

NUREG/CR-2483  
PNL-4085

---

---

# Evaluation of Simultaneous Testing Procedures for Nuclear Materials Control and Accounting

---

---

Prepared by W. M. Bowen

**Pacific Northwest Laboratory**  
Operated by  
Battelle Memorial Institute

Prepared for  
U.S. Nuclear Regulatory  
Commission



NOTICE

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights.

Available from

GPO Sales Program  
Division of Technical Information and Document Control  
U. S. Nuclear Regulatory Commission  
Washington, D. C. 20555

Printed copy price: \$5.50

and

National Technical Information Service  
Springfield, Virginia 22161

---

---

# Evaluation of Simultaneous Testing Procedures for Nuclear Materials Control and Accounting

---

---

Manuscript Completed: December 1981  
Date Published: March 1982

Prepared by  
W. M. Bowen

Pacific Northwest Laboratory  
Richland, WA 99352

**Prepared for**  
**Division of Facility Operations**  
**Office of Nuclear Regulatory Research**  
**U.S. Nuclear Regulatory Commission**  
**Washington, D.C. 20555**  
**NRC FIN B0208**



### ABSTRACT

This report presents the results of a comparative evaluation of four statistical testing procedures for use in the control and accounting of special nuclear materials. Of primary interest is a bivariate procedure that simultaneously tests ID and CID. Descriptions of the four testing procedures are presented with the necessary formulas and special considerations for their implementation. Results of a simulation study indicate the conditions under which each of the tests would provide superior protection against "trickle" diversions.



## EXECUTIVE SUMMARY

This report presents a comparative evaluation of four statistical testing procedures for use in the control and accounting of special nuclear materials. Of primary interest is a bivariate procedure which, for a given material balance period, simultaneously tests whether the inventory difference (ID) and the cumulative inventory difference (CID) differ from zero. This procedure is thought to be more powerful for detecting small continual diversions ("trickle" diversions) than the currently required test of ID only. Three other procedures are included in the study to provide an objective basis for evaluating the bivariate procedure. The four testing procedures are:

1. the bivariate test of ID and CID
2. the currently required test of ID only
3. a test of CID only
4. a simultaneous test of ID and CID using the Bonferroni inequality.

Descriptions of the four methods are presented with the necessary formulas for their implementation and illustrative numerical examples.

Special inventory record-keeping methods required to accommodate the calculation of the variance of CID are described. Other considerations relevant to the calculation of the variance of CID and the covariance of ID and CID are discussed.

A large simulation study was conducted, where sequences of ID's were generated under a number of hypothetical operating conditions. These conditions included specific diversion strategies, correlations among successive ID's, and actions taken when a significant loss is detected. Each sequence of ID's was monitored by each of the four testing procedures, and summary statistics were accumulated. The method of simulation is described, and the rationale for choosing input parameters is discussed. Criteria for comparing the four testing procedures are defined, and the results of the simulation study are presented.

Based on the results of this simulation study, the following conclusions are drawn:

1. When correlations among successive ID's are predominantly negative, the bivariate procedure (and, in fact, either of the other procedures involving CID) is considerably more powerful for detecting "trickle" diversions than the currently required test of ID only.
2. When correlations among successive ID's are predominantly positive, the currently required test is noticeably more powerful for detecting trickle diversions than the bivariate test.
3. In all cases, the currently required test has a substantially higher false alarm rate than the bivariate test.

The assumption of positive correlations among successive ID's is probably unrealistic in nuclear materials accounting. It is included here for the purpose of comparison and to illustrate conditions under which the current requirement is more sensitive to trickle losses than the bivariate procedure. It is assumed by many that successive ID's are always negatively correlated. If this is true, it would be advantageous to implement the bivariate procedure. However, since the power of the bivariate test for detecting "block" losses has not been established, the bivariate test would probably supplement the current requirement rather than replace it. Also, because a more sophisticated inventory record-keeping system is required, the bivariate procedure should only be implemented by facilities with access to a computer.



CONTENTS

ABSTRACT . . . . .	iii
EXECUTIVE SUMMARY. . . . .	v
LIST OF FIGURES. . . . .	viii
LIST OF TABLES . . . . .	ix
1.0 BACKGROUND AND OBJECTIVES . . . . .	1
2.0 BASIC CONCEPTS AND DEFINITIONS. . . . .	3
2.1 DISCUSSION OF ID AND CID . . . . .	3
2.2 MATRIX NOTATION. . . . .	5
3.0 FOUR TESTING PROCEDURES . . . . .	9
3.1 THE BIVARIATE PROCEDURE. . . . .	10
3.2 THE TEST OF ID ONLY. . . . .	14
3.3 THE TEST OF CID ONLY . . . . .	15
3.4 THE "BONFERRONI" PROCEDURE . . . . .	17
3.5 FURTHER DISCUSSION OF THE BONFERRONI AND BIVARIATE PROCEDURES . . . . .	20
4.0 ESTIMATION OF VARIANCES AND COVARIANCES . . . . .	23
4.1 GENERAL STRATEGY FOR COMPUTING $VAR(ID_k)$ . . . . .	23
4.2 DIFFICULTIES WITH CALCULATING $VAR(CID_k)$ . . . . . AND $COV(ID_k, CID_k)$	26
5.0 THE SIMULATION STUDY. . . . .	29
5.1 METHOD OF GENERATING A SEQUENCE OF ID'S. . . . .	30
5.2 CHOICE OF INPUT PARAMETERS $\mu_{10}$ AND $\Sigma_{10}$ . . . . .	32
5.2.1 Discussion of Diversion Strategies. . . . .	32
5.2.2 Selection of Variance-Covariance Matrices . . . . .	35
5.3 DECISION RULES AND CRITERIA FOR COMPARISONS. . . . .	41
5.4 RESULTS OF THE SIMULATION. . . . .	42
6.0 CONCLUSIONS . . . . .	63
REFERENCES. . . . .	65
APPENDIX A. SIMULATION COMPUTER CODE WITH USER INFORMATION. . . . .	A.1

## LIST OF FIGURES

<u>NO.</u>		<u>PAGE</u>
5.1	FIRST TIME DETECTION PROBABILITIES FOR MATRIX I <sub>10</sub> , $\rho = 0$	45
5.2	EXPECTED UNDETECTED LOSSES FOR MATRIX I <sub>10</sub> , $\rho = 0$	46
5.3	FIRST TIME DETECTION PROBABILITIES FOR MATRIX 1, $\rho = -.3$	47
5.4	EXPECTED UNDETECTED LOSSES FOR MATRIX 1, $\rho = -.3$	48
5.5	FIRST TIME DETECTION PROBABILITIES FOR MATRIX 2, $\rho = -.3$	49
5.6	EXPECTED UNDETECTED LOSSES FOR MATRIX 2, $\rho = -.3$	50
5.7	FIRST TIME DETECTION PROBABILITIES FOR MATRIX 1, $\rho = -.5$	51
5.8	EXPECTED UNDETECTED LOSSES FOR MATRIX 1, $\rho = -.5$	52
5.9	FIRST TIME DETECTION PROBABILITIES FOR MATRIX 2, $\rho = -.5$	53
5.10	EXPECTED UNDETECTED LOSSES FOR MATRIX 2, $\rho = -.5$	54
5.11	FIRST TIME DETECTION PROBABILITIES FOR MATRIX 1, $\rho = .3$	55
5.12	EXPECTED UNDETECTED LOSSES FOR MATRIX 1, $\rho = .3$	56
5.13	FIRST TIME DETECTION PROBABILITIES FOR MATRIX 2, $\rho = .3$	57
5.14	EXPECTED UNDETECTED LOSSES FOR MATRIX 2, $\rho = .3$	58
5.15	FIRST TIME DETECTION PROBABILITIES FOR MATRIX 1, $\rho = .5$	59
5.16	EXPECTED UNDETECTED LOSSES FOR MATRIX 1, $\rho = .5$	60
5.17	FIRST TIME DETECTION PROBABILITIES FOR MATRIX 2, $\rho = .5$	61
5.18	EXPECTED UNDETECTED LOSSES FOR MATRIX 2, $\rho = .5$	62

## LIST OF TABLES

<u>NO.</u>		<u>PAGE</u>
3.1	Upper $(1-\alpha)100^{\text{th}}$ Percentile Points of the Chi-Square Distribution With Two Degrees of Freedom	11
3.2	Upper $(1-\alpha/2)100^{\text{th}}$ Percentile Point of the Standard Normal Distribution	15
4.1	Computation of Net Inventory Quantities for Variance Calculations	24
A.1	Input Data For Example	A.6
A.2	Summary Statistics for First-Time Detections	A.27



## 1.0 BACKGROUND AND OBJECTIVES

Facilities licensed by the U. S. Nuclear Regulatory Commission (NRC) to handle special nuclear materials (SNM) are currently required to perform a statistical test at the end of each material balance period. The test results indicate whether differences between book and physical inventories of SNM are too large to have happened by chance due to variability in the physical inventory measurement system. Briefly speaking, the statistical test compares the total inventory difference (ID) for a particular material balance period with its limit of error (LEID), which is an appropriate multiple of the standard deviation of ID. If the ID quantity is outside the interval (-LEID, LEID), the conclusion is that the ID differs from zero by a statistically significant amount. The intent of this requirement is to enable licensees to promptly detect missing quantities of SNM and take appropriate action to resolve the cause of such discrepancies.

References (1) and (2) point out that although this required test of ID is satisfactory for detecting sudden large losses (block losses), it is not a powerful procedure for detecting small continual losses over a number of successive material balance periods (trickle losses). Statistically monitoring the cumulative inventory difference (CID) has been advocated by many, including reference (1), as being a more powerful procedure for detecting trickle losses. However, since ID and CID are correlated, if they are individually tested each period it is not possible to exactly specify an overall significance level for the tests. In fact, all that can be said with any certainty is that the overall significance level is less than or equal to the sum of the significance levels for the two individual tests. As a result, this procedure would generally be less powerful than desired.

Reference (2) addresses this problem and proposes as a possible solution a single bivariate test of the hypothesis that, for a given material balance period, ID and CID are both zero. One obvious advantage of this procedure is that the significance level of the test can be exactly specified each material balance period. Also, the results of several illustrative examples in reference (2) indicated an increased

probability of detecting trickle losses when the bivariate test was used, compared to the currently required test. However, a number of questions were raised about how the bivariate procedure should be implemented and how the detection probabilities might be influenced by, for example, changes in the covariance structure among successive ID's, differing diversion strategies, and actions taken when a loss is indicated by the test.

An inherent difficulty in applying such a statistical test to an actual process would be the identification and estimation of sources of covariance among the successive ID's. Although references (1) and (3) discuss assumptions about these covariances which would considerably simplify their estimation, these assumptions were not empirically validated and the question remains as to whether they are generally applicable to all processes.

It was recommended in reference (2) that actual inventory records from one or more operating plants should be analyzed to form an objective basis for evaluating the possible advantages and disadvantages of implementing the bivariate procedure.

The purpose of this study is to evaluate the bivariate test of ID and CID to determine whether:

1. it could be feasibly implemented in an actual plant
2. it would enhance the capability for detecting trickle losses
3. the false alarm rate could be controlled.

This evaluation is based on an analysis of inventory records from an actual plant and the results of a simulation study which compares the bivariate procedure with the current requirement and with two other testing procedures. Detailed descriptions of the four testing procedures and formulas for their implementation are provided.



## 2.0 BASIC CONCEPTS AND DEFINITIONS

### 2.1 DISCUSSION OF ID AND CID

While there may be mechanisms of process control involving identification of process anomalies that provide indications of possible inventory discrepancies, the basic quantitative control methodology is to measure the material and establish a material balance. Records are maintained of the quantities arriving at a plant (A) and the quantities removed from a plant (R) during a specified material balance period. Letting "beginning inventory" (B) and "ending inventory" (E) denote, respectively, the amount of material on hand at the beginning and at the end of the material balance period, the balance can be expressed as

$$B + A - R = E. \quad (2.1)$$

However, for a number of reasons, it is unlikely in practice that this equation will balance exactly. The result is an inventory difference (ID) which is given by

$$ID = B + A - R - E. \quad (2.2)$$

If equation (2.1) balances, the ID in equation (2.2) is zero. When equation (2.1) does not balance, the ID in equation (2.2) represents some quantity of SNM.

For clarity of the discussions to follow, a distinction must be made between the true ID and the observed ID. The true ID is the actual amount by which equation (2.1) is out of balance, excluding the effects of measurement errors (in the absence of measurement errors, the observed ID and the true ID would be identical). The true ID would be zero in an ideal situation. In practice, however, a non-zero true ID may occur, for example, because of process losses or holdups, or stolen or diverted material. The observed ID is an estimate of the true ID. Its value is affected by both the random

errors and the biases of the methods used to measure the components of the right-hand side of equation (2.2). For simplicity, these sources of variation will be referred to as measurement variability. The observed ID would also be affected by any mistakes in the recording of measurement data or in calculating the ID quantity, any unrecorded inventory, or any gross measurement errors such as uncorrected measurement bias. For simplicity, these errors will be referred to as correctable mistakes.

For the purpose of examining a sequence of material balance periods, the cumulative inventory difference (CID) will be of interest. The CID quantity is defined as the sum of the ID quantities from a specified sequence of material balance periods. If the  $k$  most recent material balance periods are included,  $CID_k$  is given by the equation

$$CID_k = \sum_{i=1}^k ID_i \quad (2.3)$$

where  $ID_i$  is the ID quantity for the  $i^{\text{th}}$  material balance period in the sequence. By definition, the beginning inventory for the  $i^{\text{th}}$  material balance period ( $B_i$ ) is the ending inventory of the  $(i-1)^{\text{st}}$  material balance period ( $E_{i-1}$ ). Thus, adding the appropriate subscripts to the right-hand side of equation (2.2) and then substituting for  $ID_i$  in equation (2.3),  $CID_k$  can be expressed as

$$CID_k = B_1 + \sum_{i=1}^k A_i - \sum_{i=1}^k R_i - E_k$$

or

$$CID_k = B_1 + CA_k - CR_k - E_k \quad (2.4)$$

where  $CA_k$  and  $CR_k$  denote, respectively, the cumulative additions and cumulative removals over the  $k$  most recent material balance periods. It is evident from equation (2.4) that  $CID_k$  has the same four basic components as an ID quantity; namely, beginning inventory, additions, removals, and

ending inventory. This analogy is useful in developing recursive formulas for computing variances and covariances.

As with individual ID's, the distinction must be made between a true  $CID_k$  and an observed  $CID_k$ . The true  $CID_k$  is the sum given by equation (2.3), where the  $ID_i$  quantities represent true ID quantities (i.e., ID quantities with no measurement variability). The observed  $CID_k$  is an estimate of the true  $CID_k$  and is given by equation (2.3) where the  $ID_i$  quantities represent observed ID quantities (i.e., ID quantities susceptible to measurement variability).

## 2.2 MATRIX NOTATION

It is convenient to introduce vector and matrix notation, which will allow concise presentations of most of the concepts and computational formulas throughout the remainder of this report. The following definitions will be useful.

Definition 1:  $I_k$  is the k-by-k identity matrix with ones for the diagonal elements and zeros for the off diagonal elements.  $I_k$  is given by

$$I_k = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & \cdot \\ \vdots & \vdots & & \ddots & \cdot \\ 0 & 0 & . & . & 1 \end{bmatrix} \quad (2.5)$$

Definition 2:  $\underline{j}_k$  is the k-element vector of ones given by

$$\underline{j}_k = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \cdot \quad (2.6)$$

Definition 3:  $L_k$  is a k-by-2 matrix whose first column has zeros in the first k-1 rows and a one in the k<sup>th</sup> row, and whose second column is the vector  $\underline{j}_k$ .  $L_k$  is given by

$$L_k = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ 1 & 1 \end{bmatrix} . \quad (2.7)$$

Definition 4: The observed ID quantity for the i<sup>th</sup> material balance period is given by

$$ID_i = \mu_i + e_i , \quad (2.8)$$

where  $\mu_i$  is the value of the true ID for period i, and  $e_i$  is assumed to be a random variable from the normal (or Gaussian) distribution with mean zero and variance  $\sigma_i^2$ . In the context of Section 2.1,  $e_i$  is the contribution of measurement variability to  $ID_i$ . It follows immediately that  $ID_i$  is from the normal distribution with mean  $\mu_i$  and variance  $\sigma_i^2$ .

Definition 5: A sequence of k successive observed ID quantities is represented by the k-element vector  $\underline{ID}_k$ , given by

$$\underline{ID}_k = \begin{bmatrix} ID_1 \\ ID_2 \\ \vdots \\ ID_k \end{bmatrix} \quad (2.9)$$

which from Definition 4 can be expressed as

$$\underline{ID}_k = \underline{\mu}_k + \underline{e}_k \quad (2.10)$$

where  $\underline{\mu}_k$  is a k-element vector with elements  $\mu_i$ , ( $i = 1, \dots, k$ ) and  $\underline{e}_k$  is a

k-element vector with elements  $e_i$ , ( $i = 1, \dots, k$ ). In addition to Definition 4, it will be assumed that the vector  $\underline{e}_k$  is from the k-variate normal distribution with a k-element mean vector of zeros and a symmetric k-by-k variance-covariance matrix  $\Sigma_k$ , given by

$$\Sigma_k = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2k} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \cdot & \cdot \\ \vdots & \vdots & \cdot & \cdot & \cdot \\ \sigma_{k1} & \sigma_{k2} & \cdot & \cdot & \sigma_k^2 \end{bmatrix} \quad (2.11)$$

where  $\sigma_i^2$  is the variance of  $e_i$ , and  $\sigma_{ij}$  is the covariance of  $e_i$  and  $e_j$  for  $i \neq j$ . It follows immediately that the vector  $\underline{ID}_k$  is from the k-variate normal distribution with a k-element mean vector  $\underline{\mu}_k$  and a k-by-k variance-covariance matrix  $\Sigma_k$ .

Definition 6: The observed value of  $CID_k$ , given by equation (2.3), can be expressed as

$$CID_k = \underline{j}'_k \underline{ID}_k = \underline{j}'_k \underline{\mu}_k + \underline{j}'_k \underline{e}_k \quad (2.12)$$

which by Definition 5 is from the normal distribution with mean given by

$$\underline{j}'_k \underline{\mu}_k = \sum_{i=1}^k \mu_i \quad (2.13)$$

and variance given by

$$\underline{j}'_k \Sigma_k \underline{j}_k = \sum_{i=1}^k \sigma_i^2 + 2 \sum_{i=1}^k \sum_{j>i}^k \sigma_{ij} \quad (2.14)$$

Definition 7: The proposed bivariate testing procedure will involve the 2-element vector  $\underline{d}_k$ , given by

$$\underline{d}_k = \begin{bmatrix} ID_k \\ CID_k \end{bmatrix} = L'_k \underline{ID}_k \quad (2.15)$$

which by Definition 5 is from the bivariate normal distribution with a 2-element mean vector given by

$$\underline{\delta}_k = L'_k \underline{\mu}_k = \begin{bmatrix} \mu_k \\ \sum_{i=1}^k \mu_i \end{bmatrix} \quad (2.16)$$

and a 2-by-2 variance-covariance matrix given by

$$V_k = L'_k \Sigma_k L_k = \begin{bmatrix} \text{Var}(ID_k) & \text{Cov}(ID_k, CID_k) \\ \text{Cov}(ID_k, CID_k) & \text{Var}(CID_k) \end{bmatrix}. \quad (2.17)$$

Note that the quantities  $ID_k$  and  $\mu_k$  are the observed value and true value, respectively, of the ID for period  $k$  and should not be confused with the vectors  $\underline{ID}_k$  and  $\underline{\mu}_k$ .



### 3.0 FOUR TESTING PROCEDURES

This section introduces the four statistical testing procedures that are compared in this study:

1. the proposed bivariate test of  $ID_k$  and  $CID_k$
2. the currently required test of  $ID_k$  only
3. a test of  $CID_k$  only
4. a simultaneous test of  $ID_k$  and  $CID_k$  using the Bonferroni inequality.

These four procedures are not the only available testing procedures and there may be others that are powerful for detecting trickle losses. However, the rationale for considering only these particular procedures follows from the discussion in Section 1.

