an entropy function as the substitute for the likelihood function. Let $P \in \Theta$ be the parameter of interest, and we propose an approach in which a belief function $Bel_P^P$ on $\Theta$ is built from the entropy function. In this case, the entropy ratio refers to a “relative plausibility,” and the contour function can thus be readily calculated as

$$pl_Y(P) = \frac{H(P)}{H(\bar{P})},$$ \quad (18)

where $P = \{p_{km}, w_{tm}\}$ in which $p_{km}$ and $w_{tm}$ are proper probability vector mass functions on the interval [0, 1] and $H(P)$ is

subject to

$$y_t = \sum_{k=1}^{3} y_{t-k} \sum_{m=1}^{M} p_{km} z_{km} + \sum_{m=1}^{M} w_{tm} v_{tm}, \quad (20)$$

$$\sum_{m=1}^{M} p_{km} = 1,$$

$$\sum_{m=1}^{M} w_{tm} = 1.$$ \quad (21)

Meanwhile, $H(\bar{P})$ is defined as

$$H(\bar{P}) = \sup_{P \in \Theta} H(P),$$ \quad (22)

and the marginal contour functions on each individual parameter are

$$pl(p_{km}) = \sup_{P \sim p_{km}} pl(P),$$

$$pl(w_{tm}) = \sup_{P \sim w_{tm}} pl(P).$$ \quad (23)

Subject to the additional constraints equations (20)-(21), this belief function is called the entropy-based belief function on $\Theta$ induced by $Y$. The corresponding plausibility function can be estimated by $pl_Y(P)$ as

$$Bel_Y^P(C) = \sup_{P \in C} pl_Y(P),$$ \quad (24)

for all $C \subseteq \Theta$. The focal sets of $Bel_Y^P$ are the level sets of $pl_Y(P)$ defined as

![Figure 1: Example data.](image_url)
where \( \omega \) is uniformly distributed in \([0, 1]\). This belief function \( Bel_0^\beta \) is equivalent to the random set induced by the Lebesgue measure \( \lambda \) on \([0, 1]\) and the multivalued mapping \( \Gamma_y \) from \([0, 1]\) to \(2^\Theta[3]\).

2.4. The Basic Method for Predictive Belief Function. To forecast future data \( Y_f = (Y_{T+1}, ..., Y_{T+h}) \), the sampling model used by Dempster [12] is introduced here. In this model, the forecast data \( Y_{T+1} \) is expressed as a function of the proper probability mass function \( P \) which is obtained by past observed data \( y = (y_1, ..., y_T) \) and an unobserved auxiliary \( S \in S \) variable with known probability measure \( \mu \) not depending on \( P \). Note that \( S \) is the sample space of \( s \):

\[
Y_f = \varphi(P, S),
\]

(26)

where \( \varphi \) is defined in such a way that the distribution of \( Y_f \) is conditional on fixed \( P \). When \( Y_f \) is a continuous random variable, equation (26) can be computed by

\[
Y_f = G_{x,y}^{-1}(S) = \inf \left\{ y_f \left| F_{x,y}(y_f) \geq S \right. \right\},
\]

(27)

where \( G_{x,y}^{-1} \) is the inverse conditional cumulative distribution function (cdf) of \( Y_f \) given \( y \) and \( s \) is uniformly distributed in \([0, 1]\). We note that, in the likelihood-based belief function case, \( G_{x,y}^{-1} \) becomes \( G_{y,x}^{-1} \) which is the inverse of some conditional cumulative distribution functions, say normal, student-\( t \), or other parametric distributions. However, in the entropy-based belief function case, we do not need any distribution assumptions on \( G_{x,y}^{-1} \). Let us compose \( \Gamma_y: [0, 1] \times S \longrightarrow 2^\Theta \) with \( \varphi \), and we get a new multivalued mapping \( \Gamma'_y \) form \([0, 1] \times S \longrightarrow 2^Y \) defined as

\[
\Gamma'_y: [0, 1] \times S \longrightarrow 2^Y,
\]

\[
(\omega, s) \longrightarrow \varphi(\Gamma_y(\omega), s).
\]

(28)

The predictive belief function \( Bel_0^\beta \) and plausibility \( Pl_0^\beta \) are then induced by the multivalued mapping \( \Gamma_y \) and \( \lambda \otimes \mu \) on \([0, 1] \times S \) as follows:

\[
Bel_0^\beta(C) = (\lambda \otimes \mu) \left\{ (\omega, s) \left| \varphi(\Gamma_y(\omega), s) \subseteq C \right. \right\},
\]

(29)

\[
Pl_0^\beta(C) = (\lambda \otimes \mu) \left\{ (\omega, s) \left| \varphi(\Gamma_y(\omega), s) \cap C \neq \emptyset \right. \right\},
\]

for all \( C \subseteq R \).

3.3. Simulation and Application: Entropy-Based Belief Function

Let us now turn our attention to examining the performance of the entropy-based belief function. In this section, the inference and forecasting methodology outlined in the previous section will be applied to the simulated data and real data. We apply the entropy-based belief function to the forecasting AR and regression models which will be presented in Sections 3.1 and 3.2, respectively.

3.1. The Entropy-Based Belief Function in Autoregressive Model Problem

3.1.1. Prediction. As we have seen in Section 2, the estimation problem is to make statements about some probability after observing some data. To describe the prediction method for the AR model, let us consider the case where the \( K = 1 \); thus, the first-order linear autoregressive AR(1) model is defined by

\[
Y_{T+1} = \beta_1 + \beta_2 y_T + \epsilon_T,
\]

(30)

where \( Y_{T+1} \) is a predicted value of \( T + 1 \) which is not yet observed and \( y_T \) is the observed value at time \( T \).

