Rejoinder to “Comments on Schoenberg et al. (2003)” by Hamid Ghorbani

I thank the author and editors for the interest in and careful reading of our 2003 paper. Unfortunately, I cannot seem to track down the code we used to make the calculations in this paper, but it seems that Dr. Hamid correctly identified that we reported $\text{AIC}/2$ rather than $\text{AIC}$ for each of the models considered. Regarding the estimate of the parameter in the half-normal distribution, Dr. Hamid’s estimate for the MLE is correct and ours was incorrect, too low by exactly a factor of $\sqrt{2}$. I am sorry we made this mistake and thank Dr. Hamid for pointing this out. Fortunately, the other estimates appear to be all right.

Regarding the other two points made by Dr. Hamid, we computed the integrated squared difference between the empirical distribution and the estimated theoretical distribution and referred to it as a Cramér–von Mises statistic, which does not strike me as highly unusual though, as Dr. Hamid notes, may not be an appropriate name. Perhaps we should have referred to this statistic instead as the integrated squared difference? I found it hard to believe that the estimation of the parameters governing $F$ would make such a big impact on the null distribution of this integrated squared difference statistic, so I re-ran the analysis now, computing the integrated squared difference between the empirical and the fitted Pareto distribution functions with respect to $dx$. My R code is attached as part of the Supplemental Material. I get a value of 0.0257 for the wildfire data, which is close to the value of 0.0223 reported in our paper. There are some choices to be made as to how exactly to approximate this integral, and these probably explain this small discrepancy. I then simulated 548 draws from a Pareto distribution with $\alpha = 0.15625\, \text{m}^2$ and $\beta = 0.610$, re-estimated the parameter $\beta$ by maximum likelihood, recalculated the integrated squared difference statistic, and repeated 10,000 times. In 13 of the 10,000 simulations the statistic was less than the observed value of 0.0257, so the estimated $p$-value is 0.0013. Thus, the Pareto is rejected, as in our 2003 paper. In the 2003 paper, we got a $p$-value of 0.0043, but Dr. Hamid finds a $p$-value of 0.26857! This discrepancy still seems mysterious to me. Perhaps the asymptotic distribution used in the pCvM function used by Dr. Hamid is not adequate for small samples from a heavy-tailed distribution like the Pareto? I am not familiar with this pCvM function so I am not sure.

Regarding Dr. Hamid’s final point, I think in 2003, we did Monte Carlo simulations by simulating draws from the null distribution repeatedly, and for each simulation, we re-estimated the QQ plot for that simulation. We did not, for each simulation, re-estimate the parameters governing the distribution in question. Is this perhaps the cause of the confusion? Again, this may be the source of very slight discrepancies from Dr. Hamid’s results, but I think it is the standard procedure. Dr. Hamid’s observation that very similar confidence bounds are obtained using a theoretical formula is interesting and comforting, and he is surely right that use of such a formula would no doubt speed up computations considerably. Again, I thank Dr. Hamid for his interest and very careful reading, and apologize for errors in our 2003 paper, especially the incorrect estimate of the half-normal standard deviation parameter.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article. 

Frederic Paik Schoenberg
Department of Statistics, University of California at Los Angeles, Los Angeles, California
Correspondence
Frederic Paik Schoenberg, Department of Statistics, University of California at Los Angeles, 8142 Math-Science Building, Los Angeles, CA 90095-1554. Email: frederic@stat.ucla.edu