Briefly, the currently required test of  $ID_k$  is not a powerful procedure for detecting trickle losses. Thus, it is desirable to either replace or supplement the current requirement with a procedure which is sensitive to trickle losses. Monitoring  $CID_k$  has been advocated as a powerful procedure for detecting trickle losses. Thus, three possible options for enhancing the capability for detecting trickle losses are:

1. require that only  $CID_k$  (rather than  $ID_k$ ) be statistically tested each period
2. test both  $ID_k$  and  $CID_k$  simultaneously using the Bonferroni inequality to specify an upper bound on the overall level of significance
3. use the bivariate procedure to test both  $ID_k$  and  $CID_k$  simultaneously at an exact specified level of significance.

Thus, detection probabilities and other characteristics of these three procedures will be compared with those of the current requirement. Although testing only  $CID_k$  each period is not recommended, these comparisons will provide a basis for evaluating and comparing the two simultaneous procedures.

Collectively, the four testing procedures require the observed values of  $ID_k$  and  $CID_k$  and estimates of the elements of the matrix  $V_k$ , given by equation (2.17). It will be assumed in this section that estimates of  $VAR(ID_k)$  and  $Var(CID_k)$  are available at each material balance period. The estimation of these variances is discussed in Section 4.

In practice, it is generally quite difficult to determine the degrees of freedom for estimates of  $Var(ID_k)$  and  $Var(CID_k)$ . Thus, these estimates are usually treated as known parameters for the purpose of statistical testing. This problem was addressed in reference (2), where it was illustrated that the adverse consequences of treating these estimates as known parameters are negligible for most practical situations in SNM control and accounting. This philosophy is adopted here, and the four testing procedures assume the variance estimates are based on sufficiently large degrees of freedom to justify treating them as known parameters.

### 3.1 THE BIVARIATE PROCEDURE

When information is available from two or more successive material balance periods, the proposed bivariate procedure tests the validity of the "null hypothesis" that the true values of both  $ID_k$  and  $CID_k$  are zero. In the notation of Section 2.2, the null hypothesis states that both elements of the vector  $\underline{\delta}_k$ , given by equation (2.16), are zero. The "alternative hypothesis" is that at least one of the elements of  $\underline{\delta}_k$  is not zero. In standard statistical notation the hypotheses can be expressed as

$$\begin{array}{l} \text{vs} \quad H_0: \underline{\delta}_k = \underline{0} \\ \quad \quad H_A: \underline{\delta}_k \neq \underline{0} \end{array} \quad (3.1)$$

where  $\underline{\delta}_k$  is given by equation (2.16) and  $\underline{0}$  is a two-element vector of zeros.

The test statistic is computed using the formula

$$\chi^2 = \underline{d}_k' V_k^{-1} \underline{d}_k \quad (3.2)$$

where  $\underline{d}_k$  is given by equation (2.15) and  $V_k^{-1}$  is the inverse of the matrix  $V_k$ , which is given by equation (2.17). The statistic  $\chi^2$  is compared with the upper  $(1-\alpha)100^{\text{th}}$  percentile point of the chi-square distribution with two degrees of freedom, denoted  $\chi_{2,(1-\alpha)}^2$ , where  $\alpha$  is the specified significance level of the test of  $H_0$  vs  $H_A$ . The decision rule is to either

1. Reject  $H_0$  in favor of  $H_A$  if  $\chi^2 \geq \chi_{2,(1-\alpha)}^2$ , or
2. Accept  $H_0$  if  $\chi^2 < \chi_{2,(1-\alpha)}^2$ .

Table 3.1 gives the value of  $\chi_{2,(1-\alpha)}^2$  for a few selected values of  $\alpha$ . More extensive tables of the chi-square distribution are available in most basic statistics texts.

TABLE 3.1. Upper  $(1-\alpha)100^{\text{th}}$  Percentile Points of the Chi-Square Distribution With Two Degrees of Freedom

$\alpha$	$\chi_{2,(1-\alpha)}^2$
.20	3.22
.10	4.61
.05	5.99
.025	7.38
.01	9.21

Although estimates of the variances of  $ID_k$  and  $CID_k$  are assumed to be available, an estimate of the covariance of  $ID_k$  and  $CID_k$  is required to complete the matrix  $V_k$  which is used in equation (3.2). A recursive method for computing this covariance estimate each period is easily developed as follows:

From equations (2.3) and (2.12),  $CID_k$  is simply the sum of the  $k$  most recent observed ID quantities. Thus, another way of expressing equation (2.3) is

$$CID_k = CID_{k-1} + ID_k \quad (3.3)$$

from which it follows that

$$\text{Cov}(ID_k, CID_k) = \text{Var}(ID_k) + \text{Cov}(ID_k, CID_{k-1}) \quad (3.4)$$

and that

$$\text{Var}(CID_k) = \text{Var}(CID_{k-1}) + \text{Var}(ID_k) + 2 \text{Cov}(ID_k, CID_{k-1}). \quad (3.5)$$

Now since estimates of the variances of  $ID_k$  and  $CID_k$  are available for each period, including the  $(k-1)^{\text{st}}$  period, an estimate of  $\text{Cov}(ID_k, CID_{k-1})$  is obtained by rearranging equation (3.5) as

$$\text{Cov}(ID_k, CID_{k-1}) = \frac{1}{2} [\text{Var}(CID_k) - \text{Var}(CID_{k-1}) - \text{Var}(ID_k)]. \quad (3.6)$$

The expression on the right-hand side of equation (3.6) is then substituted for  $\text{Cov}(ID_k, CID_{k-1})$  in equation (3.4) to obtain an estimate of  $\text{Cov}(ID_k, CID_k)$ , which is given by

$$\text{Cov}(ID_k, CID_k) = \frac{1}{2} [\text{Var}(CID_k) - \text{Var}(CID_{k-1}) + \text{Var}(ID_k)]. \quad (3.7)$$

This estimate is used to complete the matrix  $V_k$  for use in equation (3.2).

It is important to note that the recursive formula and indeed the bivariate procedure are only to be used for  $k \geq 2$ . For if  $k = 1$ ,  $CID_1 = ID_1$  and  $\text{Var}(CID_1) = \text{Var}(ID_1)$ . Clearly, then  $\text{Cov}(ID_1, CID_1) = \text{Var}(ID_1)$ , and the matrix  $V_1$  is singular (i.e.,  $V_1$  cannot be inverted). The practical interpretation is that  $ID_1$  and  $CID_1$  carry exactly the same information, so that a test of  $ID_1$  only (the current requirement) is sufficient.

One reason for emphasizing the special case  $k = 1$  is that the cumulative calculations should be restarted whenever a significant loss (or gain) of SNM is discovered. Specifically, when the null hypothesis has been rejected and the cause has been determined and then corrected, the cumulative sum of ID quantities should be restarted the following material balance period. However, if the cause of the significant test result cannot be resolved, the cumulative sum of ID quantities should not be restarted unless otherwise stipulated by the management or the NRC.

To illustrate the bivariate procedure, consider the following example. Suppose that at the end of the fourth material balance period, the following information is available

$$\left. \begin{aligned} ID_4 &= 1.1 \\ CID_4 &= 4.4 \end{aligned} \right\} \text{ or } \underline{d}_4 = \begin{bmatrix} 1.1 \\ 4.4 \end{bmatrix}$$

$$\begin{aligned} \text{Var}(ID_4) &= 1.0 \\ \text{Var}(CID_4) &= 2.8 \\ \text{Var}(CID_3) &= 2.2 \end{aligned}$$

and we wish to test  $H_0$  vs  $H_A$  as given by "equations" (3.1). The significance level of the test has been specified to be  $\alpha = .05$ .

Using equation (3.7), the covariance estimate is

$$\text{Cov}(ID_4, CID_4) = \frac{1}{2} (2.8 - 2.2 + 1.0) = .8$$

and the matrix  $V_4$  of equation (2.17) is

$$V_4 = \begin{bmatrix} 1.0 & .8 \\ .8 & 2.8 \end{bmatrix} .$$

The inverse of  $V_4$  is

$$V_4^{-1} = \begin{bmatrix} 1.2963 & -.3704 \\ -.3704 & .4630 \end{bmatrix}$$

and the test statistic, computed using equation (3.2), is

$$\chi^2 = [1.1, 4.4] \begin{bmatrix} 1.2963 & -.3704 \\ -.3704 & .4630 \end{bmatrix} \begin{bmatrix} 1.1 \\ 4.4 \end{bmatrix} = 6.95 .$$

Entering Table 3.1 with  $\alpha = .05$ , the value of  $\chi_{2,.95}^2$  is 5.99, and since

$$6.95 > 5.99$$

the null hypothesis is rejected in favor of the alternative hypothesis at the .05 level of significance.

### 3.2 THE TEST OF ID ONLY

The currently required procedure is to test at each period the null hypothesis that the true value of  $ID_k$  is zero. In the notation of Section 2.2, Definition 4, the null and alternative hypotheses are

$$\begin{array}{l} H_0: \mu_k = 0 \\ \text{vs} \\ H_A: \mu_k \neq 0 \end{array} \quad (3.8)$$

The test statistic is computed using the formula

$$Z = \frac{ID_k}{\sqrt{\text{Var}(ID_k)}} \quad (3.9)$$

Then for a specified level of significance,  $\alpha$ ,  $Z$  is compared with the lower  $(\alpha/2)100^{\text{th}}$  percentile point and the upper  $(1-\alpha/2)100^{\text{th}}$  percentile point of the standard normal distribution, denoted by  $Z_{(\alpha/2)}$  and  $Z_{(1-\alpha/2)}$ , respectively. Since the standard normal distribution is symmetric,  $Z_{(\alpha/2)} = -Z_{(1-\alpha/2)}$ .

The decision rule is to either

1. Reject  $H_0$  in favor of  $H_A$  if  $Z \leq -Z_{(1-\alpha/2)}$  or  $Z \geq Z_{(1-\alpha/2)}$ , or
2. Accept  $H_0$  if  $-Z_{(1-\alpha/2)} < Z < Z_{(1-\alpha/2)}$ .

Table 3.2 gives the value of  $Z_{(1-\alpha/2)}$  for a few selected values of  $\alpha$ . More extensive tables of the standard normal distribution can be found in most basic statistics texts.



TABLE 3.2. Upper  $(1-\alpha/2)100^{\text{th}}$  Percentile Points of the Standard Normal Distribution

$\alpha$	$Z_{(1-\alpha/2)}$
.20	1.282
.10	1.645
.05	1.960
.025	2.240
.01	2.576

To illustrate the use of this procedure, consider the example described in Section 3.1. The necessary information is

$$\begin{aligned} \text{ID}_4 &= 1.1 \\ \text{VAR}(\text{ID}_4) &= 1.0 \end{aligned}$$

and we wish to test

$$\begin{aligned} \text{vs } H_0: \mu_4 &= 0 \\ H_A: \mu_4 &\neq 0 \end{aligned}$$

at the  $\alpha = .05$  level of significance. The test statistic is

$$Z = \frac{1.1}{\sqrt{1.0}} = 1.1$$

and from Table 3.2,  $Z_{(.975)} = 1.96$ . Since  $-1.96 < 1.1 < 1.96$ , we accept  $H_0$ . There is insufficient evidence to conclude that  $\mu_4$  is not zero.

### 3.3 THE TEST OF CID ONLY

As mentioned previously, the test of  $\text{CID}_k$ , by itself, is not recommended. However, it does help to provide a basis for objectively evaluating and comparing the other three procedures.

This procedure tests at each period the null hypothesis that the true value of  $CID_k$  is zero. In the notation of Section 2.2, Definition 6, the null and alternative hypotheses are

$$\begin{array}{l}
 H_0: \sum_{i=1}^k \mu_i = 0 \\
 \text{vs} \\
 H_A: \sum_{i=1}^k \mu_i \neq 0 .
 \end{array} \tag{3.10}$$

The test statistic is computed using the formula

$$Z = \frac{CID_k}{\sqrt{\text{Var}(CID_k)}} . \tag{3.11}$$

Then for a specified level of significance,  $\alpha$ , the test proceeds exactly like the test described in Section 3.2. The decision rule is to either

1. Reject  $H_0$  in favor of  $H_A$  if  $Z \leq -Z_{(1-\alpha/2)}$  or  $Z \geq Z_{(1-\alpha/2)}$ , or
2. Accept  $H_0$  if  $-Z_{(1-\alpha/2)} < Z < Z_{(1-\alpha/2)}$ ,

where, as before,  $Z_{(1-\alpha/2)}$  is the upper  $(1-\alpha/2)100^{\text{th}}$  percentile point of the standard normal distribution.

To illustrate the use of this procedure, consider again the example data of Section 3.1. The necessary information is

$$\begin{array}{l}
 CID_4 = 4.4 \\
 \text{Var}(CID_4) = 2.8
 \end{array}$$

and we wish to test

$$\begin{array}{l}
 H_0: \sum_{i=1}^4 \mu_i = 0 \\
 \text{vs} \\
 H_A: \sum_{i=1}^4 \mu_i \neq 0
 \end{array}$$

at the  $\alpha = .05$  level of significance. The test statistic is

$$Z = \frac{4.4}{\sqrt{2.8}} = 2.63$$

and from Table 3.2,  $Z_{(.975)} = 1.96$ . Since  $2.63 > 1.96$ , we reject  $H_0$  in favor of  $H_A$ , and conclude that the true value of  $CID_4$  is larger than zero. This would be interpreted as a cumulative loss of SNM.

#### 3.4 THE "BONFERRONI" PROCEDURE

In the illustrative example of Section 3.1, the bivariate test resulted in rejection of the null hypothesis that both

$$\mu_4 \text{ and } \sum_{i=1}^4 \mu_i$$

are zero. In Sections 3.2 and 3.3 the same example information was used to perform individual tests on  $ID_4$  and  $CID_4$ . The result of the test of  $ID_4$  in Section 3.2 was that there is insufficient evidence to conclude that  $\mu_4$  differs from zero. However, the test of  $CID_4$  in Section 3.3 indicated that

$$\sum_{i=1}^4 \mu_i$$

is larger than zero. In this example, if it could be assumed that no significant test results were observed during the first three periods, the results described above would tend to indicate that "small" losses of SNM have occurred over the four periods resulting in a significant cumulative loss.

One obvious question that arises is: Why not just perform the individual tests of  $ID_k$  and  $CID_k$ , but look at the test results simultaneously? This idea seems appealing because it avoids the covariance calculation and the matrix operations necessary for the bivariate test. One major disadvantage is that since  $ID_k$  and  $CID_k$  are correlated, the exact level of significance of such a procedure cannot be determined. However, it is

possible to set an upper bound on the overall level of significance.

The Bonferroni inequality applied to this situation states that if  $ID_k$  and  $CID_k$  are individually tested as described in Section 3.2 and 3.3 at significance levels of  $\alpha_1$  and  $\alpha_2$ , respectively, then the overall level of significance,  $\alpha$ , for the two tests is bounded from above by the sum of  $\alpha_1$  and  $\alpha_2$ . That is,

$$\alpha \leq \alpha_1 + \alpha_2 \quad . \quad (3.12)$$

A very simple approach is to specify the desired overall level of significance, say  $\alpha^*$ , and then let

$$\alpha_1 = \alpha_2 = \alpha^*/2 \quad (3.13)$$

so that the overall level of significance is bounded from above by  $\alpha^*$ . That is,

$$\alpha \leq \alpha^*/2 + \alpha^*/2 = \alpha^* \quad . \quad (3.14)$$

The fourth procedure to be considered in this study uses the above concepts to simultaneously test the hypotheses stated in "equations" (3.8) and (3.10). Using slightly different notation, these hypotheses are restated here as

$$H_1: \mu_k = 0 \quad \text{vs} \quad H_2: \mu_k \neq 0 \quad (3.15)$$

and

$$H_3: \sum_{i=1}^k \mu_i = 0 \quad \text{vs} \quad H_4: \sum_{i=1}^k \mu_i \neq 0 \quad .$$

The desired overall significance level,  $\alpha^*$ , is specified, and the test proceeds as follows.

Two test statistics are computed using formulas (3.9) and (3.11). Using subscripts, these are

$$Z_1 = \frac{ID_k}{\sqrt{\text{Var}(ID_k)}} \quad (3.16)$$

and

$$Z_2 = \frac{CID_k}{\sqrt{\text{Var}(CID_k)}} .$$

The decision rule is to

1. Reject  $H_1$  in favor of  $H_2$  if  $Z_1 \leq -Z_{(1-\alpha^*/4)}$  or  $Z_1 \geq Z_{(1-\alpha^*/4)}$ ,
2. Reject  $H_3$  in favor of  $H_4$  if  $Z_2 \leq -Z_{(1-\alpha^*/4)}$  or  $Z_2 \geq Z_{(1-\alpha^*/4)}$ , or
3. Accept both  $H_1$  and  $H_3$  if both  $-Z_{(1-\alpha^*/4)} < Z_1 < Z_{(1-\alpha^*/4)}$  and  $-Z_{(1-\alpha^*/4)} < Z_2 < Z_{(1-\alpha^*/4)}$  .

That is, each of the individual tests has a significance level of  $\alpha^*/2$ , so that in accordance with equation (3.14) the overall significance level is less than or equal to  $\alpha^*$ .

To illustrate the procedure, refer again to the example data given in Section 3.1. Let the desired overall significance level be  $\alpha^* = .05$ . Then  $\alpha_1 = \alpha_2 = .025$ , and entering Table 3.1 at  $\alpha = .025$ , the value  $Z_{(.9875)} = 2.24$  is obtained. The test statistics are

$$Z_1 = \frac{1.1}{\sqrt{1.0}} = 1.1$$

and

$$Z_2 = \frac{4.4}{\sqrt{2.8}} = 2.63 .$$

Now

$$-2.24 < Z_1 = 1.1 < 2.24$$

and

$$Z_2 = 2.63 > 2.24 .$$

Thus, at an overall significance level of  $\alpha \leq .05$ , we would accept  $H_1$ , but reject  $H_3$  in favor of  $H_4$ . In other words, there is not sufficient evidence to conclude that  $\mu_4$  differs from zero, but there is strong evidence that

$$\sum_{i=1}^4 \mu_i$$

is larger than zero.

Although the overall level of significance is "controlled" in the sense that it is bounded from above, the actual level of significance will be less than desired. The result is that this procedure will tend to be more conservative (i.e., less powerful) than the bivariate procedure.

### 3.5 FURTHER DISCUSSION OF THE BONFERRONI AND BIVARIATE PROCEDURES

Reference (2) gives a detailed discussion of the differences between the Bonferroni and bivariate procedures. Primarily, the distinguishing factor is the shape of the "acceptance regions".

The acceptance region for the Bonferroni procedure is simply the interior of a rectangle drawn on the  $ID_k, CID_k$  plane, centered at  $(ID_k = 0, CID_k = 0)$ , and defined by

$$[-Z_{(1-\alpha^*/4)} \sqrt{\text{Var}(ID_k)} < ID_k < Z_{(1-\alpha^*/4)} \sqrt{\text{Var}(ID_k)}]$$

$$[-Z_{(1-\alpha^*/4)} \sqrt{\text{Var}(CID_k)} < CID_k < Z_{(1-\alpha^*/4)} \sqrt{\text{Var}(CID_k)}] .$$

If the observed point  $(ID_k, CID_k)$  were inside this rectangle, both null hypotheses  $H_1$  and  $H_3$ , given by "equations" (3.15), would be accepted. Otherwise, in accordance with the decision rule for this procedure, either  $H_1$  or  $H_3$  (or both) would be rejected. If one or both of the null hypotheses were rejected, it would be clear from the values of the test statistics which variable ( $ID_k$  or  $CID_k$ ) caused the rejection.

The acceptance region for the bivariate procedure is the interior of an ellipse, drawn on the  $ID_k, CID_k$  plane, centered at  $(ID_k = 0, CID_k = 0)$ , elongated in the  $CID_k$  direction, and "tilted" according to the sign and size of  $Cov(ID_k, CID_k)$ . The boundary of this elliptical region is defined by

$$\chi^2 = \chi_{2,(1-\alpha)}^2, \quad (3.17)$$

where  $\chi^2$  is the test statistic given by equation (3.2). If the observed point  $(ID_k, CID_k)$  is inside this region, the null hypothesis given by "equation" (3.1) is accepted. Otherwise, the null hypothesis is rejected, but it is not clear from the value of the test statistic which of the variables ( $ID_k$  or  $CID_k$ ) has caused the rejection. In order to interpret a significant result, it may be necessary to perform the individual tests on  $ID_k$  and  $CID_k$ . An advantage of using the bivariate procedure, however, is that the significance level can be exactly specified. That is, the probability that the point  $(ID_k, CID_k)$  would be inside the ellipse defined by equation (3.17), when  $\underline{\delta}_k = \underline{0}$ , is exactly  $(1-\alpha)$ .

