

Guardbanding With Confidence

**David Deaver
Fluke Corporation
Everett, Washington**

ABSTRACT:

Setting test limits different than specification limits influences the risk of accepting defective units (consumer risk) and rejecting conforming units (producer risk). Much has been written about setting limits to accomplish various strategies such as maintaining low consumer risk, equalizing consumer risk and producer risk, or minimizing the cost of faulty test decisions.

INTRODUCTION:

Ideally, one would encounter only perfect products which all perform to their specifications and no adjustments or rejection of defective units would be necessary. Close to the ideal condition would be the situation where some defective units exist but all can be easily determined by comparing to an ideal standard; or a nearly ideal standard; or a fairly good standard; or ... an available standard. There is considerable debate as to how much better a standard must be than the items being tested, the test uncertainty ratio. Good metrology, a number of years ago, required that standards be at least ten times better than the products being compared to them; a test uncertainty ratio (TUR) of 10:1. Increased performance in the products being tested, coupled with an emphasis on containing costs, has resulted in a reduction of acceptable TURs to 4:1 with some arguing that 3:1 is sufficient. There is even more debate about what to do when the desired TUR cannot be met. At a minimum, the points less than a certain TUR (usually 4:1) must be noted on the calibration report. At the other extreme, a detailed statistical uncertainty analysis must be undertaken to establish the uncertainty of the calibration. Guardbanding has been used by some to provide some middle ground. However, there are many guardband strategies that become calibration lab policy with little understanding of their effect on false test decisions; false accepts, known as consumer risk and false rejects, known as producer risk.

COMPARING STRATEGIES:

In a previous paper ¹, charts and graphs were presented which showed the risk of false test decisions as a function of TUR, confidence level and a guardband factor (K) which was defined as a multiplier of the specification limit (SL) to produce the test limit (TL); that is, $TL = K * SL$. The charts assume that the distribution of expected readings from the standards and the units under test (UUTs) are reasonably normal, have symmetric specification limits, and that systematic biases have been corrected. This paper will use that data to examine and compare the risks associated with a number of strategies being used today. The strategies are compared in Table 1 for TURs of 2:1 and slightly less than 4:1. Figure 1 plots the false accept and false reject risks for each of the strategies for a $\pm 2\sigma$ confidence interval to allow the strategies to be visually compared. Figures 2-6 chart the guardband factors for the various strategies as a function of TUR for confidence intervals from $\pm 1\sigma$ to $\pm 3\sigma$.

① $CR = CR_{4:1}$

In this strategy, the test limit is set to maintain same risk of false accepts as a 4:1 TUR would produce. By implication that 4:1 is an acceptable TUR, a false accept rate (consumer risk) of 0.8% is acceptable for a UUT and STD specified at $\pm 2\sigma$ confidence intervals. This strategy picks guardband factors which maintain this same level of consumer risk for all TURs. Calculating the risks for this strategy, however, involves solving double integrals and other painful endeavors that were undertaken in the previous paper using MathCAD® ². The graphs in that paper provide a practical means to implement this strategy. It is the one recommended by the author and is the strategy to which the others will be compared. It has been presented in the literature since 1954 by Eagle ³ and Grubbs and Coon ⁴ and was recommended to the NCSL in a paper presented by Hutchinson ⁵ in 1991.

Figure 1 shows the risk of false test decisions for the various strategies. This strategy, for a $\pm 2\sigma$ confidence interval, is shown as a horizontal line at 0.8% risk of accepting defective units. The penalty of maintaining the same risk as a 4:1 TUR is that more conforming units are rejected as shown in the graph on the right of Figure 1. At a TUR of 2:1, we could expect to reject good units at about a 4% rate when tested to the specification limits. Curve 1 shows that, when the false accept rate is maintained at 0.8%, the false reject rate would be expected to increase to nearly 7%. Curve 1 on each of the left graphs in Figures 2-6 shows the value of K as a function of TUR required to implement this strategy. Each figure is plotted for a different confidence interval. Most of the literature assumes a confidence interval of $\pm 2\sigma$. However, there is considerable variation in how conservatively products are specified so curves are presented for confidence intervals from $\pm 1\sigma$ to $\pm 3\sigma$.

$$\textcircled{2} \quad K = 1 - \frac{1}{TUR}$$

When the TUR is less than 4:1, this strategy subtracts the uncertainty of the UUT from the specification limit to obtain the test limit. For instance, if the specification of a unit to be tested is 0.1%, (assumed to have a $\pm 2\sigma$ confidence interval) and a standard with an uncertainty of 0.04% (also to $\pm 2\sigma$) were available to test it, the resulting TUR would be 2.5. Test limits would be set at $\pm 0.06\%$, or 60% of the specification limits.

