How to Make Plausibility-Based Forecasting More Accurate

Kongliang Zhu[1], Nantiworn Thianpaen[2], and Vladik Kreinovich[3]
[1]Faculty of Economics, Chiang Mai University, Thailand
email 258zkl@gmail.com
[2]Faculty of Economics, Chiang Mai University, Thailand, and Faculty of Management Sciences, Suratthani Rajabhat University, Thailand, nantiworn@outlook.com
[3]Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

1. Outline

- In recent papers, a new plausibility-based forecasting method was proposed.
- This method has been empirically successful.
- One of the steps – selecting a uniform probability distribution for the plausibility level – is heuristic.
- Is this selection optimal or a modified selection would like to a more accurate forecast?
- In this talk, we show that the uniform distribution does not always lead to (asymptotically) optimal estimates.
- We show how to modify this step so that the resulting estimates become asymptotically optimal.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

2. Need for Prediction

- One of the main objectives of science is:
  - given the available data $x_1, \ldots, x_n$,
  - to predict future values of different quantities y.
- The usual approach to solving this problem consists of two stages:
  - first, we find a *model* that describes the observed data; and
  - then, we use this model to predict the future value of each of the quantities y.
- Often, it is sufficient to have a *deterministic model*: $x_i = f_i(p)$ and $y = f(p)$ for some parameters p.
- We use the observed values to estimate p; then, we use these estimates to predict the desired future values y.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

3. Deterministic Prediction and Beyond

- This is how, e.g., solar eclipses can be predicted for centuries ahead; here:
  - parameters p includes initial locations, initial velocities, and masses of all the celestial bodies;
  - observations x_i are visible locations of celestial bodies at different moments of time.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

4. Need for Statistical Prediction

- In most practical problems, a fully deterministic prediction is not possible:
  - in addition to the parameters p,
  - both the observed values x_i and the future value y are affected by parameters z_j beyond our control,
  - parameters that can be viewed as *random*.
- Thus, we have a *probabilistic model*:

$$x_i = f(p, z_1, \ldots, z_m) \text{ and } y = f(p, z_1, \ldots, z_m),$$

where z_j are random variables.

- Usually:
  - we do not know the exact probability distribution for the variables z_i, but
  - we know a finite-parametric family of distributions that contains the actual (unknown) distribution.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

5. Need for Statistical Prediction

- For example, we may know that the distribution is Gaussian, or that it is uniform.
- Let q denote the parameter(s) that describe this distribution.
- In this case, both x_i and y are random variables whose distribution depends on all the parameters $\theta = (p, q)$:

$$x_i \sim f_{i,\theta} \text{ and } y \sim f_\theta.$$

- In this case, to identify the model:
  - we first estimate the parameters θ based on the observations $x_1, \ldots, x_n$, and then
  - we use the distribution f_θ corresponding to these parameter values to predict the values y
  - or, to be more precise, to predict the probability of different values of y.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

6. Need for a Confidence Interval

- In the statistical case, we cannot predict the *exact* value of y.
- So, it is desirable to predict the *range* of possible values of y.
- For many distributions (e.g., for normal), it is possible to have arbitrarily small and arbitrarily large values.
- In such situations, there is no *guaranteed* range of values of y.
- However, we can still try to estimate a *confidence interval*, i.e.,
  - for a given small value $\alpha > 0$,
  - an interval $[\underline{y}_\alpha, \overline{y}_\alpha]$ that contains the actual value y with confidence $1 - \alpha$.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

7. Confidence Interval (cont-d)

- In other words, we would like to find an interval for which $\text{Prob}(y \in [\underline{y}_\alpha, \overline{y}_\alpha]) \geq \alpha$.
- When we know the cdf $F(y) \stackrel{\text{def}}{=} \text{Prob}(Y \leq y)$, we can take $[\underline{y}_\alpha, \overline{y}_\alpha] = \left[F^{-1}\left(\frac{\alpha}{2}\right), F^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$.
- In general, a statistical estimate based on a finite sample is only approximate.
- Thus, based on a finite sample, we can predict the value of the parameters θ only approximately.
- Therefore, we only have an approximate estimate of the probabilities of different values of y.
- So, instead of the actual cdf $F(y)$, we only know the bounds on the cdf: $\underline{F}(y) \leq F(y) \leq \overline{F}(y)$.
- We want to select $[\underline{y}_\alpha, \overline{y}_\alpha]$ so that the probability of being outside this interval is guaranteed not be $\leq \alpha$.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