Reference (2) and most texts on multivariate statistical methods illustrate that when the acceptance regions for the Bonferroni and bivariate procedures are overlaid (i.e., one is superimposed on the other), there are regions where a point  $(ID_k, CID_k)$  could be in the acceptance region of one procedure but outside the acceptance region of the other. These regions, of course, explain the difference between the two procedures.





#### 4.0 ESTIMATION OF VARIANCES AND COVARIANCES

The identification and modeling of sources of variance and covariance in SNM control and accounting is difficult to address in a general context. This is because each plant has its own unique characteristics, anomalies and procedures, so that it is impossible to specify a step-by-step method of estimation that would be applicable to all licensees.

As part of this project, inventory records and variance calculations for five successive material balance periods from an operating plant were studied. Some of the estimation techniques used by this particular plant and some of the conclusions drawn from the study could be useful to licensees who attempt to implement either the bivariate or Bonferroni procedure. This section discusses some concepts which should be general considerations for computing  $\text{Var}(ID_k)$ ,  $\text{Var}(CID_k)$ , and  $\text{Cov}(ID_k, CID_k)$ .

##### 4.1 GENERAL STRATEGY FOR COMPUTING $\text{VAR}(ID_k)$

Many licensed facilities have established their own individual procedures for computing  $\text{Var}(ID_k)$ . Most of these procedures are computerized and are routinely carried out as part of the materials accounting process. This is true of the plant which was studied for this project. Although the explicit details of the variance calculations are too voluminous to present in this report, a brief overview of the general strategy is given.

The plant under study keeps inventory records for approximately thirty-seven types of material in connection with its "UO<sub>2</sub> process". These materials are grouped into eighteen strata based on such considerations as conversion factors, sampling, weighing and analytical methods used to evaluate the amount of UO<sub>2</sub> entering, leaving or being stored in the plant. That is, each of the thirty-seven material types is assigned to one of the eighteen strata.

Within each stratum, the quantities B, A, R and E are determined and an ID is calculated using equation (2.2). These within stratum ID quantities are summed over the eighteen strata to give the value of the overall ID quantity for the UO<sub>2</sub> process.

For the purpose of computing variances, the following quantities are defined within each stratum.

$X_{BE}$  = Amount of material common to both B and E

$X_{BR}$  = Amount of material common to both B and R

$X_{AE}$  = Amount of material common to both A and E

$X_{AR}$  = Amount of material common to both A and R

The "gross quantities" B, A, R and E are then adjusted by subtracting the appropriate "materials-in-common" quantities as illustrated in Table 4.1.

TABLE 4.1. Computation of Net Inventory Quantities for Variance Calculations

Gross Quantities:	B	A	R	E
Adjustments:	$\left\{ \begin{array}{l} -X_{BE} \\ -X_{BR} \\ -X_{AE} \\ -X_{AR} \end{array} \right.$	$\left\{ \begin{array}{l} \\ -X_{AE} \\ \\ \end{array} \right.$	$\left\{ \begin{array}{l} \\ -X_{BR} \\ -X_{AR} \\ \end{array} \right.$	$\left\{ \begin{array}{l} -X_{BE} \\ \\ -X_{AE} \\ \end{array} \right.$
Net Quantities:	B*	A*	R*	E*

The net quantities B\*, A\*, R\*, and E\* are then used for the variance calculations as follows.

Each material balance period, random and systematic "error parameters" (or variances) are determined for each of the various methods of sampling, weighing and analysis. The appropriate "error parameters", conversion factors, and net material quantities for a particular stratum are multiplied and algebraically manipulated to obtain estimates of the variances due to the random and systematic errors within each stratum. These within stratum variance estimates are then summed over the eighteen strata to obtain overall estimates of the random and systematic error variances. Finally, these overall variance estimates are summed to obtain an estimate of the variance of ID for the  $UO_2$  process. For the  $k^{th}$  period, this is  $Var(ID_k)$ .

Using the net material quantities within a stratum tends to minimize correlations due to "materials-in-common" among  $B^*$ ,  $A^*$ ,  $R^*$  and  $E^*$ . These net quantities are practically treated as if they were uncorrelated when computing variances. However, the net quantities within a stratum could be correlated due to, for example, being measured by the same technician, weighed by the same scale, etc. Also, there are definite correlations among certain of the strata. For example, one stratum might be  $UF_6$  cylinders, showing large values of  $B^*$  and  $A^*$  but small values of  $R^*$  and  $E^*$ . Whereas, a complimentary stratum would be fuel rods, showing small values of  $B^*$  and  $A^*$  but large values of  $R^*$  and  $E^*$ . That is,  $UO_2$  enters the plant in the form of a raw material and leaves the plant in the form of a finished product. Thus, the  $B^*$  and  $A^*$  quantities of the  $UF_6$  cylinders stratum are highly correlated with the  $R^*$  and  $E^*$  quantities of the fuel rods stratum.

It may not be appropriate to assume independence among the strata and among the quantities within a stratum when computing variances. Ignoring correlations, of course, can lead to either overestimating or underestimating  $Var(ID_k)$ . If  $Var(ID_k)$  is overestimated, the currently required test of  $ID_k$  could be rendered useless. If  $Var(ID_k)$  is underestimated, the test of  $ID_k$  could be overly sensitive resulting in an excessive "false alarm" rate.

#### 4.2 DIFFICULTIES WITH CALCULATING $\text{Var}(\text{CID}_k)$ AND $\text{COV}(\text{ID}_k, \text{CID}_k)$

An attempt was made to compute  $\text{Var}(\text{CID}_k)$  for periods  $k = 2, 3, 4$  and  $5$  from the inventory records of the plant under study. This experience revealed that since licensees are not required to monitor  $\text{CID}_k$ , many of their current inventory record keeping methods do not provide sufficiently detailed information to accommodate calculating  $\text{Var}(\text{CID}_k)$ . If licensees were to implement one of the testing procedures requiring  $\text{Var}(\text{CID}_k)$ , modifications of their current practices would be necessary in two major areas. Specifically, inventory record-keeping would have to be expanded to include the needed cumulative information, and methods for pooling the error parameter data over the  $k$  most recent periods would need to be developed.

As with computing  $\text{Var}(\text{ID}_k)$ , each plant is unique in some ways, so that it is not possible to spell out a generally applicable method of computing  $\text{Var}(\text{CID}_k)$ . However, there are some basic concepts, relevant to the two problem areas mentioned above, which should be considered before attempting to implement a testing procedure which involves estimating  $\text{Var}(\text{CID}_k)$ .

In Section 4.1 a general overview of the calculation of  $\text{Var}(\text{ID}_k)$  was given. The "first steps" described there are the stratification of the materials and the within stratum materials-in-common adjustments of the inventory quantities as illustrated in Table 4.1. Organizing the inventory records for the calculation of  $\text{Var}(\text{CID}_k)$  would utilize exactly the same stratification of materials as for the calculation of  $\text{Var}(\text{ID}_k)$ . However, within a stratum there would be inventory quantities for the  $k$  most recent periods. The materials-in-common adjustments of these quantities would be slightly more complicated than for the case of a single period, illustrated in Table 4.1.

To see this, consider three consecutive periods. Within a stratum, the necessary gross quantities would be:

$$B_1, A_1, R_1, B_2, A_2, R_2, B_3, A_3, R_3, E_3.$$

The reason for excluding  $E_1$  and  $E_2$ , of course, is that  $E_1 = B_2$  and  $E_2 = B_3$ . The materials-in-common adjustments are made in three "stages". First, both  $B_1$  and  $A_1$  could have materials in common with  $R_1, B_2, R_2, B_3, R_3$  and  $E_3$ . The record keeping procedure would have to supply sufficient detail to identify these materials-in-common quantities and make the adjustments. Secondly, after making the above adjustments, both  $B_2$  and  $A_2$  could have additional materials in common with  $R_2, B_3, R_3$  and  $E_3$ . These additional adjustments would be made. Finally, after the above two adjustments are made, both  $B_3$  and  $A_3$  could have additional materials in common with  $R_3$  and  $E_3$ , and these adjustments would be made. The resulting net quantities would then be used in the variance calculations.

Clearly, as the number of periods in a sequence increases, the number of materials-in-common adjustments gets quite large. However, with the availability of computers this is hardly a deterrent. The major concern with the multistage adjusting process would be preventing the same materials from being adjusted out more than once.

The next step given in Section 4.1 for computing  $\text{Var}(ID_k)$  is to multiply the net quantities with appropriate conversion factors and error parameters. The analogous step in computing  $\text{Var}(CID_k)$  requires careful consideration.

The actual plant records which were studied for this project included a different table of error parameter estimates for each of the five material balance periods. While some of the estimates were quite stable over the five periods, others fluctuated dramatically. The actual raw data from which these parameters were estimated were not included in the plant records, nor were the degrees of freedom of the estimates. Thus, statistical tests of equality of the parameters over the periods were not possible. However, this raises the question of how to best utilize the error parameter data in computing  $\text{Var}(CID_k)$ .

A possible procedure is to use the raw data to statistically test whether a particular parameter changes over the  $k$  periods being considered. If it is found to be stable, then the estimates for the  $k$  periods should be "pooled". For example, if  $k$  variance estimates,  $S_i^2$  ( $i = 1, 2, \dots, k$ )

are available, where  $S_i^2$  has  $n_i$  degrees of freedom, the pooled estimate would be

$$s^2 = \frac{\sum_{i=1}^k n_i S_i^2}{\sum_{i=1}^k n_i} \quad (4.1)$$

which has  $\sum_{i=1}^k n_i$  degrees of freedom.

If, however, a parameter is found to change over the periods then the estimates for the individual periods should not be pooled.

Whether pooled or individual parameter estimates are used, the next step is to judiciously match error parameter estimates, conversion factors and net quantities and then carry out the appropriate algebraic manipulations to compute the within-stratum variance of  $CID_k$ .

Finally, these within-stratum variance estimates are summed over the strata to obtain an estimate of the variance of the overall  $CID_k$ . This is  $Var(CID_k)$ .

If carried out correctly, the calculations will result in an estimate of  $Var(CID_k)$  that is "compatible" with the estimate of  $Var(ID_k)$  for computing  $Cov(ID_k, CID_k)$  from equation (3.7). If the estimates of  $Var(ID_k)$ ,  $Var(CID_{k-1})$  and  $Var(CID_k)$  are not compatible then  $Var(CID_k)$  [and also  $Var(CID_{k-1})$ ] has either been overestimated or underestimated. This will, of course, affect the estimate of  $Cov(ID_k, CID_k)$ . The consequences of incorrectly estimating these variances and covariances would be statistical tests that are either insensitive or overly sensitive.

## 5.0 THE SIMULATION STUDY

The discussions thus far have focused on statistically testing  $ID_k$  and/or  $CID_k$  at the end of a particular material balance period using one of the four procedures presented in Section 3. In order to evaluate the four procedures, it is desirable to monitor several processes over a number of material balance periods, where different SNM diversion strategies are in effect. However, as mentioned in Section 4, most operating plants do not keep sufficiently detailed inventory records to accommodate the calculation of  $Var(CID_k)$  and  $Cov(ID_k, CID_k)$ . Thus, existing (historical) plant inventory records are of little use. Also, it would be infeasible in terms of time and cost to enlist several "volunteer" plants to implement the more sophisticated record-keeping methods for the next five to ten years, after which a comparative study of the four procedures might be attempted. A very feasible and economical approach is to carry out a computer simulation study.

A simulation study was conducted for this project. Briefly speaking, this involved modeling a process over ten successive material balance periods by specifying an "SNM diversion strategy vector"  $\underline{\mu}_{10}$  and a variance-covariance matrix  $\underline{\Sigma}_{10}$  relating the ten successive ID's (see Definition 5, Section 2.2). Then, in accordance with equation (2.9), two thousand  $\underline{ID}_{10}$  vectors were randomly generated and each of the four procedures was used to monitor all two thousand simulated processes over the ten material balance periods. Various statistics were accumulated which were used to compute detection probabilities, average undetected losses and false alarm rates for the four procedures. This is equivalent to monitoring two thousand identical operating plants where differences in inventory records are due only to random variability in the physical inventory measurement system.

The process simulation scheme was carried out for nine different variance-covariance matrices and twenty-four diversion strategies. That is, two thousand  $\underline{ID}_{10}$  vectors were generated for each of two hundred sixteen covariance/diversion strategy combinations.

This section presents the method of simulation, the choice of input parameters, assumptions and decision rules used, and the results of the simulation study.

### 5.1 METHOD OF GENERATING A SEQUENCE OF ID'S

For the simulation study, it was necessary to be able to generate a random sample of size two thousand from the ten-variate normal distribution with specified mean vector  $\underline{\mu}_{10}$  and specified variance-covariance matrix  $\Sigma_{10}$  (see Definition 5, Section 2.2). A computer program was written which uses the following random variable generating technique.

On most computer systems, a "random number generator" is available which will generate a random sample from the uniform (0, 1) distribution (sometimes called the rectangular distribution on the unit interval). If a random sample of size two is generated, let  $u_1$  and  $u_2$  be the observed values. These can be transformed to a random sample of size two from the normal (0, 1) distribution using the formulas

$$x_1 = [-2 \log_e(u_1)]^{1/2} \cos(2\pi u_2)$$

and

(5.1)

$$x_2 = [-2 \log_e(u_1)]^{1/2} \sin(2\pi u_2).$$

That is,  $x_1$  and  $x_2$  represent the observed outcome of a random sample of size two from the normal distribution with mean zero and variance one.

It is recommended that this "sampling" be repeated several times and that the observed values be "averaged" to obtain one "random normal deviate". For this study, eight random samples of size two are generated, giving a total sample size of sixteen. Letting the sixteen observed values be denoted  $x_1, x_2, \dots, x_{16}$ , a single value is computed using the formula



$$y = \frac{\sum_{i=1}^{16} x_i}{4} \quad (5.2)$$

where  $y$  represents the observed outcome of a random sample of size one from the normal  $(0, 1)$  distribution. The "averaging" is done in an attempt to eliminate the effect of non-randomness which is a problem with some computer systems.

To generate a sample of size one from the ten-variate normal distribution with specified mean vector  $\underline{\mu}_{10}$  and specified variance-covariance matrix  $\underline{\Sigma}_{10}$ , the following technique is used. First, the above described procedure is repeated ten times to generate ten  $y$  values, denoted  $y_1, y_2, \dots, y_{10}$ . These represent a random sample of size ten from the normal  $(0, 1)$  distribution. The  $y_i$  values are arranged in a ten-element vector, given by

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_{10} \end{bmatrix} \quad (5.3)$$

Then  $\underline{y}$  is a random sample of size one from the ten-variate normal distribution with the ten-element mean vector  $\underline{0}$  and the ten-by-ten variance-covariance matrix  $I_{10}$ . This is denoted  $N_{10}(\underline{0}, I_{10})$ , where  $I_{10}$  is given by equation (2.5).

Next, let  $T$  be the ten-by-ten upper triangular matrix resulting from a Cholesky square root factorization of the specified variance-covariance matrix  $\underline{\Sigma}_{10}$ . That is,

$$\underline{\Sigma}_{10} = T'T \quad (5.4)$$

Then, the vector  $\underline{ID}_{10}$  is computed from the formula

$$\underline{ID}_{10} = \underline{\mu}_{10} + T'\underline{y} \quad (5.5)$$

This is consistent with equation (2.10), where  $\underline{e}_{10} = T'\underline{y}$ . Now, since  $\underline{y}$  is from  $N_{10}(\underline{0}, I_{10})$ , it follows immediately that  $\underline{ID}_{10}$  given by equation (5.5) is from  $N_{10}(\underline{\mu}_{10}, T'T)$ , or  $N_{10}(\underline{\mu}_{10}, \Sigma_{10})$  as desired.

## 5.2 CHOICE OF INPUT PARAMETERS $\underline{\mu}_{10}$ AND $\Sigma_{10}$

It was mentioned in Section 1 that the purpose of this project is to evaluate the bivariate testing procedure, which involves comparing it with the other three procedures. The criteria for these comparisons are: probability of detection (i.e., power), expected undetected losses, and false alarm rates. It is of interest to determine how these criteria are influenced by differing diversion strategies, changes in the covariance structure among successive ID's, and actions taken when a loss is indicated by a test. This section discusses the choice of diversion strategies and variance-covariance matrices which were included in the simulation study.

### 5.2.1 Discussion of Diversion Strategies

From Definition 5, Section 2.2, a sequence of successive observed ID quantities can be expressed as the sum of a vector of true ID quantities and a vector of random "measurement errors". In order to evaluate the four testing procedures on the basis of their ability to detect different diversion strategies, it is necessary to make the following simplifying assumptions:

1. The vector  $\underline{\mu}_{10}$ , defined by equation (2.10), represents only the amounts of SNM diverted from each of ten successive material balance periods.
2. There are no "correctable mistakes" in the inventory records.

The vector  $\underline{\mu}_{10}$  will henceforth be referred to as the "diversion strategy vector".

Suppose a goal quantity  $Q$  of SNM is to be diverted from a process over ten successive material balance periods. Then according to the above assumptions,  $\mu_i$  is the amount of SNM diverted from the  $i^{\text{th}}$  period, and

$$Q = \sum_{i=1}^{10} \mu_i \quad (5.6)$$

which, from equation (2.13), is the true value of  $CID_{10}$ .

Define the quantity  $q_i$  to be

$$q_i = \frac{\mu_i}{Q} \quad (5.7)$$

That is,  $q_i$  is the fraction of the goal quantity diverted during the  $i^{\text{th}}$  material balance period. The computer program used for the simulation (see Appendix A) reads as input a "diversion pattern vector"  $\underline{q}$ , given by

$$\underline{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_{10} \end{bmatrix} \quad (5.8)$$

and multiple values of  $Q$ . Then for a given value of  $Q$ , the diversion strategy vector  $\underline{\mu}_{10}$  is computed from the formula

$$\underline{\mu}_{10} = Q \underline{q} \quad (5.9)$$

This allows one diversion pattern to be used for multiple goal quantities. In this study,  $Q$  ranged in value from 0 to 30 units of SNM. No specific units of measure were used. Thus, for example, if a constant loss diversion strategy is of interest,  $q_i = .10$  for all  $i$ , and for a specific value of  $Q$ ,

$$\underline{\mu}_{10} = Q \begin{bmatrix} .10 \\ .10 \\ \vdots \\ .10 \end{bmatrix}$$

If the goal quantity is  $Q = 15$  units of SNM, then the strategy would be to divert  $\mu_j = 1.5$  units of SNM from each of the ten periods.

For this study, only constant loss diversion strategies were considered. However, the variance-covariance structures were selected such that for the preliminary simulation runs, two diversion patterns were studied. These arose as follows:

1. When  $\sqrt{\text{Var}(ID_j)}$  is constant over the ten periods, then the constant loss diversion pattern is equivalent to having  $q_j$  proportional to  $\sqrt{\text{Var}(ID_j)}$ . This has been recognized by many as an "optimal" diversion pattern.
2. When  $\sqrt{\text{Var}(ID_j)}$  is not constant over the ten periods, then the constant diversion pattern is less than optimal, but probably a more realistic situation. That is, the diverter is unable to predict  $\sqrt{\text{Var}(ID_j)}$ , and simply diverts equal amounts from each period.

### 5.2.2 Selection of Variance-Covariance Matrices

Deciding which variance-covariance matrices to include in the simulation study was a multistage procedure. The results of the analysis of actual plant inventory records were considered, as were assumptions made in references (1) and (3). Finally, preliminary small scale simulation runs were made and the results helped narrow the scope of the study.

It was mentioned in Section 4 that most licensees do not keep sufficiently detailed records to accommodate calculation of  $\text{Var}(\text{CID}_k)$ . This is definitely not a shortcoming of these facilities, since they are not required to monitor  $\text{CID}_k$ . However, an attempt to crudely estimate  $\text{Var}(\text{CID}_k)$  from actual plant inventory records was unsuccessful. Since individual item identifications were not available, the materials-in-common adjustments, described in Section 4.2, could not be made.

