Statistical Extreme Value Analysis Concerning Risk of Wingtip to Wingtip or Fixed Object Collision for Taxiing Large Aircraft

Prepared under a

Cooperative Research and Development Agreement Between
The Federal Aviation Administration (FAA) and The Boeing Company

Statistical Extreme Value Analysis Concerning Risk of Wingtip to Wingtip or Fixed Object Collision for Taxiing Large Aircraft*

Summary

This report describes the analysis concerning the risk of collision between two large aircraft taxiing on parallel taxiways and the risk of collision between a large taxiing aircraft and a fixed object, such as a building.

The data used for this study came from two sources, namely

1. New York's John F. Kennedy International Airport (JFK), where 747 taxiway deviation data were collected from 6/24/1999 to 2/17/2000 at two laser locations. The two lasers monitored simultaneously two parallel 75 ft straight taxiway segments with shoulder called ALPHA and BRAVO, respectively.

2. Anchorage International Airport (ANC), where 747 taxiway deviation data were collected from 9/24/2000 to 9/27/2001 at two laser locations for each of two (not parallel) 75 ft straight taxiway segments with shoulder called KILO and ROMEO, respectively.

Although these data sets were screened to capture mostly 747 deviations there is the possibility that other, similarly large aircraft, e.g., L-1011, A-330, A-340, 777, MD-11 and DC-10, were included. This possibility is stronger for the JFK data than for the ANC data. For the latter we expect most deviations to be from 747 aircraft.

These deviation data ($n_{JFK} = 2,518$ and $n_{ANC} = 9,796$ with a combined total of $n = 12,314$ deviations) were previously analyzed with regard to extreme deviations of individual aircraft from the taxiway centerline and the results are documented in reports [8] and [9]. These reports addressed the risk of an aircraft deviating at a fixed location along the taxiway beyond a certain threshold distance from the taxiway centerline.

*Data gathered and provided by the FAA. Analysis and report prepared by Fritz Scholz, The Boeing Company, June 13, 2005.

The deviation data were taken over a good part of a year at JFK and over a full year at ANC. There was no apparent seasonal or time of day effect. Taxiway centerline lights and higher vigilance under adverse conditions may have compensated for any impact from such factors.

The character of the deviation distribution at JFK and ANC is very similar, the only difference being that the JFK deviations are spread out more than the ANC deviations by a factor of 1.097. We speculate that the 10% wider spread in the JFK deviations results from the fact that the centerline lights are offset from the taxiway centerline by 12 inches at ANC and by 22 inches at JFK. This difference in offset of the centerline lights could cause wider meandering swings at JFK than at ANC because pilots tend to avoid the bumping of the nose gear wheels on the slightly protruding lights.

Taxiway Centerline Separation from Object for Wingtip to Object Clearance

Although it would have been easily possible to characterize the relation between object separation from taxiway centerline (T), wingspan (W), aircraft deviation (d) from taxiway centerline, and collision risk (p), it was not done in reports [8] and [9]. Instead it is done here in parallel with the corresponding risk of two aircraft colliding with each other at their respective wingtips while taxiing on parallel taxiways.

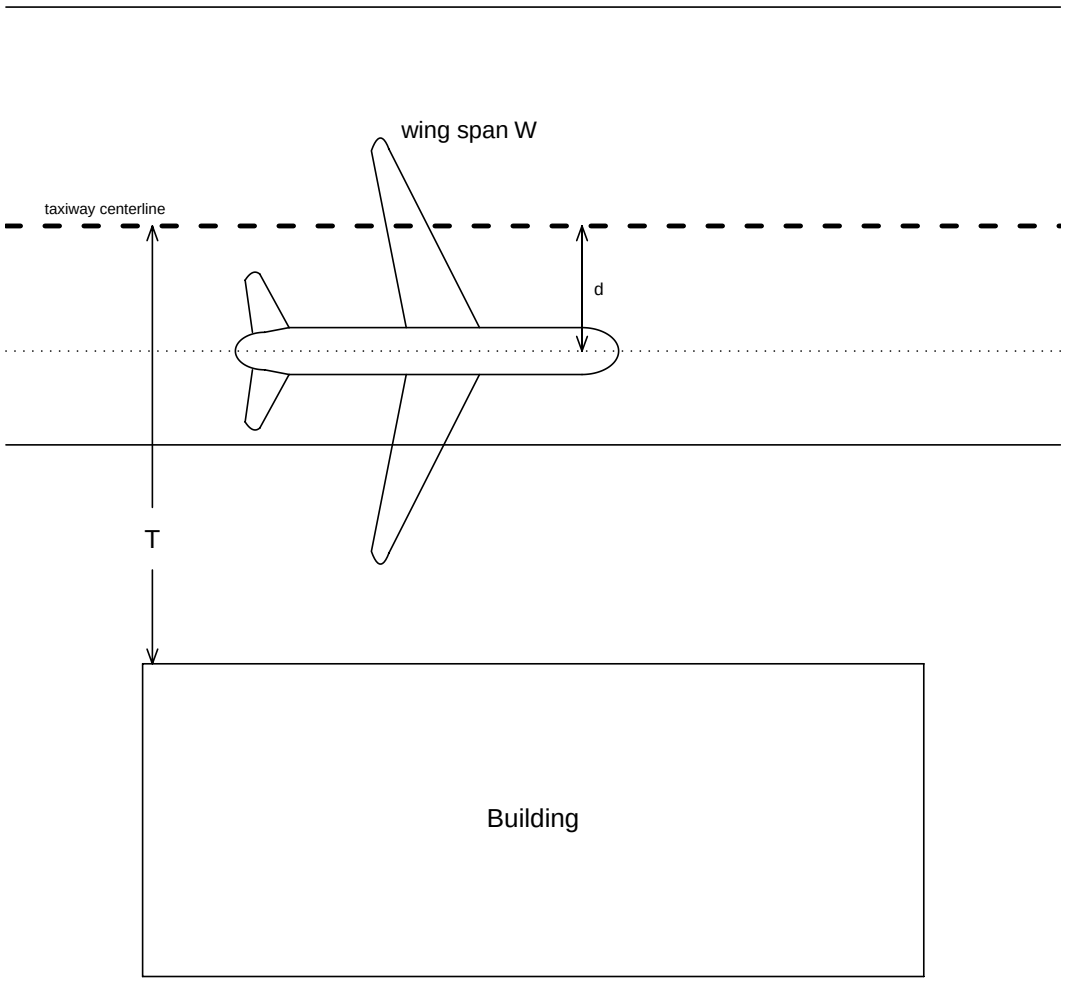
Since the aircraft centerline deviations were measured at a particular point (laser location) along the taxiway they do not reflect the maximum deviation over a straight taxiway segment. These maximum deviations tend to be more spread out than the measured point deviations at one specific location. Using such point deviation data for risk calculations would allow us to calculate the risk of deviating too far from taxiway centerline at some specific point but it would lead to some underestimation of the risk of deviating too far from taxiway centerline somewhere along the taxiway.

The potential collision situation between an aircraft wingtip and a fixed object is shown in Figure 1. Here T indicates the distance of the fixed object from the taxiway centerline, W is the wingspan of the taxiing aircraft and d the deviation of the aircraft from the taxiway centerline when first encountering the fixed object. As illustrated in Figure 1 a collision is avoided when

$$T > \frac{W}{2} + d \, .$$

In Figure 1 the fixed object is shown as the rectangular box of a building. Technically this risk assessment is affected by the same concerns that were raised previously in the context of running off the taxiway, since we deal with the lengthwise exposed section of the building. However, here one may make the case that the first exposed corner represents the point location of interest and one would think that a pilot who comes too close to that point (without colliding) will not proceed in that direction towards the building. In this we assume that the pilot is aware of the building and that it is not hidden in fog. Also, the side of the building is typically much smaller than the length of a straight taxiway segment.

Figure 1: Separation Clearance Between Aircraft and Fixed Object (Building)



Thus the distribution of maximum deviation over such a shorter length is not too much different from the deviation distribution at a point. Concerning the fog issue one can hope for greater vigilance on the part of the pilot in trying to more closely follow the taxiway centerlights.

From report [9] we extract the values in Table A which gives the estimates $\hat{\eta}_{1-p}$ and 95% upper confidence bounds $\hat{\eta}_{U,1-p}$ for the $(1-p)$-quantiles η_{1-p} of the d distribution. The $(1-p)$-quantile η_{1-p} of the d distribution is that number which is exceeded by the proportion p of all such d values. Specific values p of interest are $p = 10^{-3}, 10^{-4}, \ldots, 10^{-9}$. The separation T would then have to exceed $\eta_{1-p} + W/2$ in order to limit the collision risk to p.

Table A: $(1-p)$-Quantile Estimates and 95% Upper Bounds for the d Distribution

	one-sided exceedance risk p						
	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
estimate	6.04 ft	8.19 ft	10.49 ft	12.97 ft	15.62 ft	18.47 ft	21.53 ft
95% upper bounds	6.26 ft	8.52 ft	10.95 ft	13.56 ft	16.36 ft	19.36 ft	22.58 ft

Instead of the unknown η_{1-p} we use the estimates and 95% upper confidence bounds for η_{1-p} given in Table A. These can then be combined with various half wingspans $W/2$ to arrive at threshold estimates and upper bounds $\hat{\eta}_{1-p} + W/2$ and $\hat{\eta}_{U,1-p} + W/2$ that should be exceeded by the distance T from the taxiway centerline to the fixed object for the respectively targeted exceedance risk p. The results are summarized in Tables B and C for ANC and JFK, respectively. Note that consecutive rows for estimates and upper bounds differ by half the wingspan increment (5 ft). The rows, although deriving from each other in this way, are shown mainly for convenience.

The reason for separate tables for ANC and JFK results from the fact that the bias adjusted deviations from JFK were divided by 1.097 in order to align their scale or spread to that of the bias adjusted deviations from ANC before combining both data sets. Thus we have to multiply threshold values from Table A by 1.097 to make them apply back to the JFK situation.

Another back-adjustment that needs to be made is the adding in of biases that were removed in order to render the deviations symmetric around zero. These biases, due to parallax issues and the avoidance of taxiway centerlights, combined maximally to roughly .75 ft.

Thus, to arrive at the value 117.4 ft to the right of the 220 ft wingspan in Table C for JFK one calculates (taking 6.04 from Table A) $6.04 \times 1.097 + .75 + 220/2 = 117.4$ ft. In the calculation for Table B one omits the factor 1.097 in the calculation.

Table B: Estimates and 95% Upper Confidence Bounds to be Exceeded by the Required Separation T Between Taxiway Centerline and Object for ANC

	Estimate $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	96.8	98.9	101.2	103.7	106.4	109.2	112.3
190	101.8	103.9	106.2	108.7	111.4	114.2	117.3
200	106.8	108.9	111.2	113.7	116.4	119.2	122.3
210	111.8	113.9	116.2	118.7	121.4	124.2	127.3
220	116.8	118.9	121.2	123.7	126.4	129.2	132.3
230	121.8	123.9	126.2	128.7	131.4	134.2	137.3
240	126.8	128.9	131.2	133.7	136.4	139.2	142.3
250	131.8	133.9	136.2	138.7	141.4	144.2	147.3
260	136.8	138.9	141.2	143.7	146.4	149.2	152.3
270	141.8	143.9	146.2	148.7	151.4	154.2	157.3
280	146.8	148.9	151.2	153.7	156.4	159.2	162.3
	95% Upper Bound $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	97.0	99.3	101.7	104.3	107.1	110.1	113.3
190	102.0	104.3	106.7	109.3	112.1	115.1	118.3
200	107.0	109.3	111.7	114.3	117.1	120.1	123.3
210	112.0	114.3	116.7	119.3	122.1	125.1	128.3
220	117.0	119.3	121.7	124.3	127.1	130.1	133.3
230	122.0	124.3	126.7	129.3	132.1	135.1	138.3
240	127.0	129.3	131.7	134.3	137.1	140.1	143.3
250	132.0	134.3	136.7	139.3	142.1	145.1	148.3
260	137.0	139.3	141.7	144.3	147.1	150.1	153.3
270	142.0	144.3	146.7	149.3	152.1	155.1	158.3
280	147.0	149.3	151.7	154.3	157.1	160.1	163.3

Table C: Estimates and 95% Upper Confidence Bounds to be Exceeded by the Required Separation T Between Taxiway Centerline and Object for JFK

	Estimate $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	97.4	99.7	102.3	105.0	107.9	111.0	114.4
190	102.4	104.7	107.3	110.0	112.9	116.0	119.4
200	107.4	109.7	112.3	115.0	117.9	121.0	124.4
210	112.4	114.7	117.3	120.0	122.9	126.0	129.4
220	117.4	119.7	122.3	125.0	127.9	131.0	134.4
230	122.4	124.7	127.3	130.0	132.9	136.0	139.4
240	127.4	129.7	132.3	135.0	137.9	141.0	144.4
250	132.4	134.7	137.3	140.0	142.9	146.0	149.4
260	137.4	139.7	142.3	145.0	147.9	151.0	154.4
270	142.4	144.7	147.3	150.0	152.9	156.0	159.4
280	147.4	149.7	152.3	155.0	157.9	161.0	164.4
	95% Upper Bound $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	97.6	100.1	102.8	105.6	108.7	112.0	115.5
190	102.6	105.1	107.8	110.6	113.7	117.0	120.5
200	107.6	110.1	112.8	115.6	118.7	122.0	125.5
210	112.6	115.1	117.8	120.6	123.7	127.0	130.5
220	117.6	120.1	122.8	125.6	128.7	132.0	135.5
230	122.6	125.1	127.8	130.6	133.7	137.0	140.5
240	127.6	130.1	132.8	135.6	138.7	142.0	145.5
250	132.6	135.1	137.8	140.6	143.7	147.0	150.5
260	137.6	140.1	142.8	145.6	148.7	152.0	155.5
270	142.6	145.1	147.8	150.6	153.7	157.0	160.5
280	147.6	150.1	152.8	155.6	158.7	162.0	165.5

Parallel Taxiway Centerline Separation for Wingtip to Wingtip Clearance

The corresponding wingtip to wingtip collision situation between two aircraft is illustrated in Figure 2 which makes clear how the separation T between taxiway centerlines and the average wingspan $(W_1 + W_2)/2$ interact with the respective deviations d_1 and d_2 of the two aircraft from their respective taxiway centerlines. To avoid collision the combined deviations need to satisfy the following criterion:

$$T > \frac{W_1 + W_2}{2} + d_1 + d_2 \ .$$

From this criterion it is clear that high $(1-p)$-quantiles of the compounded deviations d_1+d_2 from both aircraft are of primary interest. The $(1-p)$-quantile of $d_1 + d_2$ is that number δ_{1-p} which is exceeded by only a proportion p of all such compounded deviations $d_1 + d_2$. Specific values p of interest are $p = 10^{-3}, 10^{-4}, \ldots, 10^{-9}$. The separation T would then have to exceed $\delta_{1-p} + (W_1 + W_2)/2$ in order to limit the collision risk to p.

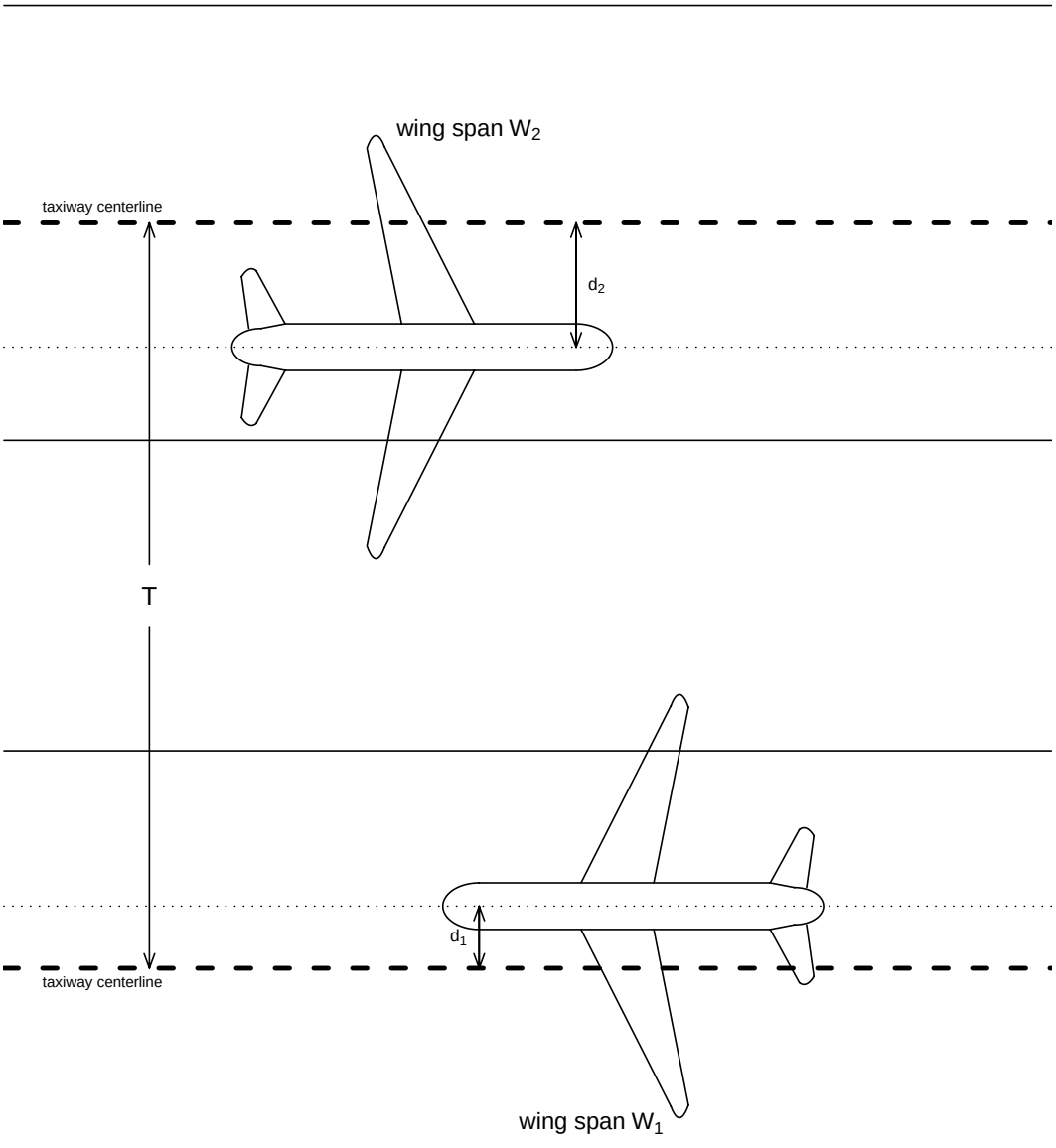
This study did not use the few observed instances at JFK where aircraft passed each other at the same time as they passed the laser. This was done to avoid registration difficulties. Although these cases might have given some evidence of possible avoidance bias, there were too few of them to get a reasonable assessment of such an effect. Leaving them out from the collision risk analysis we act conservatively, i.e., the possible bias would render the risk smaller than indicated by the analysis of deviations for which such a bias would not have been active. Two approaches were taken to deal with this sparsity or lack of direct data on $d_1 + d_2$ for passing aircraft.

First Approach

The first approach followed the previous Schiphol taxiway deviation study [1] which used the individual aircraft deviations by randomly generating combined deviation measurements $d_i + d_j$. This was done through random splitting of the sample of individual deviations into two halves and random pairing of each deviation from the first half with one from the second half, to obtain $6,157$ such pairs $d_i + d_j$. For each set of such pairs the extreme value extrapolation method from [7] was applied to the absolute values $|d_i + d_j|$ to obtain estimates and 95% confidence bounds for the $(1-p)$-quantiles of the $d_1 + d_2$ distribution for $p = 10^{-3}, 10^{-4}, \ldots, 10^{-9}$. Using the absolute values $|d_i + d_j|$ was justified by the general symmetry of the deviation histogram around zero, which entails similar symmetry for the $d_i + d_j$ distribution. Thus a $(1-2p)$-quantile of the $|d_i + d_j|$ distribution is the same as a $(1-p)$-quantile of the $d_i + d_j$ distribution.

To understand and correct for the random splitting and pairing effect this process was repeated 500 times and a combined estimate and confidence bound was then settled on in the tabulation of the results. We also investigated the effect of the variability arising from the initial random sample of $12,314$ deviations by preceding each one of the previous random splits and pairings by a bootstrap random sample of $12,314$ deviations taken with replacement from the original deviation sample. However, it was found that the variation

Figure 2: Separation Clearance Between Aircraft on Parallel Taxiways



due to random splits and pairings was the main variability factor and that the answers arising with the bootstrap sampling scheme were less conservative than just working with the original sample. Presumably this is due to the dominant effect of the random splits and pairings and also to a known deficiency of the bootstrap method when dealing with data extremes, namely that a fair proportion of the bootstrap samples don't contain the most extreme deviations from the original deviation sample.

Second Approach

The second approach looked at all pairwise sums $d_i + d_j$ with $1 \leq i < j \leq 12,314$. There are $N = 75,811,141$ such paired sums. The absolute values $|d_i + d_j|$ were then ordered from smallest to largest and the top $1,000$ and $10,000$ were plotted against their respective tail fraction rank order $(1 - .5)/N, (2 - .5)/N, \ldots (10,000 - .5)/N$ on a log scale. The resulting plot showed very regular behavior (see Figure 17 on page 26) and a quadratic was fitted by the method of least squares. This quadratic can then be used to estimate any of the $(1-p)$-quantiles for $p = 10^{-3}, 10^{-4}, \ldots, 10^{-9}$. The advantage of this second approach is that for $p = 10^{-6}$ one is within the data range of the calculated $|d_i + d_j|$ values because of the size of $N = 75,811,141$. Even for $p = 10^{-7}$ we are still within the range but here we rely on the smooth behavior of the fitted quadratic and ignore the fluttering behavior of the very extreme $|d_i + d_j|$ values.

There was not much difference in the fitted quadratics between using the $1,000$ or $10,000$ top $|d_i + d_j|$ values. For the subsequent bootstrap process we used the top 1000 values of $|d_i + d_j|$ in the fitting. In this bootstrap approach we recalculated the above quantile estimates, as obtained from the fitted quadratic, for samples of size $12,314$ drawn with replacement from the original deviation sample. This provides some sense of the variability/uncertainty of the original quantile estimates and allows us to calculate confidence bounds using bootstrap methodology.