The main idea is to estimate \( Y_{T+1} \) as a function of \( \beta \), which is viewed as an expectation of probability with \( M \) support and some pivotal variable \( S \), whose distribution does not depend on \( \beta \). We can write equivalently

\[
Y_{T+1} = \varphi_y(\beta, s) = \beta_1 + \beta_2 y_T + \tilde{F}^{-1}(s), \quad s \sim U[0, 1],
\]

(31)

where \( \tilde{F}(\cdot)^{-1} \) is the quantile function of empirical distribution based on the estimated residuals, \( \epsilon_1, ..., \epsilon_T \). Then, in the case of entropy approach, we rewrite equation (30) as

\[
Y_{T+1} = \frac{\sum_{m=1}^M P_{1m} y_{1m} + \sum_{m=1}^M P_{2m} y_{2m}}{P_T(y_T) + \tilde{F}^{-1}(s)},
\]

(32)

where \( m \) is the number of supports of \( z \) and \( v \). Consequently, in the entropy estimation, we rewrite the focal set equation (31) as

\[
Y_{T+1} = \varphi_y(P, s) = \sum_{m=1}^M P_{1m} y_{1m} + \sum_{m=1}^M P_{2m} y_{2m}(y_T) + \tilde{F}^{-1}(s).
\]

(33)

To forecast \( Y_{T+1} \mid y_T \), the predictive belief function and plausibility function can then be approximated using Monte Carlo simulation as in the following procedure (Kanjanatarakul et al. [3] and Kanjanatarakul et al. [7]):

Table 2: Estimated coefficients of the example AR(1) model.

| \( P_{k1} \) | \( P_{k2} \) | \( P_{k2} \) | \( z_{km} \) | Estimates | True |
|---|---|---|---|---|---|
| 0.2173 | 0.3179 | 0.4648 | \([-5, 0, 5]\) | 1.2370 | 1 |
| 0.4256 | 0.3255 | 0.2489 | \([-5, 0, 5]\) | -0.8831 | -1 |
(1) Draw \(\omega_i, s_i\) independently from uniform \([0, 1]\) \(N\) draws. Then, compute (or approximate) the focal set which is the interval defined by the following lower and upper bounds:

\[
\varphi_y(\Gamma_y(\omega, s)) = \left\{ \varphi(P, s) \big| p_l_y(P) \geq \omega \right\}
\]

\[
= \left[ y_l(s_i, \omega_i), y_u(s_i, \omega_i) \right],
\]

where \(y_l(s_i, \omega_i)\) and \(y_u(s_i, \omega_i)\) are the lower and upper bounds which can be computed using a constrained nonlinear optimization algorithm. This optimization problem can be solved using the nonlinear optimization:

\[
y_l(s_i, \omega_i) = \min \varphi_y(P, \omega_i, s_i),
\]

\[
y_u(s_i, \omega_i) = \max \varphi_y(P, \omega_i, s_i),
\]

which subject to

\[
pl_y(P) \geq \omega_i,
\]

\[
y_t = \sum_{m=1}^{M} p_{1m} e_{1m} + \sum_{m=1}^{M} p_{2m} e_{2m}(y_{t-1}) + \sum_{m=1}^{M} w_{im} v_{im},
\]

\[
\sum_{m=1}^{M} p_{1m} = 1,
\]

\[
\sum_{m=1}^{M} p_{2m} = 1,
\]

\[
\sum_{m=1}^{M} w_{im} = 1.
\]

(2) We can then approximate the predictive belief (Bel\(_r\)) and plausibility (pl\(_r\)) functions on \(Y_f\) as

\[
Bel_r(C) = \frac{1}{N} \#\{i \in \{1, \ldots, N\} \big| \varphi_y(\Gamma_y(\omega_i, s_i)) \subseteq C\}
\]

\[
pl_r(C) = \frac{1}{N} \#\{i \in \{1, \ldots, N\} \big| \varphi_y(\Gamma_y(\omega_i, s_i)) \cap C \neq \emptyset\}
\]

or all \(C \subseteq \Theta\). These lower and upper cdfs of the predictive belief function are approximated using \(N = 5000\) randomly generated focal sets (Algorithm 1).

### 3.1.2. Experiment Study

In this section, we conduct a simulation and experiment studies to investigate the finite sample performance of the proposed Shannon entropy-based belief methods. The simulated data are constructed from the AR model using different error distributions consisting of normal, Student-\(t\), skewed Student-\(t\), and uniform. We note that the normal and student-\(t\) are selected as symmetric distribution, while skewed student-\(t\) and uniform are typical examples of highly skewed distribution and flat distribution, respectively. In these experiments, we consider three competing likelihood functions: normal, student-\(t\), and skewed student-\(t\) likelihoods.

To compare these likelihood-based belief functions, the one-step-ahead forecast is computed from simulated data and compared with the true prediction value. We note that, as the function \(pl(y) = P(y)\) has a unique maximum \(\hat{y}\), it may be taken as a point prediction of the AR(1) model-based belief function approach. This is to say \(\hat{y}\) can be viewed as the most plausible value of \(Y\) given the past information. Thus, we need a criterion to gauge the performance of the method. Here, the mean square error (MSE) method is employed, and it can be defined as

\[
MSE = R^{-1} \sum_{r=1}^{R} (Y_{true,r} - Y_{f,r})^2,
\]

where \(R = 20\) is the number of repetitions, \(Y_{true}\) is the true value, and \(Y_f = \hat{y}\) is the predicted value.

