

Online Supplement 5: On Deriving an Ideal Cutoff Based on Simulation Results

Many of the techniques presented in the article require a cutoff against which an estimate is compared to or a value to which a correlation is set in a comparison model. The literature recommends various cutoffs between .85 and 1, which are mostly based on convention. In a rare example where a recommendation is derived based on a simulation, Voorhees et al. (2016) considered the false positive and false negative rates of the techniques used in their study and concluded that the cutoff of .85 had the best balance having large power and still an acceptable Type I error rate. Following a similar approach to Voorhees et al, we calculated a weighted average of Type I and Type II error rates.

We applied a similar logic on our data and compared the Type I (false positive) and Type II (false negative) rates for the data shown in Table 10 in the article. In this case, the false negative rates are much more important than false positive rates because a false negative implies that a discriminant validity problem went undetected. Table 1 shows the weighted averages for three different cutoffs for the two recommended techniques using four different weights. Based on the table, we would conclude that .95 is more accurate than .85 or .9, and that this conclusion is robust to different weights.

Table 1 Summary of the techniques and alternative cutoffs based on Table 10 in the article

	Cutoff	Average false positive rates	Average false negative rates	Weighted average by weight of false negatives over false positives			
				1	3	5	10
χ^2 (cut)	.85	.508	.002	.255	.129	.087	.048
	.9	.289	.009	.149	.079	.056	.035
	.95	.286	.003	.145	.074	.050	.029
CI_{CFA} (cut)	.85	.450	.002	.226	.114	.077	.043
	.9	.237	.014	.126	.070	.051	.034
	.95	.248	.002	.125	.064	.043	.025

However, this recommendation is highly sensitive to the population correlations and cutoffs used in the simulation design and thus not very useful. More specifically, we ran a set of simulations to study the robustness of the choice of cutoff on the correlations used in the simulation study. This analysis, essentially showed that if a simulation study would consider all possible correlations between 0 and 1 and not just a few specific values (e.g. .8, .85, .9, .95), an ideal cutoff would be arbitrarily close to 1 with any reasonable set of weights.

In the current approach for choosing a cutoff based on simulation results, the ideal way to do the simulations would be to consider every possible population correlation and every possible cutoff and then choose the cutoff that produces the smallest weighted average error rate. For simplicity, let's consider just non-negative population correlation $P \in [0,1]$ and positive cutoffs $C \in]0,1[$. Let's also consider just confidence intervals because it simplifies the problem. Let U be the upper limit of a confidence interval calculated from a sample. We can then define false positive (problem detected erroneously) and false negative rates (a problem not detected) as follows:

False positive rate: $\text{Mean}(U \geq C)$ when $C < P$

False negative rate: $\text{Mean}(U < C)$ when $C \geq P$

We will now use a simulation study to analyze the ideal cutoff using a weighted average of these two error rates across the different cutoffs. For simplicity and computational efficiency, we assume that the focal correlation can be calculated directly without the use of a factor analysis and use 250 as the sample size. We will simulate a million samples each from a different population where the population correlation varies uniformly between $[0,1]$ and use these samples to calculate the upper limit of a confidence interval. The distribution of U as a function of P is visualized below in Figure 1.