Synthesis

It turned out that the quantile estimates/confidence bounds obtained from the first approach were somewhat higher than those obtained under the second approach. When comparing confidence bounds for the two approaches the worst case difference was .55 ft. In a sense the two approaches confirm each other reasonably well. One cannot expect same results from the same data when using different estimation techniques. Note that the second approach did not involve the use of the extreme value extrapolation method from [7]. We opted to stay conservatively with the results from the first approach. These quantile estimates and upper bounds are shown in Table D.

These quantile estimates and upper bounds, denoted by $\hat{\delta}_{1-p}$ and $\hat{\delta}_{U,1-p}$ respectively, can then be combined with various average wingspans $(W_1 + W_2)/2$ to arrive at threshold estimates and upper bounds $\hat{\delta}_{1-p} + (W_1 + W_2)/2$ and $\hat{\delta}_{U,1-p} + (W_1 + W_2)/2$ that should be exceeded by the taxiway separation distance T for the respectively targeted exceedance risk p. This was tailored to both the ANC and the JFK situation. The reason for this distinct treatment is the same as was explained previously.

Table D: $(1-p)$-Quantile Estimates and 95% Upper Bounds for the d_1+d_2 Distribution

	one-sided exceedance risk p						
	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
estimate	7.90 ft	10.09 ft	12.12 ft	14.35 ft	16.29 ft	18.13 ft	19.84 ft
95% upper bounds	7.91 ft	10.14 ft	12.22 ft	14.54 ft	16.58 ft	18.55 ft	20.51 ft

As before another back-adjustment needs to be made for the biases that were removed in order to render the deviations symmetric around zero. These biases, due to parallax issues and the avoidance of taxiway centerlights, combined maximally to roughly .75 ft. Even though in certain situations such biases for two passing aircraft could cancel out to some extent we decided again to conservatively add in this bias for both deviations in d_1+d_2, i.e., we added 1.5 ft to the (rescaled by 1.097) estimates/bounds in Table D before adding the resulting back-adjusted estimates/bounds to the average wingspan. These results are shown in Tables E and F for ANC and JFK, respectively.

For comparison sake we also reproduce the relevant table from [1] as Table G. These results are given only for the available risks of $p = 10^{-6}, \ldots, 10^{-9}$. The separation thresholds T are remarkably close in Tables E-G even though there are several reasons that may explain the existing differences. These are detailed in the report. We prefer to emphasize here the closeness in relative terms.

In effect, the Tables E-G provide some insight as to how far these results are generalizable. Clearly the results from the three airports are different, about $3-4$ ft of each other, but in terms of the scale of the numbers given in Tables E-G this discrepancy seems small. This closeness of results from the three studies is illustrated in Figure 3 which shows the Code E Separation Standard with respect to the required separation T for a $p = 10^{-7}$ collision risk. The actual shown deviations of both aircraft (7 ft and 6 ft) are arbitrary and one should keep in mind that a 13 ft combined deviation d_1+d_2 can come about in many ways, e.g., 2 ft plus 11 ft, 4 ft plus 9 ft, -4 ft plus 17 ft, etc. Recall that a combined $p = 10^{-7}$ deviation would be 16.58 ft with 95% confidence (according to Table D). The Code E Separation Standard is much wider than that required by any of the three study results. In fact, the illustrated required separation of 232.7 ft for a collision risk $p \leq 10^{-7}$ (with 95% confidence) is still almost 30 ft less than then Code E Separation Standard of 262.5 ft.

With measurements from other airports we may see further differences, but not very much more than observed here. We do not understand all the reason behind such differences. The difference in deviation scale at ANC and JFK, speculated to be due to differing offsets of centerline lights from the taxiway centerline, invites further corroberation. At Schiphol the asymmetry of the deviations would ask for some clarification.

Figure 3: Separation Clearance Between Aircraft on Parallel Taxiways

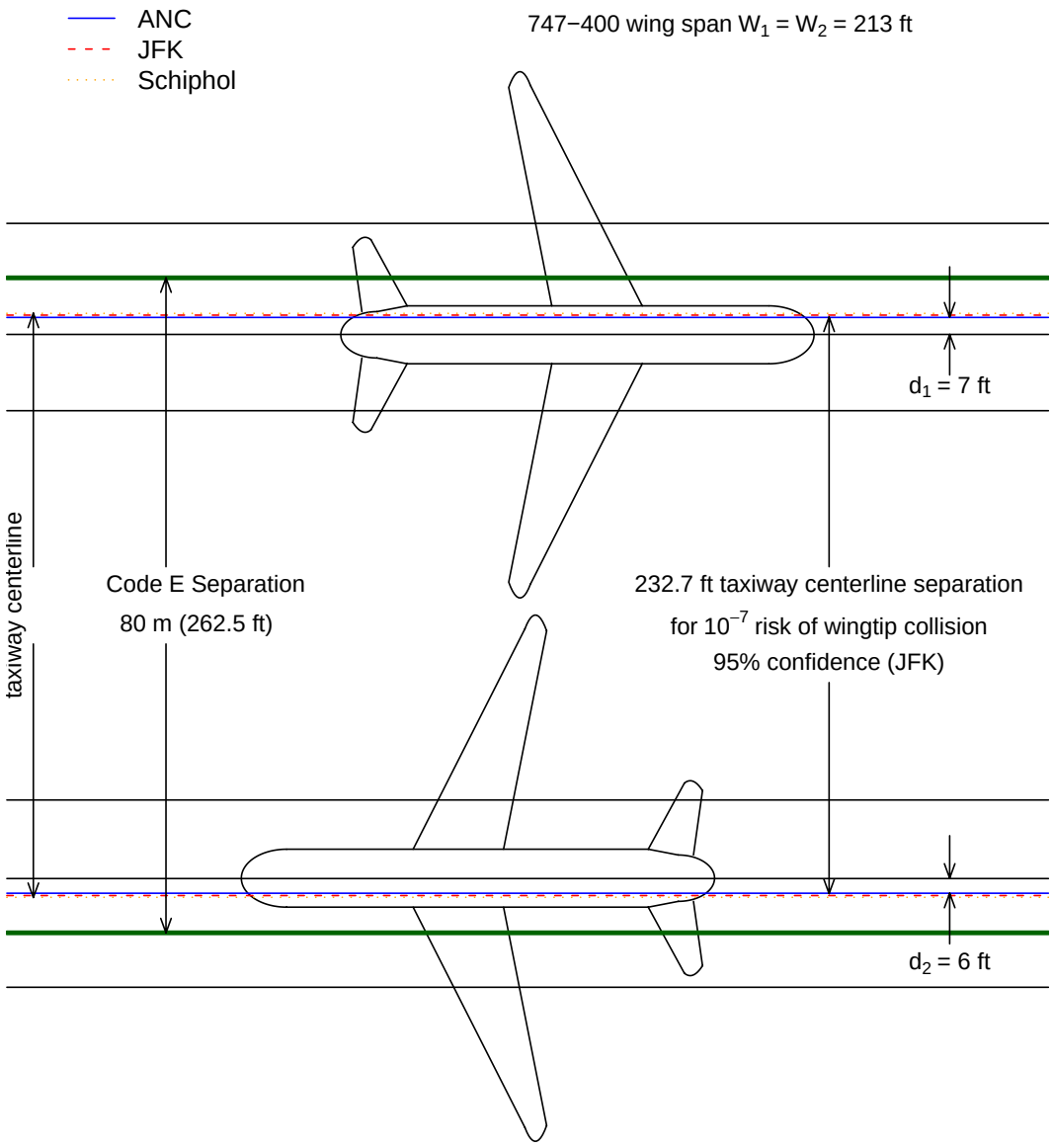


Table E: Required Separation T Between Taxiway Centerlines for ANC

wingspan (ft)	Estimate $\leq T$ (ft) collision risk p						
	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	189.4	191.6	193.8	195.9	197.8	199.7	201.4
190	199.4	201.6	203.8	205.9	207.8	209.7	211.4
200	209.4	211.6	213.8	215.9	217.8	219.7	221.4
210	219.4	221.6	223.8	225.9	227.8	229.7	231.4
220	229.4	231.6	233.8	235.9	237.8	239.7	241.4
230	239.4	241.6	243.8	245.9	247.8	249.7	251.4
240	249.4	251.6	253.8	255.9	257.8	259.7	261.4
250	259.4	261.6	263.8	265.9	267.8	269.7	271.4
260	269.4	271.6	273.8	275.9	277.8	279.7	281.4
270	279.4	281.6	283.8	285.9	287.8	289.7	291.4
280	289.4	291.8	293.8	295.9	297.8	299.7	301.4
wingspan (ft)	95% Upper Bound $\leq T$ (ft) collision risk p						
	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	189.4	191.7	193.9	196.1	198.1	200.1	202.1
190	199.4	201.7	203.9	206.1	208.1	210.1	212.1
200	209.4	211.7	213.9	216.1	218.1	220.1	222.1
210	219.4	221.7	223.9	226.1	228.1	230.1	232.1
220	229.4	231.7	233.9	236.1	238.1	240.1	242.1
230	239.4	241.7	243.9	246.1	248.1	250.1	252.1
240	249.4	251.7	253.9	256.1	258.1	260.1	262.1
250	259.4	261.7	263.9	266.1	268.1	270.1	272.1
260	269.4	271.7	273.9	276.1	278.1	280.1	282.1
270	279.4	281.7	283.9	286.1	288.1	290.1	292.1
280	289.4	291.7	293.9	296.1	298.1	300.1	302.1

Table F: Required Separation T Between Taxiway Centerlines for JFK

	Estimate $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	190.2	192.6	195.0	197.2	199.4	201.4	203.3
190	200.2	202.6	205.0	207.2	209.4	211.4	213.3
200	210.2	212.6	215.0	217.2	219.4	221.4	223.3
210	220.2	222.6	225.0	227.2	229.4	231.4	233.3
220	230.2	232.6	235.0	237.2	239.4	241.4	243.3
230	240.2	242.6	245.0	247.2	249.4	251.4	253.3
240	250.2	252.6	255.0	257.2	259.4	261.4	263.3
250	260.2	262.6	265.0	267.2	269.4	271.4	273.3
260	270.2	272.6	275.0	277.2	279.4	281.4	283.3
270	280.2	282.6	285.0	287.2	289.4	291.4	293.3
280	290.2	292.6	295.0	297.2	299.4	301.4	303.3
	95% Upper Bound $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	190.2	192.7	195.1	197.5	199.7	201.8	204.0
190	200.2	202.7	205.1	207.5	209.7	211.8	214.0
200	210.2	212.7	215.1	217.5	219.7	221.8	224.0
210	220.2	222.7	225.1	227.5	229.7	231.8	234.0
220	230.2	232.7	235.1	237.5	239.7	241.8	244.0
230	240.2	242.7	245.1	247.5	249.7	251.8	254.0
240	250.2	252.7	255.1	257.5	259.7	261.8	264.0
250	260.2	262.7	265.1	267.5	269.7	271.8	274.0
260	270.2	272.7	275.1	277.5	279.7	281.8	284.0
270	280.2	282.7	285.1	287.5	289.7	291.8	294.0
280	290.2	292.7	295.1	297.5	299.7	301.8	304.0

Table G: Required Separation T Between Taxiway Centerlines for Schiphol

	Estimate $\leq T$ (ft)				90% Upper Bound $\leq T$ (ft)			
	collision risk p				collision risk p			
wingspan (ft)	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	198.8	200.7	202.4	204.1	200.5	202.6	204.6	206.6
190	208.8	210.7	212.4	214.1	210.5	212.6	214.6	216.6
200	218.8	220.7	222.4	224.1	220.5	222.6	224.6	226.6
210	228.8	230.7	232.4	234.1	230.5	232.6	234.6	236.6
220	238.8	240.7	242.4	244.1	240.5	242.6	244.6	246.6
230	248.8	250.7	252.4	254.1	250.5	252.6	254.6	256.6
240	258.8	260.7	262.4	264.1	260.5	262.6	264.6	266.6
250	268.8	270.7	272.4	274.1	270.5	272.6	274.6	276.6
260	278.8	280.7	282.4	284.1	280.5	282.6	284.6	286.6
270	288.8	290.7	292.4	294.1	290.5	292.6	294.6	296.6
280	298.8	300.7	302.4	304.1	300.5	302.6	304.6	306.6

Acknowledgements

I thank George Legarreta of the FAA and Jerry Robinson of Boeing for their untiring efforts to get this project (COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT 01-CRDA-0164 between THE FEDERAL AVIATION ADMINISTRATION WILLIAM J. HUGHES TECHNICAL CENTER and THE BOEING COMPANY) going, and again George Legarreta for his careful review of drafts of this report. I thank Ryan King, Jim White and Pete Sparacino of the FAA William J. Hughes Technical Center for their data collection effort and Andrew Booker of Boeing for assistance concerning the Schiphol deviation data. I especially want to thank Jerry Robinson for his inspired and involved support throughout.

1 Problem Statement, Data Description, and Approach

This report deals with the risk of collision between the wingtip of a taxiing aircraft and an object to the side of a taxiway and with the risk of wingtip to wingtip collision between two passing aircraft on parallel taxiways. Of particular concern is the separation distance T between taxiway centerline and object or between the centerlines of the parallel taxiways in relation to any tolerated risk level p.

This risk analysis for wingtip to object clearance will be based directly on the extreme deviation behavior of taxiing 747 aircraft as analyzed in reports [8] and [9] and could have been done there already. Instead it is included here in conjunction with the wingtip to wingtip collison risk analysis.

Since a wingtip to wingtip collision event can only occur at a specific point, namely when the wingtips of both aircraft pass each other, we are dealing with a pointwise risk as opposed to a lengthwise risk. The latter is encountered when dealing with the risk of an aircraft veering off a taxiway over a certain taxiway segment. Such veering off is determined by the maximum absolute deviation of the aircraft over the full taxiway segment under consideration. This risk can not yet be addressed in a satisfactory manner with the currently collected data since these were obtained at fixed laser stations along the respective taxiways. These measurements do not represent the distribution of maximum deviations over a taxiway segment.

Concerning the wingtip to object clearance one may raise the same concerns as with assessing the risk of veering off the taxiway. Much will depend on the nature of the object. If the object consists of a building that runs parallel to the whole length of the straight taxiway, there would be little difference in the two situations, except that the pilot may be more aware of the building than of the taxiway edge. If the object presents a rather short front to the taxiway one may neglect the concern of lengthwise risk and treat this as a pointwise deviation issue. If the object is a building with a lengthwise exposure surface one may also argue that the first encountered point of such a building could serve as the point for analysis since any further exposure to the building should induce the pilot to move away from the building rather than towards it, see Figure 1. This argument will fail of course if the building is hidden in fog. However, in this situation the pilot may follow the centerline lights more closely than otherwise, which would negate that concern. Thus in the main we rely on the relatively short lengthwise exposure to the object so that we may treat this as a pointwise risk.

The data underlying the risk analyses come from collection efforts at two airports. At New York's John F. Kennedy International Airport (JFK) 747 taxiway deviation data were collected from 6/24/1999 to 2/17/2000 at two laser locations for each of two parallel 75 ft straight taxiway segments with shoulder, called ALPHA and BRAVO. Similar data were collected at Anchorage International Airport (ANC) from 9/24/2000 to 9/27/2001 at two straight (non-parallel) 75 ft straight taxiway segments with shoulder, called KILO and ROMEO.

Although these data sets were screened to capture mostly 747 deviations there is the possibility that other, similarly large aircraft, e.g., L-1011, A-330, A-340, 777, MD-11 and DC-10, were included. This possibility is stronger for the JFK data than for the ANC data. For the latter we expect most deviations to be from 747 aircraft.

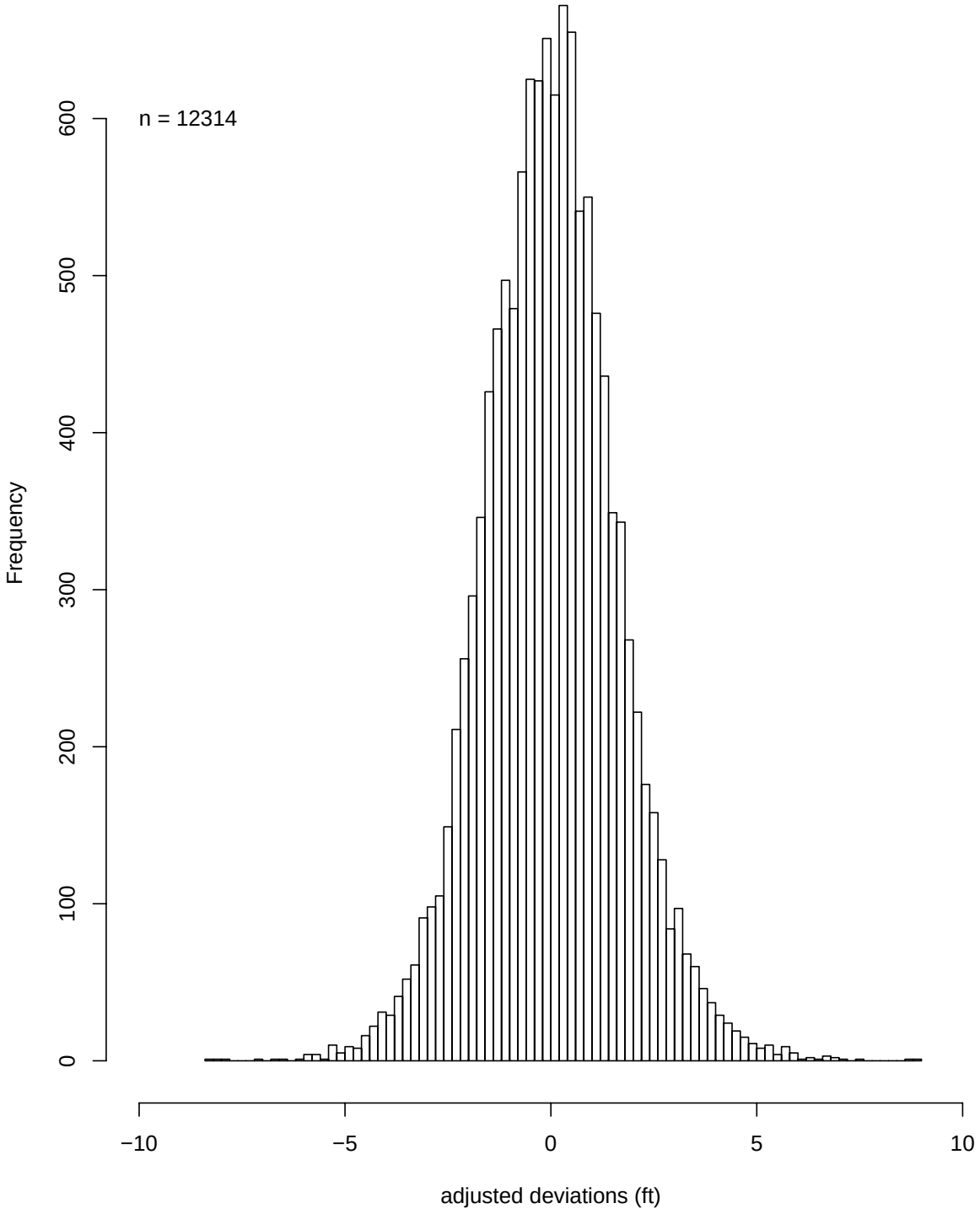
The extreme taxiway deviation behavior at the most relevant main gear location was analyzed in two previous reports [8] and [9]. For more details on the data and how they were treated we refer to these reports. In these reports the pooling of data from the two taxiways for each airport and from different headings on each taxiway was justified through bias elimination and rescaling. Two biases were identified and corrected for. The first of these biases is thought to arise from parallax issues due to the pilot's offset position relative to the aircraft centerline. This bias is of opposite sign for the two headings on each taxiway. The second bias was attributed to the taxiway centerlights which pilots tend to avoid. These centerlights run parallel to the taxiway centerline at some offset. This bias has the same sign for either heading.

The distributions of the bias adjusted ANC deviations and the bias adjusted JFK deviations (after dividing the latter by a factor of 1.097) looked so similar that their pooling was justified. The 10% wider scatter in the bias adjusted JFK deviations could be due to the fact that at JFK the centerline lights are offset by 22 inches from the centerline while at ANC the corresponding offset is 12 inches. This wider offset could cause wider aircraft meandering and thus more dispersed deviations at JFK.

The pooled adjusted deviations from both airports were found to be symmetrically distributed around zero. Hence it was possible to use the absolute deviations in gauging exceedance risks. This increases the sample size by a factor of two since extremes from both sides of the sample come to bear on the issue. Figure 4 shows a histogram of the pooled adjusted deviations. As indicated, there are $n = 12,314$ such deviations, of which $n_1 = 2,518$ and $n_2 = 9,796$ came from the JFK and ANC data sets, respectively.

We proceed by first dealing with the wingtip to object clearance risk assessment since that is a direct application of previous analysis of the pooled data. This is followed by two approaches of dealing with the risk of wingtip to wingtip collision. For this problem almost no direct data was available, i.e., deviations from taxiway centerline were hardly ever measured while aircraft were passing each other. To deal with this we developed two proxy approaches, one of which closely parallels that of the Schiphol taxiway deviation study [1] which used the extreme value methodology in [7]. The other approach estimates the convolution distribution of the combined deviations $d_1 + d_2$ directly and interpolates/extrapolates to the relevant $(1-p)$-quantiles of this distribution via a fitted quadratic on a log-scale for p. For this latter approach confidence bounds were obtained via the bootstrap method. The results from both approaches are quite comparable and complement each other.

Figure 4: Pooled Adjusted JFK and ANC Deviations



2 Separation Between Object and Centerline of Taxiway

Figure 1 on page iii shows that the criterion for clearance between a fixed object and the wingtip of a taxiing aircraft is

$$T > \frac{W}{2} + d\,,$$

where T is the closest distance between the taxiway centerline and the fixed object, W is the wingspan of the taxiing aircraft, and d is the deviation of the aircraft from the taxiway centerline.

In order to control the risk p of collision we need to have $T > W/2 + \eta_{1-p}$ where η_{1-p} is the $(1-p)$-quantile of the d-distribution, i.e., it is that value which is exceeded by a proportion p of all d-values in the d-distribution. Thus $T > W/2 + \eta_{1-p}$ means that in at most a proportion p of all cases we may violate the clearance criterion. As values of interest for p we consider $p = 10^{-3}, 10^{-4}, \ldots, 10^{-9}$.