In this Monte Carlo simulation, we consider sample size \(T = 20\) and \(T = 40\). \(R = 20\) samples are generated for each case. To make a fair comparison, the error terms are generated from \(N(0, \sigma^2 = 1)\), \(t(0, \sigma^2 = 1, \nu = 4)\), skewed \(-t(0, \alpha^2 = 1, \nu = 4, \xi = 1.5)\), and \(U(-2, 2)\). Then, the sampling experiments are based on the AR (1) model:

\[
y_t = 0.5 - 0.5 y_{t-1} + \epsilon_t.
\]

In this experiment, the first 20 observations are used to predict the 21st observation. Likewise, the first 40 observations are used to predict the 41st observation. Then, the most plausible predicted values, \(Y^{21st}\) and \(Y^{41st}\), are compared with \(Y_{true} = 0.5 - 0.5 y_T + \epsilon_{T+1}\), \(T = 20\), and \(Y_{true} = 0.5 - 0.5 y_T + \epsilon_{T+1}, T = 40\), respectively. In the case of the Shannon entropy, the support \(z_{im}\) is initially set to \((-5, 0, 5)\) and the supports for \(v_{im}\) to \((-3\sigma, 0, 3\sigma)\), where \(\sigma\) is computed from the conventional AR model. Table 3 reports the prediction error MSE of our considered estimators. Consider the performance of the proposed estimator when the errors are generated from a uniform distribution, and we found that the error of the Shannon entropy-based belief function is smaller than the other methods over the one-step-ahead forecast. In addition, the performance of this method is better when the sample size increases from 20 to 40.

Next, we investigate the performance of the proposed estimator when the errors are generated from normal, Student-\(t\), and skewed Student-\(t\) distributions and compare with the parametric likelihood-based belief function. We observe that the overall result is different in Table 3. We find that the Shannon entropy-based belief function seems to outperform the likelihood-based belief function in terms of lowest MSE of the prediction when the \(U(-2, 2)\) error distribution is assumed. When the errors are generated from other parametric distributions, we find that the likelihood-based belief function method with the correct distribution function always has the best performance. However, we notice that our Shannon entropy-based belief function performs the second-best prediction over the wide range of error distributions.
the Shannon entropy-based belief function, which performs unknown in practice, these results confirm the advantage of is obtained. However, the error distribution is sometimes

$q_{\text{he}}$ means that if the distribution is known and we estimate

likelihood-based belief function outperforms the entropy.

error is generated from known distribution, the parametric

method performs well particularly when the error is un-

\[ \text{Algorithm 1: Monte Carlo algorithm for approximating a predictive entropy-based belief function (random interval case).} \]

| \( Y_f \) | Normal likelihood-based belief function | Student-\( t \) likelihood-based belief function | Skewed Student-\( t \) likelihood-based belief function | Shannon entropy-based belief function | Renyi entropy-based belief function | Tsallis entropy-based belief function |
|-------|----------------------------------------|------------------------------------------|------------------------------------------|---------------------------------|---------------------------------|---------------------------------|
| \( T = 20 \) | \( T = 40 \) | \( T = 20 \) | \( T = 40 \) | \( T = 20 \) | \( T = 40 \) | \( T = 20 \) | \( T = 40 \) |
| \( M S E \) | \( 0.9346 (0.9541) \) | \( 1.2327 (1.8079) \) | \( 13.8763 (15.4710) \) | \( 0.5784 (0.5824) \) | \( 0.5796 (0.5831) \) | \( 0.5788 (0.5826) \) |
| \( M S E \) | \( 0.6891 (0.7552) \) | \( 1.6996 (1.6252) \) | \( 8.7598 (12.2013) \) | \( 0.2596 (0.5914) \) | \( 0.2608 (0.5948) \) | \( 0.2605 (0.5953) \) |
| \( N(0, 1) \) | \( T = 20 \) | \( T = 40 \) | \( T = 20 \) | \( T = 40 \) | \( T = 20 \) | \( T = 40 \) | \( T = 20 \) | \( T = 40 \) |
| \( M S E \) | \( 0.4601 (0.5491) \) | \( 1.2275 (1.4574) \) | \( 2.0607 (2.4733) \) | \( 0.5582 (0.7576) \) | \( 0.5688 (0.7589) \) | \( 0.5591 (0.7601) \) |
| \( M S E \) | \( 0.4995 (0.3136) \) | \( 1.4068 (1.4647) \) | \( 1.2092 (1.1063) \) | \( 0.5294 (0.3787) \) | \( 0.5301 (0.3793) \) | \( 0.5312 (0.3798) \) |
| \( T = 0, 1, 4 \) | \( T = 20 \) | \( T = 40 \) | \( T = 20 \) | \( T = 40 \) | \( T = 20 \) | \( T = 40 \) | \( T = 20 \) | \( T = 40 \) |
| \( M S E \) | \( 1.2850 (1.476) \) | \( 1.1815 (2.6127) \) | \( 1.6660 (1.8974) \) | \( 1.2191 (1.7849) \) | \( 1.2115 (1.7858) \) | \( 1.2114 (1.7857) \) |
| \( M S E \) | \( 0.7604 (0.4309) \) | \( 0.6142 (0.3654) \) | \( 0.8783 (1.1256) \) | \( 0.7538 (0.2542) \) | \( 0.7542 (0.2550) \) | \( 0.7542 (0.2550) \) |
| \( T \) is standard deviation on \( M S E \). The bold numbers present the best prediction result.

Thus, in this simulation study, we can conclude that our method performs well particularly when the error is unknown or has a uniform distribution. Nonetheless, when the error is generated from known distribution, the parametric likelihood-based belief function outperforms the entropy. This means that if the distribution is known and we estimate the model based on the true distribution, the accurate result is obtained. However, the error distribution is sometimes unknown in practice, these results confirm the advantage of the Shannon entropy-based belief function, which performs uniformly well over a wide range of error distributions.

