# Predictability, Sensitivity, and Value

Caren Marzban

Applied Physics Lab
Dept of Statistics
University of Washington, Seattle, WA 98195

Abstract

Three ideas are examined: The first, attempts to quantify a question that arises in assessing the limits of predictability; specifically, How does the relationship between skill and lead time depend on the scale over which forecasts are averaged? The second, examines basic techniques of statistical experimental design and sensitivity analysis for the purpose of improving skill over S2S time scales. And the third, re-examines the relevance of economic value as an alternative to forecast skill in assessing the goodness of forecasts.

## Idea 1:

It is known that forecast skill deteriorates with lead time.

Lead times in the S2S range appear to involve some sort of time averaging.

A VERY simple model:

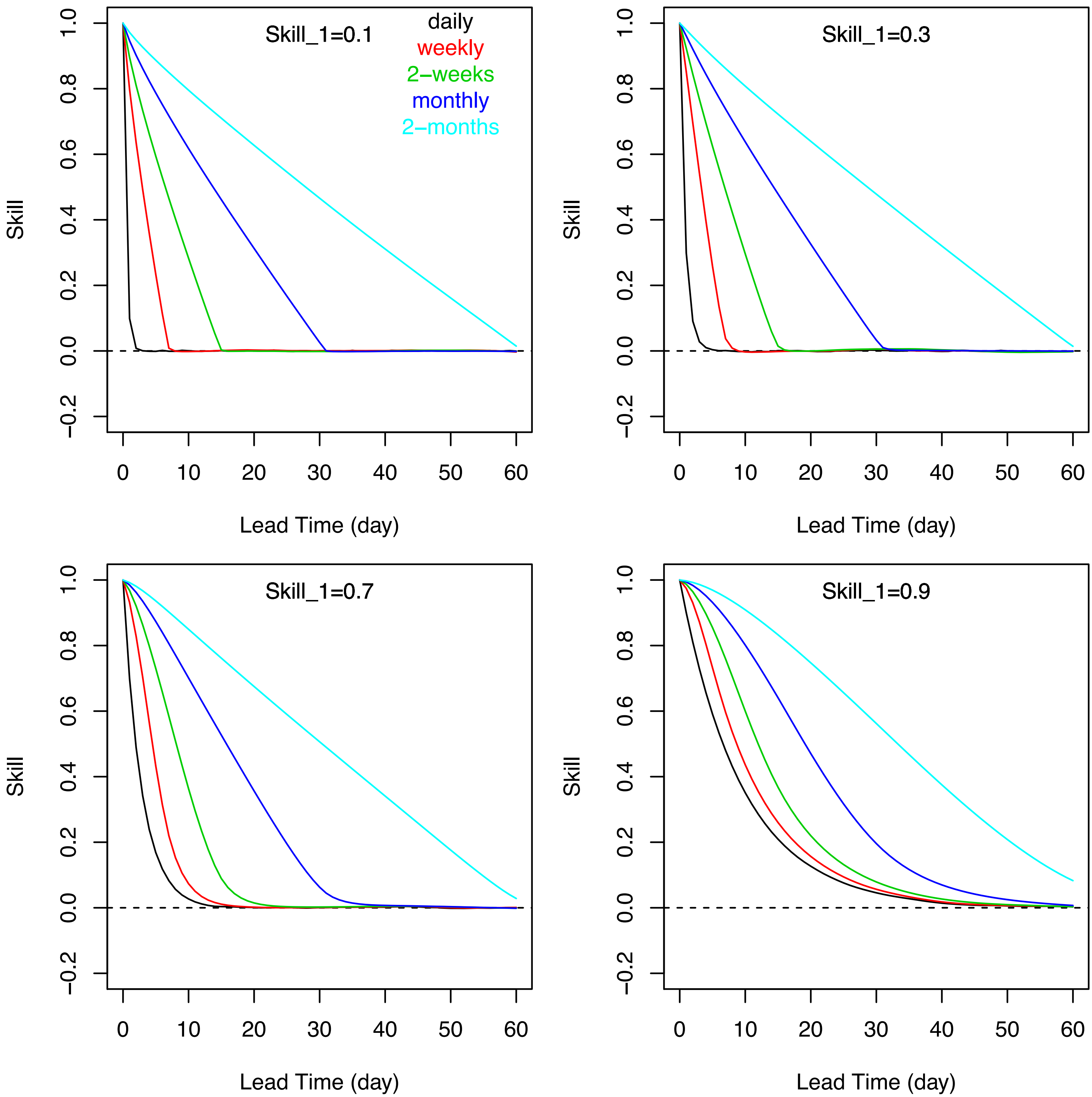
$$AR(2) \qquad x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t$$