Instead of the unknown η_{1-p} we have estimates and 95% upper confidence bounds for η_{1-p}, denoted by $\hat{\eta}_{1-p}$ and $\hat{\eta}_{U,1-p}$, respectively. They are presented in Table 1. The values in Table 1 were extracted from report [9]. The intent of 95% upper confidence bounds is to provide some conservative allowance for the sampling variability which causes variability in the estimates, i.e., different data sets collected under similar conditions would have resulted in different estimates. Such 95% upper confidence bounds are meant to exceed the target η_{1-p} with probability .95 while the estimates typically fall below and above η_{1-p} with equal chance .5.

These estimates and upper confidence bounds can then be combined with various half wingspans $W/2$ to arrive at threshold estimates and upper bounds $\hat{\eta}_{1-p} + W/2$ and $\hat{\eta}_{U,1-p} + W/2$ that should be exceeded by the distance T from the taxiway centerline to the fixed object for the respectively targeted exceedance risk p. The results are summarized in Tables 2 and 3 for ANC and JFK, respectively.

Table 1: $(1-p)$-Quantile Estimates and 95% Upper Bounds for the d Distribution

	one-sided exceedance risk p						
	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
estimate	6.04 ft	8.19 ft	10.49 ft	12.97 ft	15.62 ft	18.47 ft	21.53 ft
95% upper bounds	6.26 ft	8.52 ft	10.95 ft	13.56 ft	16.36 ft	19.36 ft	22.58 ft

Table 2: Estimates and 95% Upper Confidence Bounds to be Exceeded by the Required Separation T Between Taxiway Centerline and Object for ANC

	Estimate $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	96.8	98.9	101.2	103.7	106.4	109.2	112.3
190	101.8	103.9	106.2	108.7	111.4	114.2	117.3
200	106.8	108.9	111.2	113.7	116.4	119.2	122.3
210	111.8	113.9	116.2	118.7	121.4	124.2	127.3
220	116.8	118.9	121.2	123.7	126.4	129.2	132.3
230	121.8	123.9	126.2	128.7	131.4	134.2	137.3
240	126.8	128.9	131.2	133.7	136.4	139.2	142.3
250	131.8	133.9	136.2	138.7	141.4	144.2	147.3
260	136.8	138.9	141.2	143.7	146.4	149.2	152.3
270	141.8	143.9	146.2	148.7	151.4	154.2	157.3
280	146.8	148.9	151.2	153.7	156.4	159.2	162.3
	95% Upper Bound $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	97.0	99.3	101.7	104.3	107.1	110.1	113.3
190	102.0	104.3	106.7	109.3	112.1	115.1	118.3
200	107.0	109.3	111.7	114.3	117.1	120.1	123.3
210	112.0	114.3	116.7	119.3	122.1	125.1	128.3
220	117.0	119.3	121.7	124.3	127.1	130.1	133.3
230	122.0	124.3	126.7	129.3	132.1	135.1	138.3
240	127.0	129.3	131.7	134.3	137.1	140.1	143.3
250	132.0	134.3	136.7	139.3	142.1	145.1	148.3
260	137.0	139.3	141.7	144.3	147.1	150.1	153.3
270	142.0	144.3	146.7	149.3	152.1	155.1	158.3
280	147.0	149.3	151.7	154.3	157.1	160.1	163.3

Table 3: Estimates and 95% Upper Confidence Bounds to be Exceeded by the Required Separation T Between Taxiway Centerline and Object for JFK

	Estimate $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	97.4	99.7	102.3	105.0	107.9	111.0	114.4
190	102.4	104.7	107.3	110.0	112.9	116.0	119.4
200	107.4	109.7	112.3	115.0	117.9	121.0	124.4
210	112.4	114.7	117.3	120.0	122.9	126.0	129.4
220	117.4	119.7	122.3	125.0	127.9	131.0	134.4
230	122.4	124.7	127.3	130.0	132.9	136.0	139.4
240	127.4	129.7	132.3	135.0	137.9	141.0	144.4
250	132.4	134.7	137.3	140.0	142.9	146.0	149.4
260	137.4	139.7	142.3	145.0	147.9	151.0	154.4
270	142.4	144.7	147.3	150.0	152.9	156.0	159.4
280	147.4	149.7	152.3	155.0	157.9	161.0	164.4
	95% Upper Bound $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	97.6	100.1	102.8	105.6	108.7	112.0	115.5
190	102.6	105.1	107.8	110.6	113.7	117.0	120.5
200	107.6	110.1	112.8	115.6	118.7	122.0	125.5
210	112.6	115.1	117.8	120.6	123.7	127.0	130.5
220	117.6	120.1	122.8	125.6	128.7	132.0	135.5
230	122.6	125.1	127.8	130.6	133.7	137.0	140.5
240	127.6	130.1	132.8	135.6	138.7	142.0	145.5
250	132.6	135.1	137.8	140.6	143.7	147.0	150.5
260	137.6	140.1	142.8	145.6	148.7	152.0	155.5
270	142.6	145.1	147.8	150.6	153.7	157.0	160.5
280	147.6	150.1	152.8	155.6	158.7	162.0	165.5

The reason for separate tables for ANC and JFK is the fact that before being able to combine the bias adjusted data from ANC and JFK we had to divide the deviations from JFK by a factor of 1.097. Thus we have to multiply threshold values from Table 1 by 1.097 to make them apply back to the JFK situation. Another back-adjustment that needs to be made is the adding in of biases that were removed in order to render the deviations symmetric around zero. These biases, due to parallax issues and the avoidance of taxiway centerlights, combined maximally to roughly .75 ft.

Thus, to arrive at the value 117.4 ft to the right of the 220 ft wingspan in Table 3 for JFK one calculates (taking 6.04 from Table 1) $6.04 \times 1.097 + .75 + 220/2 = 117.4$ ft. In the calculation for Table 2 one omits the factor 1.097 in the calculation.

3 Separation of Parallel Taxiways

When aircraft pass each other on parallel taxiways (in same or opposite heading) there is the risk of wing tip collision due to sufficiently extreme deviations of either or both aircraft from their respective taxiway centerlines. Figure 2 on page viii illustrates this situation and sets the notation as it will be used in the following discussion. Much of the approach taken here parallels that of Booker [1] who addressed the same issue using data from Schiphol Airport near Amsterdam.

Since such a collision event can only occur at a specific point, namely when the wingtips of both aircraft pass each other, we are dealing with a pointwise risk as opposed to a lengthwise risk. The distinction between pointwise and lengthwise risk was discussed previously.

The reason for treating the wing tip collision issue as a pointwise risk is as follows. The maximum deviation along the taxiway segment does not matter since such maximum deviations for aircraft 1 or 2 are most likely not occurring at the time when the aircraft pass each other. The time or point of passing should have no connection with the actual deviations taking place at the point. Thus it makes sense to fix the point of passing as the point when deviation measurements are taken. Such a point of passing could be anywhere (randomly) along the taxiway. The deviations measured at these random locations will have the same distributional characteristics as deviation measurements taken when passing a fixed laser.

The underlying assumption is that the distribution of deviations from the taxiway centerline is the same at each fixed location along the taxiway. Measuring all the deviations at one location should produce the same sample (distributionally) as taking each measurement at a randomly chosen location, determined by the point of the two aircaft passing each other.

Unfortunately it is not possible or practical to collect deviation data on passing aircraft at such randomly set up laser location as suggested above. Actually, such data might show possible avoidance behavior of passing aircraft, i.e., if deviations from the centerline occur they are more likely in opposite direction as opposed to towards each other when the aircraft pass each other. We realize that this possibility would argue against our previously stipulated independence of passing location and aircraft deviation from taxiway centerline. However,

ignoring this possible avoidance effect in the approaches followed below should err on the conservative side, i.e., collision risks calculated under the assumed independence between deviations and passing location should be higher than those that might take the avoidance effect into account in some way, although how to achieve the latter is not clear.

Figure 2 on page viii shows that the quantity $d_1 + d_2$ is the varying aspect that figures in the separation analysis since the clearance C between wingtips can be expressed as

$$C = T - \frac{W_1 + W_2}{2} - (d_1 + d_2) \,, \tag{1}$$

where T denotes the distance between the parallel taxiway centerlines, W_1 and W_2 denote the wingspans of the passing aircraft and d_1 and d_2 denote the aircraft deviations from the taxiway centerlines and are viewed as positive when measured in the direction towards the other aircraft.

As far as the distribution of C is concerned the sign of the deviations d_1 and d_2 is immaterial since we established that the distribution of the pooled adjusted deviations can be considered as symmetric around zero. Thus the values d_1+d_2 have again a distribution that is symmetric around zero.

According to the above clearance criterion C, large positive sums d_1+d_2 are relevant because a collision could occur (with equal wing height) when $C < 0$, or $T - (W_1 + W_2)/2 < d_1 + d_2$. Since the distribution of $d_1 + d_2$ is symmetric around zero we can again use the distribution of absolute deviations $|d_1 + d_2|$ in assessing any exceedance risks, i.e., we can use the fact

$$\frac{P(|d_1 + d_2| > x)}{2} = \frac{P(d_1 + d_2 > x) + P(d_1 + d_2 < -x)}{2} = P(d_1 + d_2 > x) \quad \text{for } x > 0.$$

This allows us to use twice as many sample extremes in extrapolating to risk levels beyond the observed data.

4 Analysis Based on Random Splitting of the Deviation Sample

In view of the lack of direct wingtip separation data of passing aircraft we use as proxy the bias adjusted ANC and JFK deviation data (also rescaled in the case of JFK) and form sums of independent deviations $d_1 + d_2$ from which the clearance C derives according to equation (1). By splitting the original $n = 12,314$ deviations randomly into two halves and randomly pairing exactly one d_i from the first half with one d_j from the second half we can obtain $m = 6,157$ such independent sums which we then analyze in terms of their absolute values as indicated above.

The problem with this approach is that the forming of the $m = 6,157$ sums of pairs from the original sample of deviations is subject to the random splitting and random pairing of the $n = 12,314$ deviations. Thus one could get many possible answers in the extrapolation from all such samples of randomly generated sums. Which such estimate do we take?

This situation is somewhat similar to the situation where we don't have a sample but can generate any number M of samples of size m from a "known" population. For each such sample we could calculate an estimate $\hat{\theta}$ for the target θ that describes a particular aspect of the population (mean,variance, quantile, etc.). It is assumed that such estimates are reasonable, namely scatter in some form around the target θ. One could then take the average or median of all these estimates $\hat{\theta}_1, \ldots, \hat{\theta}_M$. Depending on the bias properties of these estimates one should get a fairly accurate view of the target θ. In a way (by sampling repeatedly from a known population, namely M times a sample of size m) one takes advantage of the fact that one already knows the population and thus indirectly θ. Taking the mean or median of a large number M of such estimates should thus provide a fairly accurate estimate of the "unknown" θ. This is as though one estimates θ based on $M \cdot m$ observations, a very large number.

Our situation at hand is slightly modified from the previous setup. In our case we do not take a new sample of deviations each time, we are fixed on the given sample of $n = 12,314$ deviations. However, we can do the splitting and sum-forming process randomly many times over. This way we capture the random aspects of the sum-forming process and can gauge the variability of any quantile extrapolation estimate that is due to that random aspect. Following the bootstrap approach we can go one step further and also emulate the sampling variability in the original sample of $n = 12,314$ deviations. This consists of taking a random sample with replacement from these $n = 12,314$ deviations and then use this random sample as the one that gets randomly split to form the $m = 6,157$ sums of pairs, leading to our first estimate of the extrapolated quantile of interest. This process (random sample with replacement from these $n = 12,314$ deviations and randomly splitting it to get $m = 6,157$ sums of pairs) can be repeated many times and thus we ultimately can get as many estimates of the extrapolated quantile of interest, say $\hat{\theta}_1, \ldots, \hat{\theta}_M$.

Figures 5-7 show via box plots the results from randomly splitting the same original sample of $n = 12,314$ deviations into two halves and forming $m = 6,157$ sums of pairs. This split was performed 500 times and for each set of $m = 6,157$ sums we applied the extrapolation method in [7] for estimating the $(1-p)$-quantile of the distribution of absolute sums $|d_1 + d_2|$ for $p = 2 \cdot 10^{-i},\ i = 3, \ldots, 9$ (note that Figures 5-7 only show the results for $i = 4, \ldots, 9$). These quantiles are the same as the corresponding $(1-p)$-quantiles of the distribution of sums $d_1 + d_2$ for $p = 10^{-i},\ i = 3, \ldots, 9$. This was done for each of the 500 absolute sum samples of size $m = 6,157$ and for each of the following tail depths $k = 50, 100, 150, 200, \ldots, 1950, 2000$. By tail depth we mean that the k most extreme absolute sums were used in the extrapolation step.

The midline of each box plot (see Chambers et. al. [2]) gives the median of these 500 estimates for the respective $(1-p)$-quantile of the absolute sum distribution. The bottom and top edge of each box signify the lower and upper quartile of the 500 estimates. From the box ends extend dashed lines to the so called *adjacent values*. The upper adjacent value is defined to be the largest of the 500 estimates that is less than or equal to the upper quartile plus 1.5 times the interquartile range. Correspondingly, the lower adjacent value is defined to be

the lowest of the 500 estimates that is greater than or equal to the lower quartile minus 1.5 times the interquartile range. The interquartile range is the vertical width of the box. Any observations beyond the adjacent values are shown individually.

Clearly these box plots exhibit a fair amount of scatter, i.e., much of the variability of the extrapolated quantile estimates can be attributed to the random splitting and pairing for forming the $m = 6,157$ absolute sums for each of the 500 samples. Also, note that the scatter decreases as k increases. At the same time the boxes tend to drift down. The effect of this drift is made clearer by the horizontal line drawn across all box plots. The level of this line is roughly near the point where the medians peak. The sole purpose of these lines is to indicate the drift. This clearly shows the effect of the choice of k. The more extremes we choose for extrapolation the less variability we have in the estimates. At the same time the downward bias from the center part of the absolute sum data sample is felt more strongly.

Since each of the estimates is by design to have a 50% chance of falling above (or below) its intended target it is reasonable to choose the median of these 500 estimates as the ultimate choice among the scatter of estimates for each k. The respective choices of $k = 900, 450, 400, 250, 250, 250, 250$ appear to be reasonable values to focus on for $p = 2 \cdot 10^{-i},\ i = 3, \ldots, 9$ since they seem to indicate the start of the general downward trend. The resulting estimate is shown in each respective plot in the upper right corner below the given value of p. The significance of the combined 95% upper bound below it will be addressed later.

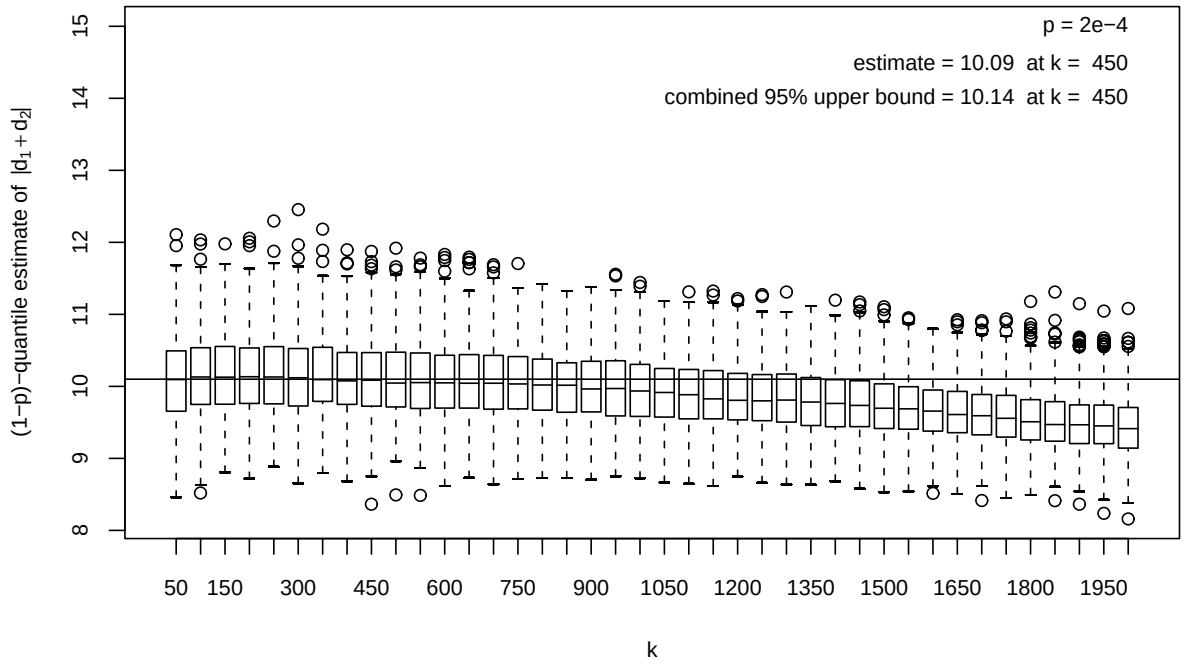
We point out that the 500 estimates are not independent of each other since the same sample was split each time. However, the dependencies between the samples should be very mild and one can expect the median to have the usual reasonable estimation properties. Conditional on the original sample of $12,314$ deviations the 500 estimates are independent of each other.

Figures 8-10 repeat the representation of Figures 5-7, but this time by first randomly choosing a bootstrap sample with replacement from the original sample of $n = 12,314$ deviations and then randomly splitting it to get $6,157$ sums. From these samples of size $6,157$ we again calculate the quantile estimates. This process is repeated 500 times. Thus the resulting 500 estimates not only capture the variability of the sample splitting and the absolute sum formation but also to some extent the sampling variation that led to the original deviation sample, i.e., we sample from the "population" given by the histogram in Figure 4 on page 3. This histogram should be a reasonable representation of the true deviation population from which the original deviation sample was drawn.

Comparing the box plots from Figures 5-7 with the corresponding ones from Figures 8-10 one sees a slightly wider scatter in the box plots of Figures 8-10. However, it seems clear that most of the added scatter is small compared to the original scatter attributed to the random splitting and absolute sum forming process. The chosen estimates are slightly lower in Figures 8-10 than the corresponding ones from Figures 5-7. This may be due to a small downward bias arising from some bootstrap troubles with extremes as is discussed below. For practical purposes this seems to be of little consequence.

We are aware that resampling from the original sample of $n = 12,314$ deviations will not be fully representative of the very extreme value behavior. For one, there is no possibility

Figure 5: The Choice of k (500 random splits, without bootstrap)



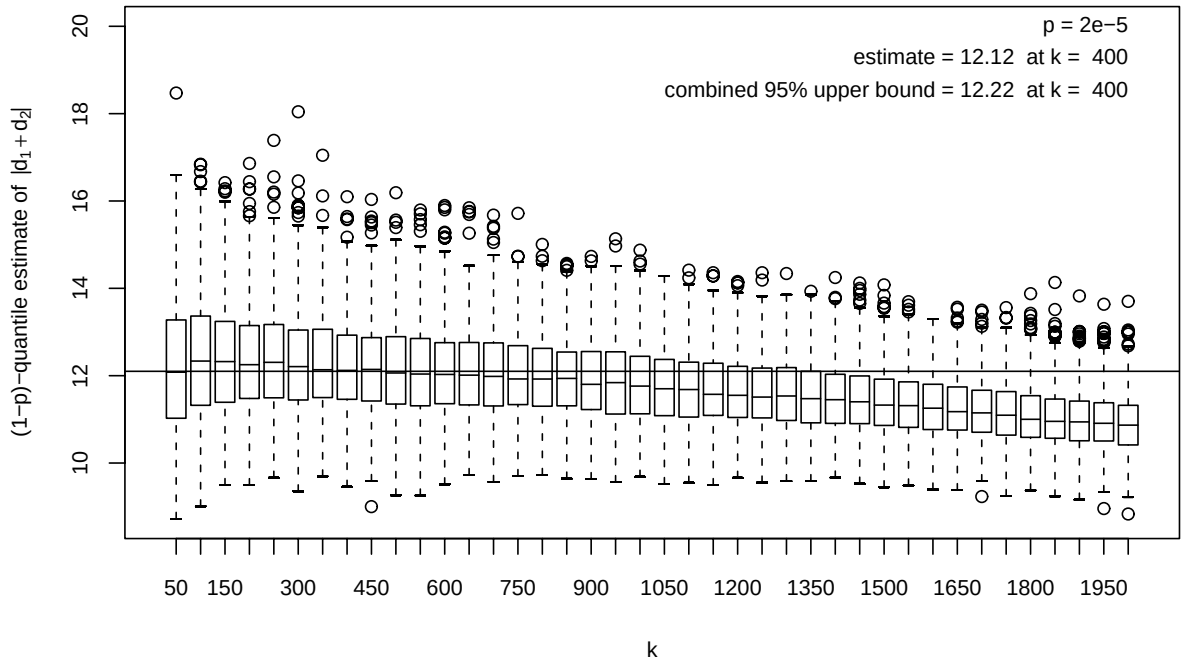
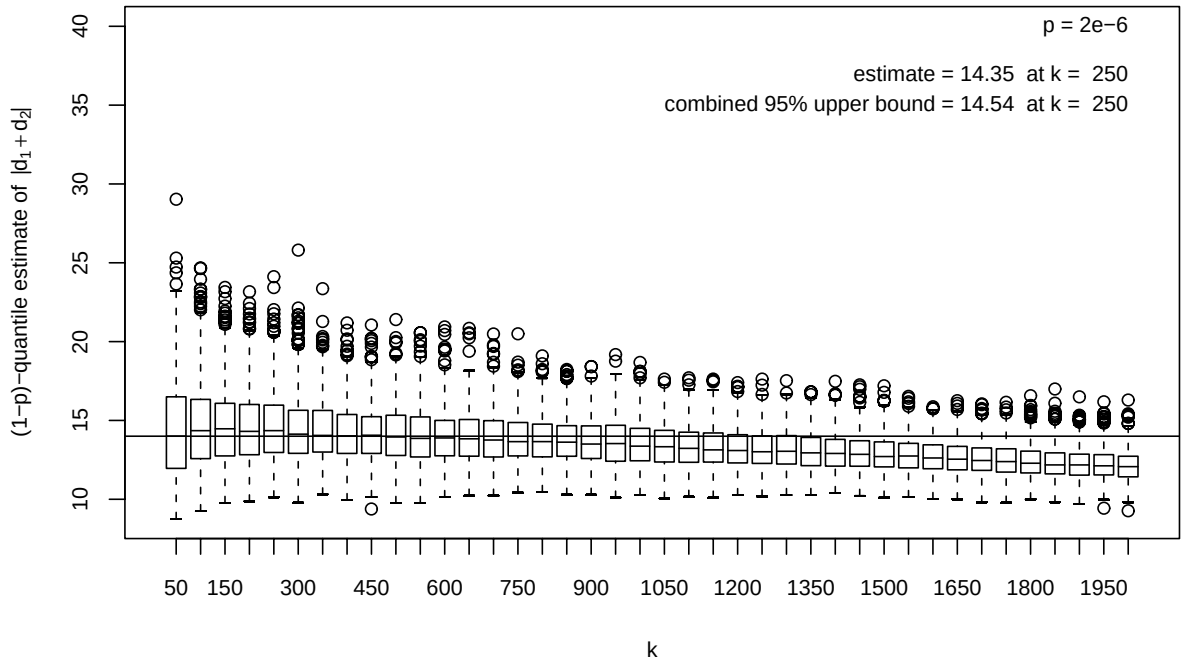