The approach taken was to assume a first-in-first-out (FIFO) inventory activity for the five material balance periods and then recompile the inventory records accordingly. Still, without individual item identifications, all of the necessary materials-in-common adjustments were not possible. The result was that  $\text{Var}(\text{CID}_k)$  was overestimated by unavoidably including covariances due to the materials-in-common quantities that were not adjusted out. Thus, when equation (3.7) was used to compute  $\text{Cov}(\text{ID}_k, \text{CID}_k)$ , the resulting estimate was inflated by the materials-in-common covariance terms.

Realizing the poor quality of these estimates of  $\text{Var}(\text{CID}_k)$  and  $\text{Cov}(\text{ID}_k, \text{CID}_k)$ , an attempt was made to estimate the variance-covariance matrix  $\Sigma_5$ , see equation (2.11). Here again the covariances were overestimated. Even so, a correlation matrix was computed using the following rule:

The correlation between any two ID's in a sequence, say  $\text{ID}_i$  and  $\text{ID}_j$ , is given by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}} \quad (5.10)$$

where from equation (2.11)  $\sigma_{ij}$  is the covariance of  $\text{ID}_i$  and  $\text{ID}_j$ , and  $\sigma_i^2$  and  $\sigma_j^2$  are the variances,  $\text{Var}(\text{ID}_i)$  and  $\text{Var}(\text{ID}_j)$ , respectively. When  $i = j$ , then

$\sigma_{ii} = \sigma_i^2$  and  $\rho_{ii} = 1$ . If this transformation is performed on all elements of the matrix  $\sum_k$  of equation (2.11), the resulting k-by-k symmetric matrix is called the correlation matrix and is given by

$$R_k = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1k} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2k} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & \cdots & 1 \end{bmatrix} \quad (5.11)$$

where  $-1 \leq \rho_{ij} \leq 1$ .

The correlation matrix  $R_5$  which was estimated from the actual plant data had the peculiar pattern

$$R_5 = \begin{bmatrix} 1 & .6 & .7 & .8 & .8 \\ .6 & 1 & .6 & .7 & .8 \\ .7 & .6 & 1 & .6 & .7 \\ .8 & .7 & .6 & 1 & .6 \\ .8 & .8 & .7 & .6 & 1 \end{bmatrix} \quad (5.12)$$

The concept of two successive ID's having a weaker correlation than two ID's which are separated by two or more periods is counter intuitive, and is quite unrealistic in this SNM inventory context. It is conjectured that this peculiar correlation pattern is evidence that the overestimation of  $\text{Var}(\text{CID}_k)$  and  $\text{Cov}(\text{ID}_k, \text{CID}_k)$  gets worse as k increases.

Although attempts to estimate  $\text{Var}(\text{CID}_k)$  quantities and covariances from the actual plant data were unsuccessful, the results of this analysis did in fact influence the final choice of variance-covariance matrices. This will be further discussed after other considerations have been presented.

Since good covariance estimates from the actual plant data were not possible, a popular assumption about the form of  $\Sigma_k$  was considered. Both references (1) and (3) advocate the assumption that two successive ID's are correlated only because the ending inventory of the earlier period is the beginning inventory of the later period. Then assuming all other inventory quantities within a material type stratum are "independent", they proceed to show that

$$\text{Cov}(\text{ID}_i, \text{ID}_{i+1}) = -\text{Var}(E_i) . \quad (5.13)$$

The independence assumption together with equation (5.13) implies a variance-covariance matrix of the form

$$\Sigma_k = \begin{bmatrix} \text{Var}(\text{ID}_1) & -\text{Var}(E_1) & 0 & 0 & \dots & 0 \\ -\text{Var}(E_1) & \text{Var}(\text{ID}_2) & -\text{Var}(E_2) & 0 & \dots & 0 \\ 0 & -\text{Var}(E_2) & \text{Var}(\text{ID}_3) & -\text{Var}(E_3) & & 0 \\ 0 & 0 & -\text{Var}(E_3) & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & -\text{Var}(E_{k-1}) & \vdots \\ 0 & 0 & \dots & \dots & -\text{Var}(E_{k-1}) & \text{Var}(\text{ID}_k) \end{bmatrix} . \quad (5.14)$$

Under these assumptions, the actual plant data were used to estimate the elements of the variance-covariance matrix  $\Sigma_5$ . Then using equation (5.10), the correlations were computed. The values of the estimated correlation between any two successive ID's ranged from -.4 to -.2.

One undesirable aspect of this assumption is that it implies that the covariance of two successive ID's is due only to materials in common with the two successive periods. This ignores the possibility of "intra-class" type correlations due to, for example, using the same personnel, instruments, techniques, etc. over several material balance periods. Thus, a more realistic correlation structure might have nonzero values of  $\rho_{ij}$  that get "weaker" (regardless of sign) as  $i$  and  $j$  get farther apart.

The next step in selecting variance-covariance matrices involved a large number of small scale simulation runs, where literally dozens of variance-covariance matrices were included. The results of this "preliminary" study combined with the considerations discussed above helped to narrow the scope of the variance-covariance matrix selection. The following conclusions were drawn:

1. When the off diagonal elements of the matrix  $R_k$ , given by equation (5.11), are predominantly negative (or positive), the resulting criteria for comparing the four procedures seem to be practically unaffected when the  $\rho_{ij}$  values are replaced with an "average" constant quantity with the same sign. This simplifies the study by the use of correlation structures with constant off diagonal elements.
2. The criteria for comparing the four testing procedures are drastically different when the  $\rho_{ij}$  are predominantly negative than when they are predominantly positive.
3. The overall criteria for comparing the four testing procedures did not change noticeably when the values of  $\sqrt{\text{Var}(ID_k)}$ , estimated from the plant data, were replaced by an "average constant" value. This is probably due to the stability of the  $\sqrt{\text{Var}(ID_k)}$  quantities over the five material-balance periods in the plant records.



Based on these considerations, the following correlation structures were included in the simulation study.

Correlation Matrix 1:

$$R_{10} = \begin{bmatrix} 1 & \rho & 0 & 0 & \dots & 0 \\ \rho & 1 & \rho & 0 & \dots & 0 \\ 0 & \rho & 1 & \rho & & \cdot \\ 0 & 0 & \rho & 1 & & \cdot \\ \vdots & & & & \ddots & \cdot \\ 0 & \dots & & & \rho & 1 \end{bmatrix} \quad (5.15)$$

Correlation Matrix 2:

$$R_{10} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^9 \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^8 \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^7 \\ \rho^3 & \rho^2 & \rho & 1 & & \cdot \\ \cdot & \cdot & \cdot & & \ddots & \cdot \\ \cdot & \cdot & \cdot & & & \rho \\ \rho^9 & \rho^8 & \rho^7 & \dots & \rho & 1 \end{bmatrix} \quad (5.16)$$

Both structures were used in the simulation study with values of  $\rho$  ranging from  $-.5$  to  $.5$ .

For a specified number (say  $M$ ) of material balance periods, the computer program used for the simulation (see Appendix A) will read the lower triangle of a correlation matrix  $R_M$  and then read the values of  $\text{Var}(ID_i), i = 1, \dots, M$ . Elements of the variance-covariance matrix  $\sum_{i,j}^M$  are then computed by rearranging

terms in equation (5.10) to obtain

$$\sigma_{ij} = \rho_{ij} \sqrt{\sigma_i^2 \sigma_j^2}$$

or

$$\sigma_{ij} = \rho_{ij} \sqrt{\text{Var}(\text{ID}_i) \text{Var}(\text{ID}_j)} \quad (5.17)$$

The program also reads a specified diversion pattern vector  $\underline{q}$ , as given by equation (5.8), and a specified number of goal quantities  $Q$ . Then for a particular value of  $Q$ , the diversion strategy vector  $\underline{\mu}_M$  is computed using equation (5.9).

At this point in the program, the distribution of the vector  $\underline{\text{ID}}_M$  is defined as  $N_M(\underline{\mu}_M, \Sigma_M)$  and the procedure described in Section 5.1 is used to randomly generate  $\underline{\text{ID}}_M$  vectors from this distribution. Although  $M$  is a user option of the program,  $M = 10$  was used in the simulation study for this project.

The specific parameters used in the simulation study were:

1. Correlation matrices 1 and 2, given by equations (5.15) and (5.16), respectively, with  $\rho$  taking the values  $-.5, -.3, 0, .3, .5$ . Since matrices 1 and 2 are both the identity matrix  $I_{10}$  when  $\rho = 0$ , a total of nine distinct correlation matrices resulted.
2.  $\sqrt{\text{Var}(\text{ID}_i)} = 1$ , for all  $i$ . This implies that the unit of measure for this study is  $\sqrt{\text{Var}(\text{ID}_i)}$ , which is assumed constant for all periods.
3. The diversion pattern is  $\underline{q}' = [.1, .1, .1, .1, .1, .1, .1, .1, .1, .1]$  and values of  $Q$  are 0, 5, 6, 7, ..., 23, 24, 25, 30, 35. That is, twenty-four goal quantities are included.

It may seem that a very restrictive set of input parameters has been chosen for the simulation study. However, the results of the preliminary study showed that the desired comparisons among the four testing procedures can be made when only these parameters are used, and the results of these comparisons can be interpreted quite generally.

### 5.3 DECISION RULES AND CRITERIA FOR COMPARISONS

Since each simulated process will be monitored over ten material balance periods, the results will depend on the action taken when a statistical test indicates a significant loss. The output from the computer program used in this study allows the user to evaluate the results under either of two decision rules.

Decision Rule 1: Given a diversion strategy vector  $\underline{\mu}_{10}$ , if a statistically significant outcome is observed at any period, it is assumed that the diversion strategy has been detected and corrected. Thus, under this decision rule, only first-time detection statistics are accumulated for the four testing procedures over the 2000 simulations.

Decision Rule 2: Given a diversion strategy vector  $\underline{\mu}_{10}$ , if a statistically significant outcome is observed at period  $i$ , the calculation of CID is restarted with period  $i + 1$ , but the diversion strategy is assumed not to have been corrected. Tabulation of the results under this assumption requires a ten-by-ten contingency table, where columns represent periods and rows represent restarts of the process. Any row or set of rows can be examined. This allows the flexibility of investigating different options for responding to statistically significant outcomes. This also allows examination of columns when it is of interest to investigate the results at a particular period.

Due to the voluminous nature of the computer output under decision rule 2, it is not possible to present such detailed results in this report. However, under decision rule 1, it is very convenient to compute and display summary statistics which require very little space but give sufficient detail to accommodate meaningful comparisons among the four testing procedures. The following criteria are used to make the comparisons.

First-Time Detection Probability: For a given diversion strategy vector  $\underline{\mu}_{10}$ , this is the probability of detecting the diversions for the first time at or before the tenth period. This is a measure of the power of a testing procedure. Thus, high probabilities are desirable.

False Alarm Rate: This is simply the "first-time detection probability", when the goal quantity is zero (i.e., when no material is being diverted). It is desirable to control the false alarm rate at an acceptable low level.

Expected Undetected Losses: This is another measure of the power of a testing procedure. To see this, suppose an amount  $\mu_i$  is diverted from each period. If detected at the  $k^{\text{th}}$  period, the diverter would have successfully diverted material from the previous  $k-1$  periods. That is, a cumulative amount

$$L = \sum_{i=1}^{k-1} \mu_i \quad (5.18)$$

was diverted and was not detected.

Thus, if a particular testing procedure is more powerful than the others for detecting constant ("trickle") diversions, it should (on the average) detect the losses sooner than the other methods. In other words, the cumulative undetected loss should be smaller (on the average) when the most powerful testing procedure is used than when any of the other procedures are used.

#### 5.4 RESULTS OF THE SIMULATION

This section presents the results of the simulation study. This is simply a series of graphs showing the results from 2000 simulations of each of the 216 processes defined by the nine correlation matrices in combination with the twenty-four goal quantities. For each of the nine correlation matrices there are two graphs. One graph shows first-time detection probabilities for the four testing procedures, plotted against the goal quantities. The second graph shows expected undetected losses for the four testing procedures plotted against the goal quantities.

For convenience in plotting and discussing the results, the four testing procedures will be labeled as follows:

- Procedure 1: the bivariate procedure
- Procedure 2: the test of  $ID_k$  only
- Procedure 3: the test of  $CID_k$  only
- Procedure 4: the Bonferroni procedure

As a starting and reference point, the case  $\rho = 0$  is considered first. In this case the correlation matrices given by equations (5.15) and (5.16) are both the identity matrix  $I_{10}$ , given by equation (2.5). That is, there are no correlations among the ten ID's. Figure 5.1 shows the first-time detection probabilities for the four procedures on the vertical axis with  $\underline{\mu}_i$ , the amount diverted from each period, on the horizontal axis. Recall that  $\mu_i$  is constant over the ten periods, so that  $\mu_i = Q/10$  (i.e.,  $\mu_i$  is 10% of the goal quantity). Examining Figure 5.1, Procedure 2 has the highest false alarm rate but is slightly less powerful than the other procedures when  $.7 \leq \mu_i \leq 1.5$ .

Looking now at Figure 5.2, the expected undetected losses are noticeably smaller for Procedure 3 when  $.7 \leq \mu_i \leq 2.5$ . Procedures 1 and 4 are indistinguishable and have expected losses that are just slightly lower than Procedure 2.

Thus, if there were no correlations among the successive ID's, monitoring  $CID_k$  only offers the best protection against trickle losses.

Consider next Matrix 1, given by equation (5.15), with  $\rho = -.3$ . Figures 5.3 and 5.4 reveal that Procedure 2 has the highest false alarm rate, but is considerably less powerful than the other procedures when  $.5 \leq \mu_i \leq 1.8$ . Procedures 1 and 4 are practically indistinguishable and are only slightly less powerful than Procedure 3 which is most powerful. Figures 5.5 and 5.6 are the results when Matrix 2, given by equation (5.16), is used with  $\rho = -.3$ . The results are practically identical with those in Figures 5.3 and 5.4.

Figures 5.7 and 5.8 show the results for Matrix 1 when  $\rho = -.5$ . Clearly, Procedures 1, 3 and 4 are significantly more powerful than Procedure 2 when  $\mu_i \leq 2$ . Note also that Procedure 2 has a higher false alarm rate than the other procedures. Figures 5.9 and 5.10 show the results for Matrix 2 when  $\rho = -.5$ . There are some differences between these and Figures 5.7 and 5.8 due to the positive elements  $\rho^2$ ,  $\rho^4$ , etc. in Matrix 2. These positive elements tend to slightly decrease the power of Procedures 1, 3 and 4.

Figures 5.11 and 5.12 show the results for Matrix 1 when  $\rho = .3$ . Figures 5.13 and 5.14 show the results for Matrix 2 when  $\rho = .3$ . There are really no noticeable differences between results for the two matrices. In both cases, Procedure 2 has a much higher false alarm rate than the others but is slightly more powerful than the others when  $\mu_i \leq 1.0$ .

Figures 5.15 and 5.16 show the results for Matrix 1 when  $\rho = .5$ . Actually there is very little change from the results when  $\rho = .3$ . However, there has been a slight upward shift in the expected undetected losses for all procedures.

Figures 5.17 and 5.18 show the results for Matrix 2 when  $\rho = .5$ . Here, the nonzero (positive) off diagonal elements have caused Procedures 1, 3 and 4 to be noticeably less powerful than in the previous examples. This is especially evident in Figure 5.18 showing a general upward shift in expected undetected losses. In this case, Procedure 2 is more powerful than any of the others when  $\mu_i \leq 2$ . When  $\mu_i > 2$ , Procedures 2 and 3 are slightly more powerful than Procedures 1 and 4.