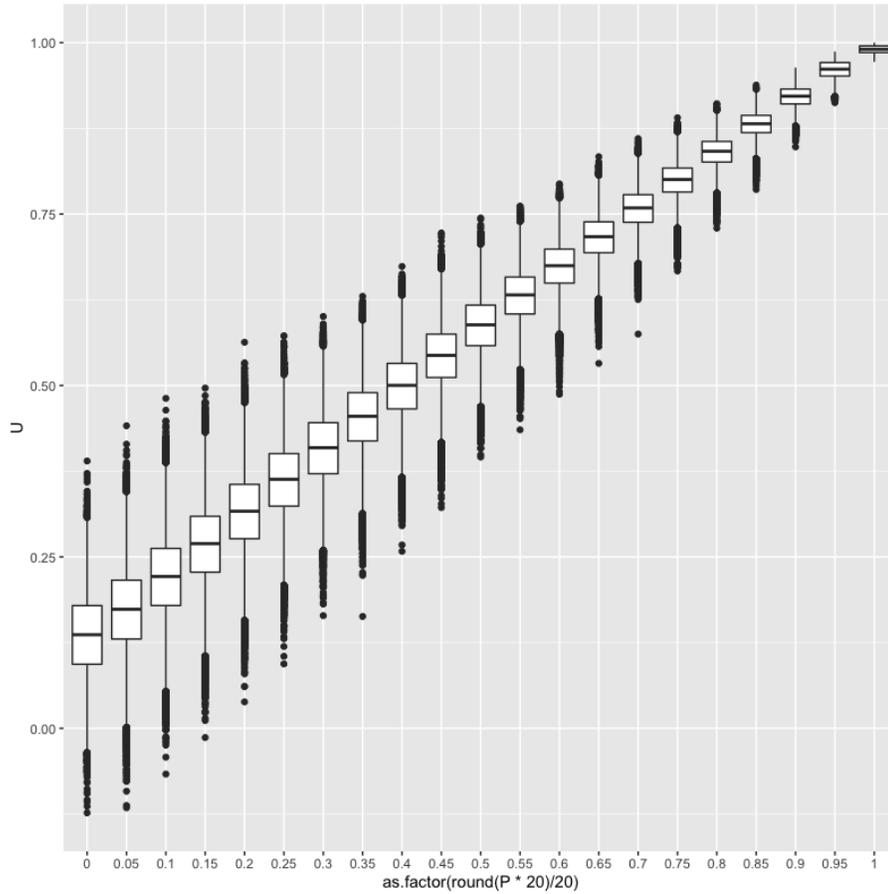


Figure 1 Distribution of U as a function of P

As expected, the variance of U decreases when P increases and U converges to P. We will now calculate the false positive and false negative rates for 999 different values of C from 0.001 to 0.999. The false positive rates are shown below in Figure 2.

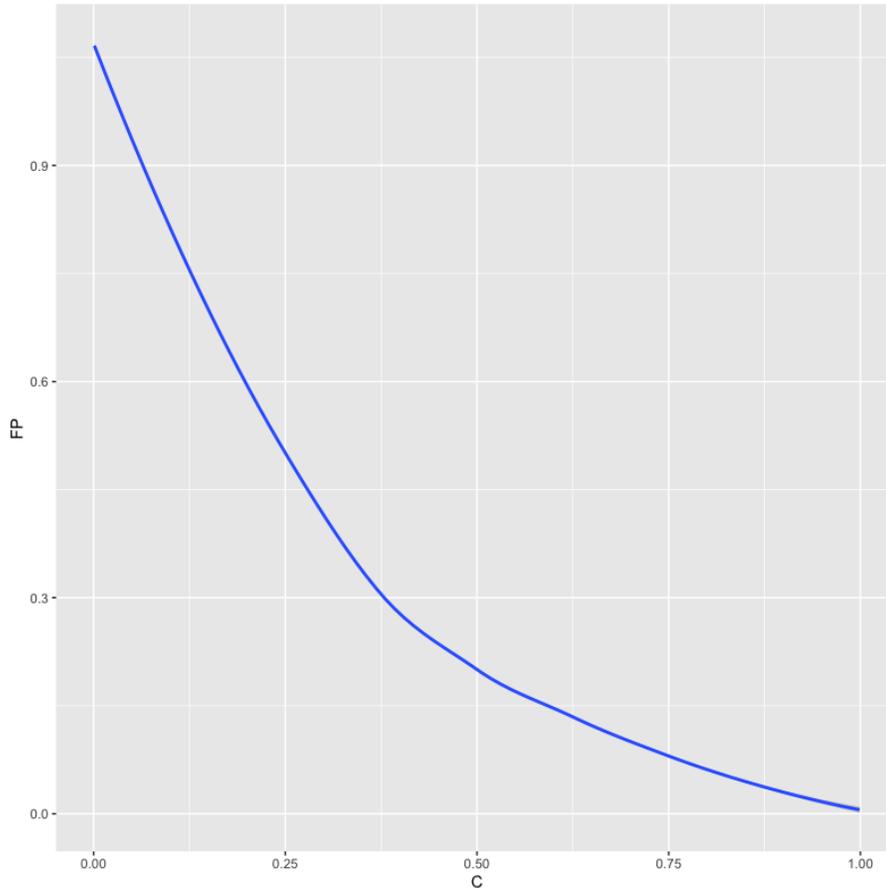


Figure 2 False positive rate as a function of C

This finding makes a lot of sense. If our C is arbitrarily close to 0, the population value needs to be arbitrarily close to zero as well for $P < C$ to hold. In this case a proper CI will contain the population value 95% of the time thus producing a 95% false positive rate. If our C is arbitrarily close to 1, then FP rate will converge to zero because the CIs get narrower and narrower as the estimate approaches 1 and will thus not contain values that are arbitrarily close to 1. If we were concerned just on the false positive rate, an ideal cutoff would thus be as close to 1 as possible.

Figure 3 below shows the false negative rates calculated similarly.