8. Confidence Interval (cont-d)

- We can take $F(\underline{y}_\alpha) \le \dfrac{\alpha}{2}$ for all possible $F(x)$.
- This can be achieved for $\underline{y}_\alpha = (\overline{F})^{-1}\left(\dfrac{\alpha}{2}\right)$.
- Similarly, we can take $\overline{y}_\alpha = (\underline{F})^{-1}\left(1 - \dfrac{\alpha}{2}\right)$.
- *Plausibility-based forecasting:* we start by forming a *likelihood function.*
- We assume that the probability density function corresponding to each observation x_i has the form $f_{i,\theta}(x_i)$.
- We assume that x_i are independent, so

$$L_x(\theta) = \prod_{i=1}^{n} f_{i,\theta}(x_i).$$

- $L_x(\theta)$ is normally used to find the *maximum likelihood* estimate $\hat{\theta}$: $L_x\left(\hat{\theta}\right) = \max\limits_{\theta} L_x(\theta)$.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

9. Plausibility Approach (cont-d)

- Instead, we use $L_x(\theta)$ to define the *plausibility function*

$$\mathrm{pl}_x(\theta) = \frac{L_x(\theta)}{\sup\limits_{\theta'} L_x(\theta')} = \frac{L_x(\theta)}{L_x\left(\hat{\theta}\right)}.$$

- Based on $\mathrm{pl}_x(\theta)$, we define, for each $\omega \in [0,1]$, a *plausibility region* $\Gamma_x(\omega) = \{\theta : \mathrm{pl}_x(\theta) \geq \omega\}$.
- We represent y as $g(\theta, z)$, where z is uniform on $[0,1]$ and $g(\theta, z) = F_\theta^{-1}(z)$.
- We compute the belief and plausibility of each set A of possible values of θ: $\mathrm{Bel}(A) = \mathrm{Prob}(g(\Gamma_x(\omega), z) \subseteq A)$,

$$\mathrm{Pl}(A) = \mathrm{Prob}(g(\Gamma_x(\omega), z) \cap A \neq \emptyset).$$

- Here, ω, z are uniformly distributed on $[0,1]$.
- Then, we compute $\underline{F}(y) = \mathrm{Bel}((-\infty, y])$, $\overline{F}(y) = \mathrm{Pl}((-\infty, y])$, and the confidence intervals.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

10. Remaining Problem

- The new approach has led to interesting applications.
- However, it is not clear why we select a uniform distribution for ω.
- Yes, this ω-distribution sounds like a reasonable idea:
  - we know that ω is located on the interval $[0, 1]$,
  - we do not know which values ω are more probable and which are less probable,
  - so we select a uniform distribution.
- However, we are not just making reasonable estimates.
- We want to make predictions with confidence.
- Maybe the above interval is only an approximation.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

11. Remaining Problem

- Maybe the above interval is only an approximation; in this case:
  - by selecting a different probability distribution for ω,
  - we can make the resulting forecasting more accurate.
- This is the problem that we will be analyzing in this talk.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

12. Let Us Consider the Simplest Possible Case

- Let us consider the simplest case when:
  - we have only one parameter $\theta = \theta_1$,
  - the predicted value y simply coincides with the value of this parameter: $g(\theta, z) = \theta$, and
  - the likelihood $L_x(\theta)$ is continuous and decreasing as we move away from the ML estimate $\hat{\theta}$:
    - $L_x(\theta)$ strictly increases for $\theta \le \hat{\theta}$;
    - $L_x(\theta)$ strictly decreases for $\theta \ge \hat{\theta}$.
- The first two conditions are really restrictive.
- However, monotonicity holds in the overwhelming majority of practical situations.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