Furthermore, it would be more interesting to consider other entropy measures such as Renyi entropy [13] and Tsallis entropy [14]; we thus introduce these two entropy-based belief functions to investigate the performance of various entropy measures. We note that Renyi entropy is the generalization of Shannon entropy while Tsallis entropy is nonlogarithmic and is obtained through the joint generalization of the averaging procedures and the concept of information gain. These two entropies are indexed by an order \( d \). The Renyi entropy is formulated as
while Tsallis entropy can be defined as

\[ H^T_d(p) = c - \frac{\sum_{k=1}^{K} \sum_{m=1}^{M} \log p^{d}_{km}}{1 - d}, \]  

(42)

where the value of \( c \) is a positive constant and depends on the particular units used. For simplicity, we set \( c = 1 \) and order \( d \) (see [15]). Note that both Renyi and Tsallis entropies become the Shannon entropy as a special case when \( d = 1 \) and \( c = 0 \), respectively.

The results of these three entropies for comparison are shown in the last three columns of Table 1, and it is evident that the Shannon entropy-based belief function still performs the best prediction performance in this simulation data. However, the performances of these three entropies are not much different in terms of MSE.

**Example 2.** Application to predicting GDP growth by the AR model.

In this example, the real data are also used to investigate the performance of our method. The dataset considered, derived from the Thomson Reuter Datastream and World Bank database, consist of yearly data, from the ending of 1995 to 2018, of Gross Domestic Product (GDP). The examined data series is transformed into the growth rate and plotted in Figure 2. Then, we perform the Augmented Dickey–Fuller unit roots test on GDP growth and determine the lagging order number of the AR model according to Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The ADF-statistic result is -4.337, which is greater than the critical values at 10% significance level and means that GDP is stationary. Then, the lag selection for the AR model is shown in Table 4. The result shows that the AR model with lag 1 provides the lowest AIC and BIC.

To evaluate the performance of prediction, the forecast evaluation period is from 2014 to 2018. We perform a recursive forecasting to update GDP growth year by year, totally 5 years, which give us enough data points to evaluate the out-of-sample forecast performance. For this evaluation period, we compute the root mean square forecast error (RMSE) and mean square error (MAE) of forecasts at each forecast horizon (from one to five years ahead). We can compute the RMSE and MAE as in the following:

\[
MAE = \frac{1}{T} \sum_{i=1}^{T} |e_i|,  
\]

\[
RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^{T} e^2_i},  
\]

(43)

where \( T = 5 \) and \( e = Y_{true} - Y_f \).

The forecast comparison of the AR(1) model with the likelihood-based approach and the entropy-based approach, as well as the conventional frequentist method, performed in out-of-sample is reported in Table 5. The out-of-sample forecasts made for the 2014–2018 period are reflected by the most plausible prediction values and their large upper-lower prediction intervals. The true prediction values are shown in the first column. The estimated conditional expectation as well as the frequentist 95% prediction intervals are shown in column 2. Columns 3–7 show the most plausible predictions and quantile intervals for the normal, student-\( t \), skewed student-\( t \), and the three entropy-based belief functions, respectively.

We can now consider the performance of different methods. First, we evaluate the entropy-based method regarding the prediction value. Second, we compare the performance regarding the range of the prediction intervals.

In the first evaluation, the smaller the \( MAE \) and \( RMSE \), the superior the corresponding model’s performance. According to Table 5, we learn that all three entropy-based belief functions outperform the other likelihood-based methods. We can observe that the values of \( RMSE \) and \( MAE \) of entropy-based belief functions are a bit smaller than those of the likelihood-based belief functions. Again, the Shannon entropy still provides lower \( RMSE \) and \( MAE \) compared to the other two entropies. We expect that if the real data has a distribution and the assumption of the likelihood-based belief function is not correctly specified, a less accurate prediction is obtained. In addition, small sample size may affect the accuracy of the estimated parameters. Hence, in this study, we conclude that entropy-based belief function for AR (1) performs well in the out-of-sample forecasts relative to other competing methods.

In the second evaluation, for the prediction intervals for Thai GDP growth, the methods suggest reliable prediction intervals around the predicted value. We can observe that all prediction interval methods for the AR (1) model are reliable as the true values exist along with the range of the intervals, except for the 95% frequentist approach. The prediction interval based on the 95% frequentist for Thai GDP growth in 2014 does not cover the true value. This result reveals that the conventional frequentist approach may not provide the appropriate prediction value.

Regarding the prediction intervals, we can calculate the average of the prediction intervals for making a comparison. The results portray the difference between the methods exists. The frequentist method has a very narrow mean of the prediction intervals, while our entropy-based method performs the widest. We suspect that our approach relies on the empirical distribution assumption on the error distribution and leads to a larger simulated error in the Monte Carlo algorithm for approximating a predictive entropy-based belief function process. However, this experiment shows that our method produces an acceptable range for prediction intervals as the true values always lie in the intervals.

We now use the entropy-based belief function method to predict Thai economic growth in the future. The prediction is very important because of its relevance to the current policy debate on whether Thai GDP growth will be up or
down. The forecasts made for 2019 is based on scant data, which is reflected by large upper-lower cdf intervals (Figure 3). Figure 3 presents the lower and upper cumulative density functions (cdfs) for the prediction problem obtained from (black solid line) and (pink solid line). We also plot the vertical blue dotted line corresponding to the most plausibility value together with the quantile intervals for the GDP growth in 2019. We can observe that the value of growth rate is around 0.03961 (3.961%) in 2019 and the prediction intervals 95% ranges (0.0630 – 0.1124) (black dashed line) and (0.0396 – 0.0558) (pink dashed line).