Figure 6: The Choice of k (500 random splits, without bootstrap)



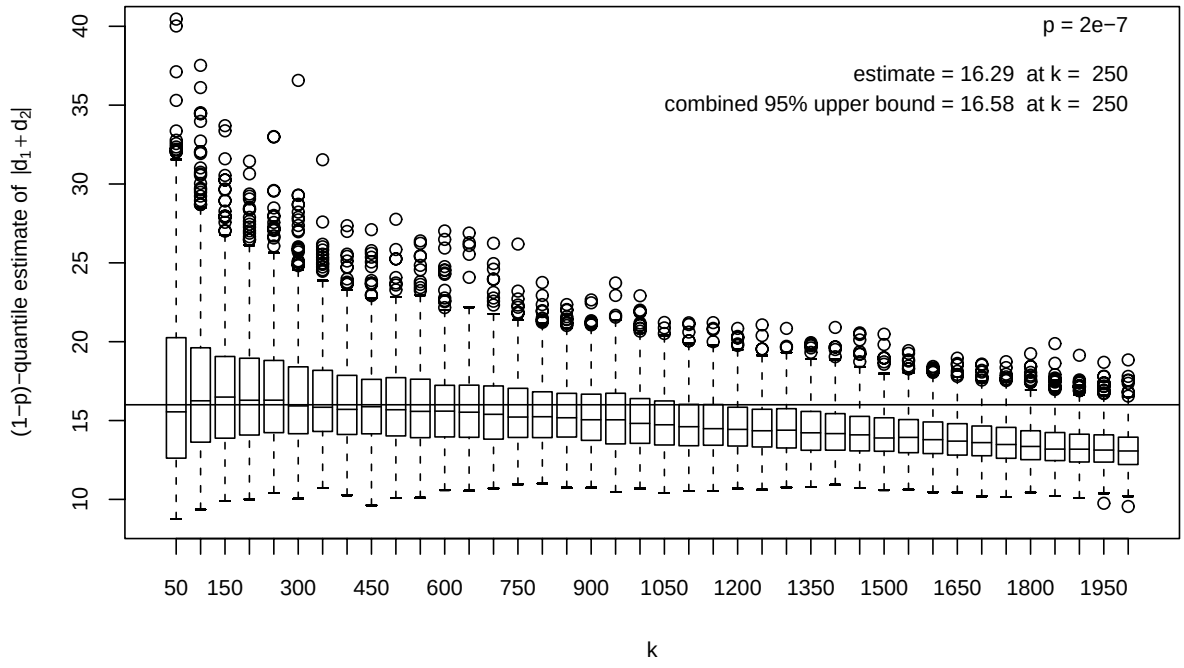
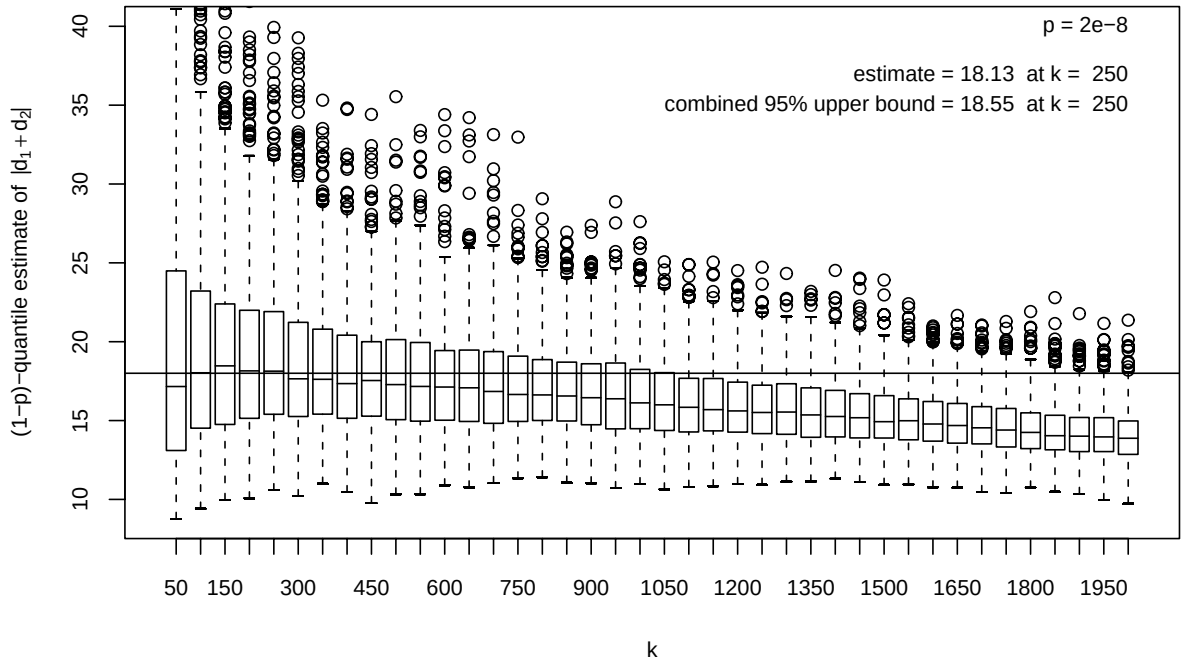


Figure 7: The Choice of k (500 random splits, without bootstrap)



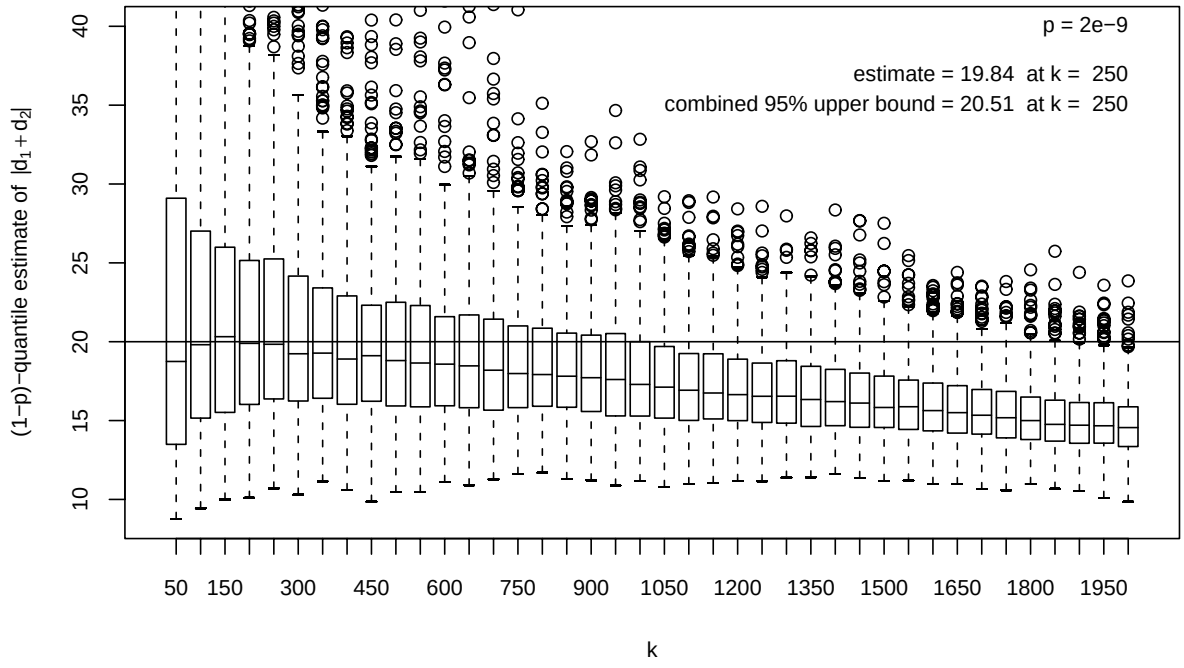
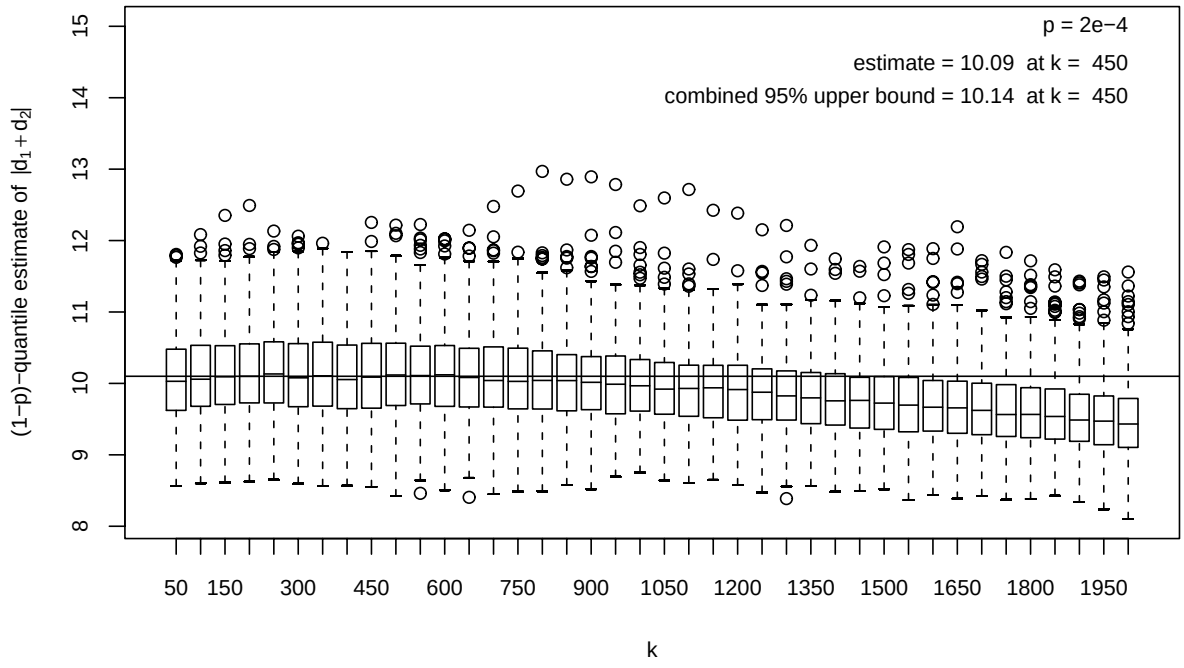


Figure 8: The Choice of k (500 random splits, with bootstrap)



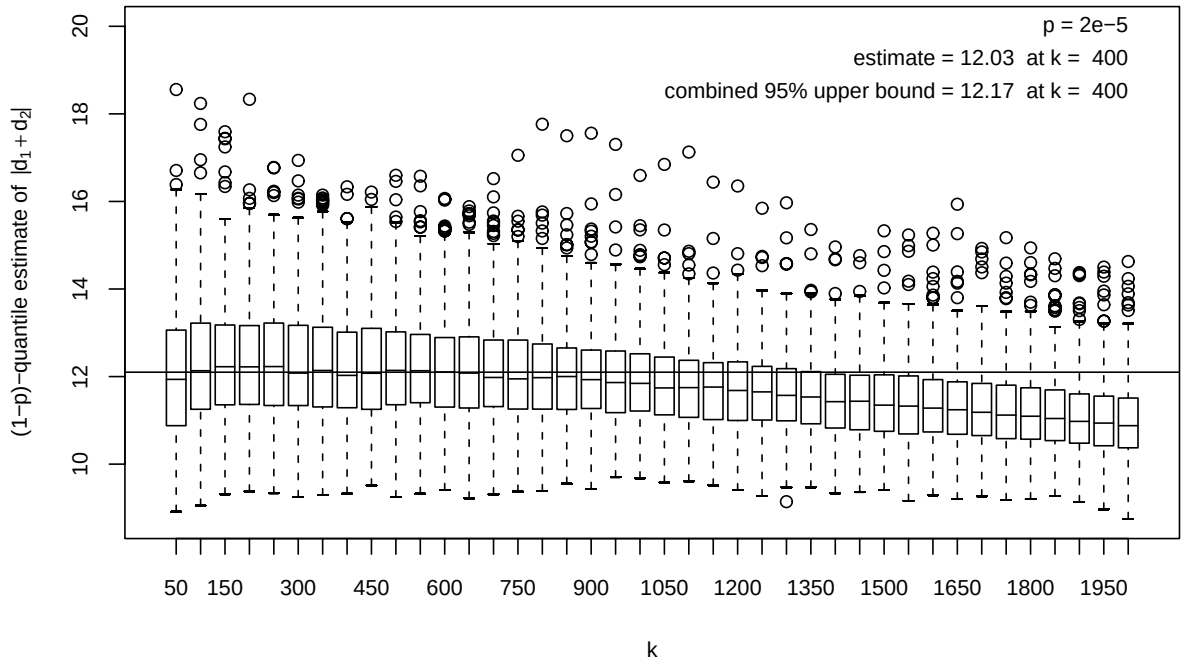
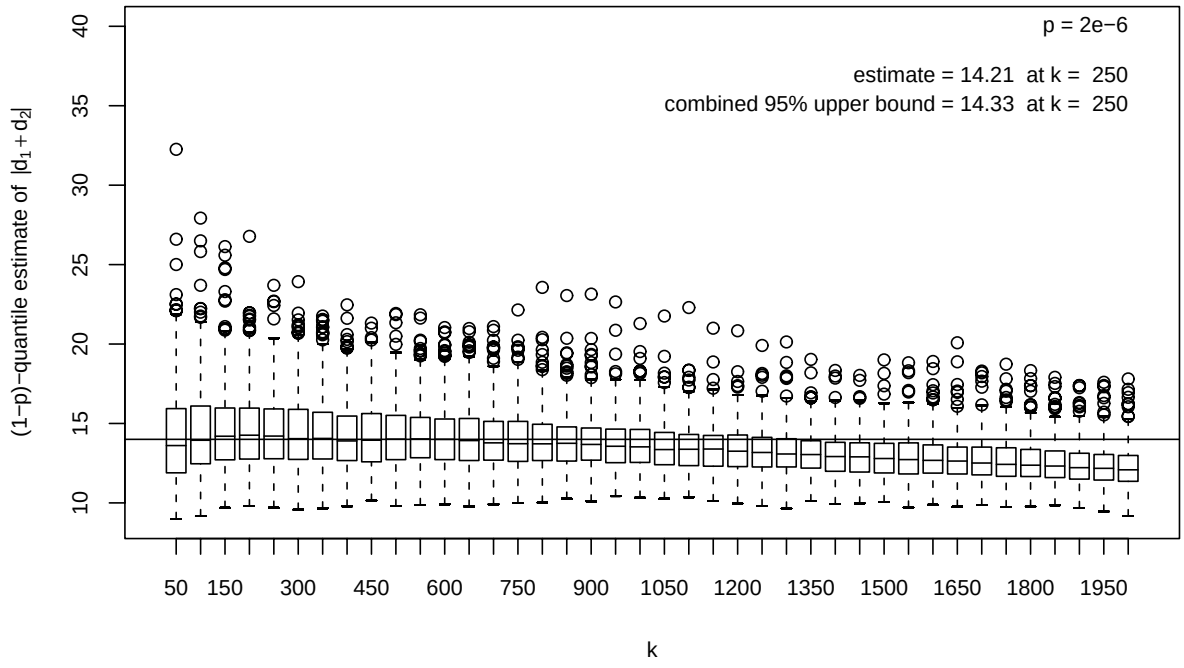


Figure 9: The Choice of k (500 random splits, with bootstrap)



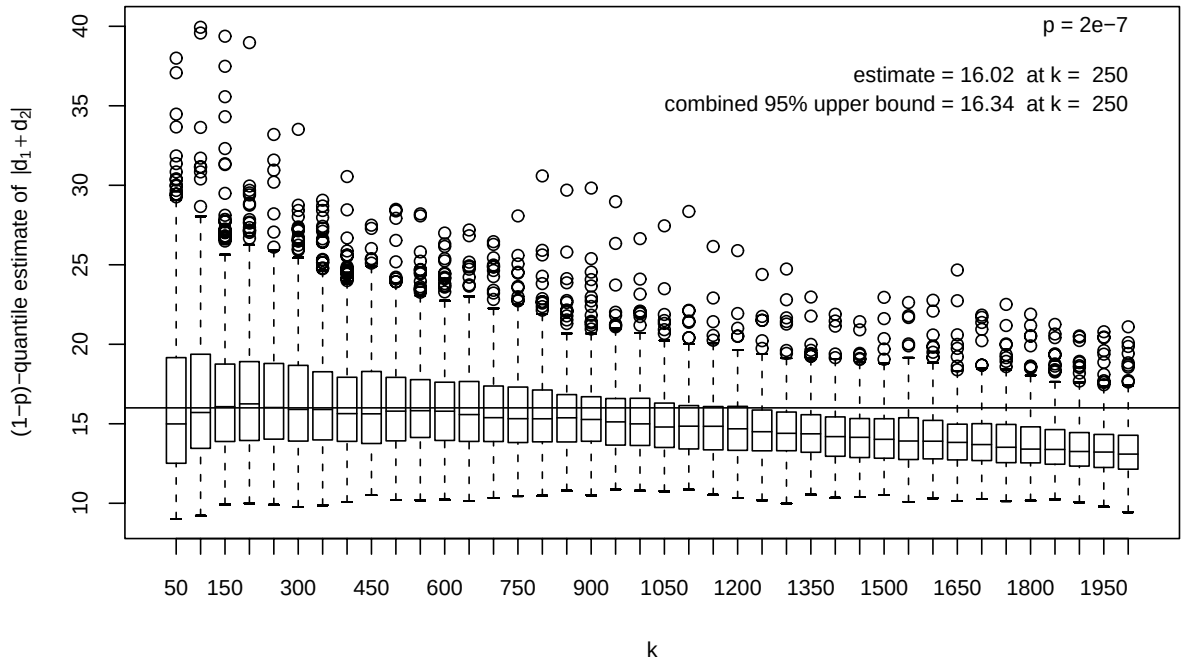
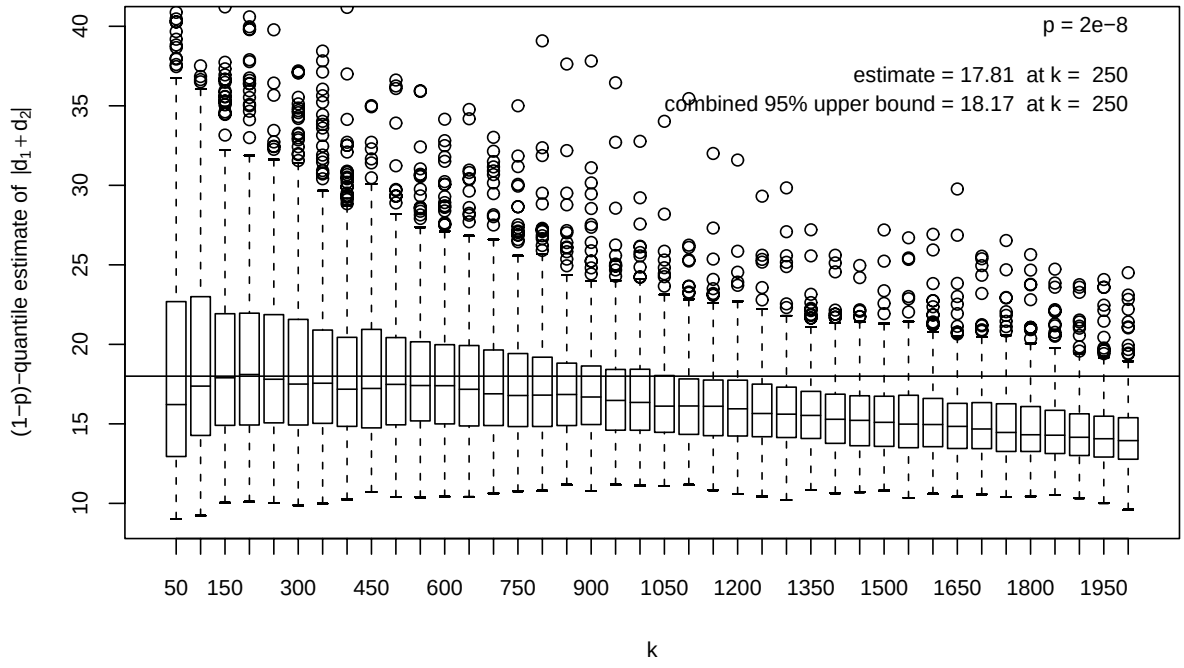
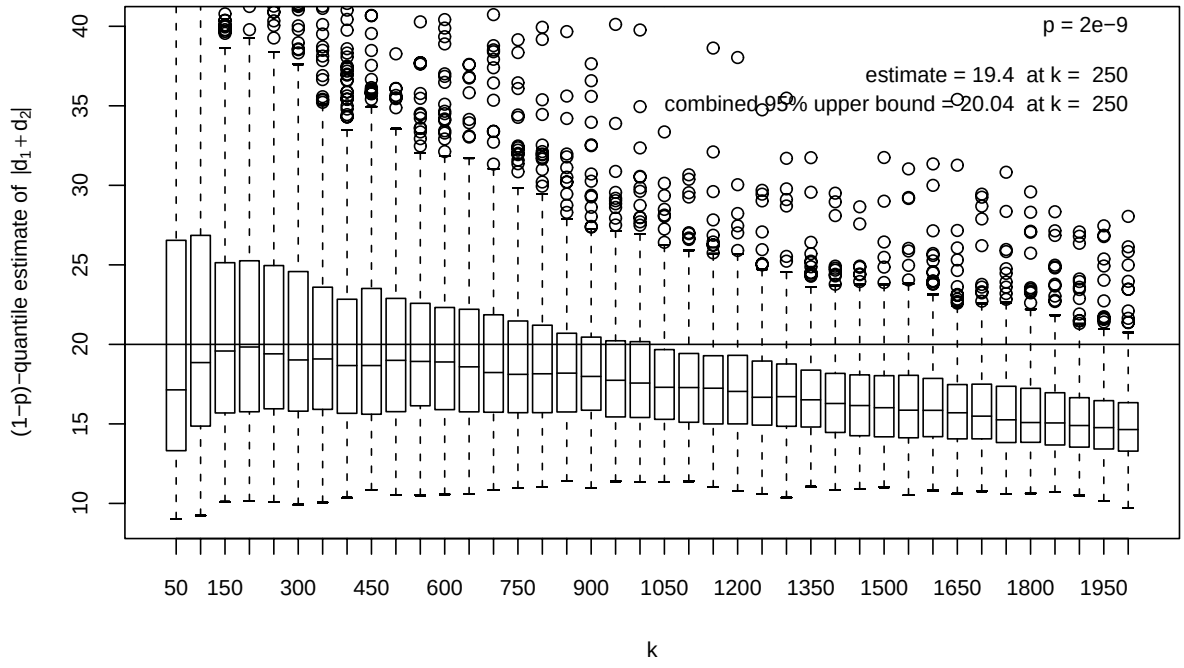


Figure 10: The Choice of k (500 random splits, with bootstrap)





of getting observations more extreme than the ones in the original sample. Also, some of the resamples may not even contain some of the very extremes of the original sample. This is known to cause trouble with the bootstrap method when using just the very extremes to estimate distribution end points. However, our extrapolation method downplays the effect of these very extremes by using a weighted least squares procedure on the k most extreme observed absolute sums.