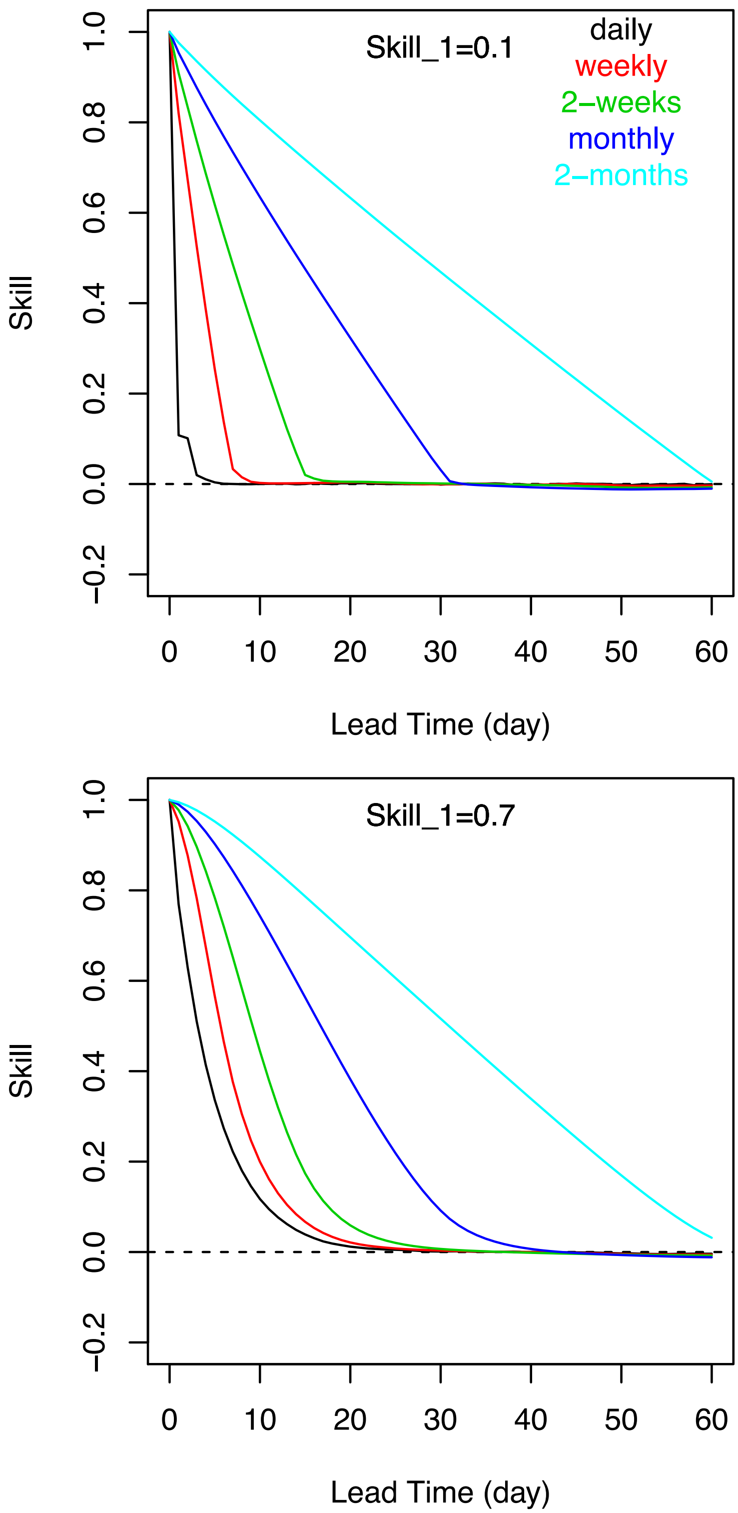
Autocorrelation function:
correlation (i.e., accuracy) vs. lag (i.e., lead time).

Running-mean: average [temperature] within a sliding window.