Furthermore, to show the accuracy of our estimation in this study, we show the contour plots of plausibility (pl(θ)) in two-dimensional parameter subspaces of each method, as

Table 4: Lag selection for the AR model.

|       | AR (1) | AR (2) | AR (3) | AR (4) |
|-------|--------|--------|--------|--------|
| AIC   | −104.41| −102.69| −100.75| −100.43|
| BIC   | −101.00| −98.14 | −95.076| −93.61 |

Table 5: Out-of-sample forecasting performance.

|       | TRUE | Y′_AR | Y′_norm | Y′_sstd | Y′_shannon | Y′_Bely | Y′_tsallis |
|-------|------|-------|---------|---------|------------|---------|------------|
| Forecast 2014 | [0.0357, 0.0367, 0.0331, 0.0367] | 0.0042 | 0.0039 | 0.0030 |
| [P_L, (−∞, qL^U_{0.05})], Bel_L, (−∞, qL^U_{0.05})] | 0.009 | | | |
| | | | | |
| Forecast 2015 | [0.0331, 0.0328, 0.0338, 0.0348] | 0.0404 | 0.0419 | 0.0410 |
| [P_L, (−∞, qL^U_{0.05})], Bel_L, (−∞, qL^U_{0.05})] | 0.030 | | | |
| | | | | |
| Forecast 2016 | [0.0340, 0.0352, 0.0338, 0.0362] | 0.0398 | 0.0401 | 0.0407 |
| [P_L, (−∞, qL^U_{0.05})], Bel_L, (−∞, qL^U_{0.05})] | 0.033 | | | |
| | | | | |
| Forecast 2017 | [0.0341, 0.0344, 0.0339, 0.0353] | 0.0395 | 0.0385 | 0.0386 |
| [P_L, (−∞, qL^U_{0.05})], Bel_L, (−∞, qL^U_{0.05})] | 0.039 | | | |
| | | | | |
| Forecast 2018 | [0.0340, 0.0334, 0.0344, 0.0357] | 0.0395 | 0.0390 | 0.0405 |
| [P_L, (−∞, qL^U_{0.05})], Bel_L, (−∞, qL^U_{0.05})] | 0.041 | | | |

*norm = normal likelihood, std = student-t likelihood, sstd = skewed student-t likelihood, and Y_f is the point prediction.*
shown in the Appendix. These contour plots show two-dimensional slices of the contour function, where one of the three parameters is fixed to its maximum normal likelihood, two of the four parameters are fixed to its maximum student-t likelihood, three of the five parameters are fixed to its maximum skewed student-t likelihood, and \((z+v)\times m\) - 2 of the \((z+v)\times m\) probability parameters are fixed to its maximum entropy. The results of these contour plots show that the estimated parameters of each method are likely to reach the maximum plausibility, as presented by the red cross. Consider the contour plots of \(p_{y}(\theta)\) for our proposed method, entropy-based belief functions. We can summarize the process graphically by plotting the optimization on a contour plot of the plausibility \(p_{y}(P)\). We can observe that, for example, the estimated probabilities \(p_{11}\) and \(p_{12}\) reach a maximum plausibility \(p_{y}(P)\) at the red cross.

### 3.2. The Entropy-Based Belief Function in the Regression Problem

#### 3.2.1. Regression Problem

Consider the simple linear regression model:

\[
y_t = \sum_{k=1}^{K} \beta_k x_{kt} + \epsilon_t, \quad t = 1, \ldots, T,
\]

where \(y_t\) is observed dependent variable at time \(t\), \(x_{kt}\) is an independent variable \(k\) at time \(t\), and \(\epsilon_t\) is the error that is not assumed to be any distribution when the entropy-based approach is used.

#### 3.2.2. Prediction with Certain Inputs

Similar to the AR model, the prediction with linear regression can be easily computed using the method described in Section 3.1.2. Let \(Y_{T+1}\) be the future value which cannot be observed at time \(T+1\) and \(x_{kT}\) be the observed inputs or covariates at time \(T\):

\[
Y_{T+1} = \sum_{k=1}^{K} \beta_k x_{kT} + \epsilon_T,
\]

where \(Y_{T+1}\) is a predicted value which is not yet observed and \(y_T\) is the observed value at time \(T\). We can write equivalently,

\[
\varphi_{y}(\beta, s) = \sum_{k=1}^{K} \beta_k x_{kT} + \hat{F}^{-1}(s), \quad s \sim U[0, 1].
\]

Under the entropy approach, we rewrite equations (45)-(46) as

\[
Y_{T+1} = \varphi_{y}(P, s) = \sum_{k=1}^{K} x_{kT} \sum_{m=1}^{M} p_{km} z_{km} + \hat{F}^{-1}(s).
\]

To forecast the future observation conditional on all information of the observed inputs at time \(T\), \(Y_{T+1} | x_{kT}\), the predictive belief function and plausibility function can then be approximated using Monte Carlo simulation as described in Section 3.1.2. We, then, compute the maximum and minimum of the linear regression function \(\varphi_{y}(P, s)\) subject to \(p_{y}(P) \geq \omega_{j}\) for minimum and maximum problems. The additional constraints are

\[
\begin{align*}
\sum_{m=1}^{M} p_{km} &= 1, \\
\sum_{m=1}^{M} \omega_{jm} y_{Tm} &= 1.
\end{align*}
\]

**Example 3.** As an example, we illustrate the proposed method by applying them to the data from the operation of a plant for the oxidation of ammonia to nitric acid, measured on 21 consecutive days. The data are analyzed by Brownlee [16] in a simple regression setting, and it is already available within R software. The dataset consists of 4 variables, namely, air flow to the plant (AIR), cooling water inlet temperature (WATER), acid concentration (percentage minus 50 times 10 (ACID)), and ammonia lost (percentage times 10) (LOSS). We consider LOSS as the dependent variable, and AIR, WATER, and ACID as independent variables. Thus, we consider the following linear regression model:

\[
LOSS_t = \beta_0 + \beta_1 AIR_t + \beta_2 WATER_t + \beta_3 ACID_t + \epsilon_t.
\]