13. Let Us Analyze the Simplest Possible Case

- In this case, we want bounds containing θ with a given confidence α, i.e., we want a confidence interval.
- In this case, the set $\{\theta : \mathrm{pl}_x(\theta) \geq \omega\}$ is an interval $[\theta^-, \theta^+]$, where $\mathrm{pl}_x(\theta^\pm) = \omega$.
- In these terms, the condition $\{\theta : \mathrm{pl}_x(\theta) \geq \omega\} = [\theta^-, \theta^+] \subseteq (-\infty, y]$ means $\theta^+ \leq y$.
- Due to monotonicity, $\theta^+ \leq y \Leftrightarrow \omega = \mathrm{pl}(\theta^+) \geq \mathrm{pl}(y)$.
- When ω is uniformly distributed on $[0, 1]$, then, for all z, we have $\mathrm{Prob}(\omega \geq z) = 1 - z$.
- In particular, for $z = \mathrm{pl}_x(y)$, we have $\underline{F}(y) = 1 - \mathrm{pl}_x(y)$.
- Thus, the upper endpoint $\overline{\theta}_\alpha$ of the confidence interval is the value for which $\mathrm{pl}_x(\overline{\theta}_\alpha) = \dfrac{\alpha}{2}$.
- Similarly, for $\underline{\theta}_\alpha$, we have $\mathrm{pl}_x(\underline{\theta}_\alpha) = \dfrac{\alpha}{2}$.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

14. The Simplest Possible Case (cont-d)

- So, $[\underline{\theta}_\alpha, \overline{\theta}_\alpha] = \left\{\theta : \mathrm{pl}_x(\theta) \geq \dfrac{\alpha}{2}\right\}$.
- In terms of $L_x(\theta)$, $\mathrm{pl}_x(\theta) \geq \dfrac{\alpha}{2} \Leftrightarrow \dfrac{L_x(\theta)}{L_x\left(\hat{\theta}\right)} \geq \dfrac{\alpha}{2} \Leftrightarrow$

$$\ln\left(L_x(\theta)\right) \geq \ln\left(L_x\left(\hat{\theta}\right)\right) - (\ln(2) + |\ln(\alpha)|).$$

- On the other hand, according to Wilks's theorem, the confidence interval is the set of all θ for which

$$\ln\left(L_x(\theta)\right) \geq \ln\left(L_x\left(\hat{\theta}\right)\right) - \frac{1}{2} \cdot \chi^2_{1,1-\alpha}.$$

- $\chi^2_{1,1-\alpha}$ is defined in terms of the χ^2_1-distribution, the square of $n \sim N(0,1)$: $\mathrm{Prob}(n^2 \leq \chi^2_{1,1-\alpha}) = 1 - \alpha$.
- This is different from the above estimate, so we need to modify the plausibility-based method.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

15. How to Modify Plausibility-Based Forecasting: Analysis

- The problem comes since we use a (heuristically selected) unform distribution for ω.
- We thus need to find an alternative distribution, for which $\overline{F}(y) = \text{Prob}(\omega \le p_x(y)) = \dfrac{\alpha}{2}$.
- For the Wilks's bound,

$$p_x(y) = \frac{L_x(\theta)}{L_x\left(\hat{\theta}\right)} = \exp\left(-\frac{1}{2}\cdot\chi^2_{1,1-\alpha}\right), \text{ so}$$

$$\text{Prob}\left(\omega \le \exp\left(-\frac{1}{2}\cdot\chi^2_{1,1-\alpha}\right)\right) = \frac{\alpha}{2}.$$

- By definition of $\chi^2_{1,1-\alpha}$, $\text{Prob}\left(n^2 \le \chi^2_{1,1-\alpha}\right) = 1-\alpha$, so

$$\text{Prob}\left(n^2 \ge \chi^2_{1,1-\alpha}\right) = (1-(1-\alpha)) = \alpha.$$

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

16. Analysis (cont-d)

- The inequality $n^2 \geq \chi^2_{1,1-\alpha}$ occurs in two equally probable situations:
  - when n is positive and $n \geq \sqrt{\chi^2_{1,1-\alpha}}$ and
  - when n is negative and $n \leq -\sqrt{\chi^2_{1,1-\alpha}}$.
- Thus, the probability of each of these two situations is equal to $\dfrac{\alpha}{2}$; in particular, we have:

$$\text{Prob}\left(n \leq -\sqrt{\chi^2_{1,1-\alpha}}\right) = \frac{\alpha}{2}.$$