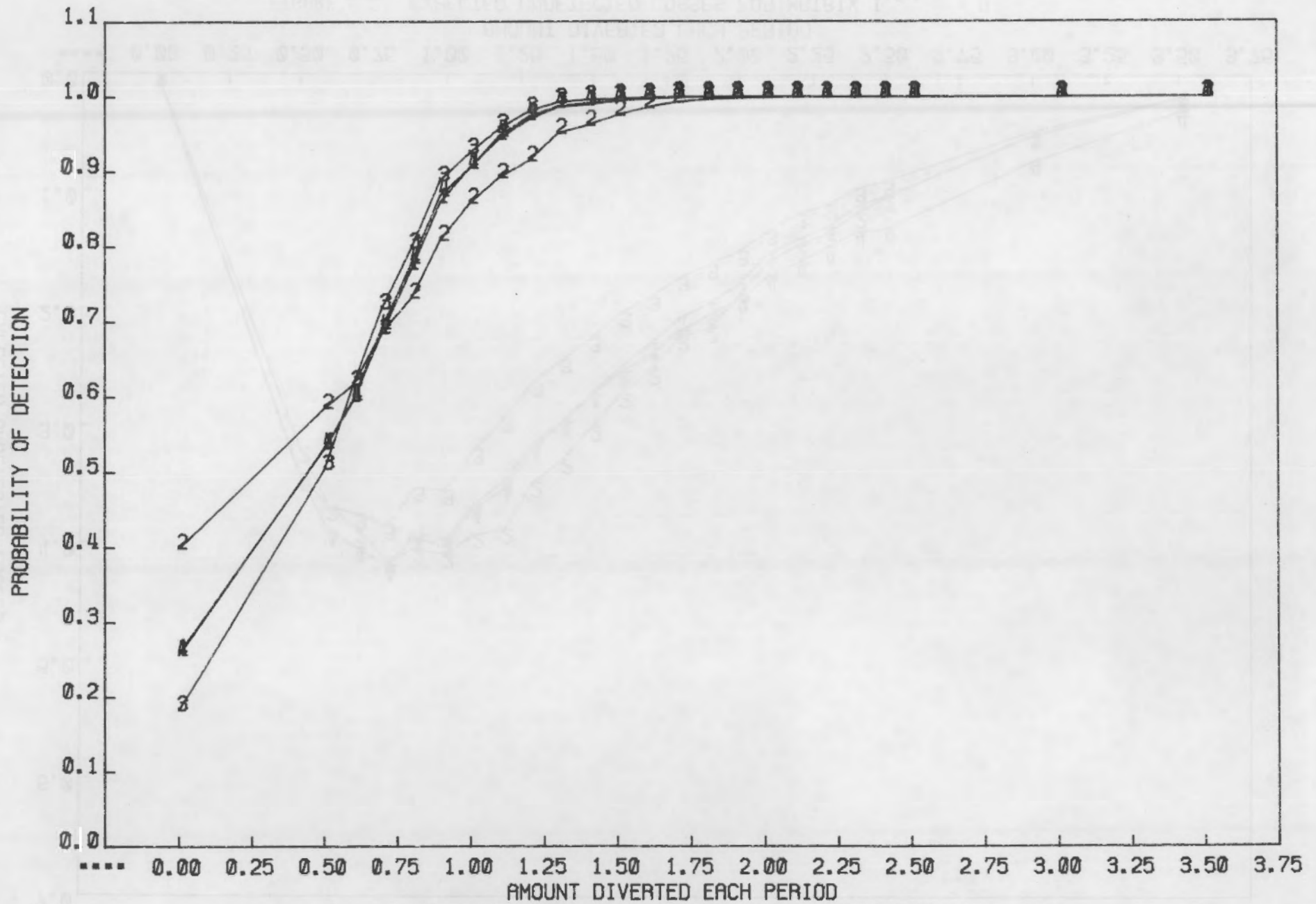


FIGURE 5.1 FIRST TIME DETECTION PROBABILITIES FOR MATRIX  $I_{10}$ ,  $\rho = 0$



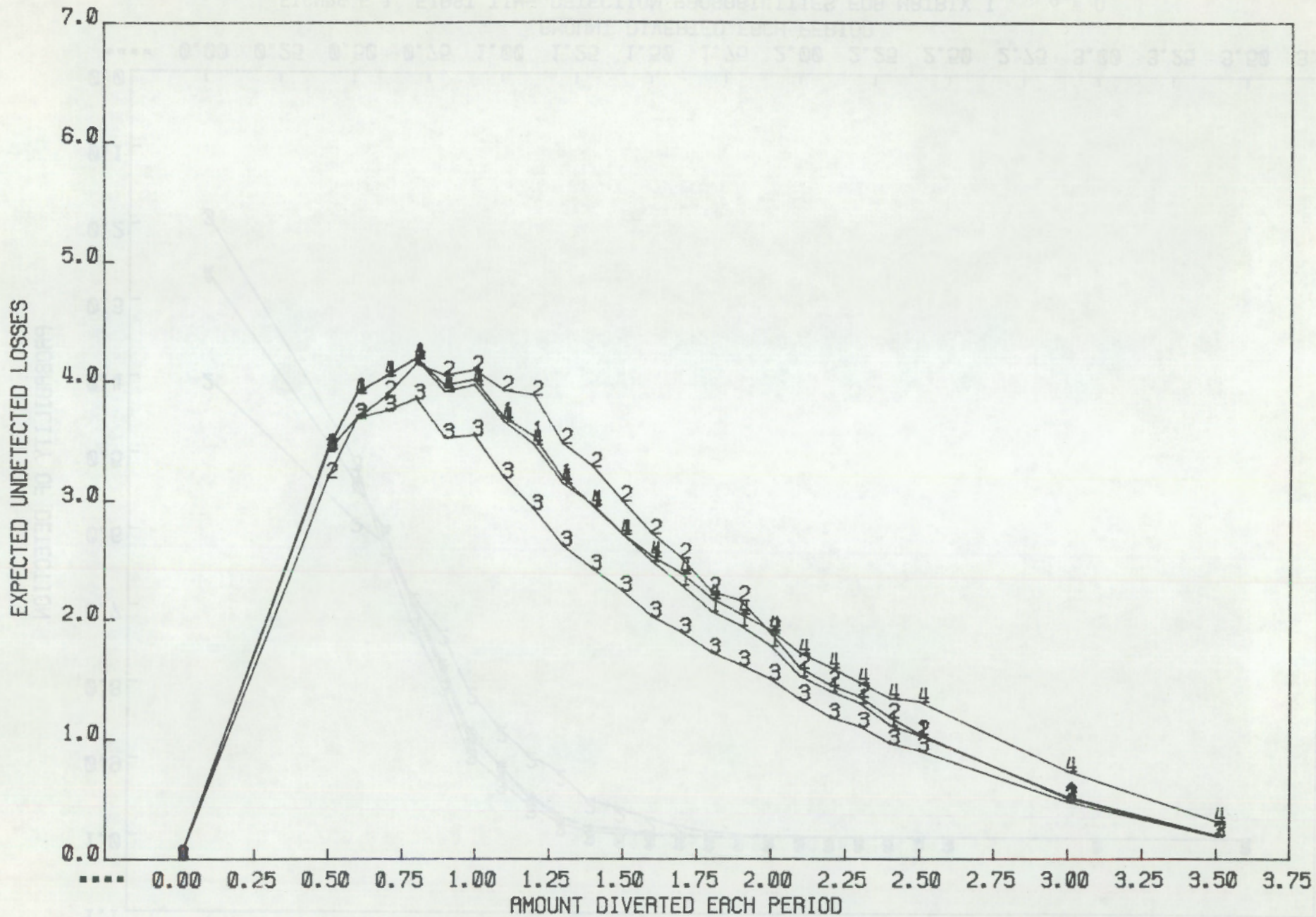


FIGURE 5.2 EXPECTED UNDETECTED LOSSES FOR MATRIX  $I_{10}$ ,  $\rho = 0$



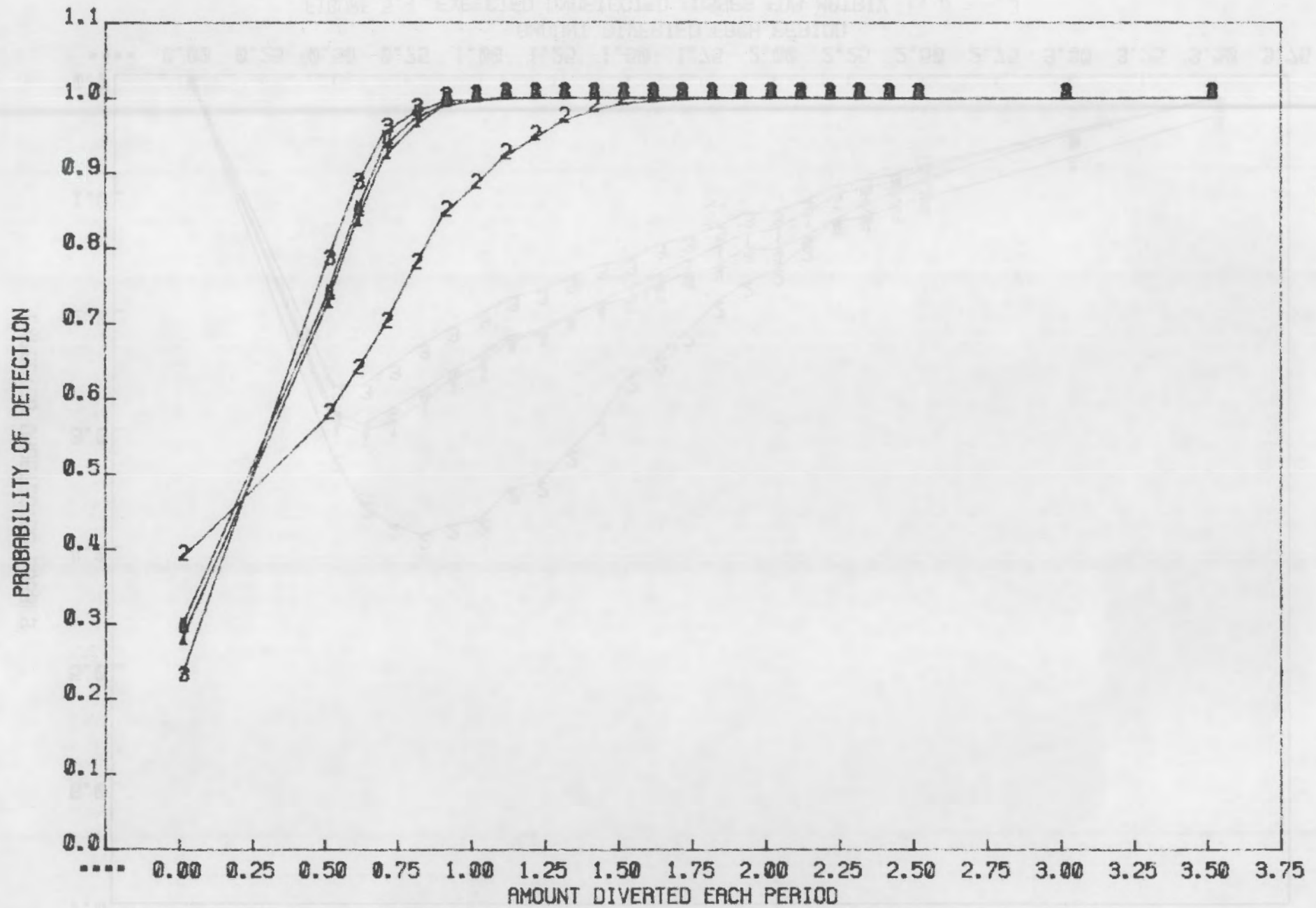


FIGURE 5.3 FIRST TIME DETECTION PROBABILITIES FOR MATRIX 1,  $\rho = -.3$

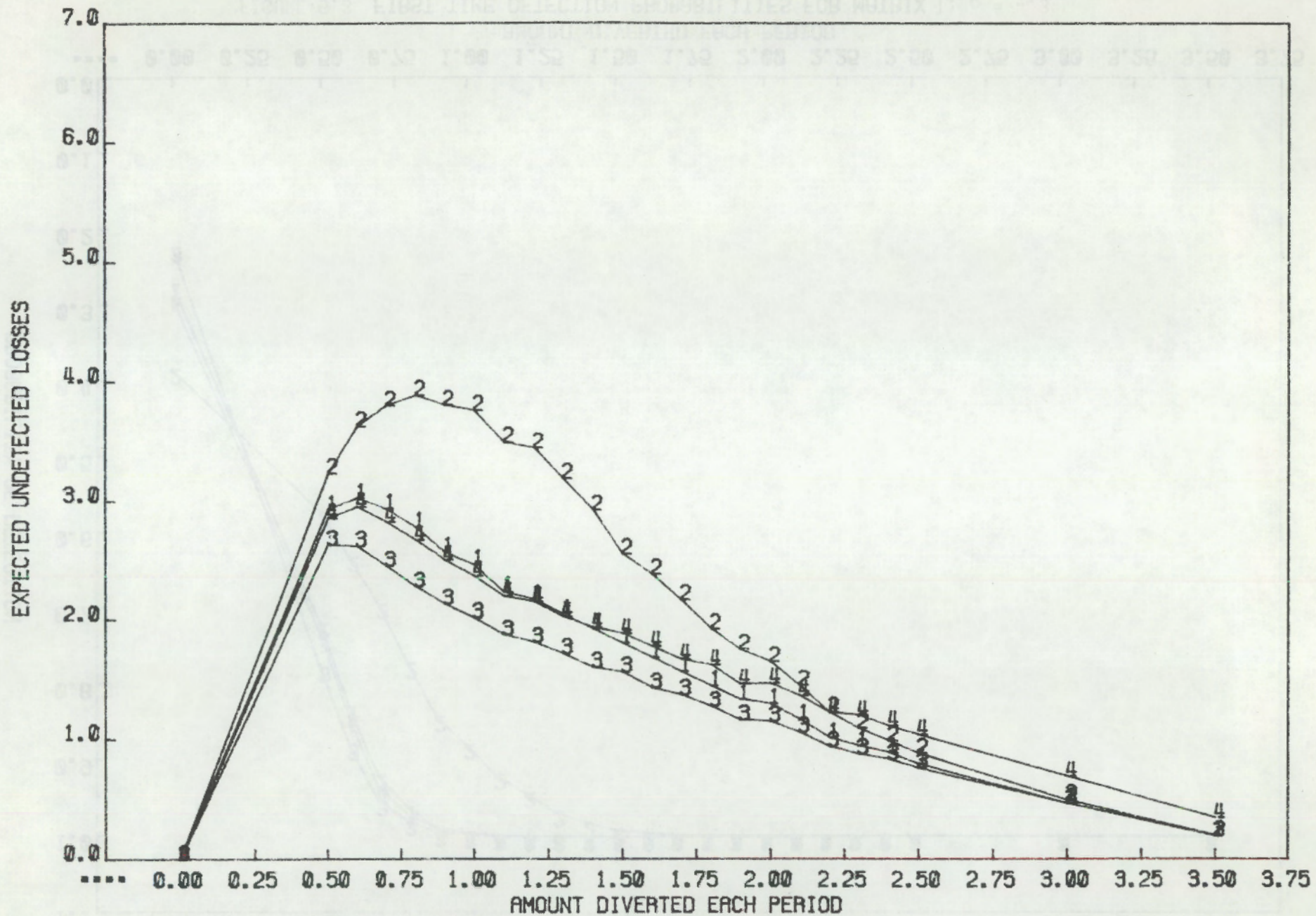


FIGURE 5.4 EXPECTED UNDETECTED LOSSES FOR MATRIX 1,  $\rho = -.3$



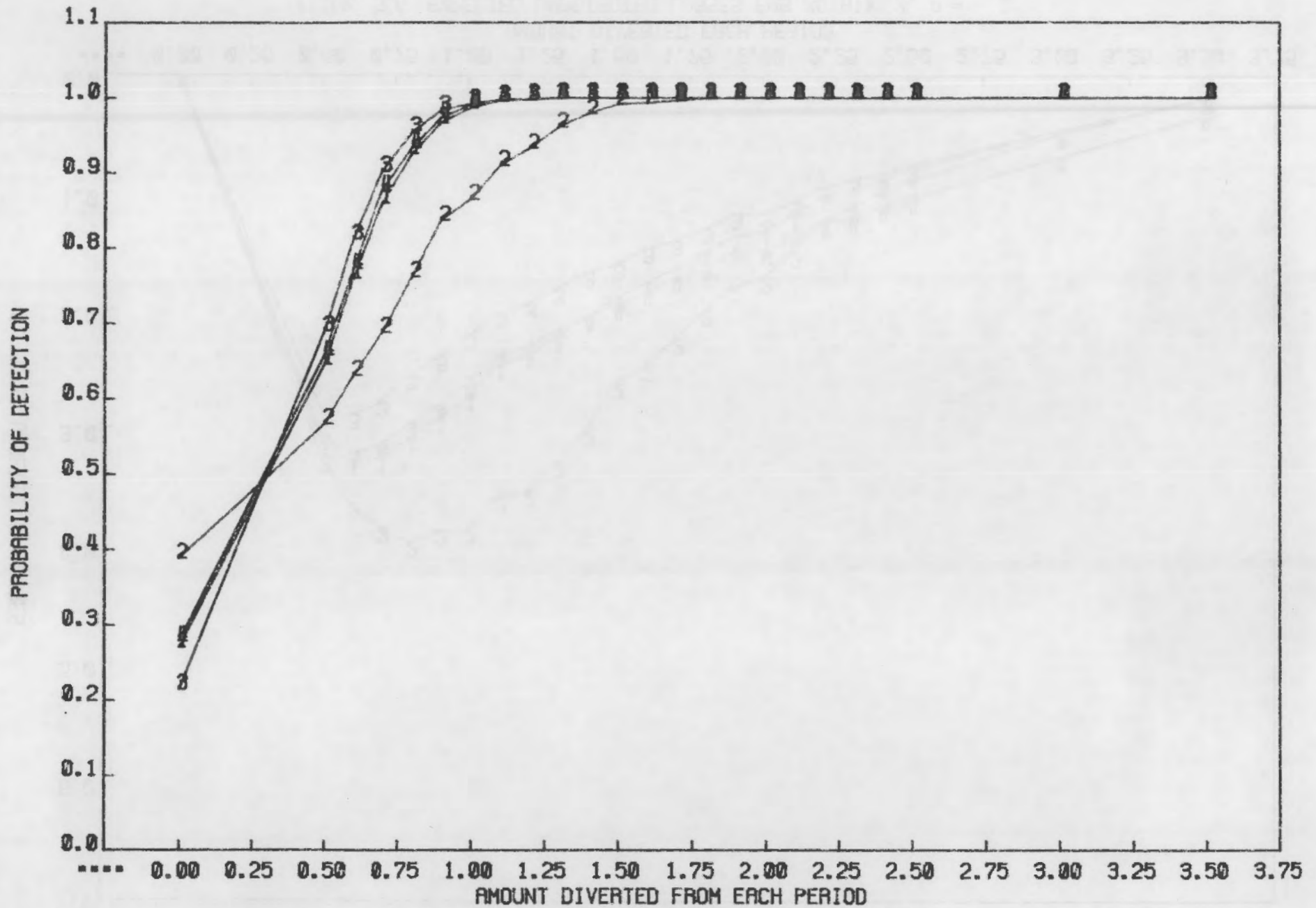


FIGURE 5.5 FIRST TIME DETECTION PROBABILITIES FOR MATRIX 2,  $\rho = -.3$

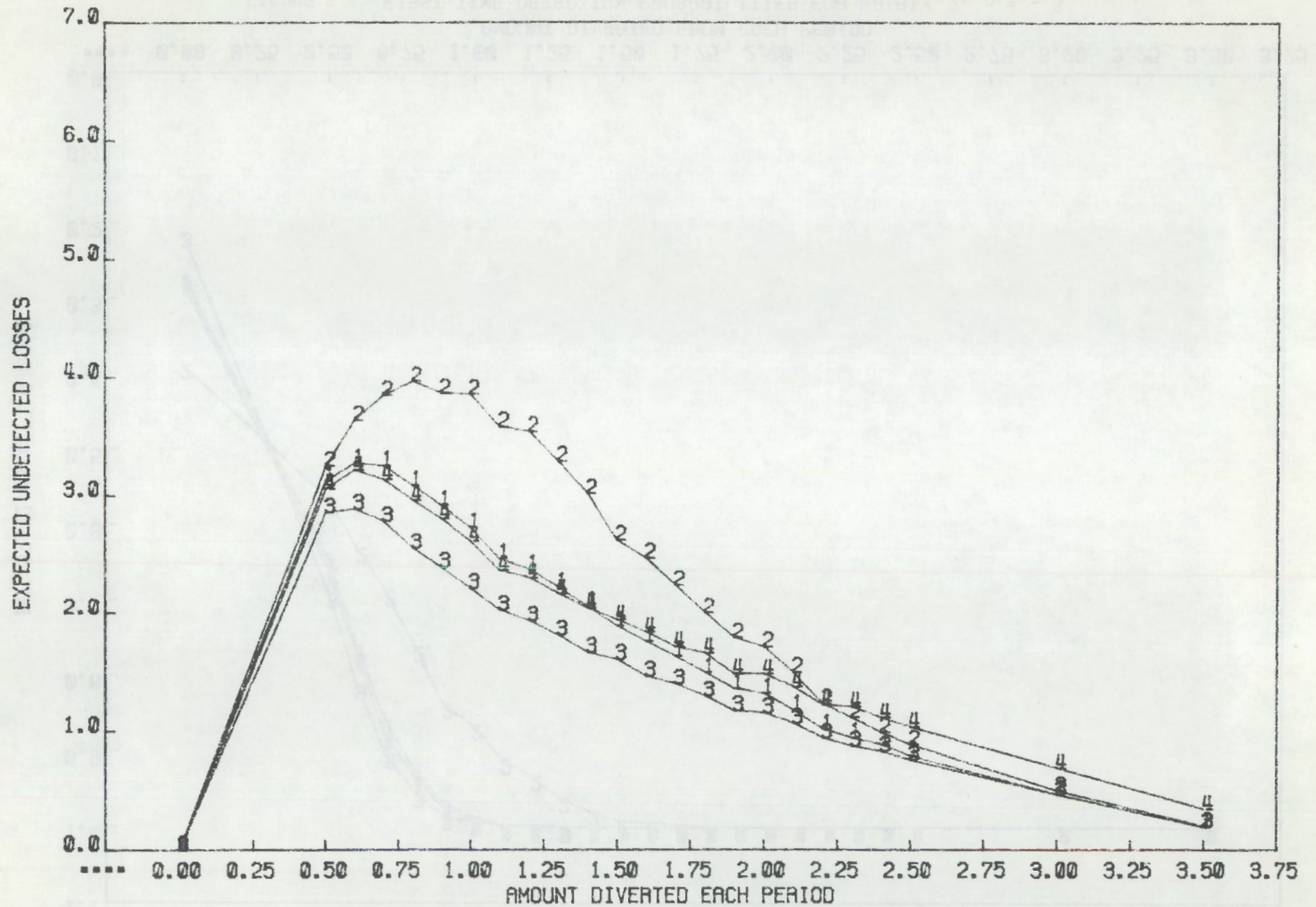


FIGURE 5.6 EXPECTED UNDETECTED LOSSES FOR MATRIX 2,  $\rho = -.3$

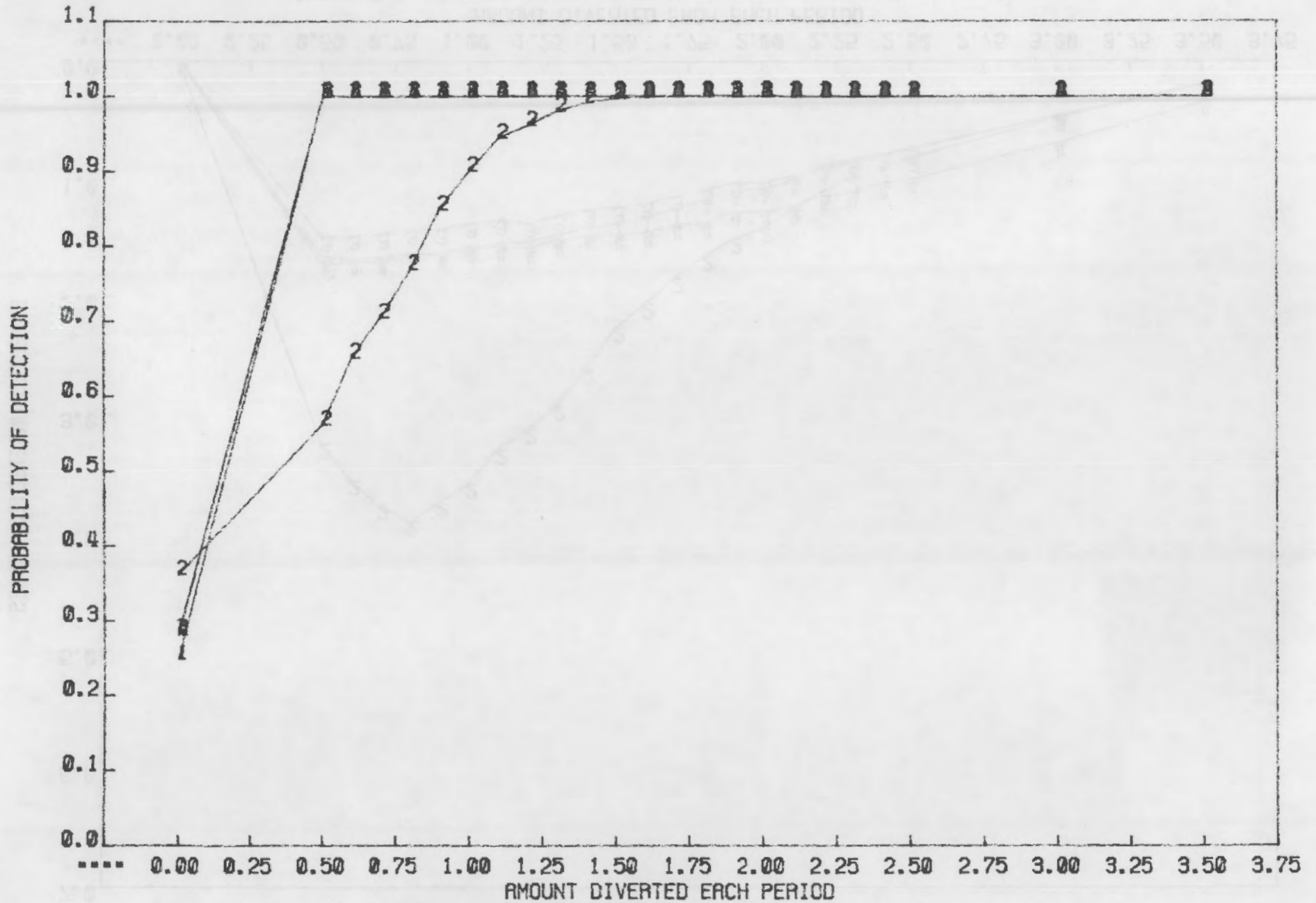


FIGURE 5.7 FIRST TIME DETECTION PROBABILITIES FOR MATRIX 1,  $\rho = -0.5$



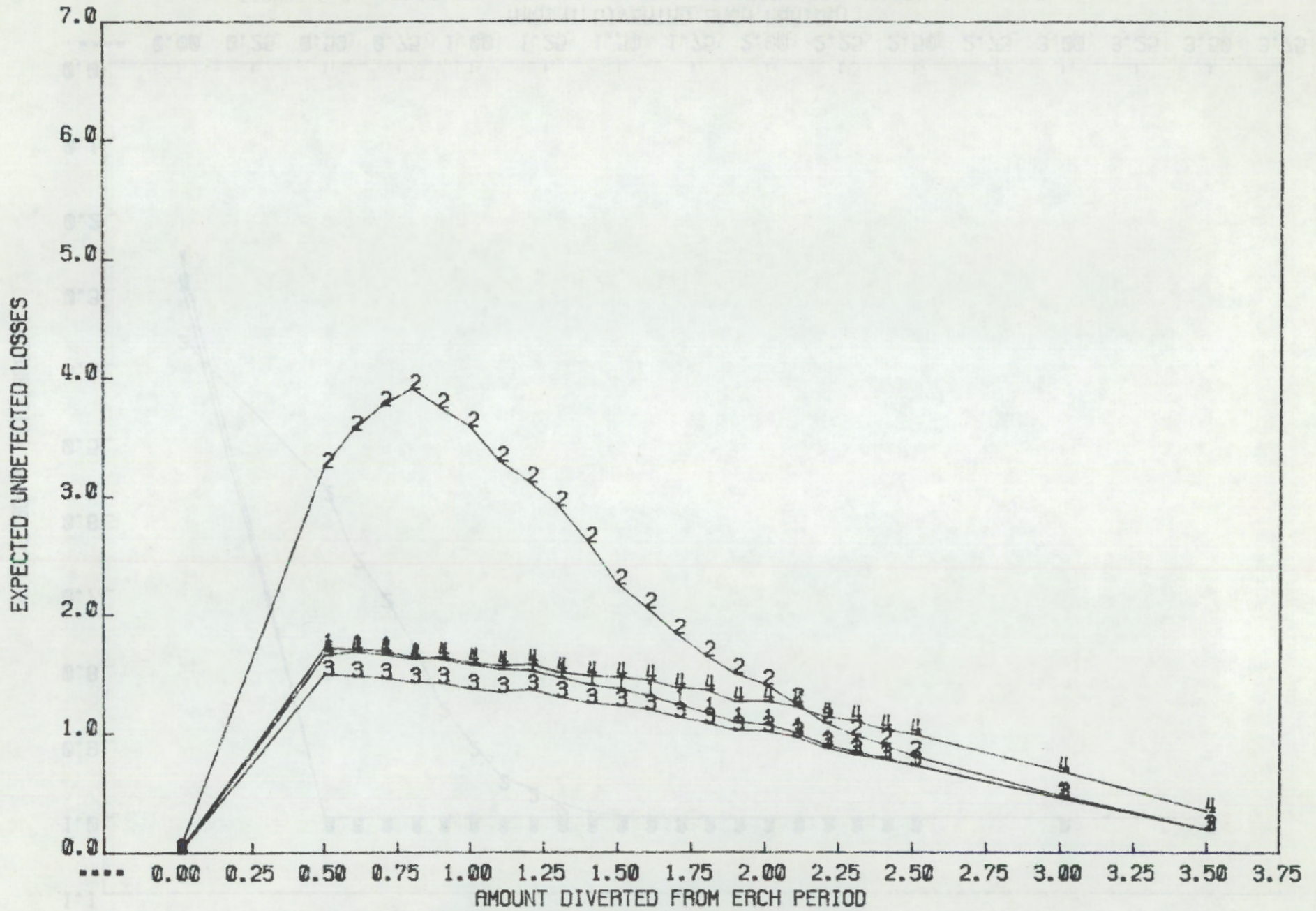


FIGURE 5.8 EXPECTED UNDETECTED LOSSES FOR MATRIX 1,  $\rho = -.5$

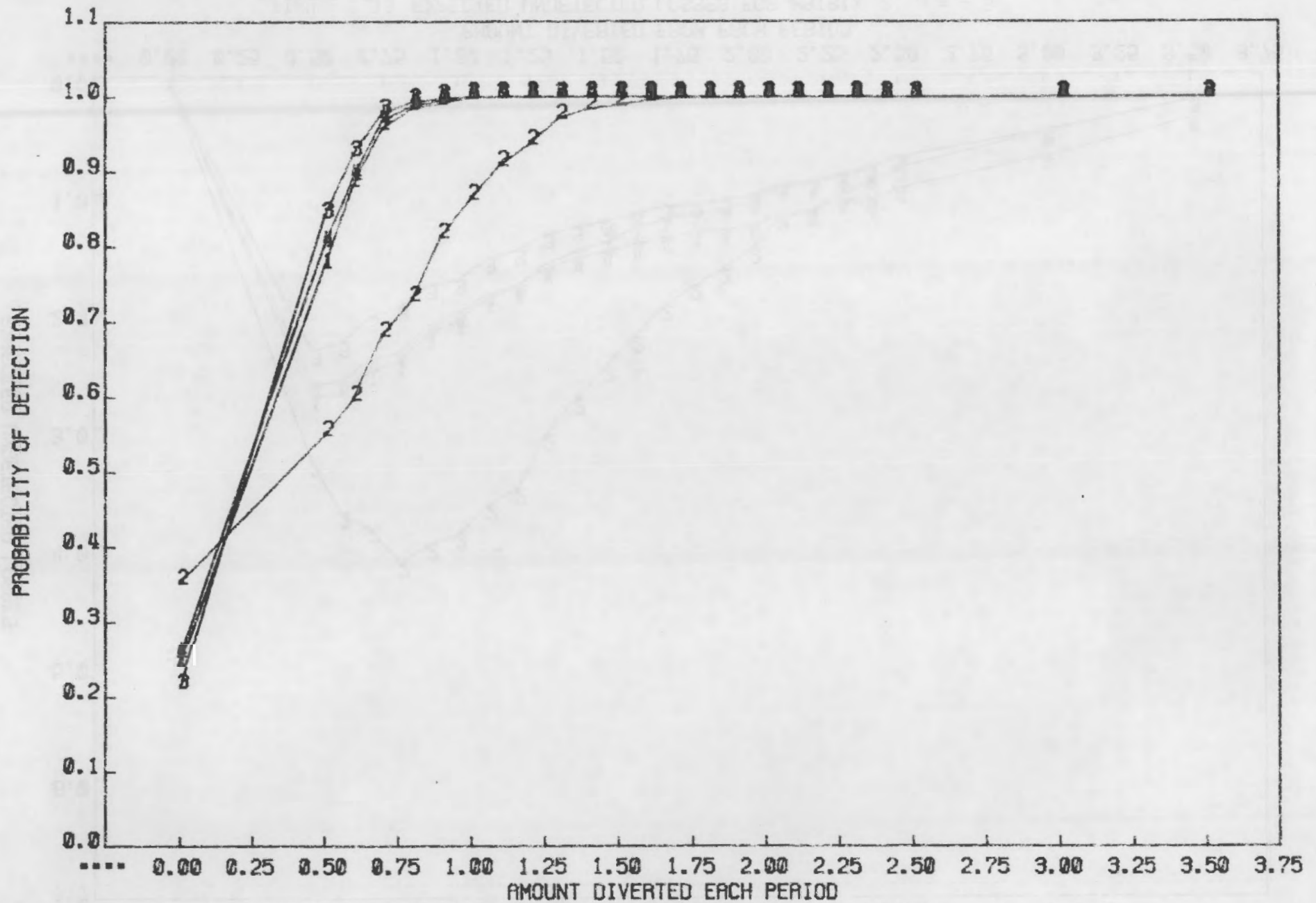


FIGURE 5.9 FIRST TIME DETECTION PROBABILITIES FOR MATRIX 2,  $\rho = -.5$



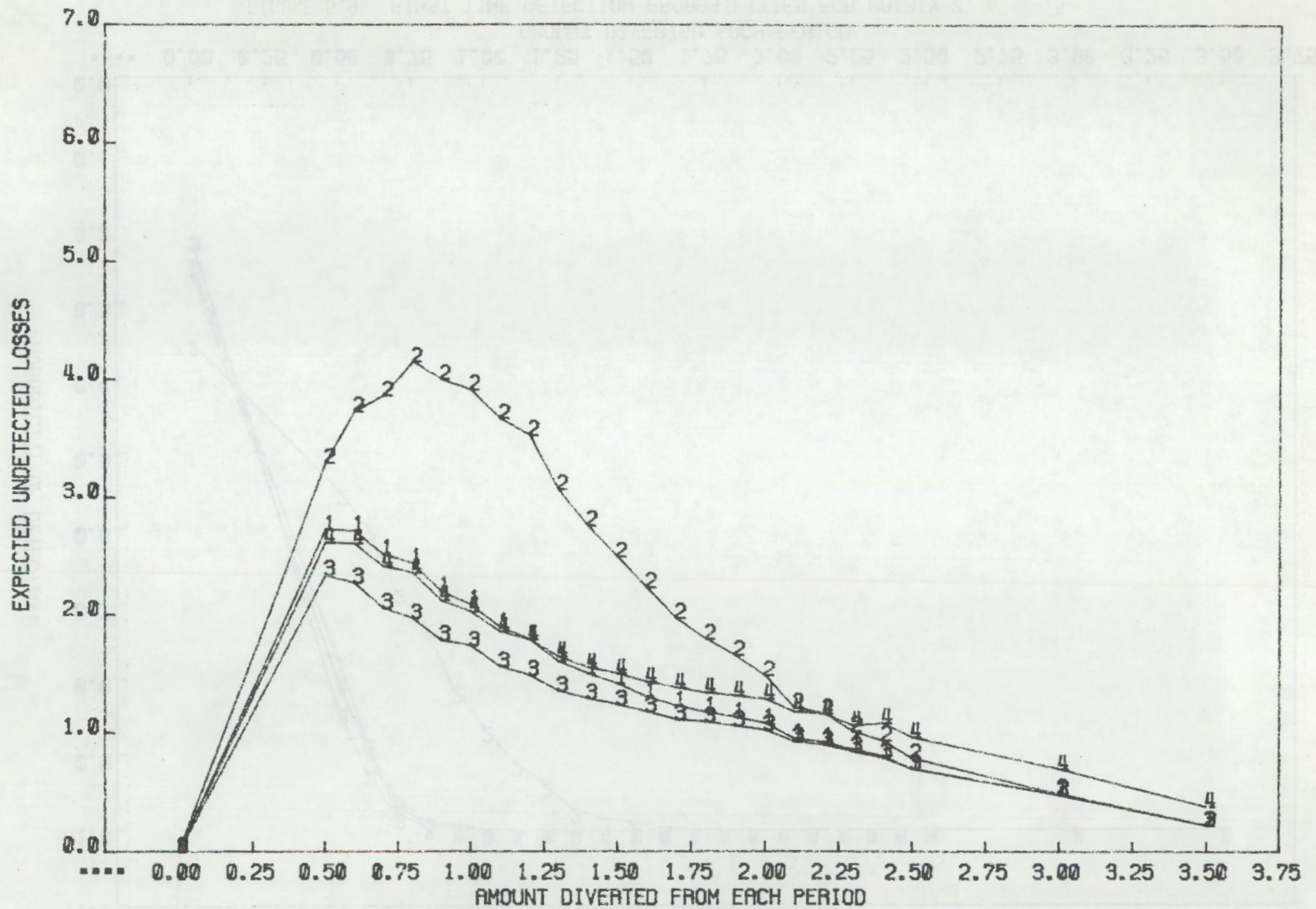


FIGURE 5.10 EXPECTED UNDETECTED LOSSES FOR MATRIX 2,  $\rho = -.5$



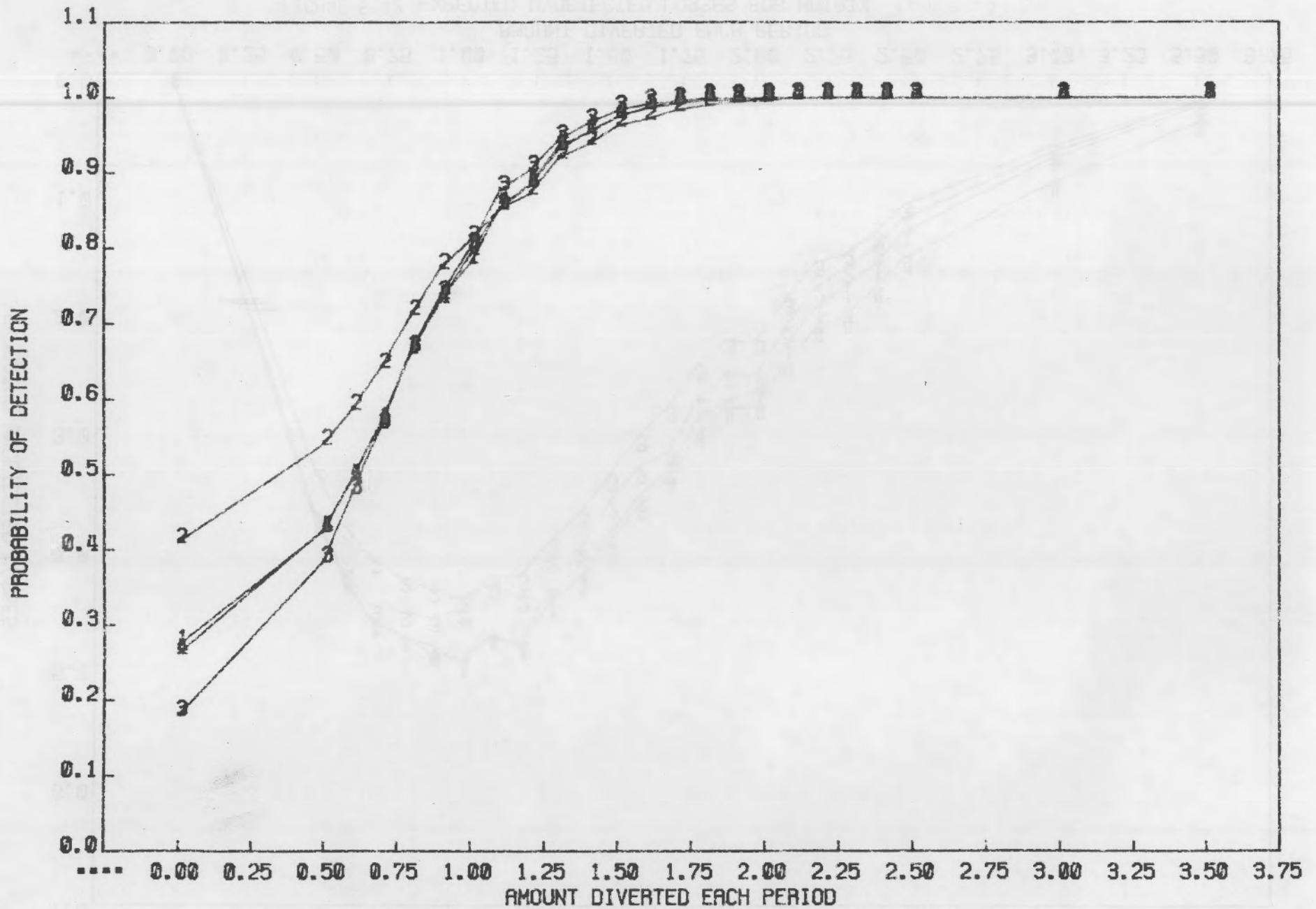


FIGURE 5.11 FIRST TIME DETECTION PROBABILITIES FOR MATRIX 1,  $\rho = .3$

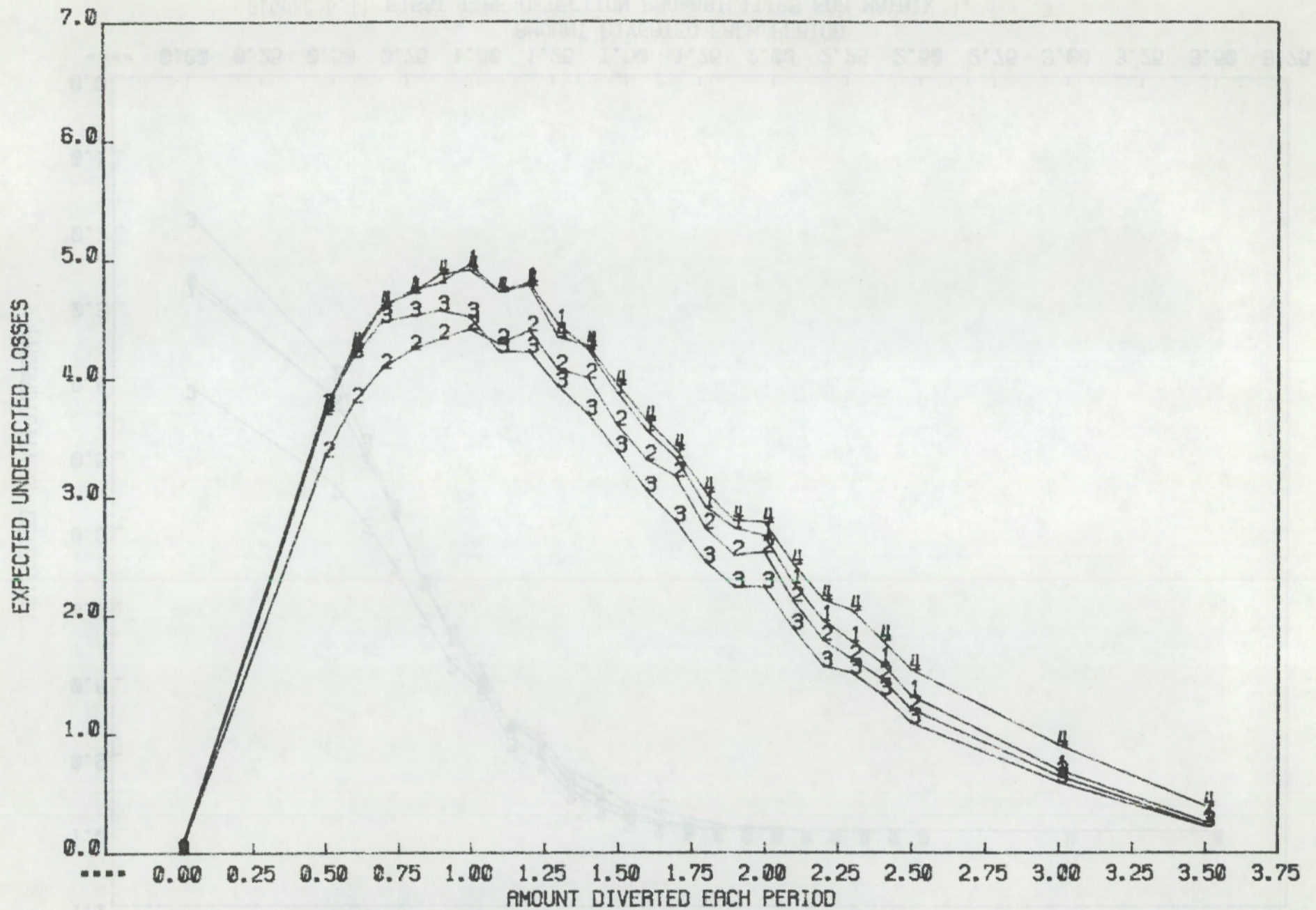


FIGURE 5.12 EXPECTED UNDETECTED LOSSES FOR MATRIX 1,  $\rho = .3$



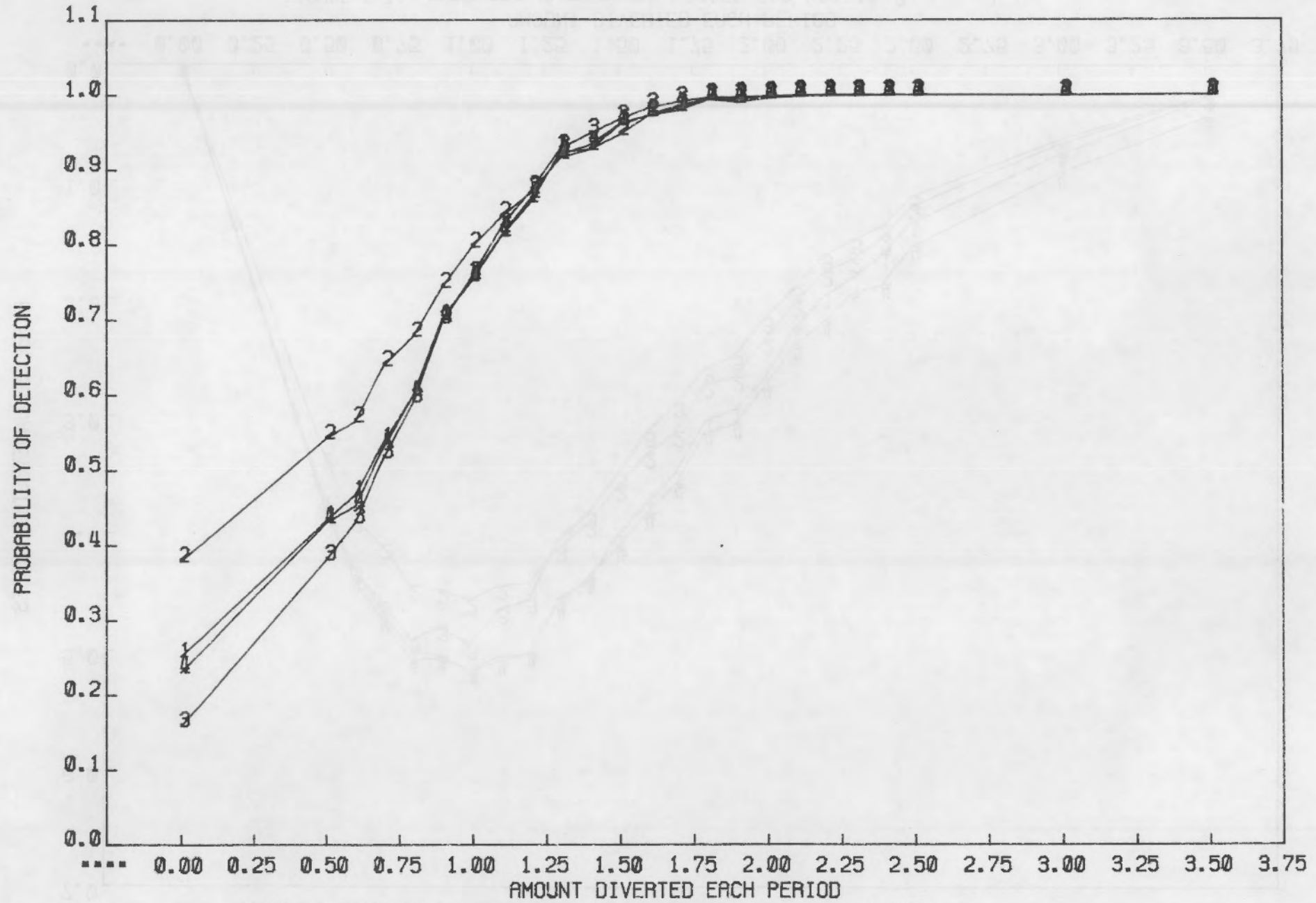


FIGURE 5.13 FIRST TIME DETECTION PROBABILITIES FOR MATRIX 2,  $\rho = .3$

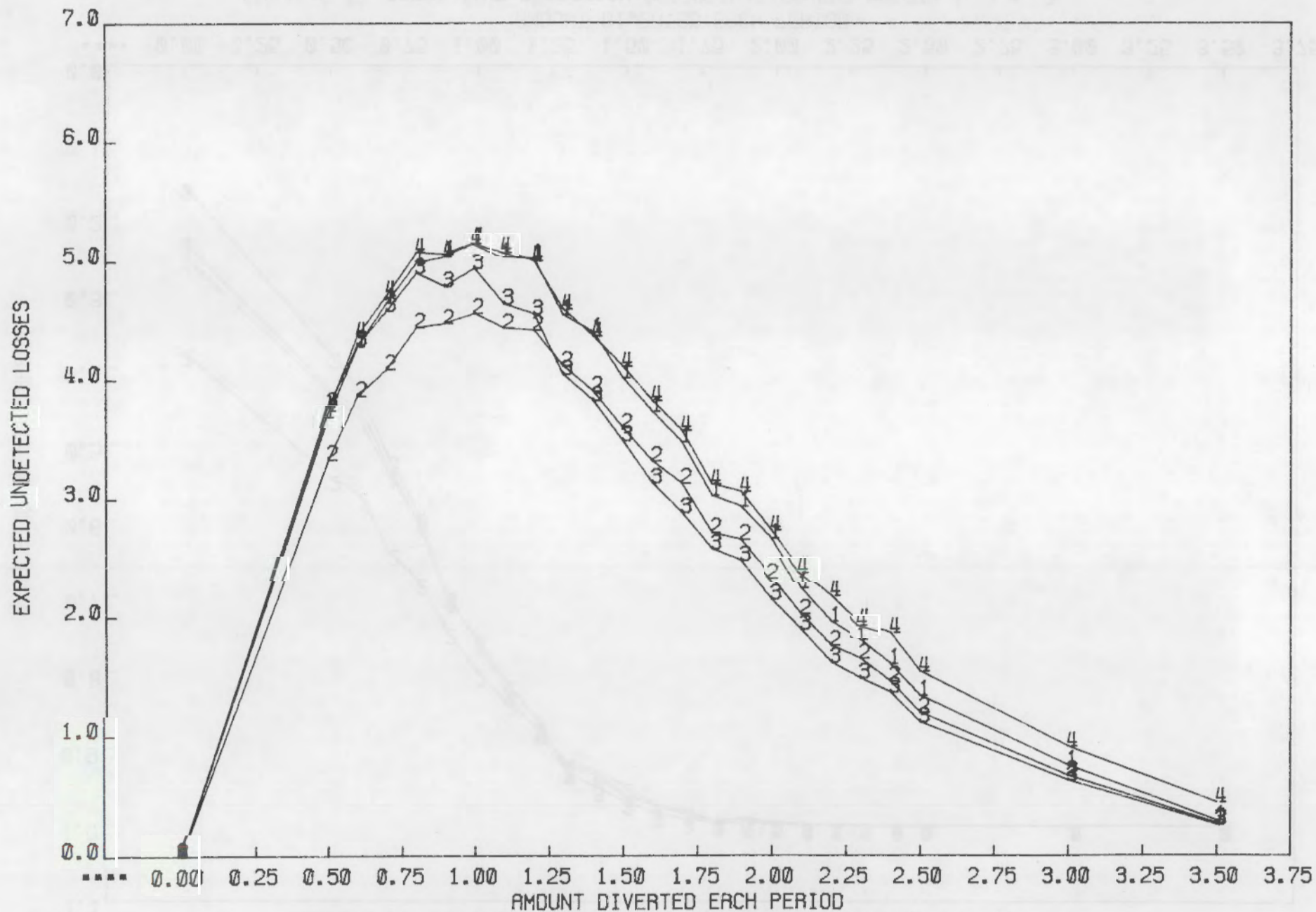


FIGURE 5.14 EXPECTED UNDETECTED LOSSES FOR MATRIX 2,  $\rho = .3$

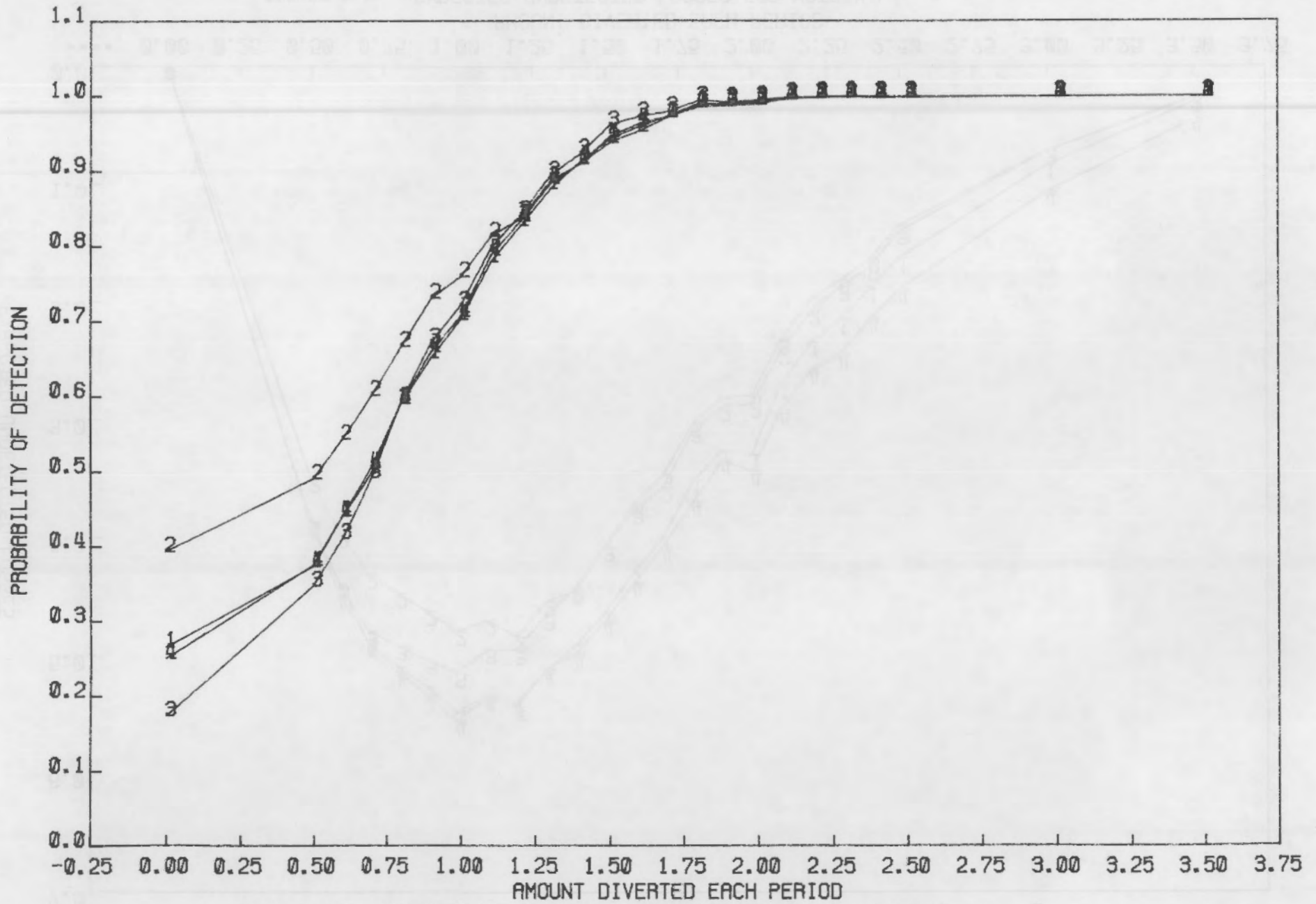


FIGURE 5.15 FIRST TIME DETECTION PROBABILITIES FOR MATRIX 1,  $\rho = .5$

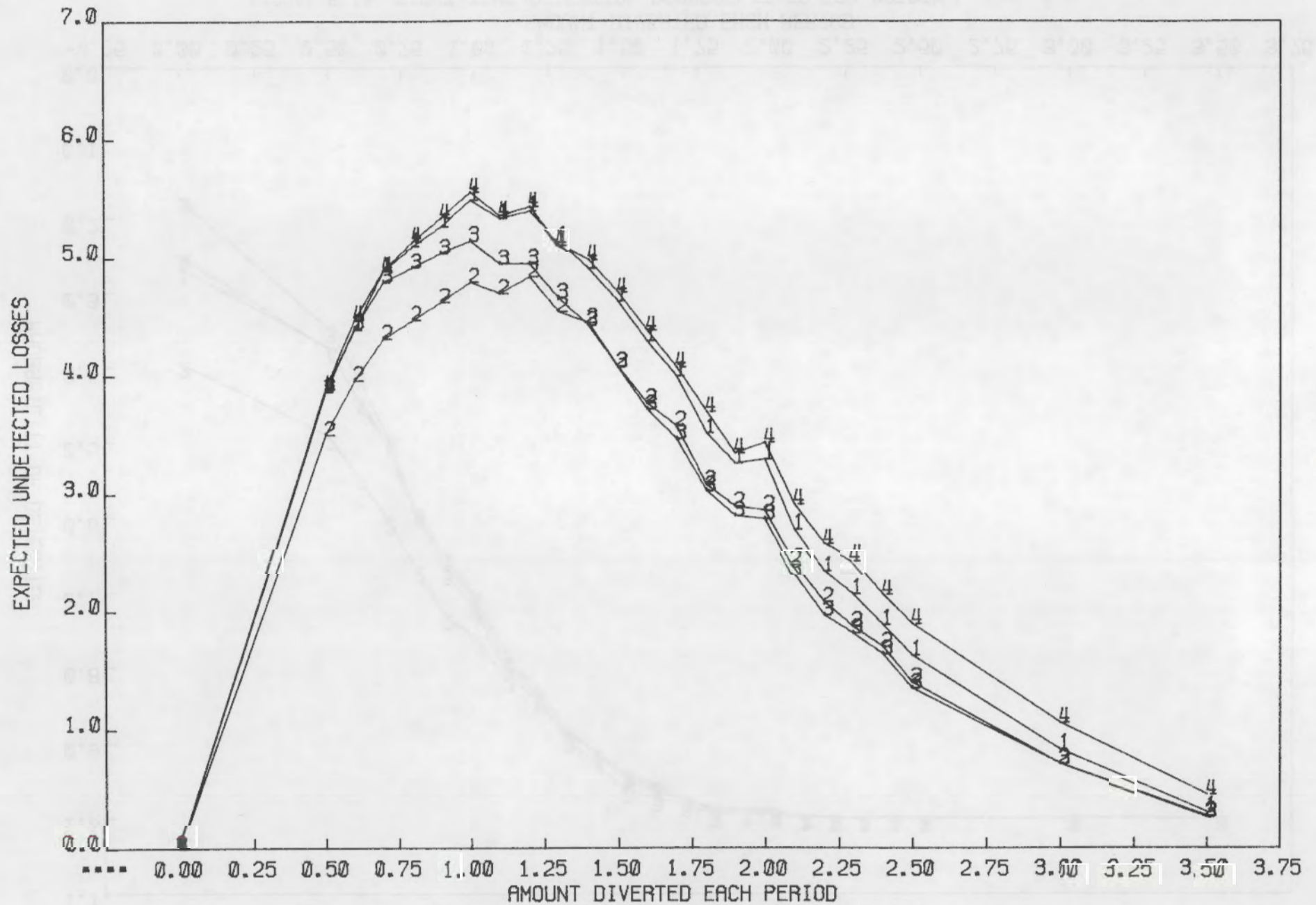


FIGURE 5.16 EXPECTED UNDETECTED LOSSES FOR MATRIX 1,  $\rho = .5$



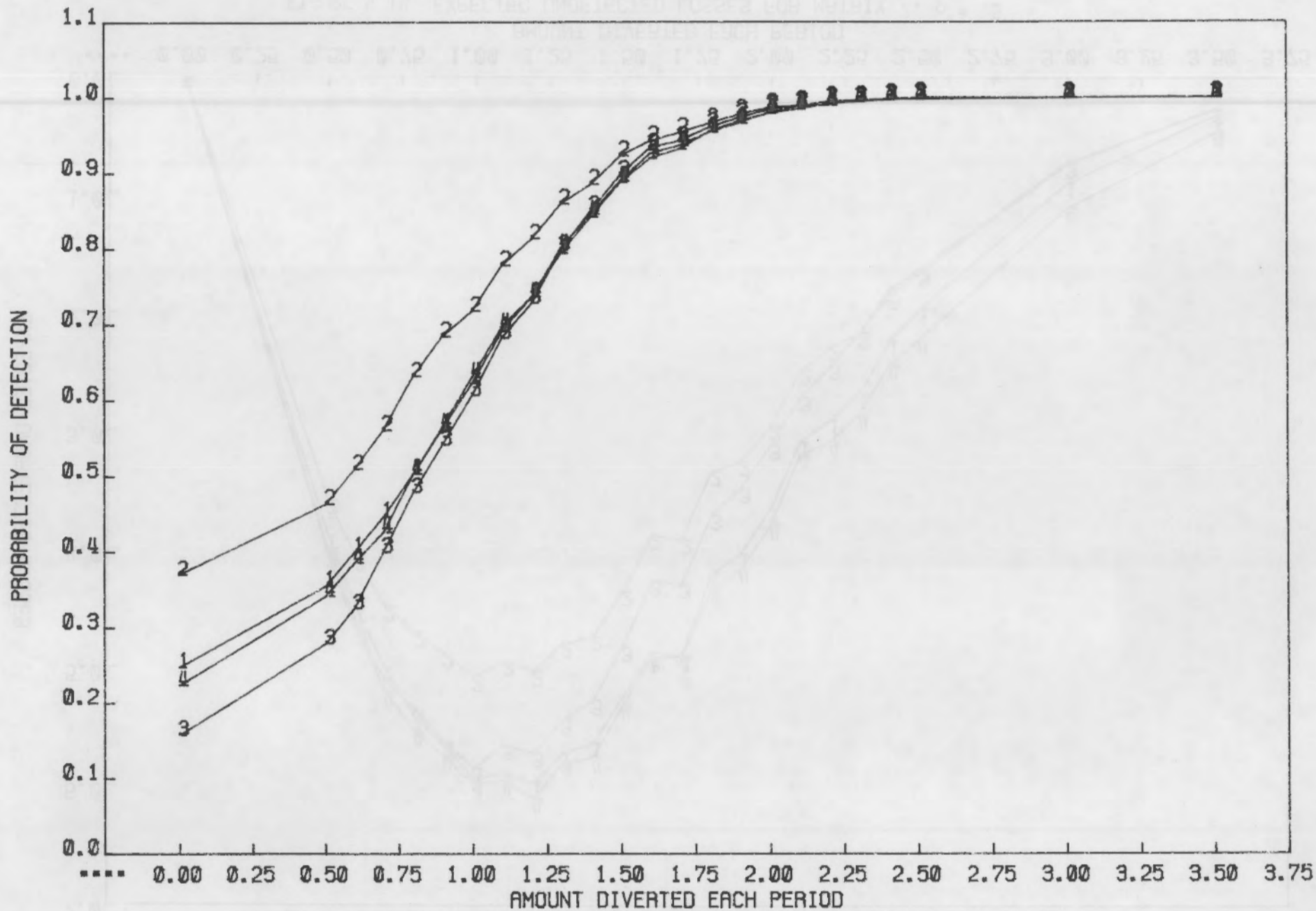


FIGURE 5.17 FIRST TIME DETECTION PROBABILITIES FOR MATRIX 2,  $\rho = .5$

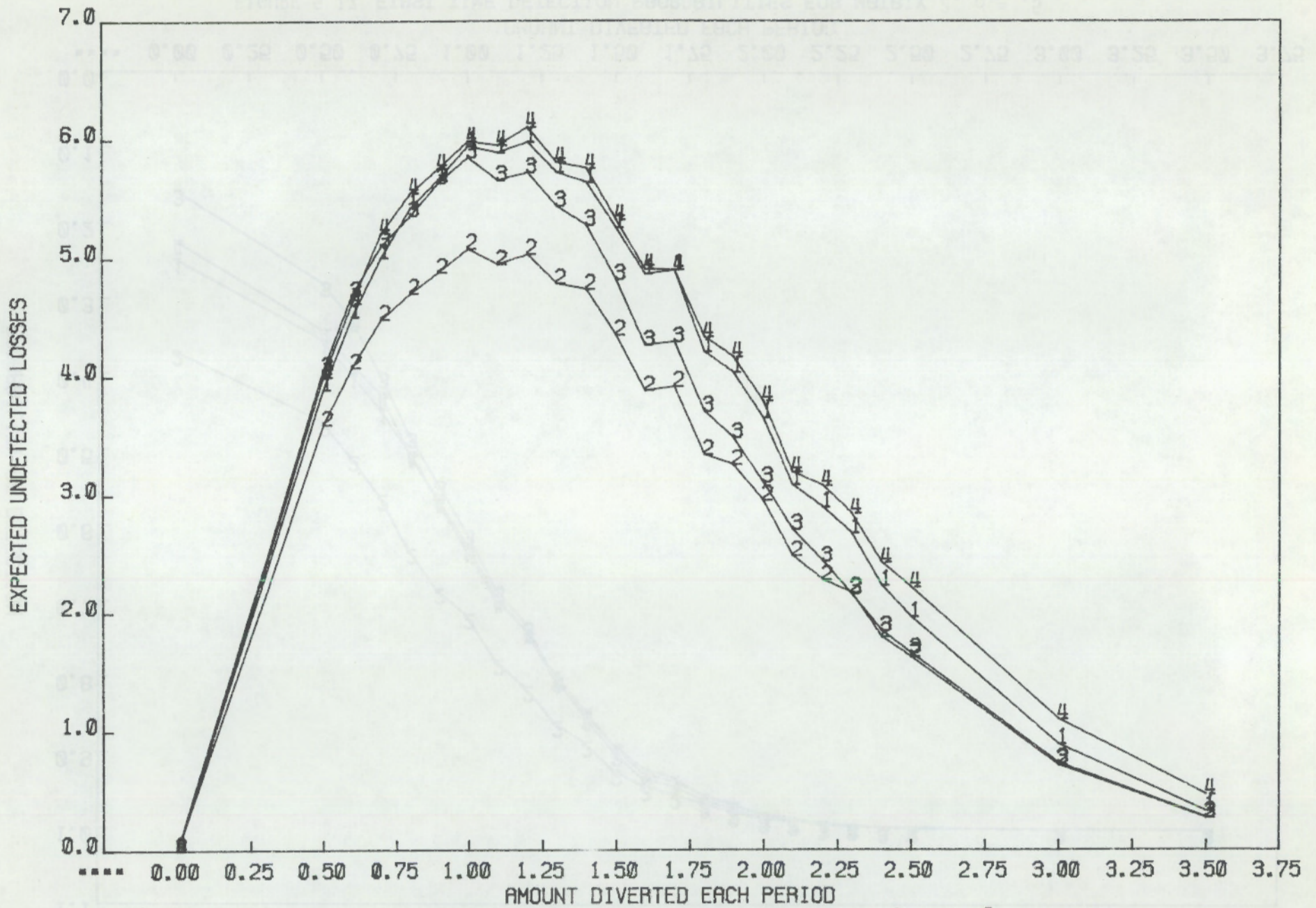


FIGURE 5.18 EXPECTED UNDETECTED LOSSES FOR MATRIX 2,  $\rho = .5$



## 6.0 CONCLUSIONS

The results of the simulation study, presented in Section 5.4, illustrate that when the correlations among successive ID's are predominantly less than or equal to zero, the bivariate and Bonferroni procedures are considerably more powerful for detecting trickle losses than the currently required test of ID only. However, when correlations among the successive ID's are predominantly greater than zero, all tests which use  $\text{Var}(\text{CID}_k)$  are generally less powerful than the test of ID only.

This is because, from equation (2.14),  $\text{Var}(\text{CID}_k)$  includes the sum of the covariances among the  $k$  ID's. When the covariances are predominantly negative,  $\text{Var}(\text{CID}_k)$  is reduced giving a more powerful test, and when the covariances are predominantly positive,  $\text{Var}(\text{CID}_k)$  is increased giving a less powerful test. In all cases considered, the currently required test of ID only had the highest false alarm rate.

Thus, if the assumptions made in references (1) and (3) are at all realistic, the correlation structure for successive ID's should be similar to either Matrix 1 or Matrix 2, equations (5.15) and (5.16), respectively, with  $\rho < 0$ . In this case, there is a definite advantage in implementing the bivariate procedure.

From a practical point of view, however, only licensees who have access to computerized recordkeeping systems should attempt to implement the bivariate procedure. This is because of the more sophisticated inventory recordkeeping which is required to accommodate the calculation of  $\text{Var}(\text{CID}_k)$ .

The simulation results for the bivariate and Bonferroni procedures are practically indistinguishable. Thus, it may seem that the Bonferroni procedure would be preferable, since it avoids the estimation of  $\text{Cov}(\text{ID}_k, \text{CID}_k)$  and the matrix calculations which are necessary for implementation of the bivariate procedure. However, both procedures require the more sophisticated recordkeeping and the calculation of  $\text{Var}(\text{CID}_k)$ . Thus, using equation (3.7), the estimation of  $\text{Cov}(\text{ID}_k, \text{CID}_k)$  is trivial and should not be a consideration in selecting one method over the other. Also, calculations involving two-element vectors and two-by-two matrices are quite simple, even if done by hand. Therefore, since the significance level can be exactly specified for the bivariate procedure but not for the Bonferroni procedure, the bivariate procedure is preferred.