The principal benefit of this strategy is that is easily calculated. It suffers from being quite conservative from a false accept perspective at the expense of rejecting a high number of conforming units and a large discontinuity at the 4:1 threshold where it is invoked. With a TUR of 4, the UUT may be tested at the specification limit; however, at a TUR of 3.999, the test limit must be reduced to 75% of the specification limit. Table 1 shows the huge producer risk penalty that is expected by this strategy to offset a consumer risk that is 4 to 5 times less than that of a TUR of 4:1.

$$\textcircled{3} \quad K = 1.25 - \frac{1}{TUR}$$

When the TUR is less than 4, the test limit is set to 1.25 times the specification limit minus the uncertainty of the standard. This strategy, NCSL's Recommended Practice RP-10, improves on some of the limitations of the 2nd method. Still easily calculable, it is continuous at a 4:1 TUR and does not require near the producer risk penalty, especially for the higher TURs. K ranges from 92% to 100% of the specification limit for TURs from 3 to 4 and a $\pm 2\sigma$ confidence interval.

$$\textcircled{4} \quad K = \sqrt{1 - \frac{1}{TUR^2}}$$

In the rss strategy, the test limit is determined by taking the square root of the specification limit squared less the square of the uncertainty of the standard. Used by the Fluke Corporation and others for a number of years, this strategy has only a slight false reject penalty over the constant risk strategy. It is fairly easy to calculate and has fairly constant consumer risk for TURs of 1.5 to 4. For a confidence interval of $\pm 2\sigma$, it has a false accept risk of about 0.6% as compared to the 0.8% risk for the first strategy of maintaining the same risk as 4:1.

⑤ $CR = CR_{3:1}$

This strategy is the same as the first strategy except that guardband factors are selected to maintain the risk the same as for a TUR of 3. Curves for this alternative are provided for comparison and in anticipation that some labs will be using 3:1 as a minimum acceptable TUR.

⑥ Minimize: $CR + PR$

Ultimately, the purchaser of equipment must bear the cost of all false test decisions; both those due to consumer risk (CR) and those due to producer risk (PR), assuming that the producer plans to stay in business. This strategy selects guardband factors which minimize the total risk, the sum of CR and PR. Grubbs and Coon derived the condition necessary for this situation and showed the interesting result. To minimize the total risk of false test decisions, the test limits are set *outside* the specification limits! As can be seen from the curves in Figure 1 and Figure 4, for a confidence interval of $\pm 2\sigma$ and a TUR of nearly 4, the test limits would be set about 6% greater than the specification limit. To minimize the risk of making a false test decision for an instrument having a specification of 0.5%, the test limits would be set to $\pm 0.53\%$. This alternative is not very palatable to most metrologists, however, and the argument is immediately raised that the consequences of a false accept is generally much greater than for a false reject.

ECONOMIC CONSIDERATIONS:

Generally, when asked how much more a false reject costs than a false accept, the reply is something like, "Quite a bit more." However, if we can quantify "Quite a bit more.", we can determine the guardband factors which minimize the cost weighted risk. Grubbs and Coon derived the equation from which these K values can be determined. Converted to the nomenclature used by this paper, values of K which minimize the weighted risk are those which satisfy the condition:

$$\frac{1}{\sqrt{2\pi}} \int_{L \frac{R^2(1-K)+1}{\sqrt{R^2+1}}}^{\infty} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{L \frac{R^2(1+K)+1}{\sqrt{R^2+1}}}^{\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{F + 1}$$

Where: K is the guardband factor

R is the test uncertainty ratio (TUR)

$\pm L$ is the confidence interval expressed in number of sigmas

F is the cost factor; how much more a false accept costs than a false reject

This is a fairly formidable equation but can be evaluated quite easily using the Solve feature of MathCAD to find the values of K which satisfy the equality. This was done for cost factors of 1, 2, 5, 10, 20, 50, and 100 and the results are plotted on the right side of Figures 2-6. F=1 is, of course, the same as the 6th strategy, resulting in setting test limits outside the specification limits.

Using an economic model is probably the optimum way of picking guardband factors to minimize the consequences of false test decisions. In the author's opinion, however, it is not likely to be widely accepted because of the difficulty in coming to agreement on the relative cost of false test decisions.