Confessions:
lag $\neq$ lead time
S2S forecasts $\neq$ window-averages
Real data $\neq$ AR
Correlation $\neq$ skill

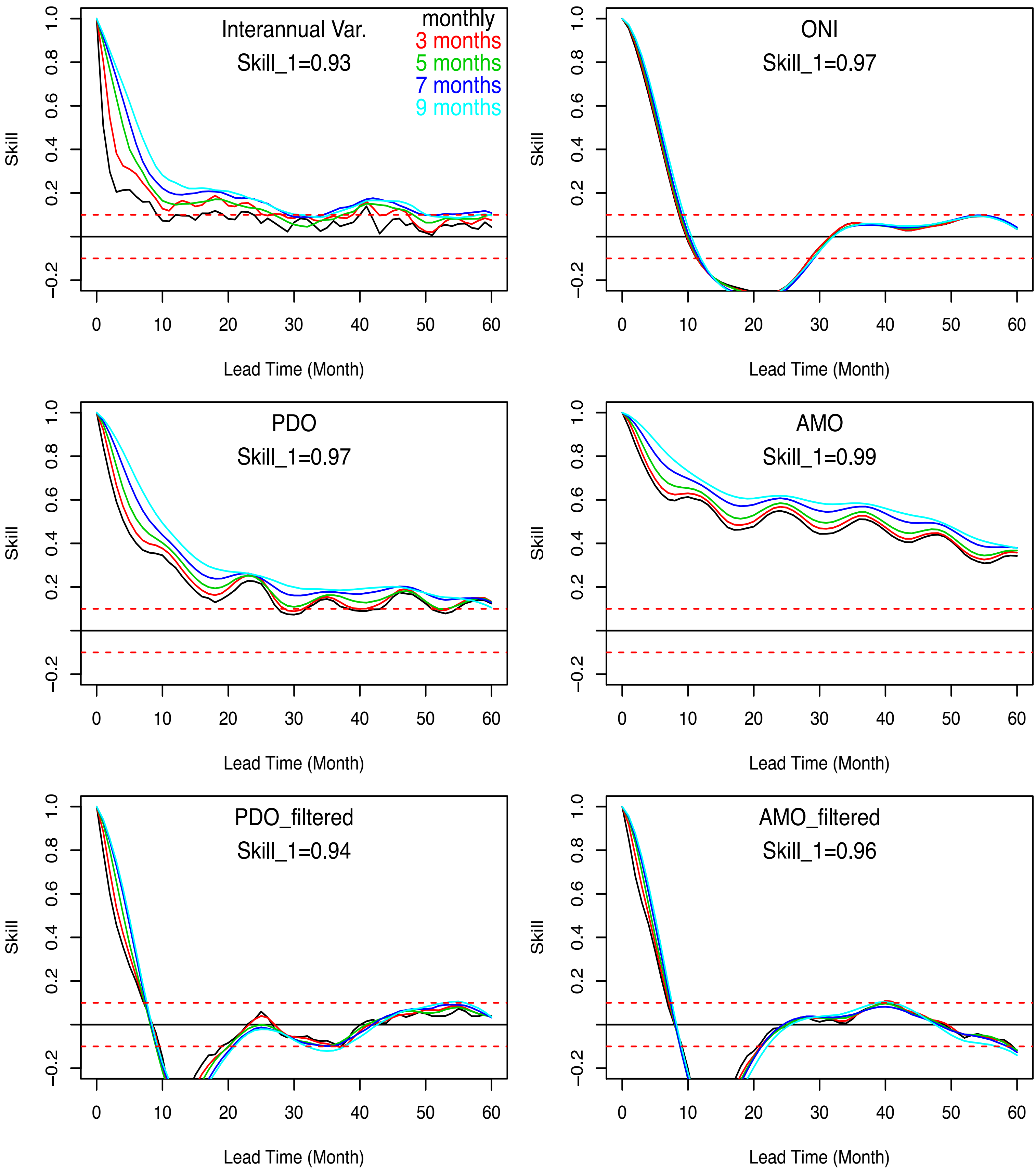
"Conclusion:" This framework may be useful for determining what sort of averaging is needed to maintain skill over longer lead times.


Skill_1=0.1
daily
weekly
2−weeks
monthly
2−months
Skill_1=0.3
Skill_1=0.7
Skill_1=0.9
Skill
Lead Time (day)
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60


Skill_1=0.1
daily
weekly
2−weeks
monthly
2−months
Skill_1=0.3
Skill_1=0.7
Skill_1=0.9
Skill
Lead Time (day)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

# Idea 1: Continued

Data (courtesy of Philippe Tissot, Texas A&M Corpus Christi).



# Idea 2

Marzban, Sandgathe, et al. (2018): objects/spatial structure $\leftarrow$ model params.
Boyle, et al. (2014): The parametric sensitivity of CAM5's MJO

Graeco-Latin Square Designs good for continuous model parameters.

Some models involve categorical (e.g., on/off) "switches."