It also appears that $k = 650$ might have been a reasonable choice for $p = 2 \cdot 10^{-6}$ and $p = 2 \cdot 10^{-7}$. However, we chose to stay with a common choice of $k = 250$ for $p = 2 \cdot 10^{-i}$, for $i = 6, 7, 8, 9$ since it seems to be reasonable for all four of these quantiles. The effect on the actual median would have been minimal either way.

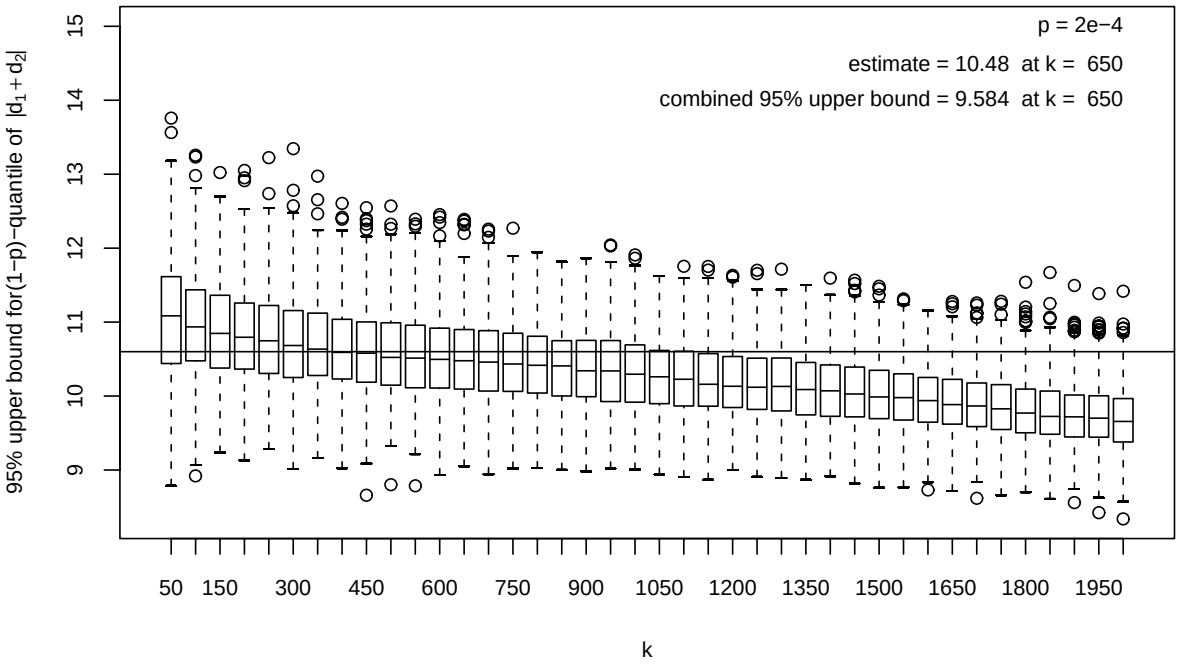
Figures 11-13 and 14-16 repeat the representation of Figures 5-7 and 8-10, but this time in terms of the 95% upper confidence bounds for the respective $(1 - p)$-quantiles. These bounds are generally higher and more scattered than the estimates as is to be expected. It is not immediatley obvious which one of these 500 upper confidence bounds to choose as final answer. The median of these 500 upper bounds looks quite attractive but its actual confidence level is no longer .95 since only 5% of the actual upper bounds are expected to fall below their target. From that one can calculate the chance of the median upper bound falling below the same target as

$$\gamma = \sum_{i=251}^{500} \binom{500}{i} .05^i .95^{500-i} = 10^{-421} \ ,$$

i.e., we would basically have a 100% upper confidence bound when using the median. A way of choosing among these 500 upper confidence bounds that maintains the 95% confidence level (approximately) is to sort these 500 upper bounds from smallest to largest and take the 34^{th} smallest of them as a combined 95% upper bound for the $(1-p)$-quantile. The rationale for this is given in Appendix A. These combined 95% upper confidence bounds show the peculiar behavior that they are below the corresponding combined estimates. This can be seen when comapring the "estimate" values in the upper right corner of the plots in Figures 5-10 with the corresponding values of the "combined 95% upper bound" in Figures 11-16. This phenomenon is not particular to the estimation/confidence bound procedure at hand as is illustrated with a different example in Appendix A.

For this reason we use a different method of obtaining 95% upper confidence bounds based on the sorted 500 estimates, i.e., we take the 269^{th} smallest of those 500 estimates as our 95% upper confidence bound. These are shown in Figures 5-7 and 8-10 in the upper right corner of each plot and are labeled as "combined 95% upper bound." The rationale for this is again given in Appendix A.

Figure 11: The Choice of k (500 random splits, without bootstrap)



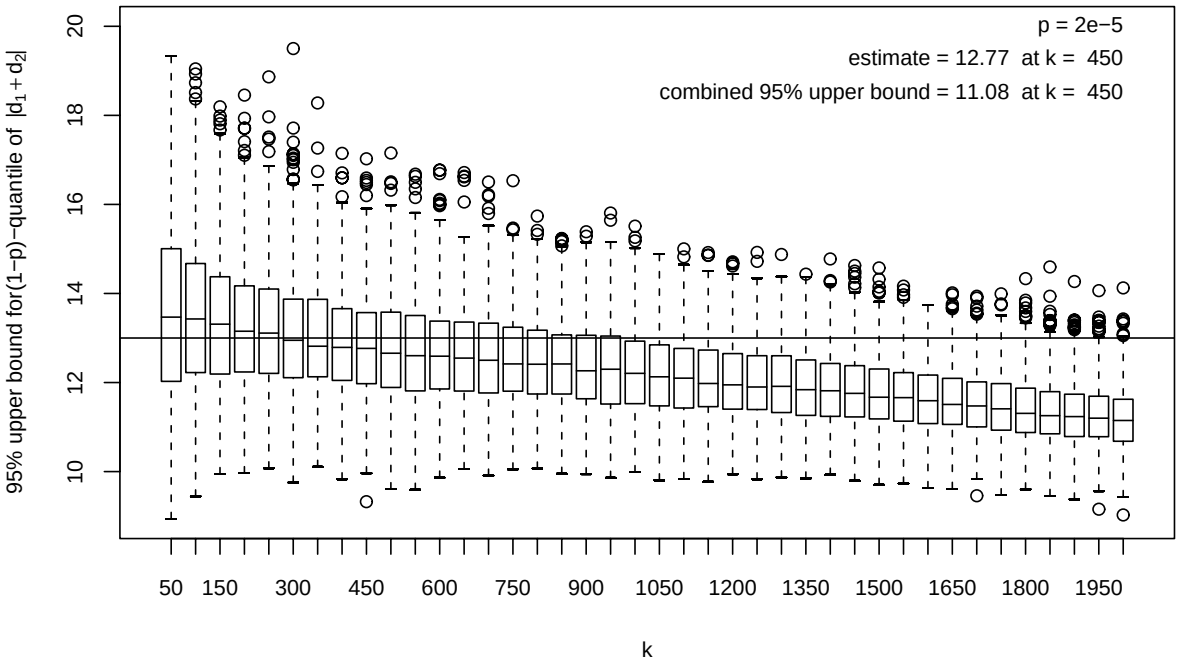
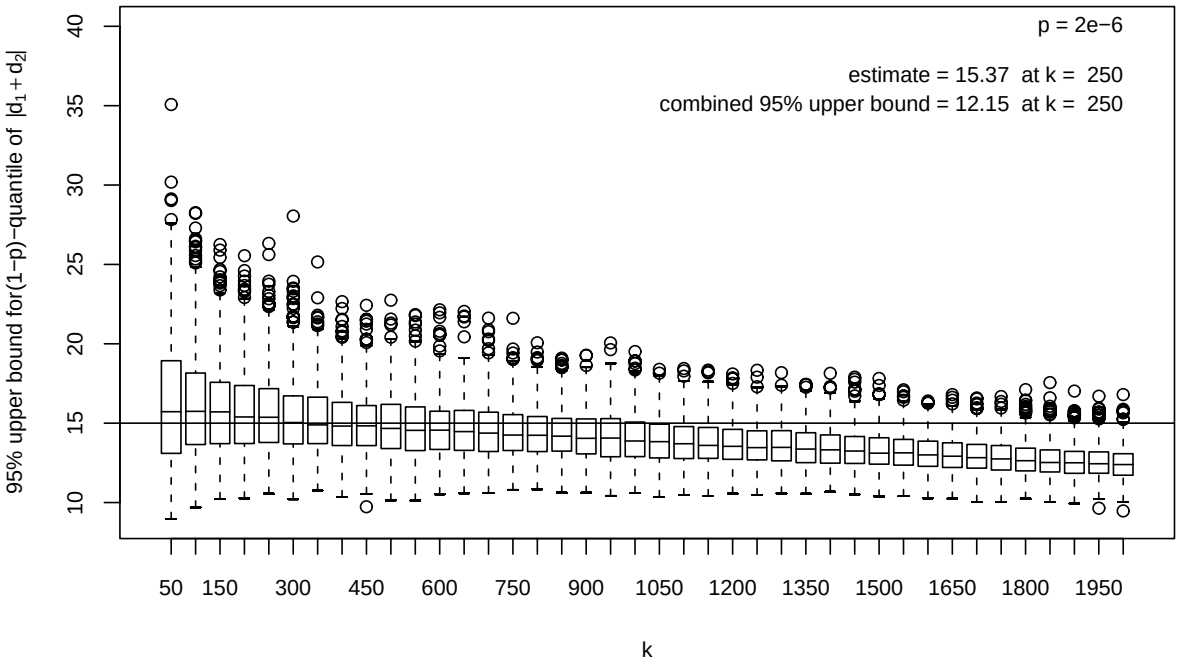


Figure 12: The Choice of k (500 random splits, without bootstrap)



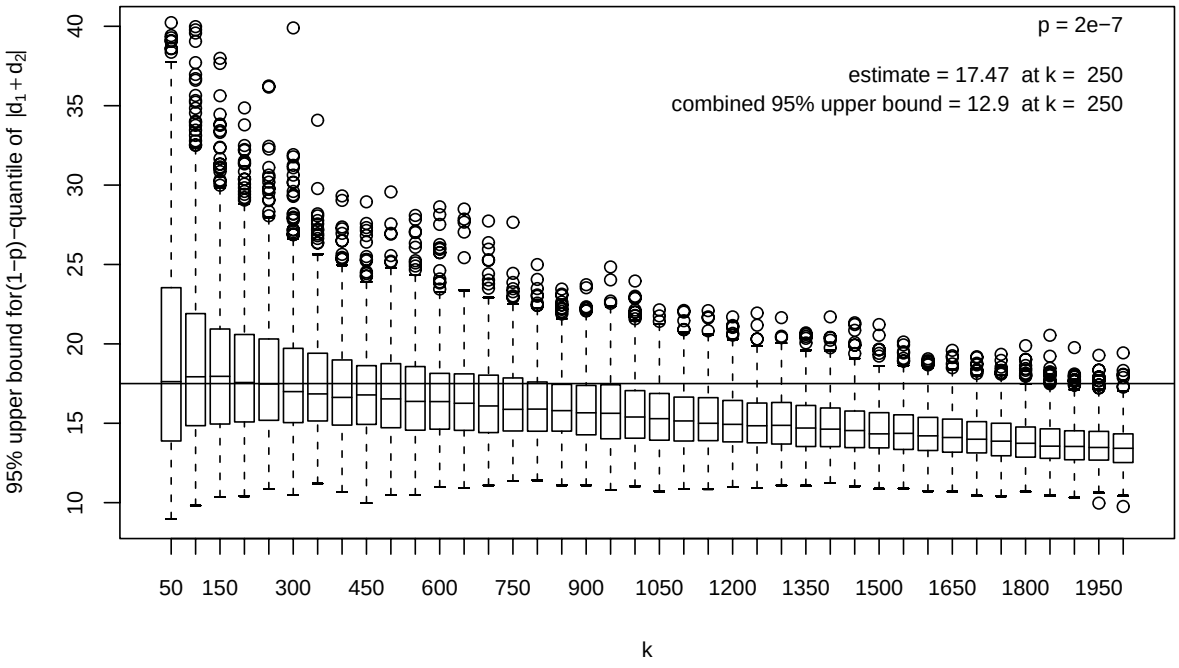
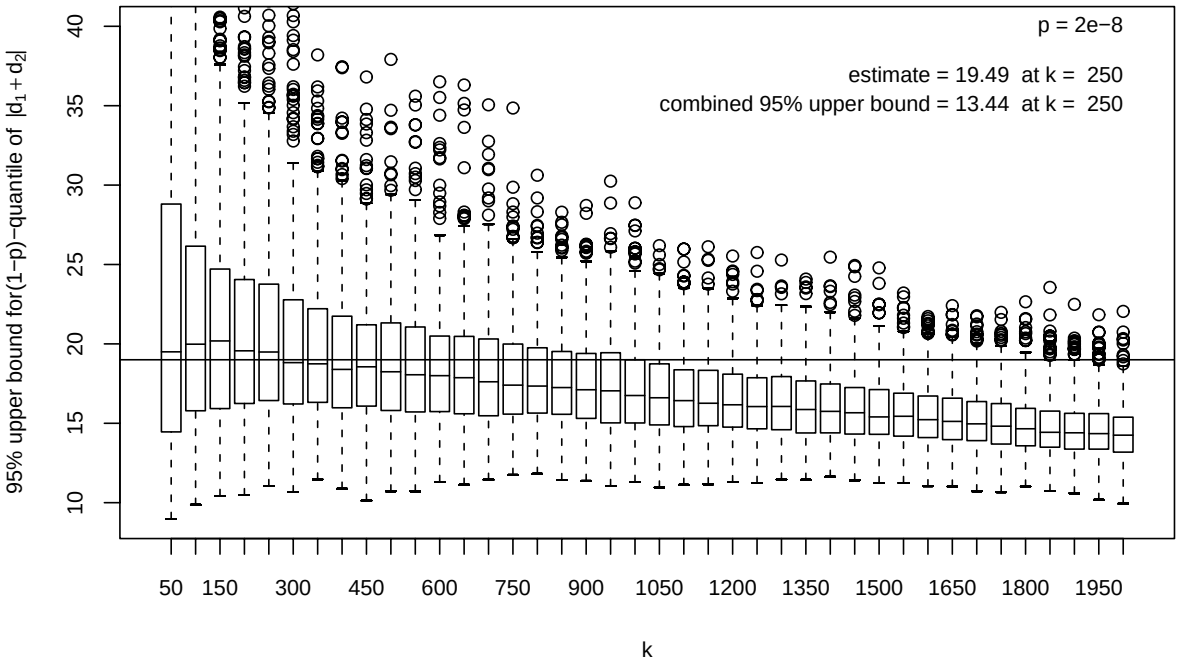


Figure 13: The Choice of k (500 random splits, without bootstrap)



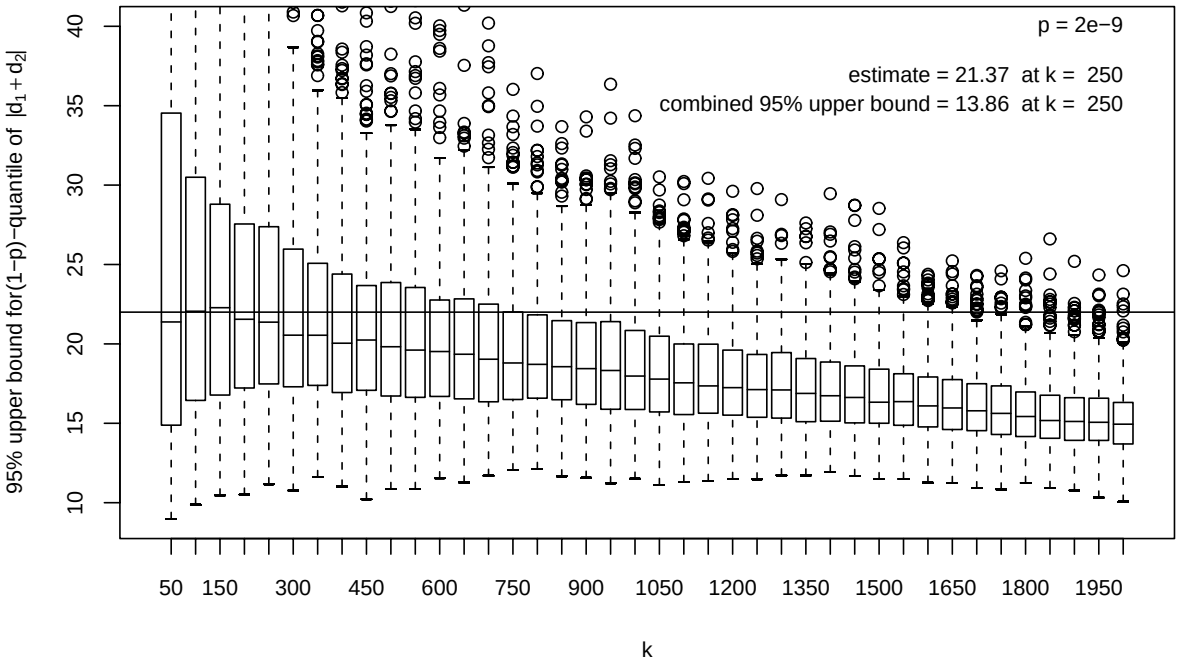
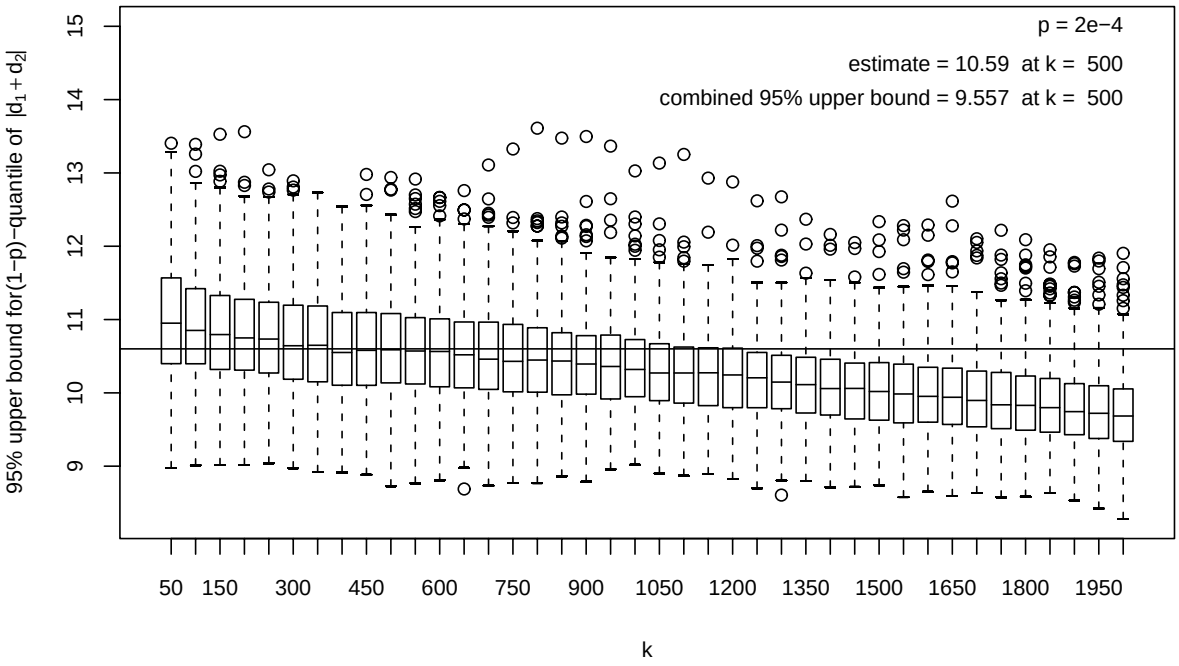


Figure 14: The Choice of k (500 random splits, with bootstrap)