CONCLUSION:

Table 1, below, provides a few points for comparison between the various guardband strategies identified by their alternative number.

Strategy		TUR = 4-			TUR = 2		
		K	CR	PR	K	CR	PR
①	CR=CR _{4:1}	1.0	0.8%	1.5%	0.91	0.8%	6.6%
②	1-1/TUR	0.75	0.02%	10%	0.5	0.03%	33%
③	1.25-1/TUR	1.0	0.8%	1.5%	0.75	0.3%	14%
④	RSS	0.97	0.6%	2%	0.86	0.63%	8.2%
⑤	CR=CR _{3:1}	-----	-----	-----	0.95	1.0%	5.4%
⑥	Min. CR+PR	1.06	1.4%	0.7%	-----	-----	-----

Table 1 Risk of False Test Decisions at TUR = 4- and TUR = 2

There are a number of papers, consultants and software packages which present this information with considerably more precision and rigor and with far fewer assumptions as to the shape of the distributions and symmetry of the test limits and specification limits. The goal of this paper, however, is to provide a straightforward, graphical means to allow a number of the more common guardband strategies to be compared on the basis of the risk of false test decisions. The data is presented to aid in evaluating current cal lab practices and to help in considering other guardband strategies. Curves are also provided for an economic model which will help to optimize the guardband factors for cases when the relative costs of false accepts and rejects can be determined. Assuming that a TUR of 4:1 represents an acceptable risk, the author would recommend the use of a constant risk strategy in selecting guardbands when TURs of 4:1 cannot be maintained to avoid an undue rate of false rejects.

REFERENCES

- [1] Deaver, David, "Maintaining Your Confidence (In a World of Declining Test Uncertainty Ratios)", 1993 NCSL Workshop & Symposium, pp. 133-153
- [2] MathCAD is a registered trademark of MathSoft, Inc. The author has no interest nor affiliation with MathSoft except that of a satisfied user of the MathCAD software program.
- [3] Eagle, Alan R., "A Method for Handling Errors in Testing and Measuring", Industrial Quality Control, March, 1954, pp. 10-15
- [4] Grubbs, Frank E. and Coons, Helen J., "On setting Test Limits Relative to Specification Limits", Industrial Quality Control, March, 1954, pp. 15-20
- [5] Hutchinson, Bill, "Setting Guardband Test Limits to Satisfy MIL-STD-455662A Requirements", 1991 NCSL Workshop & Symposium, pp. 305-309