By using the Maximum likelihood (ML) and GME estimations, we can obtain the estimated parameters, standard errors, and \(p\) values. Table 6 provides the estimated results of our ML and GME estimates altogether with some basic statistics (standard errors and \(p\) values) and the plausibilities \(p(\beta_k = 0)\) (see [7]). However, in the case of GME estimation, the computation of the plausibilities \(p(\beta_k = 0)\) is different. Recall that, in the entropy approach, the estimated parameter of the model cannot be obtained directly from the
estimation, but it is computed as the summation of all expected support points. Thus, it would be reasonable to have
\[ p(\sum_{m=1}^{M} P_{km} e_{km} = 0). \]
We know that if
\[ p_{k1} = p_{k2} = \cdots = p_{kM} = 1/M, \] then
\[ \sum_{m=1}^{M} P_{km} e_{km} = 0. \]
Hence, we turn our attention to test \( p(\sum_{m=1}^{M} P_{km} e_{km} = 1/M). \) In our approach, suppose \( M = 3 \) and we have \( \hat{\beta}_k = 0, \) whenever \( p_{k1} = p_{k3}. \) To compute this, we need to maximize the entropy under the usual constraints and additional constraint \( p_{k1} = p_{k3} = 1/3. \) Therefore, we can further compute the
\[ p(\hat{\beta}_k = 0) = p(\hat{p}_{k1} = \hat{p}_{k3}). \] In this test, we test whether
\[ p(\hat{p}_{k1} = \hat{p}_{k3}) \] is less than \( p(\hat{\beta}_k = \hat{\beta}_3) = p(\hat{p}_{k1}, \hat{p}_{k2}, \hat{p}_{k3}). \)
Let \( P_{km} \) be the optimal probability of variable \( k \) at support \( m. \) If
\[ p(\hat{p}_{k1} = \hat{p}_{k3}) < p(\hat{p}_{k1}, \hat{p}_{k2}, \hat{p}_{k3}), \] the hypothesis of the coefficient \( \hat{\beta}_k \) equal to zero is rejected, indicating that this coefficient is significant.

Figure 4 shows the marginal contour functions \( p(\hat{\beta}_k = 0) \) for parameter ACID obtained from the likelihood-based approach, and we can observe that the \( p(\hat{\beta}_3 = 0) \) is greater than 0.10, indicating that the acid concentration variable is said to have insignificant effect to ammonia lost. In Figure 5, the marginal contour functions of \( p(\hat{\beta}_k = 0) = p(\hat{p}_{k1} = \hat{p}_{k3}) \) are plotted for parameters AIR (\( k = 1 \)), WATER (\( k = 2 \)), and ACID (\( k = 3 \)). We can observe that

\[
\begin{align*}
\text{Figure 4: Marginal contour functions for the coefficients of variables AIR obtained from (a) normal likelihood, (b) student-t likelihood, and (c) skewed student-t likelihood.} \\
\text{Figure 5: Marginal plausibility of probabilities (Shannon entropy) plotted in blue lines and the plausibility of } p(\hat{p}_{k1}, \hat{p}_{k2}, \hat{p}_{k3}) \text{ plotted in red horizon line. Similar plots are also obtained for Renyi and Tsallis entropies.}
\end{align*}
\]

\[
\begin{align*}
\text{Example 4. Continuing Example 3, we then consider the task of prediction problem. We consider the ammonia lost as the dependent variable and only AIR and WATER as the inputs since they show a significant effect on LOSS variable. We have re-estimated the model, and the result is provided in Table 7 and Figure 6. We can observe that the similar significant result of the estimated parameters is obtained for all methods. We also make a prediction of LOSS when the inputs are known at time } T; \text{ hence, the prediction equation is based on}
\end{align*}
\]

\[
\text{LOSS}_T = \hat{\beta}_0 + \hat{\beta}_1 \text{AIR}_T + \hat{\beta}_2 \text{WATER}_T + \epsilon_T, 
\]

with \( \epsilon_T \) which is assumed to be empirical distribution. We can write, equivalently,
Table 6: Regression results.

|    | norm      | std       | sstd      | Shannon entropy | Renyi entropy | Tsallis entropy |
|----|-----------|-----------|-----------|-----------------|---------------|----------------|
|    | Estimate  | Estimate  | Estimate  | Estimate        | Estimate      | Estimate       |
| Intercept | $-43.7037$ (0.0000) | $-43.3522$ (0.0000) | $-41.3097$ (0.0000) | $-38.6949$ (0.0000)*** | $-38.6950$ (0.0000)*** | $-38.7344$ (0.0000)*** |
| AIR | 0.8891 (0.0000) | 0.8980 (0.0000) | 0.8605 (0.0000) | 0.7894 (0.0000)*** | 0.7890 (0.0000)*** | 0.7317 (0.0000)*** |
| WATER | 0.8166 (0.0049) | 0.7823 (0.0043) | 0.8154 (0.0079) | 1.1918 (0.0000)*** | 1.1918 (0.0000)*** | 1.2315 (0.0000)*** |
| ACID | $-0.1071$ (0.3361) | $-0.1094$ (0.3114) | $-0.1155$ (0.1733) | $-0.1821$ (0.2631) | $-0.1821$ (0.2630) | $-0.1828$ (0.2644) |

() is $p$ value, [] is $Pr(\beta_j = 0)$, and *** is $Pr(\hat{p}_{k1} = \hat{p}_{k3}) < Pr(\hat{p}_{k1}, \hat{p}_{k2}, \hat{p}_{k3})$. For the case of entropy, the support $z_{im}$ is initially set to $(-50, 0, 50)$ and the supports for $\nu_{tm}$ to $(-3\sigma, 0, 3\sigma)$, where $\sigma$ is computed from the conventional LS estimation.