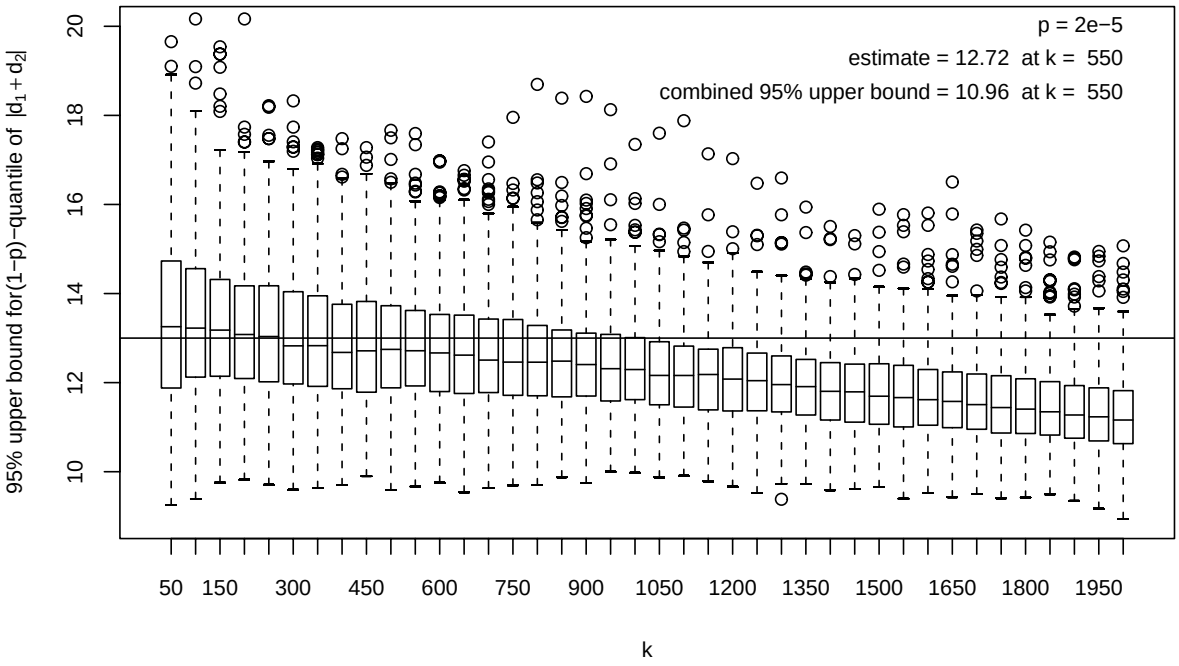
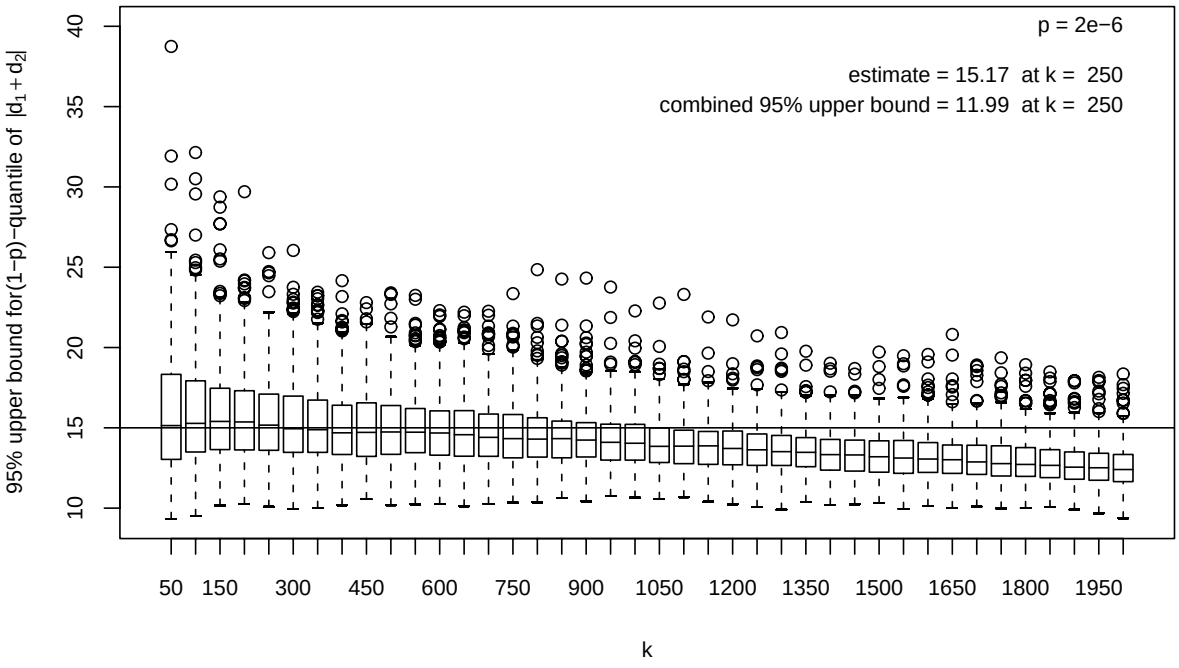


Figure 15: The Choice of k (500 random splits, with bootstrap)



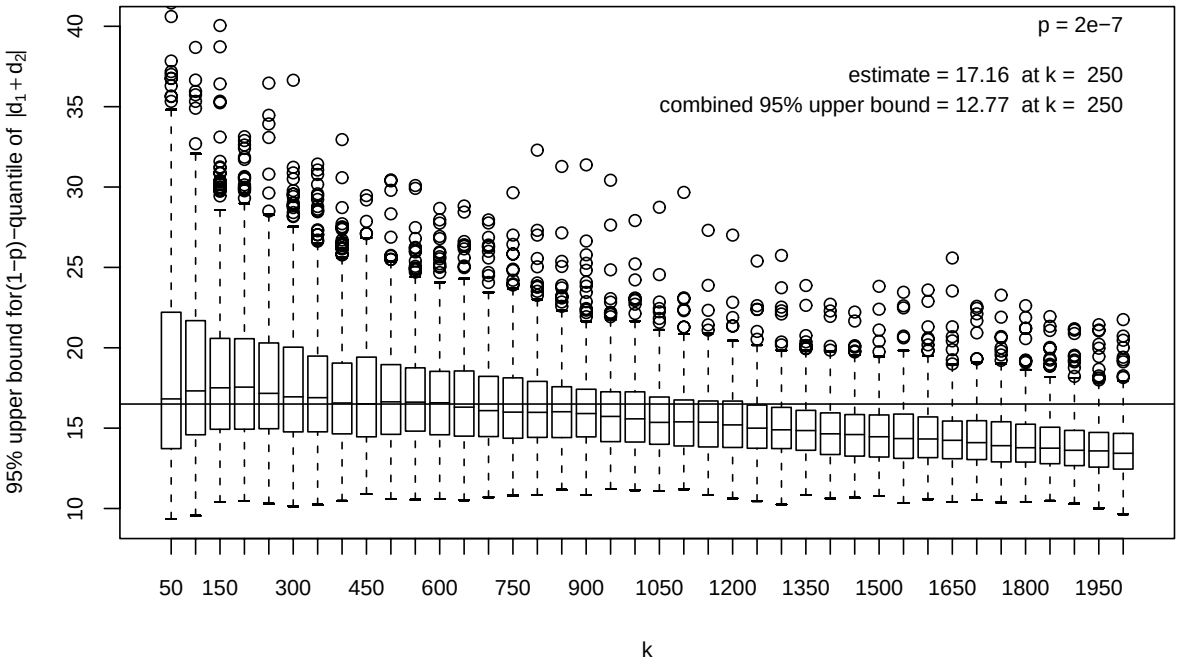
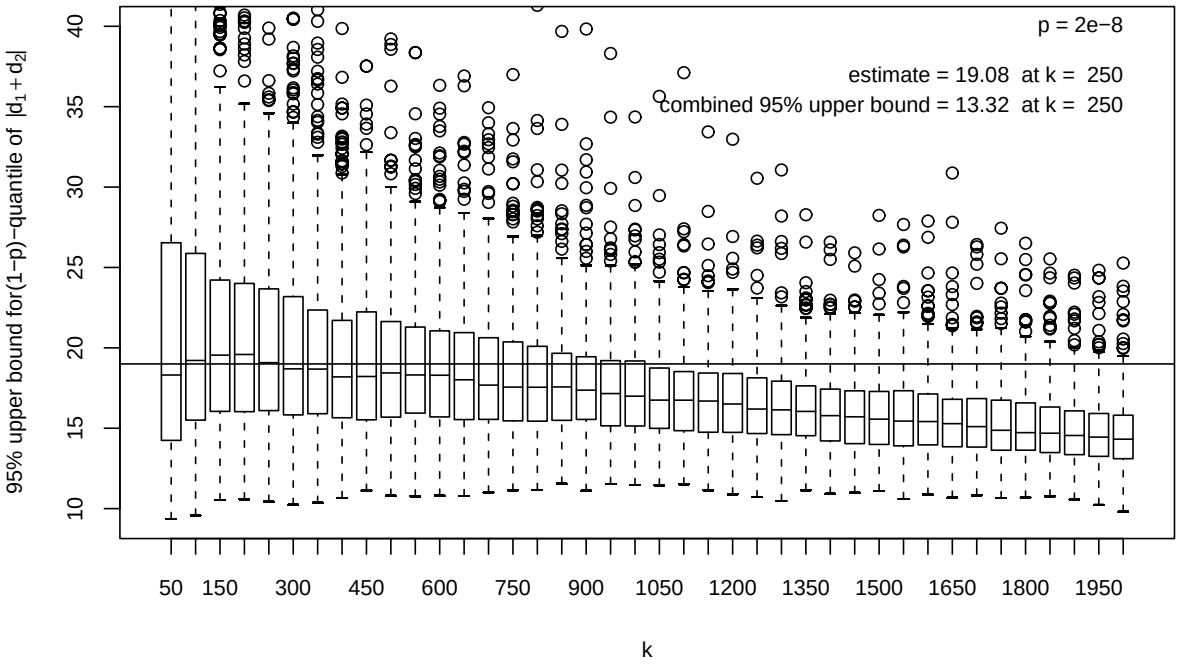


Figure 16: The Choice of k (500 random splits, with bootstrap)



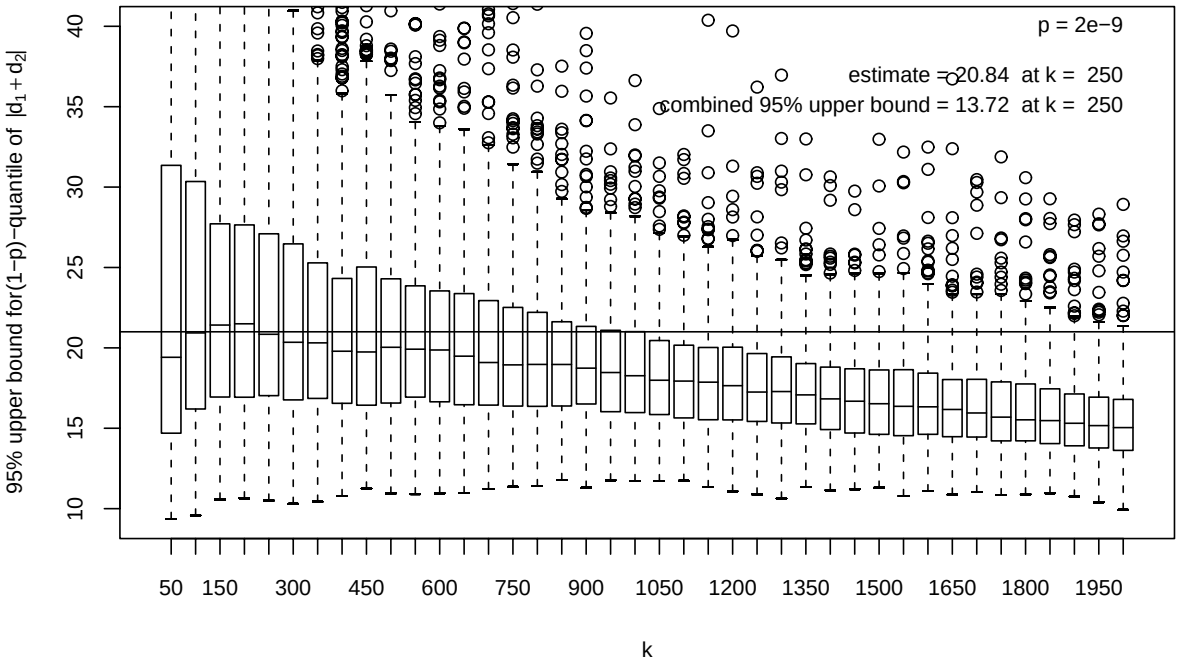


Table 4: $(1-p)$-Quantile Estimates and 95% Upper Bounds for the $d_1 + d_2$ Distribution Based on Random Splitting of Deviation Sample

	one-sided exceedance risk p						
	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
estimate	7.90 ft	10.09 ft	12.12 ft	14.35 ft	16.29 ft	18.13 ft	19.84 ft
95% upper bounds	7.91 ft	10.14 ft	12.22 ft	14.54 ft	16.58 ft	18.55 ft	20.51 ft

When comparing the estimates or 95% upper confidence bounds for the same $(1-p)$-quantile obtained with just random splitting of the same original deviation sample (without resampling from that sample first) with those that also incorporate the sampling variation from the original deviation sample one sees that the former results come out slightly higher (more conservative) than the latter. This may seem paradoxical in view of the fact that the second method (with resampling) compounds two forms of variation, namely the sampling variation of the deviations and the sampling variation of forming the paired sums. This paradox may have its explanation partly in the dominance of the variation arising from the random splitting and pairing and partly in the earlier discussed problem, namely that a good proportion of such resamples may miss some of the extremes of the original sample. For that reason we opted to use the more conservative numbers which are summarized in Table 4. We point out that the medians of the 95% upper bounds are not much higher ($.43 - .89$ ft) than the combined 95% upper bounds presented in Table 4. This small discrepancy and the fact that the medians have practically 100% coverage is reassuring support for the values in Table 4.

5 Using the Distribution of All Absolute Sums $|d_i + d_j|$

A more direct way of estimating the distribution function of the convolution distribution is to use all pairwise absolute sums $|d_i + d_j|$ with $1 \le i < j \le n = 12,314$ and study their distribution. For $n = 12,314$ there are $N = 75,811,141$ such pairs (d_i, d_j) with $i < j$ for which to form such absolute sums. However, we point out that these values of $|d_i + d_j|$ are not independent of each other since many pairs share a common d_i. Hence it is not as though we have a random sample of size N from the $|d_1 + d_2|$ distribution.

We can estimate the cumulative distribution function $G(t) = P(|d_i + d_j| \le t)$ of this $|d_i + d_j|$ distribution using the following unbiased estimator

$$\hat{G}_n(t) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} I_{[|d_i + d_j| \le t]}$$

where $I_A = 1$ whenever the statement A is true and $I_A = 0$ otherwise.

We calculated all such absolute sums $|d_i + d_j|$ and sorted them from smallest to largest. We then plotted these values against respective tail fractions p_ℓ on a log scale. If B_ℓ is the ℓ^{th} largest of these N absolute sums $|d_i + d_j|$ then the corresponding tail fraction exceeding this ℓ^{th} largest value is roughly $(\ell - .5)/N$. When several of these absolute sums $|d_i + d_j|$ are tied at a common value one would count through them and not assign a common value of ℓ to that value. On average it will not make much difference whether one does this or uses a more refined method.

Since N is so large one can see with actual data how far out such absolute sums $|d_i + d_j|$ can be for fairly small values of $(\ell - .5)/N$. We plotted the $10,000$ most extreme values of $|d_i + d_j|$ against their respective exceedance tail fraction on a logarithmic scale as shown in Figure 17. The behavior appears to be quite smooth although not quite linear.

We fitted a quadratic to these $10,000$ points (shown as a solid curve) and show the top 50 with more emphasis. The remainder of the points are almost indistinguishable from the quadratic fit. The emphasized points are expected to flutter somewhat due to the typical tail behavior of extremes. The fitted quadratic is as follows:

$$\hat{y}_p = \alpha_0 + \alpha_1 \log(p) + \alpha_2 \log(p)^2 \ ,$$

where $\alpha_0 = 1.88710941$, $\alpha_1 = -1.07362540$, and $\alpha_2 = -0.01309460$. We also fitted a quadratic to the top $1,000$ points (shown as a dashed curve) with corresponding coefficients $\alpha_0 = 1.67135058$, $\alpha_1 = -1.09748919$, and $\alpha_2 = -0.01366266$. There does not seem to be much difference between the two fits which supports the extrapolation quality of either curve.

Here $\hat{y}_p$ should be interpreted as an estimate of that value y_p that is exceeded by $100p\%$ of all absolute sum values $|d_1 + d_2|$. Conversely, for given y one can find $\hat{p} = \hat{p}(y)$ such that $y = \hat{y}_{\hat{p}}$, i.e. inverting the fitted quadratic $\hat{y}_p$ we find the estimated proportion $\hat{p}(y)$ of absolute sum values $|d_1 + d_2|$ that exceed the given value of $y > 0$. This $\hat{p}(y)$ is given as

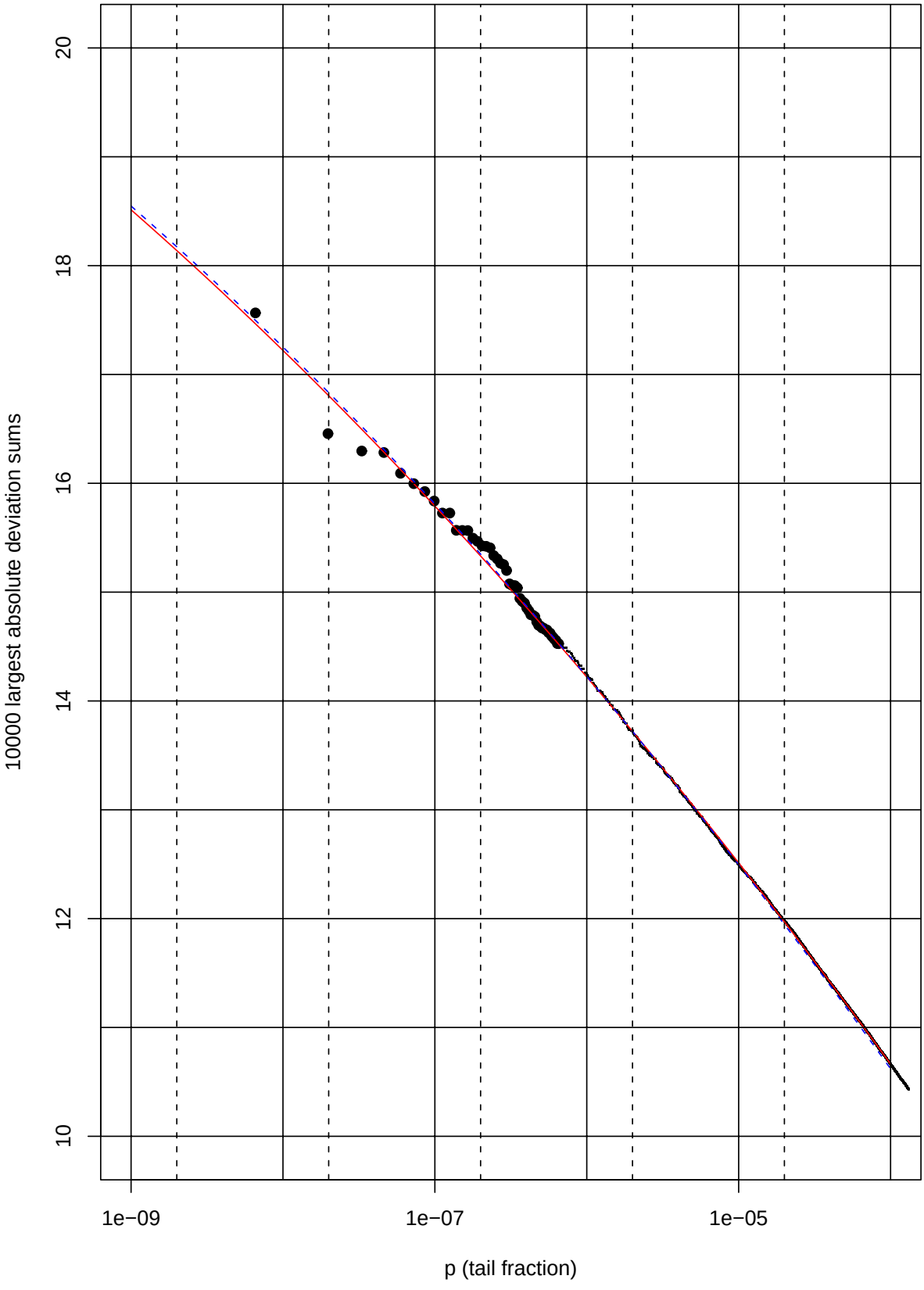
$$\hat{p}(y) = \exp\left(-\frac{\alpha_1}{2\alpha_2} + \sqrt{\frac{y - \alpha_0}{\alpha_2} + \frac{\alpha_1^2}{4\alpha_2^2}} \ \right) \ .$$

This estimate $\hat{p}(y)$ is not exactly the same as $1 - \hat{G}_n(y)$, but it comes very close, except for the very extreme values of $y > 0$. While $1 - \hat{G}_n(y)$ gives the true sample proportion of $|d_i + d_j|$ values exceeding y, the estimate $\hat{p}(y)$ is a smoothed version of it, arising from inverting the smoothed quadratic quantile estimate $\hat{y}_p$.

Given the previously mentioned symmetry property of the $d_1 + d_2$ distribution we can consider y_{2p} to be that point which is exceeded by $100p\%$ of all sum values $d_1 + d_2$. Similarly, $p_1(y) = p(y)/2$ is the proportion of sum values $d_1 + d_2$ that exceed the given value of $y > 0$. $\hat{y}_{2p}$ and $\hat{p}_1(y) = \hat{p}(y)/2$ are the corresponding estimates of y_{2p} and $p_1(y)$, respectively.

Again it is of interest to understand the uncertainty in the estimates $\hat{y}_{2p}$, i.e., how much would these estimates fluctuate when used on different samples of the same size. Appendix B details some theory pertaining to the estimate $\hat{G}_n(t)$ and its quantile inverse function and their

Figure 17: Extreme $|d_1 + d_2|$-Values Versus Their Tail Fraction



respective uncertainty assessments. Unfortunately the calculations for getting confidence bounds become quite prohibitive.

Thus we opted for the bootstrap approach which is computationally still quite demanding but more manageable. We sampled with replacement the original deviation sample of size $n = 12,314$ to get a new sample of size $n = 12,314$. From it we calculated another estimate $\hat{y}^\star_{2p,1}$ based on fitting a quadratic to the 1000 highest values of $|d_i + d_j|$. This is our first bootstrap estimate and the superscript * indicates the bootstrap nature of this estimate.