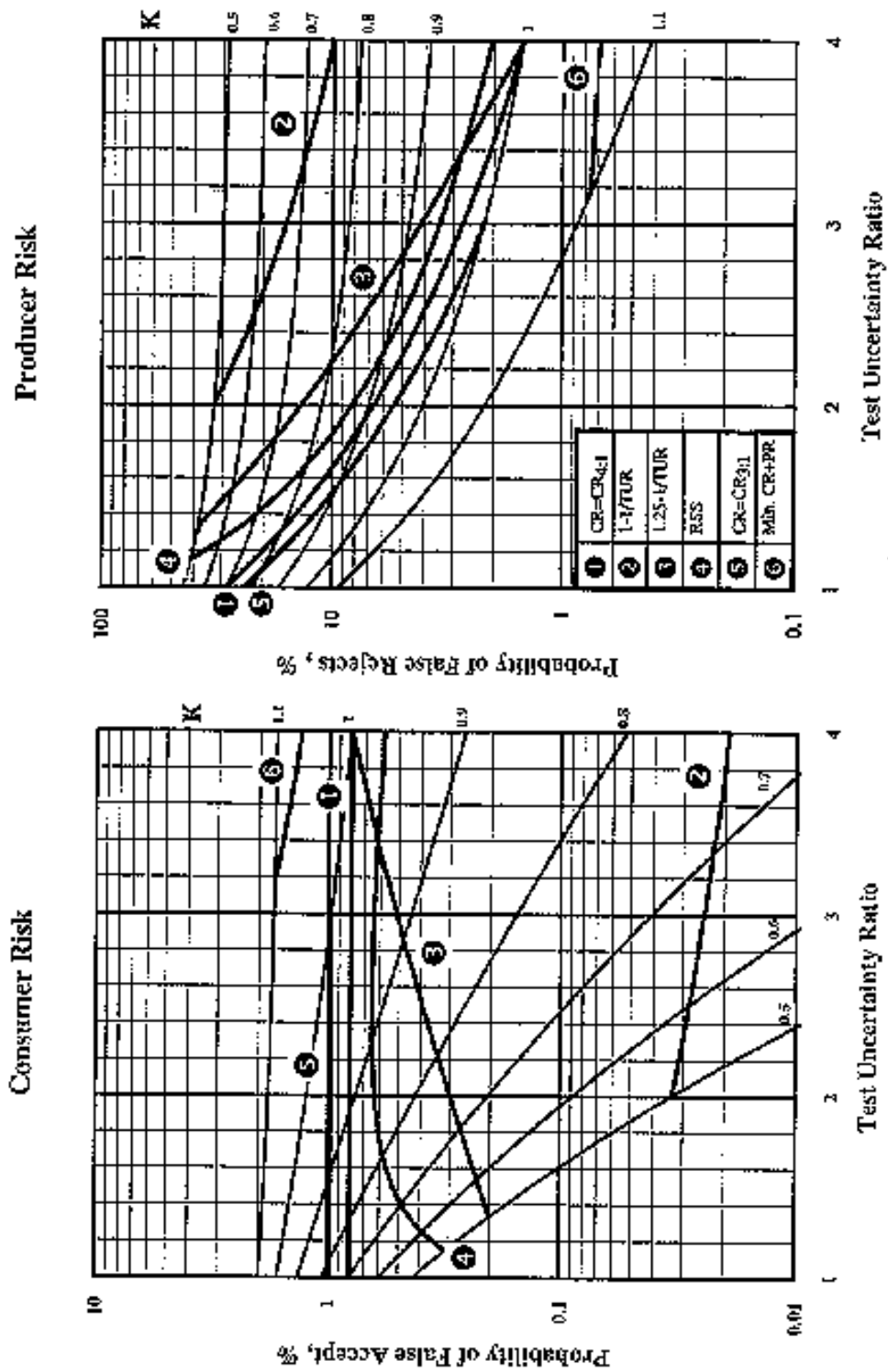


Figure 1 False Test Decisions for Six Strategies, $SL = \pm 2\sigma$