Imagine $k$ switches (NMME, ...).
Full factorial design requires $2^k$ runs.
Magic: main effects can be estimated with VERY few runs.

E.g.
$2^8 = 256$ runs
$2^{8-4} = 16$ runs

| X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 |
|---|---|---|---|---|---|---|---|
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 |
| -1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 |
| 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 |
| -1 | -1 | 1 | -1 | 1 | 1 | 1 | -1 |
| 1 | -1 | 1 | -1 | 1 | -1 | -1 | 1 |
| -1 | 1 | 1 | -1 | -1 | 1 | -1 | 1 |
| 1 | 1 | 1 | -1 | -1 | -1 | 1 | -1 |
| -1 | -1 | -1 | 1 | 1 | 1 | -1 | 1 |
| 1 | -1 | -1 | 1 | 1 | -1 | 1 | -1 |
| -1 | 1 | -1 | 1 | -1 | 1 | 1 | -1 |
| 1 | 1 | -1 | 1 | -1 | -1 | -1 | 1 |
| -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 |
| 1 | -1 | 1 | 1 | -1 | 1 | -1 | -1 |
| -1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

E.g.,
$2^{15} = 32,768$ runs
$2^{15-11} = 16$ runs

"Conclusion:" Experimental design may be useful for "fine-tuning" current models for S2S time-scales.

# Idea 3

In addition to skill, ask about Value.

Knowing value places bounds on the necessary skill.

Indeed, value and skill can be "reversed."

The Basics of Value
(Marzban, C. 2012: Displaying economic value)

Important quantities:
1) Cost/Loss Ratio
2) Prioir/Climatological probability of event.

$$\begin{array}{cc} & \text{Action} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ \text{Event} \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0 & C \\ L & L_m \end{pmatrix} \end{array} .$$

$$V = \frac{\text{Expected savings from forecasts}}{\text{Expected savings from perfect forecasts}} = \frac{min(E_0, E_1) - E_f}{min(E_0, E_1) - E_p} ,$$

$$= \frac{min(p, \frac{C}{L} - p(\frac{C-L_m}{L})) - p - (1-p)(F)\frac{C}{L} + p(H)(1 - \frac{L_m}{L})}{min(p, \frac{C}{L} - p(\frac{C-L_m}{L})) - p\frac{L_m}{L}}$$

$H$ = Hit rate, $F$ = False alarm rate.

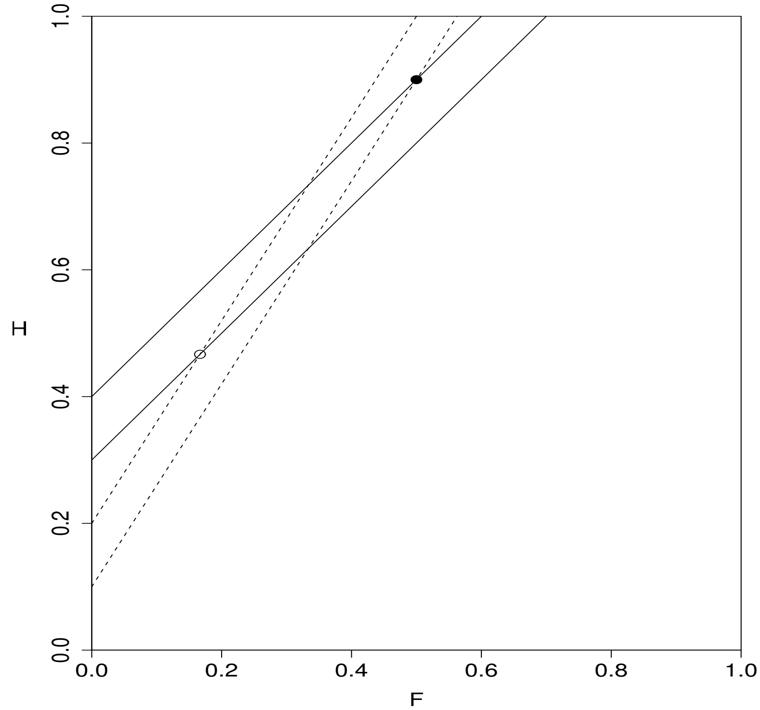



Figure 1. Reversal of quality and value. Solid parallel lines: $TSS = 0.3$ and 0.4, dashed lines $V = 0.1$, and 0.2. The forecasts corresponding to the filled circle have higher quality, but lower value than the forecasts associated with the open circle.

"Conclusion:" Especially in rare/extreme/S2S events, value is (more) important than skill.