Then we repeated this bootstrap simulation another 999 times to get altogether 1000 such estimates $\hat{y}^\star_{2p,1}, \ldots, \hat{y}^\star_{2p,1000}$. Their histograms for $p = 10^{-3}, 10^{-4}, \ldots, 10^{-9}$ are shown in Figure 18-19 together with the estimates $\hat{y}_{2p}$ for the same values of p. These histograms provide a reasonable sense of the variability that affects the estimates. Also shown in these histograms are 95% upper confidence bounds for the respective target quantiles. The construction of these confidence bounds is detailed below.

There are several different bootstrap methods for constructing confidence bounds. The one employed here proceeds as follows. If the distribution of $\hat{y}_{2p} - y_{2p}$ were known one could determine its .05-quantile $x_{.05}$ with the property

$$P(\hat{y}_{2p} - y_{2p} \geq x_{.05}) = .95 \quad \text{or} \quad P(\hat{y}_{2p} - x_{.05} \geq y_{2p}) = .95 \ ,$$

i.e., $\hat{y}_{2p} - x_{.05}$ would serve as a 95% upper confidence bounds for y_{2p}.

Unfortunately we don't know the distribution of $\hat{y}_{2p} - y_{2p}$. Conceptually one could estimate $x_{.05}$ by repeatedly (say M times) obtaining a sample of size $n = 12,314$ from the true cumulative distribution function F of the deviations d_i, sorting all $N = 75,811,141$ possible absolute paired sums $|d_i + d_j|$ from this sample of size n, fitting again a quadratic to the top 1000 of such $|d_i + d_j|$ values, and from that obtain an new estimate $\hat{y}_{2p}$ and thus a new $\hat{y}_{2p} - y_{2p}$ each time. Doing this M times would yield a sample $\hat{y}_{2p,1} - y_{2p}, \ldots, \hat{y}_{2p,M} - y_{2p}$ from which the .05-quantile would serves as estimate of $x_{.05}$. This estimate would be quite accurate if M is sufficiently large, say $M = 1000$.

In this conceptual approach it is assumed that F is known, in part because we want to sample from it and also because through F we can get the distribution function G of $|d_1+d_2|$ which in turn yields the quantile y_{2p} that was subtracted in each of the terms $\hat{y}_{2p,\nu} - y_{2p}$, $\nu = 1, \ldots, M$.

Instead we use the distribution of $\hat{y}^\star_{2p} - \hat{y}_{2p}$ as proxy and determine its .05-quantile $\hat{x}_{.05}$ as an estimate of $x_{.05}$. In doing this we replace the unknown $F(y)$ distribution by the estimate $\hat{F}_n(y)$ and the unknown y_{2p} by its estimate $\hat{y}_{2p}$. Here the estimate $\hat{F}_n(y) = \{$proportion of $d_1, \ldots, d_n$ which are $\leq y\}$ is the empirical distribution function of the original deviation sample and sampling from it is the same as sampling with replacement from the original deviation sample $d_1, \ldots, d_n$.

If $\hat{y}^\star_{2p}(.05)$ denotes the .05-quantile of the 1000 bootstrap estimates $\hat{y}^\star_{2p,1}, \ldots, \hat{y}^\star_{2p,1000}$ we get

$$\hat{x}_{.05} = \hat{y}^\star_{2p}(.05) - \hat{y}_{2p} \qquad \text{and thus} \qquad \hat{y}_{2p} - \hat{x}_{.05} = \hat{y}_{2p} - [\hat{y}^\star_{2p}(.05) - \hat{y}_{2p}]$$

as 95% upper confidence bounds for y_{2p}.

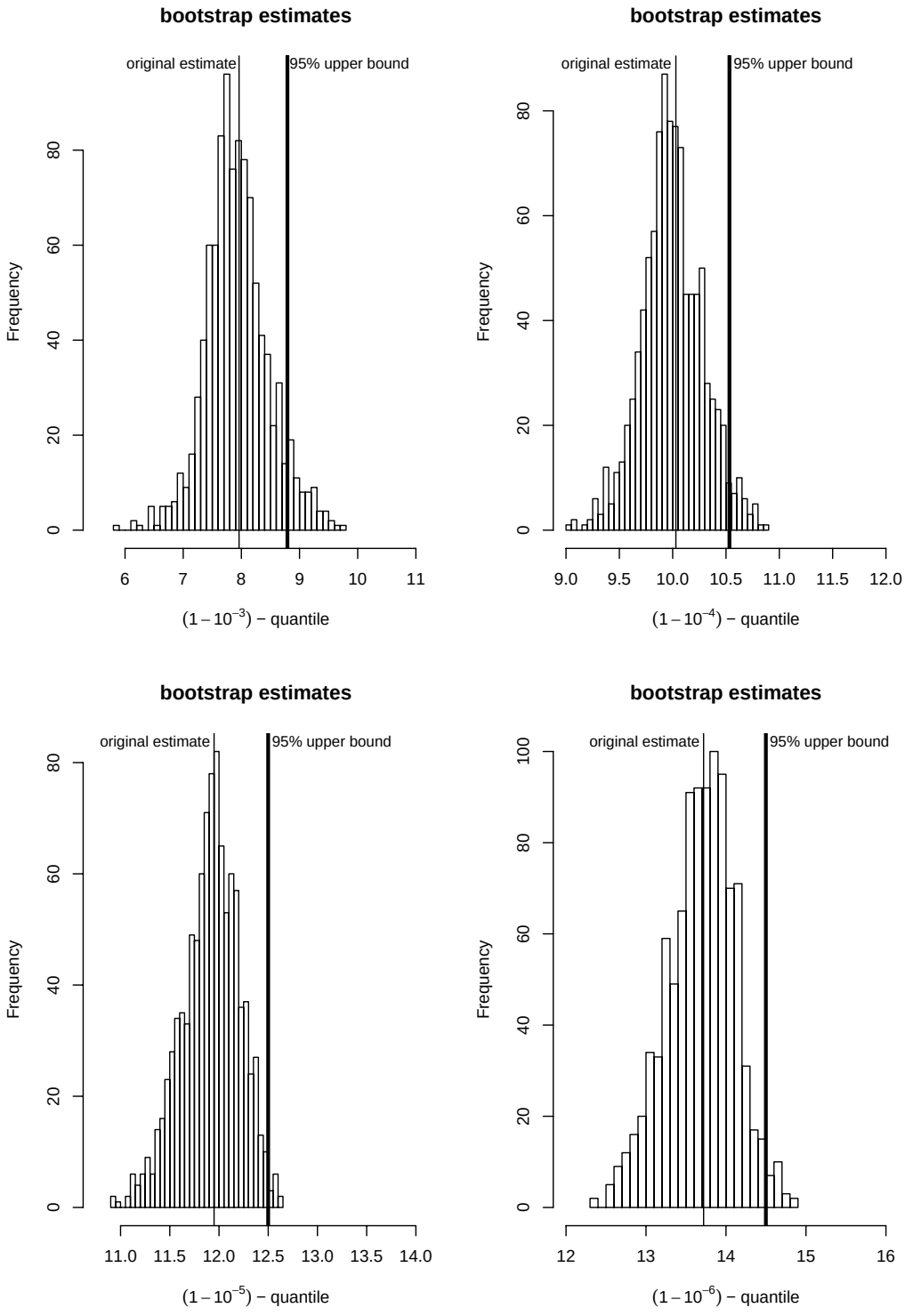


Figure 18: Bootstrap 95% Upper Confidence Bounds for $(1-p)$-Quantiles

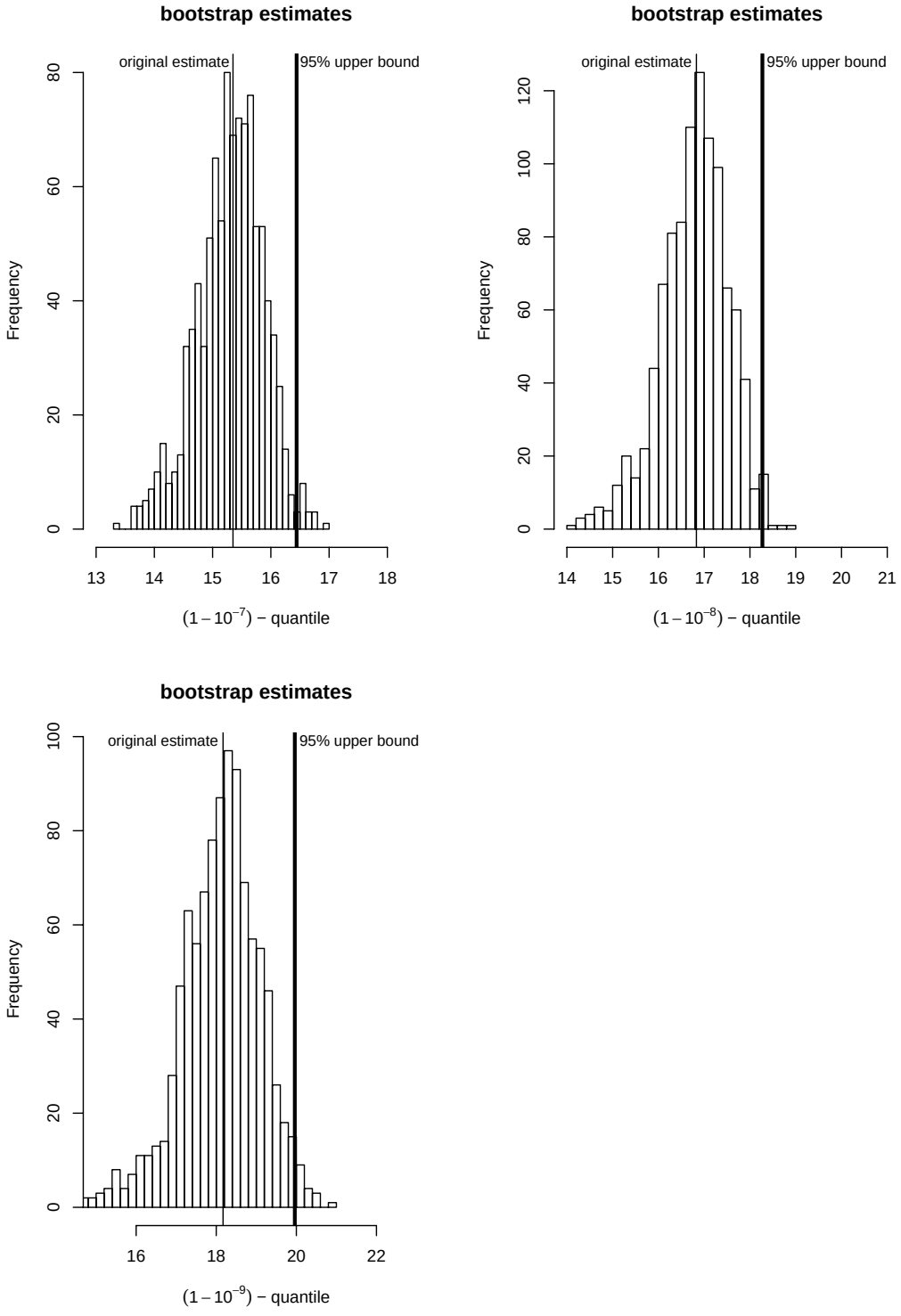


Figure 19: Bootstrap 95% Upper Confidence Bounds for $(1-p)$-Quantiles

It is these 95% upper confidence bounds that are shown in Figure 18-19 together with the original estimates as they are given in Table 5 for the one-sided risks $p = 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}$.

Table 5: Estimates and 95% Upper Bounds for the $(1-p)$-Quantiles of $d_1 + d_2$
Based on the Quadratic Fit to 1000 Largest $|d_i + d_j|$-Values Versus log(Tail Fraction)

	one-sided exceedance risk p						
	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
estimate	7.96 ft	10.03 ft	11.95 ft	13.72 ft	15.35 ft	16.83 ft	18.17 ft
95% upper bounds	8.79 ft	10.53 ft	12.50 ft	14.50 ft	16.44 ft	18.27 ft	19.96 ft

When comparing the estimates and confidence bounds in Tables 4 and 5 we find remarkable closeness in the respective confidence bounds, differing from each other by only .04 ft – .88 ft, while the respective estimates show somewhat larger discrepancies of .06 ft – 1.67 ft. We thus decided conservatively to proceed with the results from Table 4 for any further risk assessments.

6 Separation Thresholds for Parallel Taxiways

From the clearance criterion (1) it is evident that the centerline separation T of parallel taxiways should exceed the $(1-p)$-quantile of $(W_1 + W_2)/2 + d_1 + d_2$ or T should exceed $(W_1 + W_2)/2 + \delta_{1-p}$, where δ_{1-p} is the $(1-p)$-quantile of the $d_1 + d_2$ distribution. Instead of the unknown δ_{1-p} we use its estimate or conservatively a 95% upper confidence bound for δ_{1-p}, namely $\hat{\delta}_{1-p}$ or $\hat{\delta}_{U,1-p}$, respectively. To avoid potential collison at the indicated risk p we would then require

$$\frac{W_1 + W_2}{2} + \hat{\delta}_{1-p} < T \qquad \text{or} \qquad \frac{W_1 + W_2}{2} + \hat{\delta}_{U,1-p} < T \; .$$

Recall that we removed certain biases from the deviation data before proceeding with the extreme value analyses. This bias maximally was about .75 ft and its contribution to $d_1 + d_2$ is thus maximally 1.5 ft. To allow for this bias correction we conservatively add it back in when considering the entries to Table 4. Thus the back-corrected estimate of the $(1 - 10^{-6})$-quantile for $d_1 + d_2$ would be $14.35 + 1.5 = 15.85$ ft. Combining this with an average wingspan $(W_1 + W_2)/2 = 220$ ft we arrive at an estimated separation threshold value of $235.85 \approx 235.9$ ft, which is the estimate entry in the 10^{-6} column off to the right of the average wingspan of 220 ft in Table 6. The remaining values of the table are calculated analogously.

Note that Table 6 indicates that the separation thresholds apply to ANC. The reason for this is that we had to divide the adjusted JFK deviations by 1.097 before we were able to pool them with the adjusted ANC deviations. To make our quantile estimates and confidence bounds applicable to the JFK situation we have to multiply them by 1.097 before we add in the bias correction of 1.5 ft and adding that to the average wingspan. For example, the back-corrected estimate of the $(1-10^{-6})$-quantile for d_1+d_2 would be $14.35 \times 1.097 + 1.5 = 17.24 \approx 17.2$ ft which then gets added to the average wingspan, say 220 ft, to arrive at the estimate 237.2 ft, the entry in the 10^{-6} column to the right of 220 ft in Table 7. The remaining values of the table are calculated correspondingly. Note that the rows in these tables just differ by the average wingspan increment (10 ft) from row to row and they are given just for convenience of easier lookup.

The differences between corresponding estimates from Table 6 and 7 range from $1.3-1.9$ ft, while for the confidence bounds they range from $1.4-1.9$ ft.

Table 8 shows the corresponding results obtained from the Schiphol data as presented in [1]. The estimated values in Table 5 are approximately $.8-1.6$ ft higher than the values at JFK as given in Table 7, while the confidence bounds are $2.6-3$ ft higher.

Given the differences between the JFK and ANC separation thresholds one should not be surprised by further differences between the results from JFK and Schiphol. Whether one should look for explanations other than that we deal with different airports is open to debate.

We offer a few possibilities that might explain differences between Schiphol on one side and JFK and ANC on the other:

1. The Schiphol data set was smaller, $n = 7,855$, while the ANC-JFK data set consisted of $n = 12,314$ deviations.

2. Looking at the normal probability plot of the Schiphol data it is apparent that the data don't look symmetric around zero or any other value. It is not clear why this is the case. For the ANC and JFK data we saw symmetry around some bias offset that was explained by parallax effects and avoidance of taxiway centerlights. The Schiphol study did not account for any biases even though a bias (a non-zero mean deviation) was recognized.

3. For the extreme value analysis in the Schiphol study the absolute values were used without bias correction, i.e., the longer tail behavior on one side was mixed with the shorter tail behavior on the other side. In a sense one "averages" the extreme value behavior from both sample tails. It is not clear what effect that has.

4. In the Schiphol study simultaneous 90% confidence bounds were obtained while for JFK and ANC the confidence bounds were at a 95% confidence level for individual coverage. One could make a case that these two differences more or less cancel each other out.

Table 6: Required Separation T Between Taxiway Centerlines for ANC

	Estimate $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	189.4	191.6	193.8	195.9	197.8	199.7	201.4
190	199.4	201.6	203.8	205.9	207.8	209.7	211.4
200	209.4	211.6	213.8	215.9	217.8	219.7	221.4
210	219.4	221.6	223.8	225.9	227.8	229.7	231.4
220	229.4	231.6	233.8	235.9	237.8	239.7	241.4
230	239.4	241.6	243.8	245.9	247.8	249.7	251.4
240	249.4	251.6	253.8	255.9	257.8	259.7	261.4
250	259.4	261.6	263.8	265.9	267.8	269.7	271.4
260	269.4	271.6	273.8	275.9	277.8	279.7	281.4
270	279.4	281.6	283.8	285.9	287.8	289.7	291.4
280	289.4	291.8	293.8	295.9	297.8	299.7	301.4
	95% Upper Bound $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	189.4	191.7	193.9	196.1	198.1	200.1	202.1
190	199.4	201.7	203.9	206.1	208.1	210.1	212.1
200	209.4	211.7	213.9	216.1	218.1	220.1	222.1
210	219.4	221.7	223.9	226.1	228.1	230.1	232.1
220	229.4	231.7	233.9	236.1	238.1	240.1	242.1
230	239.4	241.7	243.9	246.1	248.1	250.1	252.1
240	249.4	251.7	253.9	256.1	258.1	260.1	262.1
250	259.4	261.7	263.9	266.1	268.1	270.1	272.1
260	269.4	271.7	273.9	276.1	278.1	280.1	282.1
270	279.4	281.7	283.9	286.1	288.1	290.1	292.1
280	289.4	291.7	293.9	296.1	298.1	300.1	302.1

Table 7: Required Separation T Between Taxiway Centerlines for JFK

	Estimate $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	190.2	192.6	195.0	197.2	199.4	201.4	203.3
190	200.2	202.6	205.0	207.2	209.4	211.4	213.3
200	210.2	212.6	215.0	217.2	219.4	221.4	223.3
210	220.2	222.6	225.0	227.2	229.4	231.4	233.3
220	230.2	232.6	235.0	237.2	239.4	241.4	243.3
230	240.2	242.6	245.0	247.2	249.4	251.4	253.3
240	250.2	252.6	255.0	257.2	259.4	261.4	263.3
250	260.2	262.6	265.0	267.2	269.4	271.4	273.3
260	270.2	272.6	275.0	277.2	279.4	281.4	283.3
270	280.2	282.6	285.0	287.2	289.4	291.4	293.3
280	290.2	292.6	295.0	297.2	299.4	301.4	303.3
	95% Upper Bound $\leq T$ (ft)						
	collision risk p						
wingspan (ft)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
180	190.2	192.7	195.1	197.5	199.7	201.8	204.0
190	200.2	202.7	205.1	207.5	209.7	211.8	214.0
200	210.2	212.7	215.1	217.5	219.7	221.8	224.0
210	220.2	222.7	225.1	227.5	229.7	231.8	234.0
220	230.2	232.7	235.1	237.5	239.7	241.8	244.0
230	240.2	242.7	245.1	247.5	249.7	251.8	254.0
240	250.2	252.7	255.1	257.5	259.7	261.8	264.0
250	260.2	262.7	265.1	267.5	269.7	271.8	274.0
260	270.2	272.7	275.1	277.5	279.7	281.8	284.0
270	280.2	282.7	285.1	287.5	289.7	291.8	294.0
280	290.2	292.7	295.1	297.5	299.7	301.8	304.0

Table 8: Required Separation T Between Taxiway Centerlines for Schiphol

average wingspan	Estimate $\leq T$ (ft)				90% Upper Bound $\leq T$ (ft)			
	collision risk p				collision risk p			
	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
(ft)	T (ft)							
220	238.8	240.7	242.4	244.1	240.5	242.6	244.6	246.6
230	248.8	250.7	252.4	254.1	250.5	252.6	254.6	256.6
240	258.8	260.7	262.4	264.1	260.5	262.6	264.6	266.6
250	268.8	270.7	272.4	274.1	270.5	272.6	274.6	276.6
260	278.8	280.7	282.4	284.1	280.5	282.6	284.6	286.6
270	288.8	290.7	292.4	294.1	290.5	292.6	294.6	296.6
280	298.8	300.7	302.4	304.1	300.5	302.6	304.6	306.6

5. The confidence bounds in the Schiphol study were obtained as a mean of all the confidence bounds under 500 different random splits. As indicated before (when taking the median instead of the mean) this is bound to produce a much higher confidence level than intended.

6. The different design requirements followed by the United States and those followed by the Netherlands may also be a contributing factor. For example, the FAA requires taxiway edge markings in addition to taxiway centerline markings while ICAO does not require the edges marked.

Appendix A: Combining Confidence Bounds

Suppose we have n independent and identically distributed random variables $X_1, \ldots, X_N$ with $P(X_i \geq \theta) = \gamma$, where θ is an unknown parameter of interest. We may think of the X_i as upper confidence bounds for θ with confidence level γ. For simplicity assume that the distribution of X_i is continuous.

Since we will be switching between various values of γ ($\gamma = .5$ anf $\gamma = .95$) we write on occasion more explicitly $X_1(\gamma), \ldots, X_N(\gamma)$ to avoid confusion when appropriate.

We would like to find a function $Y = f(X_1, \ldots, X_N)$ such that $P(Y \geq \theta) = \gamma$. Trivially one could take $Y = X_i$ for some prespecified i, but that does not improve the confidence bound. Furthermore, it does not resolve the question of which i to choose. One should be able to do better.

If $\gamma = .5$ one may have no problem in accepting $Y = \text{median}(X_1, \ldots, X_N)$ as the new and improved estimate for θ. However, for $\gamma \neq .5$ the median of $X_1, \ldots, X_N$ does not seem to have the correct coverage.

For example, for $\gamma = .95$ it seems that such a median is closer to θ in some distributional sense but its coverage is much higher than the targeted .95. Taking the .05-quantile of $X_1, \ldots, X_N$ would seem even closer to θ. In fact, one could consider that an estimate of θ.