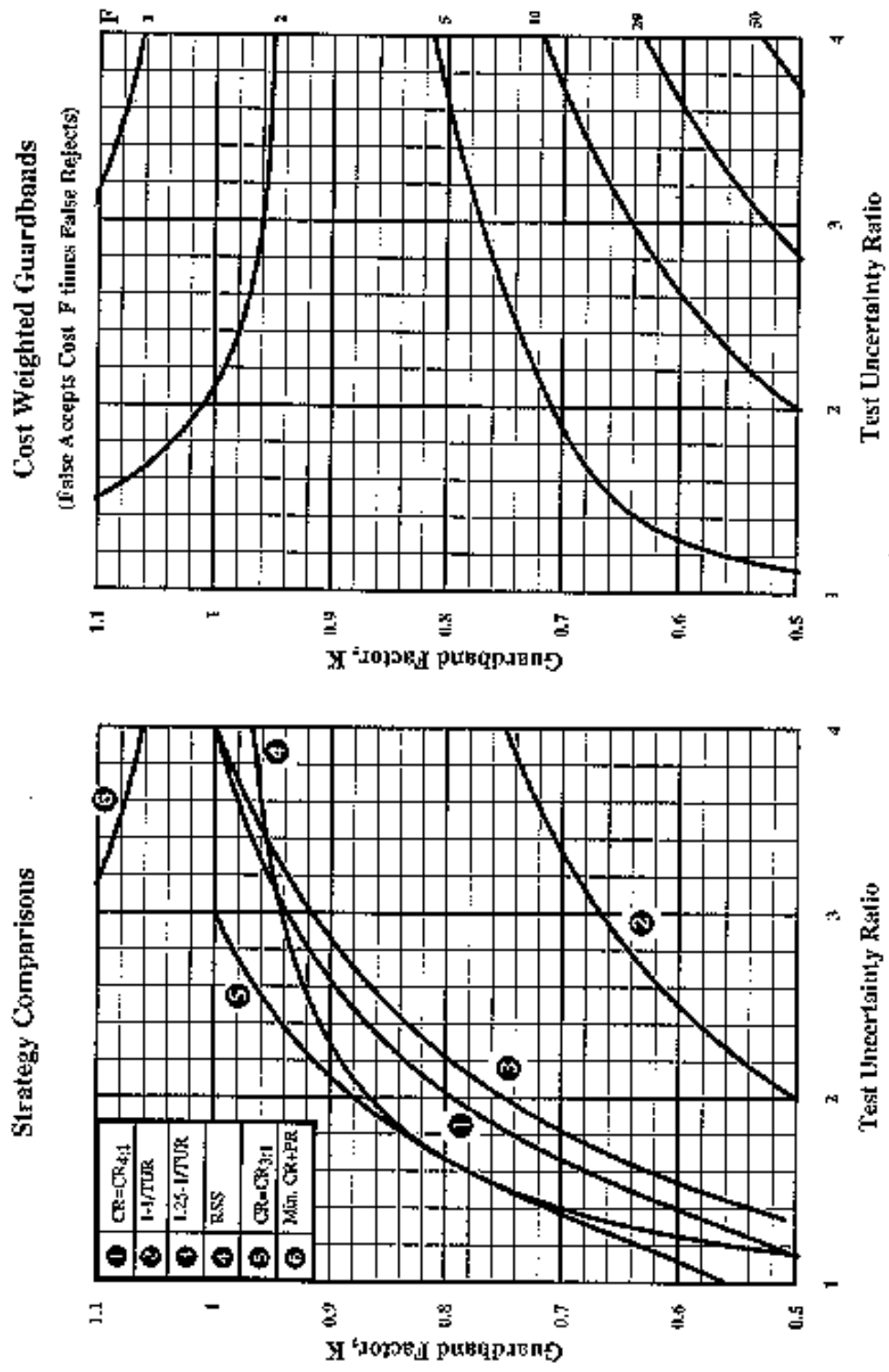
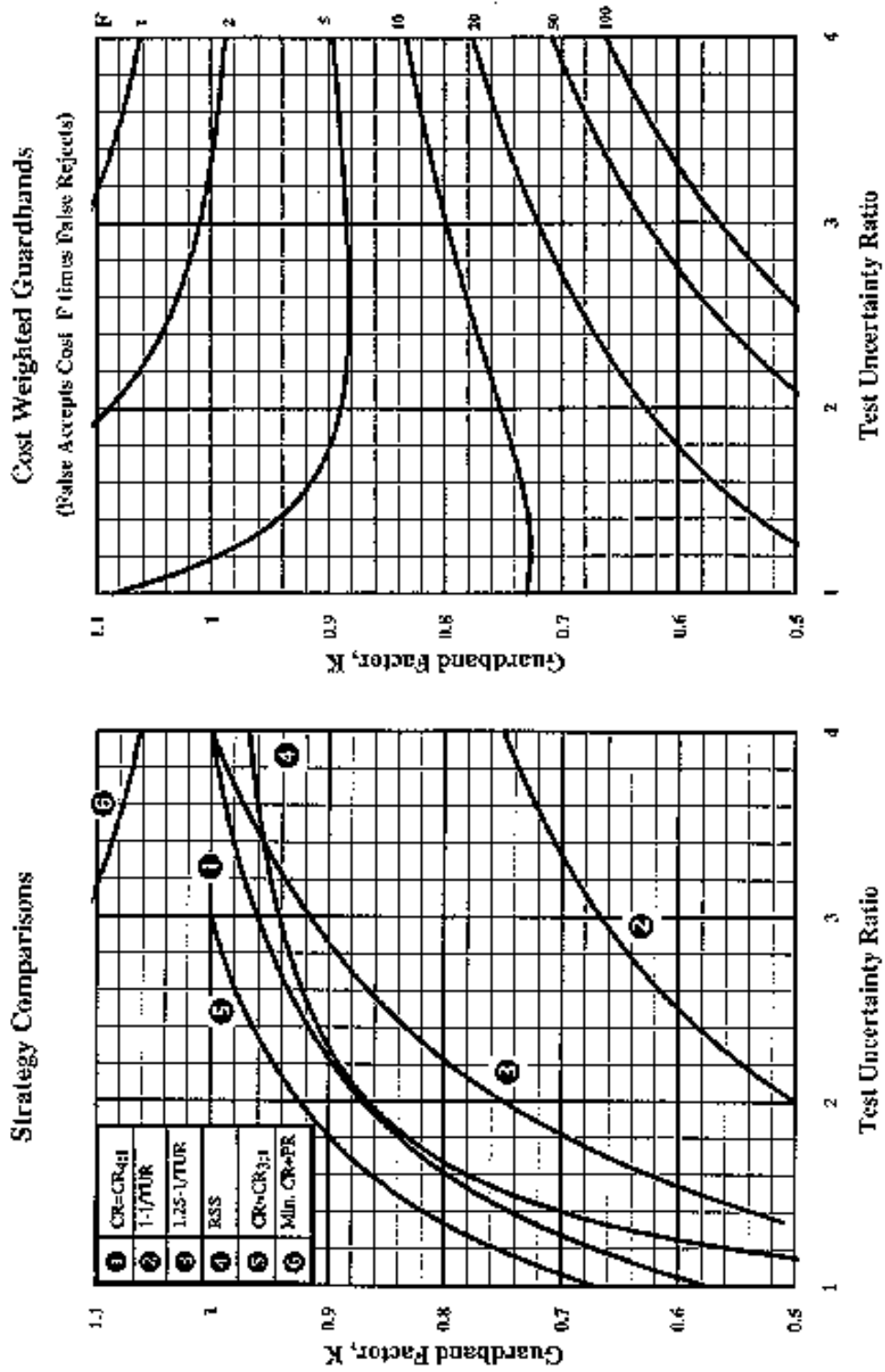