![Figure 6: Continued.](image-url)
Figure 6: Lower and upper cdfs for the prediction problem in linear regression. (a) Normal. (b) Student-t. (c) Skewed Student. (d) Shannon Entropy. (e) Renyi Entropy. (f) Tsallis Entropy.

Table 7: Regression results of the reduced model with only two inputs.

|                | norm  | std   | Estimate              | Shannon entropy | Renyi entropy | Tsallis entropy |
|----------------|-------|-------|-----------------------|-----------------|---------------|-----------------|
| Intercept      | -57.0752 (0.0000) | [0.0000] | -51.8325 (0.0000)*** | -41.4330 (0.0000)*** | -50.9705 (0.0000)*** |
| AIR            | 0.8629 (0.0000) | [0.0000] | 0.7633 (0.0000)*** | 0.7561 (0.0000)*** | 0.7029 (0.0000)*** |
| WATER          | 0.8033 (0.0233) | [0.0000] | 0.7831 (0.0079)     | 1.1479 (0.0000)*** | 1.1479 (0.0000)*** | 1.1358 (0.0000)*** |
| Expected prediction | 26.5564 <19.1543, 27.0138 <8.3351, 24.5375 <18.4833, 23.5138 <17.9744, 26.1581 <18.1415, 24.0054 <17.0348 | 9.5375 | 8.5138 | 11.581 | 9.0054 |
| Prediction bias | 11.3564 | 12.0138 | 9.5375 | 8.5138 | 11.581 | 9.0054 |

<> is the prediction interval, () is p value, [ ] is \( P(\beta_j = 0) \), and *** is \( P(\hat{p}_1 = \hat{p}_2 = \hat{p}_3) < pl(\hat{p}_{1:2:3}) \). For the case of Entropy, the support \( z_{im} \) is initially set to \((-100, 0, 100)\) and the supports for \( r_{im} \) to \((-3\sigma, 0, 3\sigma)\), where \( \sigma \) is computed from the conventional LS estimation.

Figure 7: Marginal contour functions \( pl_j(\theta) \) in two-dimensional parameter subspaces: normal likelihood.
Figure 8: Marginal contour functions $p l_j (\theta)$ in two-dimensional parameter subspaces: Student-t likelihood.

Figure 9: Continued.
\[
\text{LOSS}_T^* = \beta_0 + \beta_1 \text{AIR}_T + \beta_2 \text{WATER}_T + \phi^{-1}(s) \quad (52)
\]

The predictive belief function on \(\text{LOSS}_T^*\) can then be approximated using the methods described in Section 3.2.2. Figure 6 displays the lower and upper cdfs of the various predictive belief functions, approximated using \(N = 5000\) randomly generated focal sets. The prediction value of \(\text{LOSS}_{T+21}^*\) is plotted by the vertical blue dotted line as well as the \(\alpha\) quantile intervals with \(\alpha = 0.05, 0.95\). We then compare the prediction performance based on prediction bias. The true value of \(\text{LOSS}_{T+21}^*\) is 15; thus, we calculate prediction bias from \([\text{LOSS}_{T+21}^* - 15]\). According to the last row of Table 7, we can observe that the Shannon entropy-based belief function method outperforms the others as the lowest bias is obtained. Moreover, we find that the prediction interval of the Shannon entropy method provides a reliable interval. Although the 0.05 quantile interval does not cover the true value, the lower bound of its interval is closer to the true value = 15. Therefore, to confirm the reliability of the method, we extend the lower quantile interval of the Shannon entropy-based method to be \(\alpha = 0.01\) and we can observe that the true observed data is contained in the 0.01 lower quantile interval (green dashed line).

4. Conclusion

In recent studies, forecasting is an important tool for decision-making and strategic planning. Describing the uncertainty of the forecasts accurately is thus a very important issue. The approach developed in this paper is to model estimation uncertainty using a belief function. The belief function \(\text{Bel}_\theta\) is just another piece of evidence. It can be combined with the likelihood belief function and the joint belief function by Dempster’s rule. The combined belief function is then, as before, marginalized on \(\phi^{-1}(p, s)\) to get the upper and lower predictive belief function.

As the information of the distribution function is generally unknown, we thus have a concern that the estimation and prediction problems that are solved in likelihood function-based belief function framework might have a bias if the observed data are not normally distributed. Therefore, this study takes advantage of the entropy approach to construct a belief function and illustrate these solutions under the autoregressive (AR) and regression contexts. Specifically, we replace the parametric likelihood function with the entropy measure (Shannon, Renyi, and Tsallis) to derive the entropy-based belief function and thus the model misspecification is avoided since no distributional assumption is made.

To validate the performance of our proposed method, the simulation study is conducted. The results reveal that our method performs remarkably well in comparison with a host of competing methods. We would like to highlight that if the distribution is known and we estimate the model based on the true distribution, the accurate result is obtained. However, the error distribution is sometimes unknown in practice, these results confirm the advantage of the Shannon entropy-based belief function, which performs uniformly well over a wide range of error distributions.