We may think of θ as the .05-quantile of $X_1(.95), \ldots, X_N(.95)$ and thus use an order statistics based .95 upper confidence bound, i.e., for appropriate i take the i^{th} smallest $X_{(i)}(.95)$ of the N values $X_1(.95), \ldots, X_N(.95)$. When $N = 500$ we find

$$P(X_{(34)}(.95) \geq \theta) = \sum_{j=0}^{34-1} \binom{500}{j} .05^j .95^{500-j} = .9546$$

while

$$P(X_{(33)}(.95) \geq \theta) = \sum_{j=0}^{33-1} \binom{500}{j} .05^j .95^{500-j} = .9336 \, .$$

Thus $X_{(34)}(.95)$ would be a candidate solution for our problem. However, it seems possible that such a .95 upper confidence bound could actually wind up below the estimate obtained via the median of $X_1(.5), \ldots, X_N(.5)$.

To illustrate and investigate this issue we chose the situation where we have a sample $Y_1, \ldots, Y_n$ of size $n = 20$ from an exponential distribution with mean $\theta = 1$. A $100\gamma\%$ upper bound for θ is obtained as

$$X(\gamma) = \frac{2n\bar{Y}}{\chi^2_{2n,1-\gamma}} \quad \text{where} \quad \bar{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j \, .$$

We can do this for $N = 500$ such samples of size $n = 20$ and obtain $X(\gamma)$ for $\gamma = .5$ and $\gamma = .95$, respectively, using the same $\bar{Y}$ for either case, i.e., get

$$X_1(.5), \ldots, X_{500}(.5) \quad \text{and} \quad X_1(.95), \ldots, X_{500}(.95) \, .$$

We then compute the median$(X_1(.5), \ldots, X_{500}(.5))$ and $X_{(34)}(.95)$ and plot them against each other. Repeating this $M = 1000$ times gives us the scatter plot shown in Figure 20. Note that the 500 samples of size 20 are the same behind each plot point, i.e., the same 500 averages $\bar{Y}_1, \ldots, \bar{Y}_{500}$ are used. Thus it is a bit surprising that in 4.3% of the cases we see that the combined 95% confidence upper bound falls below the combined median (based on the same 500 samples of size 20).

Just as we used a particular order statistic of $X_1(.95), \ldots, X_{500}(.95)$ to get a 95% upper bound for θ we can also use a particular order statistic of $X_1(.5), \ldots, X_{500}(.5)$ to get a 95% upper bound for θ. Since

$$P(X_{(269)}(.5) \geq \theta) = \sum_{j=0}^{269-1} \binom{500}{j} .5^j .5^{500-j} = .9511$$

$$P(X_{(268)}(.5) \geq \theta) = \sum_{j=0}^{268-1} \binom{500}{j} .5^j .5^{500-j} = .9413$$

we can use $X_{(269)}(.5)$ as 95% upper bound for θ (with actual confidence level .9511). Since median$(X_1(.5), \ldots, X_{500}(.5)) \leq X_{(268)}(.5)$ we do not have the peculiar behavior observed previously. Figure 21 illustrates the results for the same data underlying Figure 20. Not only are these alternate 95% upper confidence bounds better behaved relative to the combined medians, but they also are more tightly clustered toward the target $\theta = 1$. The coverages are about right.

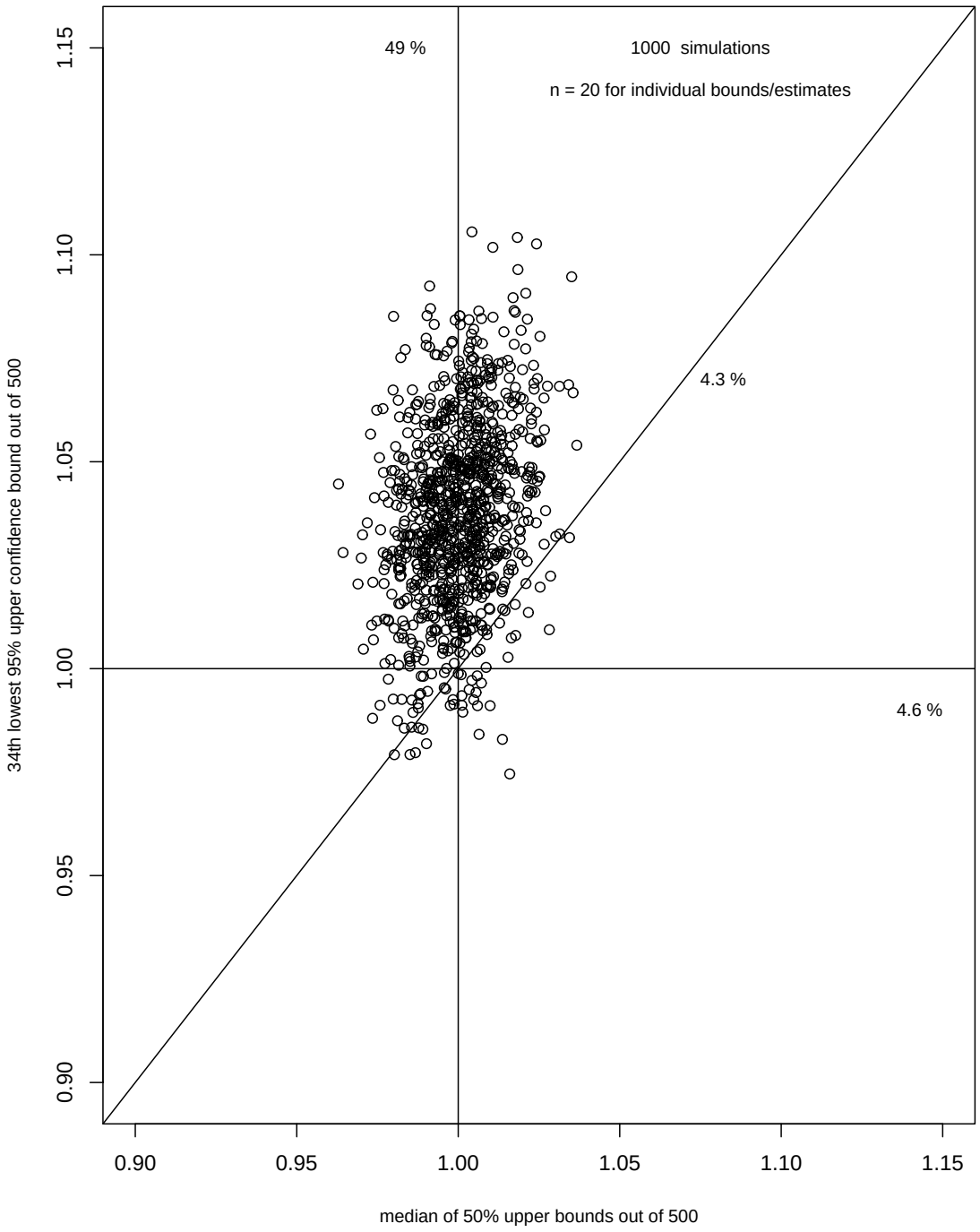


Figure 20: Comparison of Combined Estimates and 95% Confidence Bounds

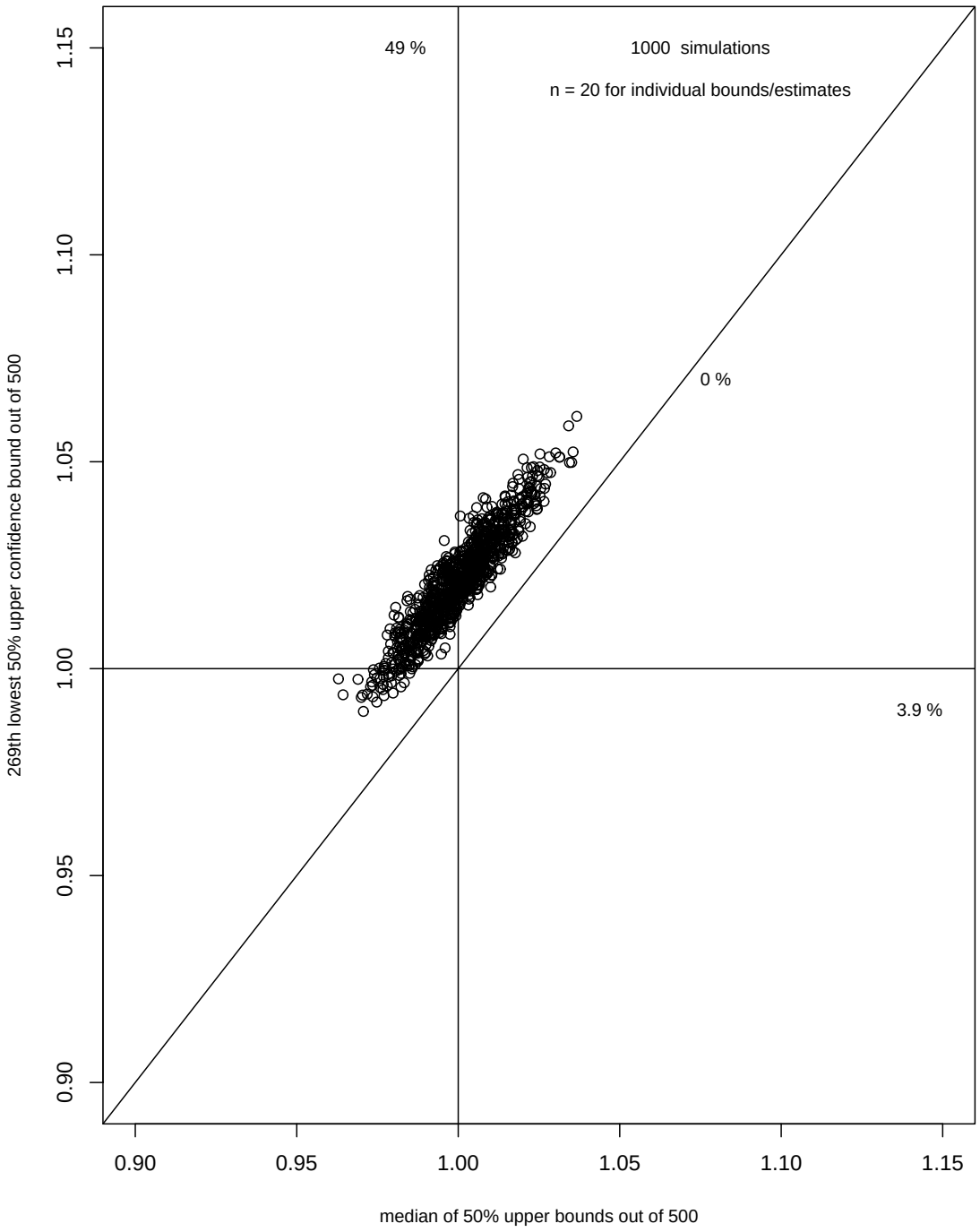


Figure 21: Comparison of Combined Estimates and Alternate 95% Confidence Bounds

Appendix B: Estimating the CDF of a Convolution

Suppose $X_1, \ldots, X_n$ is a random sample from the cumulative distribution function (cdf) F with density $f(x) = F'(x)$. We are interested in estimating the cdf of the absolute convolution of two such X's, i.e., $G(x) = P(|X_1 + X_2| \leq x)$. The natural U-statistic estimator for $G(x)$ is

$$\hat{G}_n(x) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} I_{[|X_i + X_j| \leq x]} \, .$$

This estimator is not only unbiased but has the smallest variance among all unbiased estimators of $G(x)$. While the empirical cdf $\hat{F}_{a,n}$ of the sample of absolute values $|X_1|, \ldots, |X_n|$ has atoms $1/n$ we see that $\hat{G}_n$ has atoms of much smaller size. This seems to indicate that it is easier to estimate more extreme probabilities/quantiles of G with $\hat{G}_n$ than one is able to estimate corresponding quantiles of $F_a(x) = P(|X| \leq x)$ with $\hat{F}_{a,n}$. For example, for $n = 100$ it would be difficult to estimate probabilities $F_a(x) = .999$ with $\hat{F}_{a,n}(x)$ while it would seem possible to estimate $G(x) = .999$ with $\hat{G}_n(x)$.

According to U-statistic theory [6] one has that

$$\sqrt{n} \left(\hat{G}_n(x) - G(x) \right) \longrightarrow \mathcal{N}(0, 4\zeta_1(x)) \qquad \text{as} \quad n \to \infty \, ,$$

as long as

$$\zeta_1(x) = P(|X_1 + X_2| \leq x, |X_1 + X_3| \leq x) - [P(|X_1 + X_2| \leq x)]^2 > 0 \, .$$

Here one could estimate the unknown $\zeta_1(x)$ again by an appropriate U-statistic as will be shown later. The mean and standard deviation of $\hat{G}_n(x)$ are

$$G(x) \quad \text{and} \quad \frac{2\sqrt{\zeta_1(x)}}{\sqrt{n}} \qquad \text{respectively} \, .$$

To see whether there is any practical meaning behind this consider the case where $n = 100$ and $F = \Phi$ is the standard normal distribution function. Take $x = |x|_{.999} = 3.29$ so that $P(|X| \leq x) = .999$. To address the same probability with G one would need to take $y = |x|_{.999}\sqrt{2}$ to have $G(y) = .999$.

For this $y = |x|_{.999}\sqrt{2}$ the standard deviation of $\hat{G}_n(x)$ is

$$\frac{2\sqrt{\zeta_1(y)}}{n} = \frac{2\sqrt{\zeta_1(3.29\sqrt{2})}}{10} = \frac{2\sqrt{4.104 \cdot 10^{-5}}}{10} = 0.00128 \, .$$

This standard deviation, when added to the target value of .999, could easily push us beyond 1, i.e., the uncertainty of $\hat{G}_n(x)$ would be quite large. However, one may try to finesse this issue by working with the asymptotic normality of the log-odds-ratio transformed version

Table 9: Simulated Estimates of $\hat{G}_n(x)$

0.993737	0.995152	0.995758	0.99596	0.996566
0.99697	0.99697	0.997172	0.997172	0.997172
0.997576	0.997576	0.997576	0.997778	0.99798
0.99798	0.998182	0.998182	0.998182	0.998182
0.998384	0.998384	0.998586	0.998586	0.998586
0.998788	0.998788	0.998788	0.998788	0.998788
0.998788	0.99899	0.99899	0.99899	0.99899
0.99899	0.999192	0.999192	0.999192	0.999394
0.999394	0.999394	0.999394	0.999596	0.999596
0.999596	0.999596	0.999596	0.999596	0.999596
0.999596	0.999596	0.999596	0.999596	0.999596
0.999798	0.999798	0.999798	0.999798	0.999798
0.999798	0.999798	0.999798	0.999798	0.999798
0.999798	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

of $\hat{G}_n(x)$, i.e., use $\log[\hat{G}_n(x)/(1-\hat{G}_n(x))]$. The above limiting normality for $\hat{G}_n(x)$ would transform to

$$\sqrt{n}\left(\log[\hat{G}_n(x)/(1-\hat{G}_n(x))] - \log[G(x)/(1-G(x))]\right) \longrightarrow \mathcal{N}\left(0, \frac{4\zeta_1(x)}{[G(x)(1-G(x))]^2}\right) .$$

To investigate the feasibility of such a finessed approach we simulated 100 samples of size $n = 100$ each and computed $\hat{G}_n(x)$ for $x = 3.29\sqrt{2}$, so that $G(x) = .999$. The sorted 100 values of $\hat{G}_n(x)$ are shown in Table 9 and from the large number of degenerate $\hat{G}_n(x) = 1$ values it becomes clear that taking the log-odds-ratio transformation would not be a consistent option, since $1 - \hat{G}_n(x) = 0$ would lead to division by zero in the odds ratio.

It should be said though that in contrast to this the estimate $\hat{F}_{a,n}(3.29)$ (with $F_a(3.29) = .999$) would have resulted most likely in $\hat{F}_{a,n}(3.29) = 1$ each time.

Estimating the Quantiles of a Convolution

The p-quantile x_p of the convolution cdf $G(x)$ is defined as

$$x_p = \inf \{x : G(x) \geq p\} \ .$$

Since F has density f it follows that G has density g with

$$g(x) = \int_{-\infty}^{\infty} [f(x-y) + f(-x-y)] f(y)\, dy$$

and thus that G is continuous and we have $G(x_p) = p$.
A natural estimator of x_p is

$$\hat{x}_p = \inf \left\{ x : \hat{G}_n(x) \geq p \right\} \ .$$

From this definition we have

$$\hat{x}_p \leq y \quad \Longleftrightarrow \quad \hat{G}_n(y) \geq p$$

and thus (using $G(x_p) = p$)

$$\begin{aligned} P\left(\sqrt{n}\,(\hat{x}_p - x_p) \leq t\right) &= P\left(\hat{x}_p \leq x_p + \frac{t}{\sqrt{n}}\right) \\ &= P\left(\hat{G}_n\left(x_p + \frac{t}{\sqrt{n}}\right) \geq p\right) \\ &= P\left(\sqrt{n}\left[\hat{G}_n\left(x_p + \frac{t}{\sqrt{n}}\right) - G(x_p)\right] \geq 0\right) \ . \end{aligned}$$

Writing

$$\begin{aligned} &\sqrt{n}\left[\hat{G}_n\left(x_p + \frac{t}{\sqrt{n}}\right) - G(x_p)\right] \\ &\qquad = \sqrt{n}\left[\hat{G}_n\left(x_p + \frac{t}{\sqrt{n}}\right) - G\left(x_p + \frac{t}{\sqrt{n}}\right)\right] + \sqrt{n}\left[G\left(x_p + \frac{t}{\sqrt{n}}\right) - G(x_p)\right] \end{aligned}$$

and since G has density g we can conclude that

$$\sqrt{n}\left[G\left(x_p + \frac{t}{\sqrt{n}}\right) - G(x_p)\right] \to g\,(x_p)\, t$$

for all t. Because of the continuity of F one can also obtain (see the arguments in [5] p. 380) that

$$\sqrt{n}\left[\hat{G}_n\left(x_p + \frac{t}{\sqrt{n}}\right) - G\left(x_p + \frac{t}{\sqrt{n}}\right)\right] \longrightarrow \mathcal{N}(0, 4\zeta_1(x_p)) \qquad \text{as} \quad n \to \infty \ ,$$

just as was previously claimed without the $t/\sqrt{n}$ term. Slutsky's theorem then implies

$$\sqrt{n}\left[\hat{G}_n\left(x_p+\frac{t}{\sqrt{n}}\right)-G(x_p)\right]\longrightarrow Y\sim\mathcal{N}(g(x_p)t,4\zeta_1(x_p))\qquad\text{as}\quad n\to\infty$$

and thus

$$P\left(\sqrt{n}\,(\hat{x}_p-x_p)\le t\right)\longrightarrow P(Y\ge 0)=\Phi\left(\frac{g(x_p)t}{2\sqrt{\zeta_1(x_p)}}\right)\ .$$

When $g(x_p)>0$ this means that

$$\sqrt{n}\,(\hat{x}_p-x_p)\longrightarrow\mathcal{N}\left(0,\frac{4\zeta_1(x_p)}{g^2(x_p)}\right)\qquad\text{as}\quad n\to\infty$$

or $\hat{x}_p$ is approximately normally distributed with mean x_p and standard deviation

$$\tau_p=\frac{2\sqrt{\zeta_1(x_p)}}{\sqrt{n}g(x_p)}\ .$$

One can estimate τ_p by estimating $\zeta_1(x_p)$, using a U-statistic estimator as follows. Using a symmetrized kernel

$$\psi(X_i,X_j,X_k,x_p)=I_{[|X_i+X_j|\le x_p,|X_i+X_k|\le x_p]}+I_{[|X_j+X_i|\le x_p,|X_j+X_k|\le x_p]}+I_{[|X_k+X_i|\le x_p,|X_k+X_j|\le x_p]}$$

we estimate $\zeta_1(x_p)$ by

$$\hat{\zeta}_1(\hat{x}_p)=\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n}\psi(X_i,X_j,X_k,\hat{x}_p)-\hat{G}_n^2(\hat{x}_p)\ .$$

As estimate $\hat{g}(\hat{x}_p)$ for $g(x_p)$ we take a local, numerical derivative of $\hat{G}_n(x)$ at $x=\hat{x}_p$.
From this one then obtains an estimate

$$\hat{\tau}_p=\frac{2\sqrt{\hat{\zeta}_1(\hat{x}_p)}}{\sqrt{n}\hat{g}(\hat{x}_p)}$$

for τ_p. As approximate $100\gamma\%$ confidence interval for x_p one can take

$$\left[\hat{x}_p+\Phi^{-1}\left(\frac{1-\gamma}{2}\right)\hat{\tau}_p\ ,\ \hat{x}_p-\Phi^{-1}\left(\frac{1-\gamma}{2}\right)\hat{\tau}_p\right]$$

with $\Phi^{-1}(q)$ denoting the q-quantile of the standard normal distribution.
For our application we have $n=12,314$ and thus $\binom{n}{2}=75,811,141$ which makes it quite formidable to calculate $\hat{G}_n(x)$ and to find $\hat{x}_p$. Furthermore, in order to estimate $\zeta(x_p)$ we would need to evaluate $\binom{n}{3}=311,128,922,664$ terms, which would be quite difficult if not impossible with our computing equipment.

References

[1] Booker, A. (1995), "Statistical Analysis of Aircraft Deviations from Taxiway Centerline," Boeing Information & Support Services.

[2] Chambers, J.M., Cleveland, W.S., Kleiner, B. and Tukey, P.A. (1983), *Graphical Methods for Data Analysis*, Wadsworth, Belmont, California.

[3] FAA Advisory Circular 150/5300-13, Airport Design (Table 4-1)

[4] ICAO ANNEX 14, Volume I, Aerodrome Design and Operations, Chapter 3, Paragraph 3.8.3.

[5] Lehmann, E.L. (1999), *Elements of Large-Sample Theory*, Springer, New York.

[6] Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.

[7] Scholz, F.W. (1995), "Nonparametric Tail Extrapolation," *ISSTECH-95-014*, ISS Technology, Boeing Information

[8] Scholz, F.W. (2003), "Statistical Extreme Value Analysis of ANC Taxiway Centerline Deviations for 747 Aircraft."
`http://www.faa.gov/arp/publications/reports/747anc.pdf`

[9] Scholz, F.W. (2003), "Statistical Extreme Value Analysis of JFK Taxiway Centerline Deviations for 747 Aircraft."
`http://www.faa.gov/arp/publications/reports/JFKReport101703.pdf`