Figure 2 Guardband Locations for Six Strategies, $SL = \pm 1\sigma$



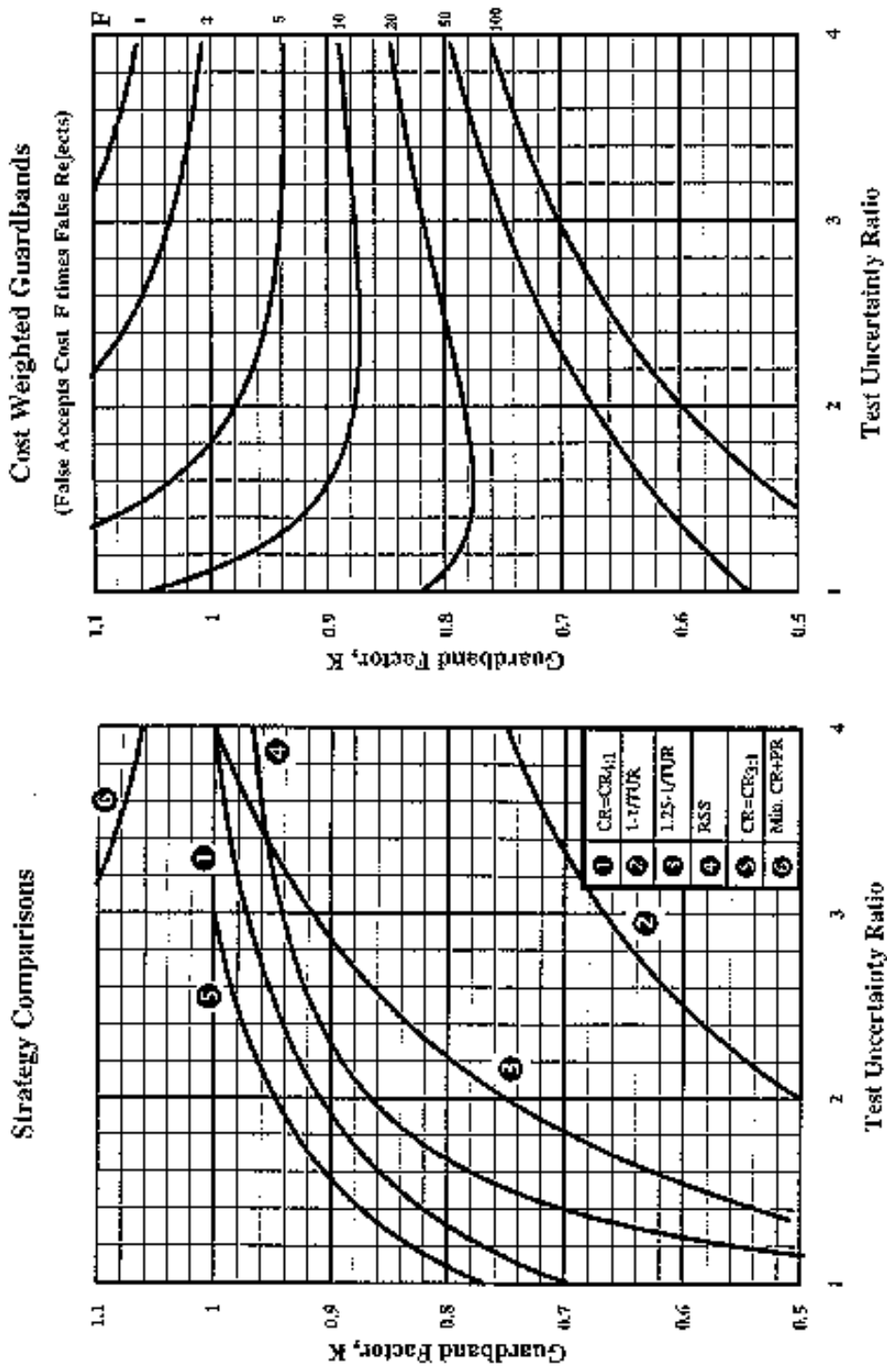