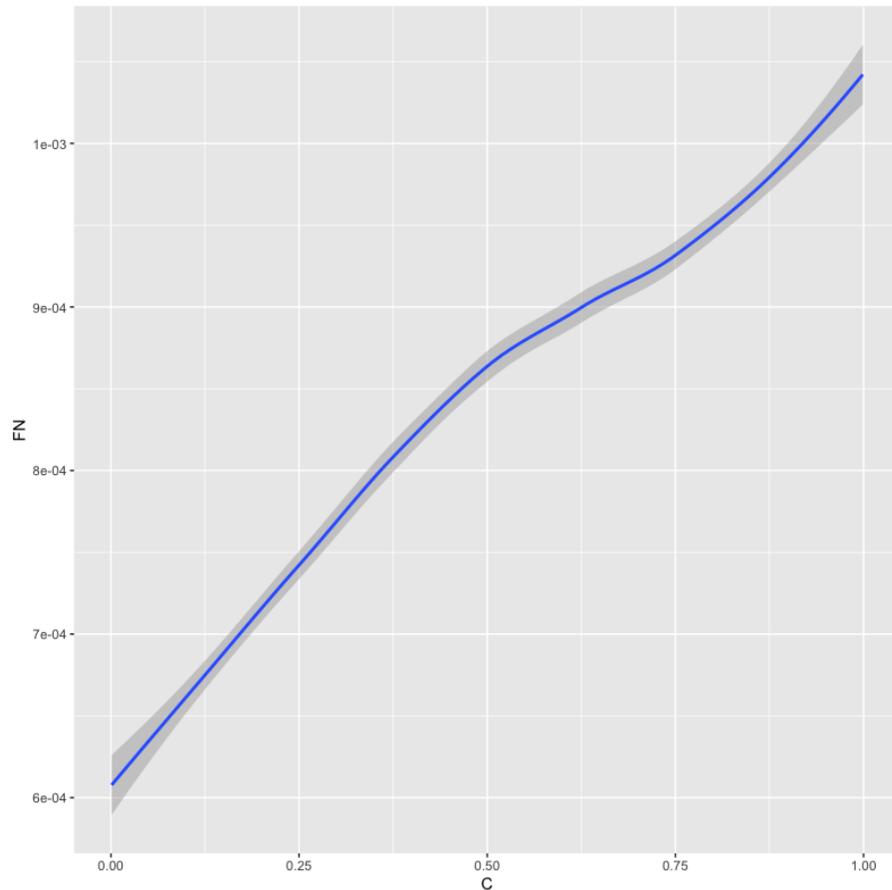


Figure 3 False negative rates as a function of C

Here we can minimize the error rate by setting the cutoff close to zero. If we just compare the two rates, we can see that the variance of false positive rates is much greater than the variance of false negative rates. Thus, a weighted average of the two is mostly determined by false positives unless the weight for false negatives is very large.

Figure 4 shows the weighted averages for weights $W=10$, $W=100$, and $W=1000$. The last two panels focus on $C \in [0.5, 1]$ because this range is more informative than the full range. With weights 10 and 100, the best cutoff is to choose a value as close to 1 as possible. If we set the weight to 1000, the optimal cutoff is somewhere around .85 to .90. However, using a weight of 1000 is probably not a reasonable choice. With even larger weights, the ideal cutoff converges toward zero.

