Figure 1

The Two Rival Probability Score Functions

*Log score (as in MLE) = thin line*
*Exponential score = thick line*

Note how the exponential function rewards extreme probabilities when they are correct much more than the log score
Figure 2

In-sample measure of $L(Y | \beta^{mle}) - L(Y | \beta^c)$

Mean $pr($bankruptcy$) = 0.03195 = long$ run empirical frequency$ 
2000 repeats$

Figure 3

In-sample measure of $S(Y | \beta^{mle}) - S(Y | \beta^c)$

Mean $pr($bankruptcy$) = 0.1067 
2000 repeats
**Figure 4**

Distribution of $p_{\text{mle}}$ (MLE probabilities)

![Histogram of $p_{\text{mle}}$](image)

Mean $pr(\text{bankruptcy}) = 0.0314$

500 repeats

**Figure 5**

Distribution of $p_{\text{exp}}$ (EXP probabilities)

![Histogram of $p_{\text{exp}}$](image)

Mean $pr(\text{bankruptcy}) = 0.1061$

500 repeats
Figure 6

Out-of-sample measure of $S(Y \mid \beta^{\text{mle}}) - S(Y \mid \beta^{\text{exp}})$

Mean $\text{pr(bankruptcy)} = 0.034$
500 repeats

Figure 7

Out-of-sample measure of $S(Y \mid \beta^{\text{mle}}) - S(Y \mid \beta^{\text{exp}})$

Mean $\text{pr(bankruptcy)} = 0.111$
500 repeats

Only 24% of bootstrap repeats generate higher EU (out-of-sample) under exponential fitting than under MLE