Figure 4 Guardband Locations for Six Strategies, $SL = \pm 2\sigma$

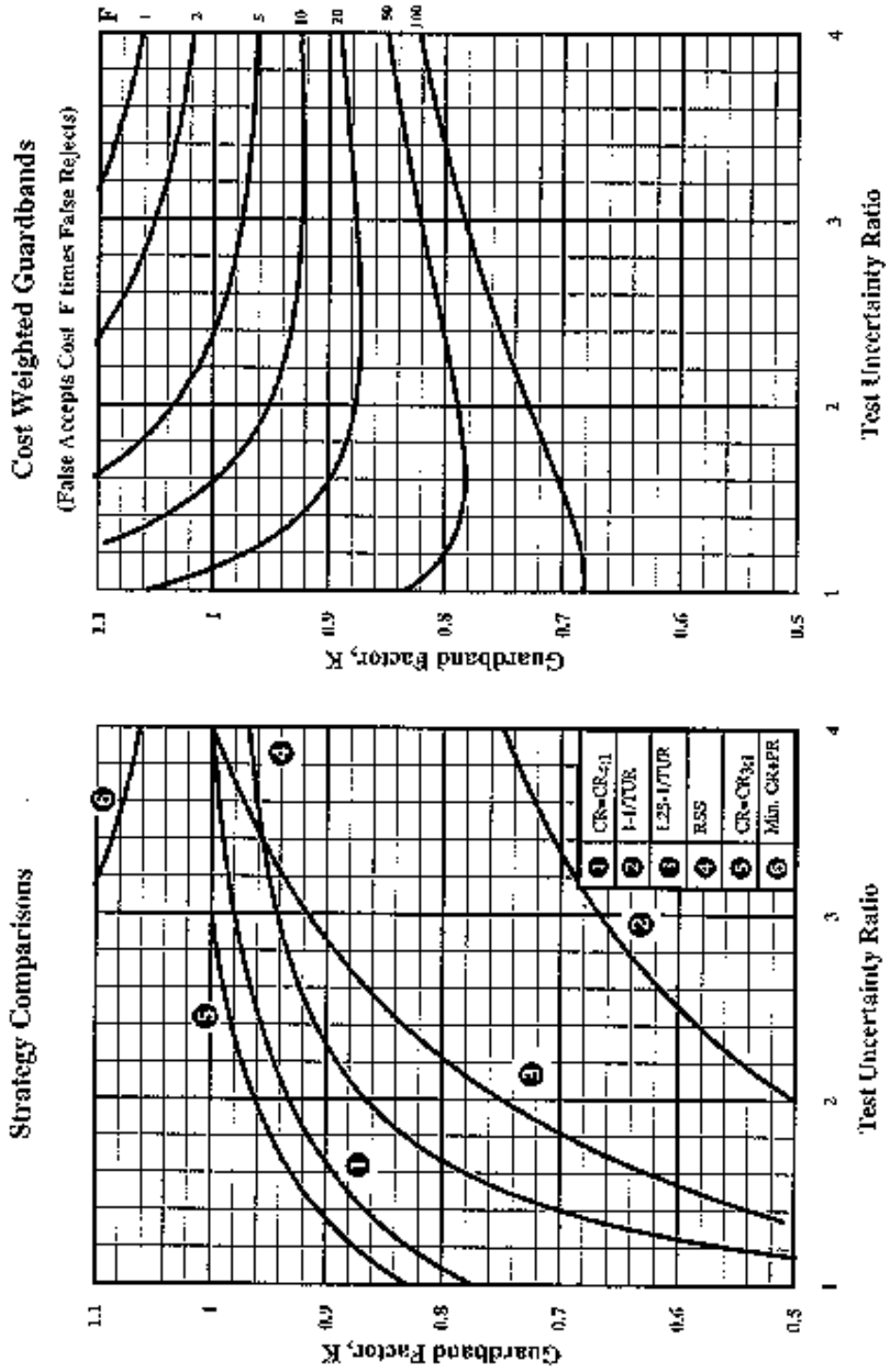


Figure 5 Guardband Locations for Six