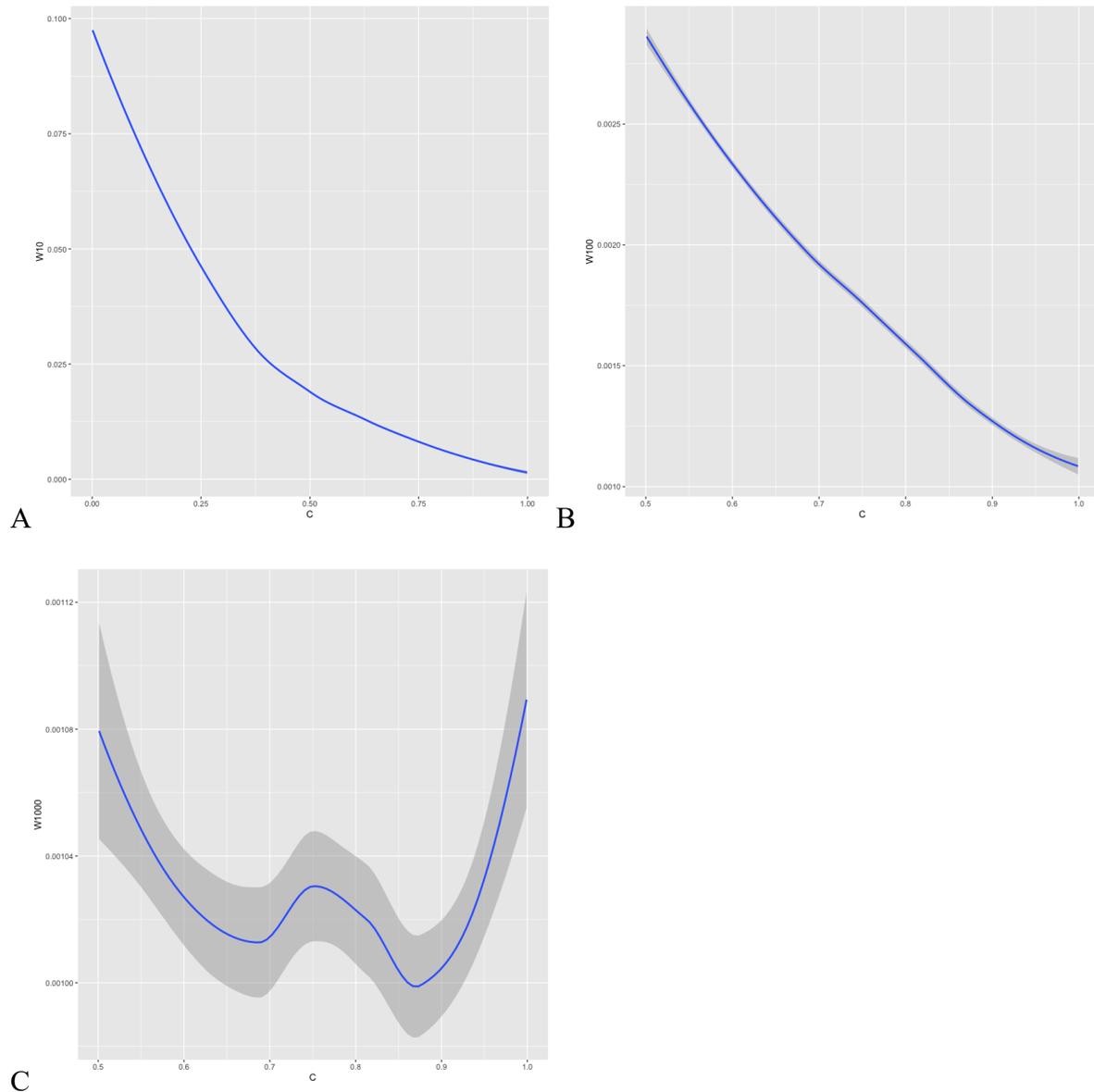


Figure 4 The weighted average of false positive and false negative rates as a function of weights given for false negative results (A: $W=10$, B: $W=100$, C: $W=1000$)

References

- Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2016). Discriminant validity testing in marketing: An analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science*, 44(1), 119–134. <https://doi.org/10.1007/s11747-015-0455-4>

R Code for the Figures

```
library(MASS)
library(parallel)
library(tidyverse)

options(mc.cores = detectCores())

data <- do.call(rbind,mclapply(seq(0,1,length.out = 1000000),
  function(P){
    d <- mvrnorm(250,c(0,0), matrix(
      c(1,P,
        P,1),2,2))
    r <- cor.test(d[,1],d[,2])
    c(P, r$estimate, r$conf.int)
  }))

data <- tibble(P = data[,1],
  r = data[,2],
  L = data[,3],
  U = data[,4])

ggplot(data = data) +
  geom_boxplot(mapping = aes(x = as.factor(round(P*20)/20), y = U))

# Error rates data

errorRatesData <- do.call(rbind,
  lapply(1:999/1000,
    function(C){
      d <- data %>%
        mutate(H0 = C<P,
          positive = U>=C,
          negative = 1-positive) %>%
        group_by(H0) %>%
        summarize(FP = mean(positive),
          FN = mean(negative))
      ret <- c(C=C,FP = d$FP[1],d$FN[2])
    }))

errorRatesData <- as.data.frame(errorRatesData)
colnames(errorRatesData) <- c("C", "FP", "FN")

errorRatesData <- errorRatesData %>%
  mutate(W10 = (FP + 10*FN)/11,
    W100 = (FP + 100*FN)/101,
    W1000 = (FP + 1000*FN)/1001,)
```

```
ggplot(data = errorRatesData) +  
  geom_smooth(mapping = aes(x = C, y = FN))
```

```
ggplot(data = errorRatesData) +  
  geom_smooth(mapping = aes(x = C, y = FP))
```

```
ggplot(data = errorRatesData) +  
  geom_smooth(mapping = aes(x = C, y = W10))
```

```
ggplot(data = filter(errorRatesData, C > .5)) +  
  geom_smooth(mapping = aes(x = C, y = W100))
```

```
ggplot(data = filter(errorRatesData, C > .5)) +  
  geom_smooth(mapping = aes(x = C, y = W1000))
```