## REFERENCES

1. Wincek, M. A., Stewart, K. B. and Piepel, G. F., Statistical Methods for Evaluating Sequential Material Balance Data. NUREG/CR-0683(PNL-2920), National Technical Information Service, Springfield, VA, April 1979.\*
2. Merrill, J. A. and Brouns, R. J., "A Statistical Approach to Simultaneous Tests for Both Short Term and Long Term Material Losses", Nuclear Materials Management, 8, Proceedings Issue of the 20th Annual Meeting, Institute of Nuclear Materials Management, July 1979.
3. Jaech, J. L., Statistical Methods in Nuclear Material Control. TID-26298, National Technical Information Service, Springfield, VA, December 1973.

\*Available for purchase from the NRC/GPO Sales Program, U.S. Nuclear Regulatory Commission, Washington, DC 20555, and/or the National Technical Information Service, Springfield, VA 22161.



APPENDIX A

SIMULATION COMPUTER CODE WITH  
USER INFORMATION

## APPENDIX A

### SIMULATION COMPUTER CODE WITH USER INFORMATION

This Appendix presents the FORTRAN computer program which was used for the simulation study discussed in Section 5. User instructions are provided which specify the necessary input data. A worked example illustrates use of the program and interpretation of the output. Finally, a complete listing of the computer code is given.

The program is "self-contained" except that the sub-function RNOR calls a uniform random deviate generator RANF which is system-supplied. Most computer systems have a system-supplied random deviate generator, but the names tend to vary. Thus, the user should check into this, and possibly alter sub-function RNOR before attempting to run this program.

This program is dimensioned to handle up to ten material balance periods. If it is desirable to consider larger problems, the arrays and matrices could be redimensioned, but the various input and output formats would have to be cautiously altered to allow for the wider input and output fields.

Descriptions of and format specifications for the necessary input data cards are now given. If a data file is to be used rather than cards, each card would correspond to one line of the data file. All integers are assumed to be right-justified, and all real numbers must either include a decimal point or be right-justified.

#### Card 1: Format is (I3)

Col. 1-3, NDAT = number of variance-covariance/diversion pattern combinations to be studied.

Note: for each of the NDAT problems to be studied, a sequence of input data cards is necessary as follows.

#### Card 2: Format is (4I4, 4F10.0)

Col. 1-4, N = number of material balance periods.

Col. 5-8, NMU = number of goal quantities to be specified.

Col. 9-12, NIT = number of iterations (or simulated processes) desired.

Col. 13-16, IVAR = 0 if correlation matrix will be used as variance-covariance matrix

= 1 if  $\text{Var}(ID_i)$  quantities are to be randomly generated, or

= 2 if  $\text{Var}(ID_i)$  quantities are to be input.

Col. 17-26, DF = degrees of freedom for chi-square distribution used to generate  $\text{Var}(ID_i)$  quantities when IVAR = 1 option is specified. Can be left blank if IVAR = 0 or 2.

Col. 27-36, ZSCORE = upper  $1-\alpha/2$  percentile of the standard normal distribution (used as critical value for testing procedures 2 and 3).

Col. 37-46, CHI2 = upper  $1-\alpha$  percentile of the chi-square distribution with 2 degrees of freedom (used as critical value for bivariate procedure).

Col. 47-56, ZSC02 = upper  $1-\alpha/4$  percentile of the standard normal distribution (used as critical value for Bonferroni procedure).

The next sequence of N cards gives the rows of the lower triangle of the correlation matrix  $R_N$ , see equation (5.11). Let  $\rho_{ij}$  be the  $ij^{\text{th}}$  element of  $R_N$ . The format for each card is (NF10.0).

Card 3: First row of lower triangle of  $R_N$

Col. 1-10,  $\rho_{11}$

Card 4: Second row of lower triangle of  $R_N$