Strategies, SL = ±2.5σ

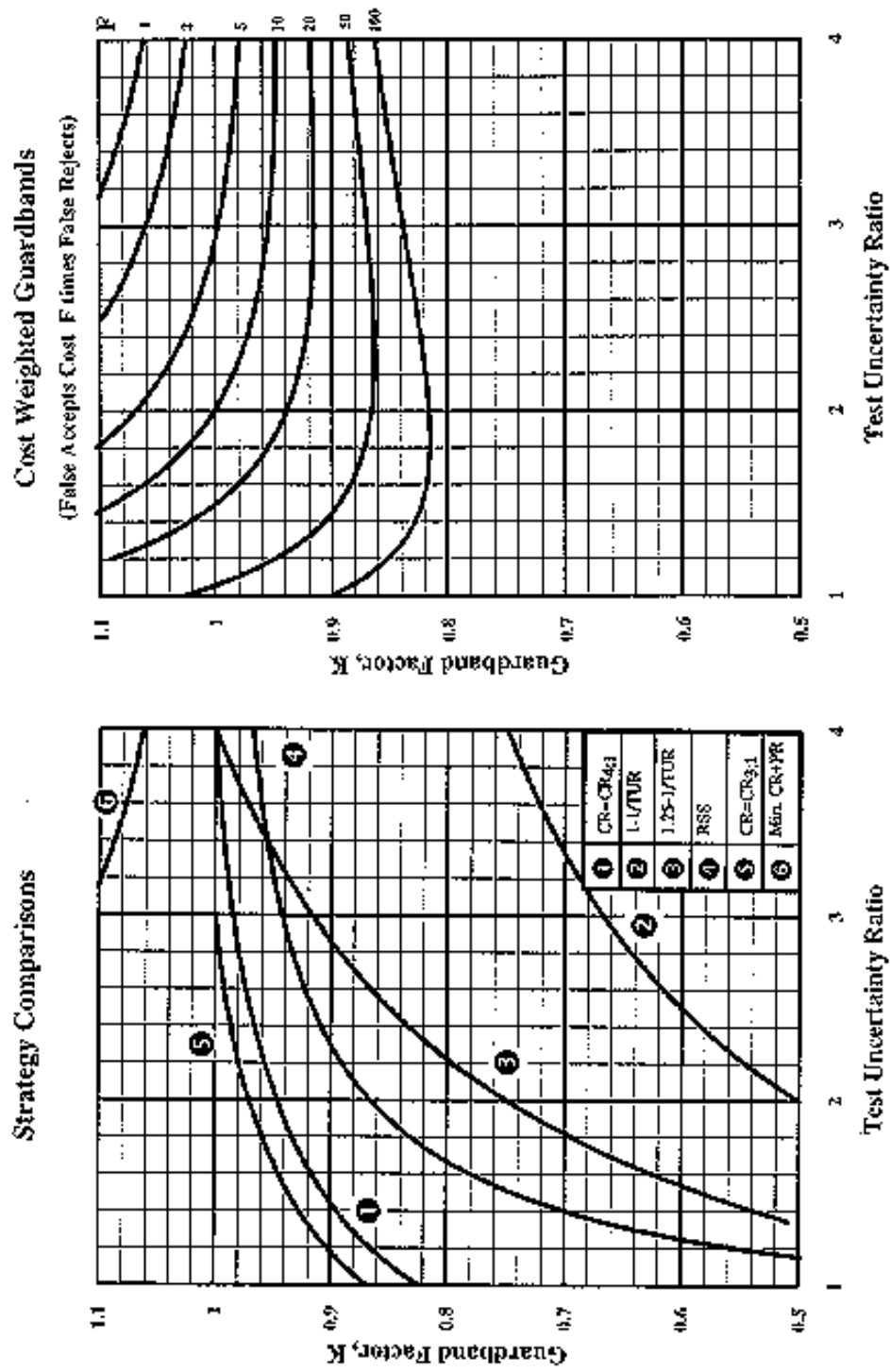


Figure 6 Guardband Locations for Six Strategies, $SL = \pm 3\sigma$