- Let us transform the desired inequality to this form.
- The inequality $\omega \leq \exp\left(-\dfrac{1}{2} \cdot \chi^2_{1,1-\alpha}\right)$ is equivalent to

$$-\sqrt{-2\ln(\omega)} \leq -\sqrt{\chi^2_{1,1-\alpha}}.$$

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

17. Analysis (cont-d)

- The inequality $\omega \le \exp\left(-\frac{1}{2}\cdot\chi^2_{1,1-\alpha}\right)$ is equivalent to

$$-\sqrt{-2\ln(\omega)} \le -\sqrt{\chi^2_{1,1-\alpha}}.$$

- Thus, the desired inequality is equivalent to

$$\text{Prob}\left(-\sqrt{-2\ln(\omega)} \le -\sqrt{\chi^2_{1,1-\alpha}}\right) = \frac{\alpha}{2}.$$

- By definition of χ^2, this equality is attained if we have $n = -\sqrt{-2\ln(\omega)}$.
- In this case, $-2\ln(\omega) = n^2$, hence $\omega = \exp\left(-\frac{n^2}{2}\right)$.
- This is the distribution for ω that we should use.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

18. What Is the Probability Density Function (pdf) of This Distribution?

- If $X \sim \rho_X(x)$, then $\text{Prob}(f(x) \le Y \le f(x+dx)) =$

$$\text{Prob}(x \le X \le x + dx) = \rho_X(x) \cdot dx.$$

- Here, $f(x+dx) = f(x) + f'(x) \cdot dx = y + f'(x) \cdot dx$, so

$$\text{Prob}(f(x) \le Y \le f(x+dx)) = \text{Prob}(y \le Y \le y + f'(x) \cdot dx) =$$

$$\rho_Y(y) \cdot |f'(x)| \cdot dx, \text{ so } \rho_Y(y) = \frac{\rho_X(x)}{|f'(x)|}.$$

- In our case, $f(x) = \exp\left(-\dfrac{x^2}{2}\right)$, hence

$$\rho(\omega) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{2 \cdot |\ln(\omega)|}} = \frac{1}{2 \cdot \sqrt{\pi} \cdot \sqrt{|\ln(\omega)|}}.$$

- The value $\ln(\omega)$ is changing very slowly, so $\omega \approx U(0,1)$.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

19. Resulting Recommendations

- First, we define the likelihood function $L_x(\theta)$ and then find its largest possible value $L_x\left(\hat{\theta}\right) = \max\limits_{\theta} L_x(\theta)$.
- Then, we define the plausibility function as

$$\mathrm{pl}_x(\theta) = \frac{L_x(\theta)}{\sup\limits_{\theta'} L_x(\theta')} = \frac{L_x(\theta)}{L_x\left(\hat{\theta}\right)}.$$

- Based on this plausibility function, we define, for each real number $\omega \in [0, 1]$, a plausibility region

$$\Gamma_x(\omega) = \{\theta : \mathrm{pl}_x(\theta) \geq \omega\}.$$

- We then represent a probability distribution for y as $y = g(\theta, z)$ for an auxiliary variable $z \sim U(0, 1)$.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

20. Resulting Recommendations (cont-d)

- Based on these regions, we can compute, for each set A:

$$\mathrm{Bel}(A) = \mathrm{Prob}(g(\Gamma_x(\omega), z) \subseteq A),$$

$$\mathrm{Pl}(A) = \mathrm{Prob}(g(\Gamma_x(\omega), z) \cap A \neq \emptyset).$$

- Here, $z \sim U(0,1)$ and $\omega = \exp\left(-\dfrac{n^2}{2}\right)$, where $n \sim N(0,1)$.
- After that, we compute $\underline{F}(y) = \mathrm{Bel}((-\infty, y])$ and $\overline{F}(y) = \mathrm{Pl}((-\infty, y])$.
- Then, for any $\alpha > 0$, we predict that, with confidence $1 - \alpha > 0$,

$$y \in \left[(\overline{F})^{-1}\left(\frac{\alpha}{2}\right), (\underline{F})^{-1}\left(1 - \frac{\alpha}{2}\right)\right].$$

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .

21. Acknowledgments

- We acknowledge the support of the Center of Excellence in Econometrics, Chiang Mai University.
- This work was also supported in part:
  - by the National Science Foundation grants:
    - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
    - DUE-0926721, and
  - by an award from the Prudential Foundation.

Need for Prediction
Deterministic . . .
Need for Statistical . . .
Plausibility Approach . . .
Remaining Problem
Let Us Consider the . . .
How to Modify . . .
What Is the . . .
Resulting . . .