Col. 1-10,  $\rho_{21}$

Col. 11-20,  $\rho_{22}$   
} additional cards as needed  
.  
.

Card N + 2  $N^{\text{th}}$  row of lower triangle of  $R_N$

Col. 1-10,  $\rho_{N1}$

Col. 11-20,  $\rho_{N2}$

⋮

Col.  $[1 + 10(N-1)]-10N$ ,  $\rho_{NN}$

Next, the  $\text{Var}(ID_j)$  quantities are input if the  $\text{IVAR} = 2$  option was specified on Card 2.

Card N + 3: Format is (NF10.0)

Col. 1-10,  $\text{Var}(ID_1)$   
Col. 11-20,  $\text{Var}(ID_2)$   
.  
.  
Col.  $[1 + 10(N-1)]-10N$ ,  $\text{Var}(ID_N)$

Note: If  $\text{IVAR} = 0$  or  $\text{IVAR} = 1$ , omit card N + 3.

Card N + 4: Diversion pattern vector, defined in equation (5.7),  
format is (NF10.0).

Col. 1-10,  $q_1$   
Col. 11-20,  $q_2$   
.  
.  
.

Next, there are NMU (specified on card 2) cards, each having one goal quantity. The format for each card is F10.0.

Card N + 5:

Col. 1-10,  $Q_1$   
.  
.  
.

Card N + 4 + NMU

Col. 1-10,  $Q_{\text{NMU}}$

The sequence of cards, beginning with Card 2, is then repeated for each covariance/diversion pattern combination. That is, there are NDAT (specified on Card 1) such sequences of cards.



### Example

It is of interest to compare the four testing procedures for monitoring 500 identical processes over five material balance periods under the following operating conditions.

First, it is assumed that the correlation structure among the five successive ID's is given by

$$R_5 = \begin{bmatrix} 1 & -.4 & 0 & 0 & 0 \\ -.4 & 1 & -.4 & 0 & 0 \\ 0 & -.4 & 1 & -.4 & 0 \\ 0 & 0 & -.4 & 1 & -.4 \\ 0 & 0 & 0 & -.4 & 1 \end{bmatrix} \quad (A.1)$$

The  $\text{Var}(\text{ID}_i)$  quantities to be input are

$$\begin{aligned} \text{Var}(\text{ID}_1) &= .53 \\ \text{Var}(\text{ID}_2) &= .56 \\ \text{Var}(\text{ID}_3) &= 1.75 \\ \text{Var}(\text{ID}_4) &= 1.02 \\ \text{Var}(\text{ID}_5) &= 1.42 \end{aligned} \quad (A.2)$$

Two diversion patterns are of interest. A constant loss pattern, given by

$$q_1' = [.2 \ .2 \ .2 \ .2 \ .2], \quad (A.3)$$

and a pattern where the amount diverted from the  $i^{\text{th}}$  period is proportional to  $\sqrt{\text{Var}(\text{ID}_i)}$ , or in this case,

$$q_2' = [.15 \ .15 \ .26 \ .20 \ .24], \quad (A.4)$$

For each diversion pattern, three goal quantities are of interest.

They are

$$\begin{aligned} Q_1 &= 0, \\ Q_2 &= 8, \\ Q_3 &= 12. \end{aligned} \tag{A.5}$$

The  $\alpha = .05$  significance level will be used for all testing procedures.

The above problem has two covariance/diversion pattern combinations. That is, the variance-covariance matrix constructed from the information in equations (A.1) and (A.2) will be used with each of the two diversion pattern vectors given by equations (A.3) and (A.4). Thus, card 1 input is NDAT = 2.

Card 2: Quantities are

N = 5 material balance periods  
NMU = 3 goal quantities  
NIT = 500 simulated processes  
IVAR = 2 indicates variances will be input  
DF = 0 or blank, since IVAR = 2  
ZSCORE = 1.96  
CHI2 = .5.99  
ZSC02 = 2.24

Cards 3-7: Lower triangle of  $R_5$ , equation (A.1).

1.0				
-.4	1.0			
0	-.4	1.0		
0	0	-.4	1.0	
0	0	0	-.4	1.0

Card 8:  $\text{Var}(ID_i)$  quantities, equations (A.2)

.53	.56	1.75	1.02	1.42
-----	-----	------	------	------

Card 9: Diversion pattern  $q_1$ , equation (A.3)

.2 .2 .2 .2 .2

Cards 10-12: Goal quantities, equation (A.5)

0.0  
8.0  
12.0

Cards 13-23: These are a repeat of cards 2-12, except that Card 20 is the diversion pattern  $q_2$ , given by equation (A.4).

The actual input data are displayed in TABLE A.1.

TABLE A.1. Input Data for Example

002				
0005000305000002 0.0			1.96	5.99 2.24
1.0				
-.4	1.0			
0.0	-.4	1.0		
0.0	0.0	-.4	1.0	
0.0	0.0	0.0	-.4	1.0
.53	.56	1.75	1.02	1.42
.2	.2	.2	.2	.2
0.0				
8.0				
12.0				
0005000305000002 0.0			1.96	5.99 2.24
1.0				
-.4	1.0			
0.0	-.4	1.0		
0.0	0.0	-.4	1.0	
0.0	0.0	0.0	-.4	1.0
.53	.56	1.75	1.02	1.42
.15	.15	.26	.2	.24
0.0				
8.0				
12.0				

The actual output from the program begins on page A.9 with the lower triangle of the correlation matrix and the  $\text{Var}(ID_i)$  quantities which were input. The variance-covariance matrix is printed, followed by the transpose of the Cholesky matrix which is given as  $T'$  in equations (5.4) and (5.5).