Furthermore, our entropy-based belief function is also applied to real data. In the first real application model, we conduct the out-of-sample forecasts from the year 2014 to 2018 and the results are displayed in Table 5. We find that the entropy-based belief function for AR (1) exhibits good performance in the out-of-sample forecasts relative to the other competing methods. We then apply our method to predict the growth rate of GDP in 2019 and find that the growth rate of the Thai GDP is around 3.961%. In the second real application, we apply our method to predict the future value using the linear regression with certain inputs. We can observe that the estimates under the likelihood-based methods and entropy-based methods exhibit slight differences, as expected. The forecasting bias of our entropy-based belief is lower than those via likelihood-based belief, in particular the Shannon entropy. This suggests that the entropy-based belief seems to produce more accurate estimates of the prediction under the regression context as well. From these results, it is encouraging that the use of entropy-based belief offers a better alternative in the prediction problem.

In brief, this study proposes the entropy-based belief function as an alternative method in place of likelihood predictive distributions. The method is very general and can
be used with any parametric distributions. However, as the computational complexity is quite high, we may face the higher computational cost in the large datasets. We leave the performance of the computation of our method to future work. In addition, for the further studies, to make a more accurate prediction, this approach may be extended to combine a random interval encoding expert assessments of the prediction from several government agencies, investors, business enterprises, and private/public institutions. Furthermore, the forecasting performance of entropy-based belief function could be improved by increasing the order $d$ in the Renyi and Tsallis entropies. Therefore, we suggest the future study to conduct the higher order of Renyi and Tsallis entropies. Finally, it deserves comparing our approach with the posterior distribution-based belief function in the Bayesian context and Deng entropy-based belief function to confirm the performance of our entropy-based method.

**Appendix**

**A. Out-of-Sample Prediction: Optimization Evaluation**

A.1. Normal Likelihood-Based Belief Function Contour Plot. Normal likelihood-based belief function contour plot of marginal contour functions $p_{θ}(θ)$ in two-dimensional parameter subspaces is shown in Figure 7.

A.2. Student-t Likelihood-Based Belief Function Contour Plot. Student-t likelihood-based belief function contour plot of marginal contour functions $p_{θ}(θ)$ in two-dimensional parameter subspaces is shown in Figure 8.

A.3. Skewed Student-t Likelihood-Based Belief Function Contour Plot. Skewed Student-t likelihood-based belief function contour plot of marginal contour functions $p_{θ}(θ)$ in two-dimensional parameter subspaces is shown in Figure 9.

**Data Availability**

In this study, we use the simulated data to show the performance of our model, and the simulation processes are already explained in the paper. For the real data analysis section, the data can be freely collected from Thomson Reuters Datastream and R-package. However, the data are available from the corresponding author upon request (woraphon.econ@gmail.com).

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**Acknowledgments**

This research was supported by CMU Junior Fellowship Program. It was also supported by the Center of Excellence in Econometrics, Faculty of Economics, at Chiang Mai University. The authors thank Prof. Theirry Deneux and three reviewers for valuable comments to improve this paper.

**References**

[1] J. Mitchell, “Where are we now? The UK recession and nowcasting GDP growth using statistical models,” *National Institute Economic Review*, vol. 209, pp. 60–69, 2009.

[2] K. Lahiri and G. Monokrousos, “Nowcasting US GDP: the role of ISM business surveys,” *International Journal of Forecasting*, vol. 29, no. 4, pp. 644–658, 2013.

[3] O. Kanjanatarakul, S. Sriroonchitta, and T. Deneux, “Forecasting using belief functions: an application to marketing econometrics,” *International Journal of Approximate Reasoning*, vol. 55, no. 5, pp. 1113–1128, 2014.

[4] A. P. Dempster, “New methods for reasoning towards posterior distributions based on sample data,” *The Annals of Mathematical Statistics*, vol. 37, no. 2, pp. 355–374, 1966.

[5] G. A. Shafer, *Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, USA, 1976.

[6] H. T. Nguyen, “On random sets and belief functions,” *Journal of Mathematical Analysis and Applications*, vol. 65, no. 3, pp. 531–542, 1978.

[7] O. Kanjanatarakul, T. Deneux, and S. Sriroonchitta, “Prediction of future observations using belief functions: a likelihood-based approach,” *International Journal of Approximate Reasoning*, vol. 72, pp. 71–94, 2016.

[8] T. Deneux, “Likelihood-based belief function: justification and some extensions to low-quality data,” *International Journal of Approximate Reasoning*, vol. 55, 2014.

[9] N. Thianpaen, J. Liu, and S. Sriroonchitta, “Time series forecast using AR-belief approach,” *Thai Journal of Mathematics*, vol. 14, no. 3, pp. 527–541, 2016.

[10] N. Chakpitak, W. Yamaka, and S. Sriroonchitta, “Comparing linear and nonlinear models in forecasting telephone subscriptions using likelihood based belief functions,” in *Proceedings of the International Conference of the Thailand Econometrics Society*, pp. 363–374, Chiang Mai, Thailand, August 2018.

[11] E. T. Jaynes, “Information theory and statistical mechanics,” *Physical Review*, vol. 106, no. 4, pp. 620–630, 1957.

[12] A. P. Dempster, “The Dempster-Shafer calculus for statisticians,” *International Journal of Approximate Reasoning*, vol. 48, no. 2, pp. 365–377, 2008.

[13] A. Rényi, *On Measures of Entropy and Information*, Hungarian Academy of Sciences, Budapest, Hungary, 1961.

[14] C. Tsallis, “Possible generalization of Boltzmann-Gibbs statistics,” *Journal of Statistical Physics*, vol. 52, no. 1-2, pp. 479–487, 1988.

[15] A. Golan and J. M. Perloff, “Comparison of maximum entropy and higher-order entropy estimators,” *Journal of Econometrics*, vol. 107, no. 1-2, pp. 195–211, 2002.

[16] K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John Wiley & Sons, New York, NY, USA, 1965.