Page A.10 begins with the diversion strategy vector  $\underline{\mu}' = Q \underline{q}'$  where  $Q$  is a goal quantity and  $\underline{q}'$  is the diversion pattern input by the user. This particular example vector corresponds to the goal quantity  $Q = 0$ , and the results will be used to assess "false alarm" frequencies and rates. This is followed by a table, which is labeled "RESULTS FOR METHOD 1". The interpretation of this table is as follows:

1. The first row gives a frequency distribution for first-time detections. The first five elements give the number of times a first-time detection occurred at each of the five periods.  
The sixth element is the total number of first-time detections which occurred during the five periods.  
The seventh element is the probability of a first-time detection occurring at some time during the five periods.
2. The second row gives the first-time detection probabilities for each of the five periods. Specifically, in this example, 25 of the 500 simulated processes (or 5%) were rejected for the first time during periods 1. Of the remaining 475 processes, 15 (or 3.16%) were rejected for the first time during periods 2. And so on.
3. The third row gives a "frequency distribution" for second-time detections. In this example, of the 25 processes which were rejected for the first time during period 1, 2 were rejected for the second time at period 2, and so on. That is, if a process is restarted after a first time rejection, but the cause of the rejection has not been resolved, it could be rejected for a second time during a later period.

4. The fourth, fifth, and sixth rows give "frequency distributions" for third-time, fourth-time, and fifth-time detections, respectively.
5. Note that the sixth column (i.e. sixth elements of rows 1, 3, 4, 5, 6 and 7) gives row totals.
6. The seventh and eighth rows give column totals and probabilities respectively. In this example, a total of 16 of the 500 processes (or 3.2%) were rejected during the third period. The sixth element of the seventh row gives the overall total number of rejections, but this may include multiple rejections of some processes and must be interpreted cautiously.

The table on page A.11 is labeled "METHOD 1 FREQUENCY DISTRIBUTION OF UNDETECTED LOSSES". Undetected losses are defined by equation (5.18), and the table is self explanatory.

For each of the four testing procedures, two tables, similar to those discussed above are constructed for each goal quantity. To conserve space, only the results for the first diversion pattern and the goal quantities  $Q = 0$  and  $Q = 8$  are presented.

THE LOWER TRIANGLE OF THE CORRELATION MATRIX

1.0000				
-0.4000	1.0000			
0.0000	-0.4000	1.0000		
0.0000	0.0000	-0.4000	1.0000	
0.0000	0.0000	0.0000	-0.4000	1.0000

THE VAR(ID) QUANTITIES FOR THE N PERIODS

0.5300	0.5600	1.7500	1.0200	1.4200
--------	--------	--------	--------	--------

THE VARIANCE-COVARIANCE MATRIX FOR THE 5 SUCCESSIVE ID QUANTITIES

0.5300	-0.2179	0.0000	0.0000	0.0000
-0.2179	0.5600	-0.3960	0.0000	0.0000
0.0000	-0.3960	1.7500	-0.5344	0.0000
0.0000	0.0000	-0.5344	1.0200	-0.4814
0.0000	0.0000	0.0000	-0.4814	1.4200

THE TRANSPOSE OF THE CHOLESKY MATRIX

0.7280				
-0.2993	0.6859			
0.0000	-0.5774	1.1902		
0.0000	0.0000	-0.4490	0.9047	
0.0000	0.0000	0.0000	-0.5321	1.0662

THE RESULTS FROM 500 ITERATIONS FOR THE DIVERSION STRATEGY  
 0.0000 0.0000 0.0000 0.0000 0.0000

RESULTS FOR METHOD 1: ROWS REPRESENT RESTARTS  
 COLUMNS REPRESENT PERIODS WHERE HO WAS REJECTED

25.0000	15.0000	13.0000	12.0000	10.0000	75.0000	0.1500
0.0500	0.0316	0.0283	0.0268	0.0230		
0.0000	2.0000	3.0000	3.0000	7.0000	15.0000	
0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

COLUMN TOTALS AND PERCENTAGES

25.0000	17.0000	16.0000	15.0000	18.0000	91.0000	
0.0500	0.0340	0.0320	0.0300	0.0360		

METHOD 1

FREQUENCY DISTRIBUTION OF UNDETECTED LOSSES

SIZE OF UNDETECTED LOSS	FREQUENCY OF OCCURRENCE
-------------------------	-------------------------

0.0	25.0
-----	------

0.0	15.0
-----	------

0.0	13.0
-----	------

0.0	12.0
-----	------

0.0	10.0
-----	------

0.0	425.0
-----	-------

AVERAGE UNDETECTED LOSS = 0.0000

VARIANCE OF UNDETECTED LOSSES = 0.0000

STANDARD DEVIATION OF UNDETECTED LOSSES = 0.0000



RESULTS FOR METHOD 2: ROWS REPRESENT RESTARTS  
 COLUMNS REPRESENT PERIODS WHERE HO WAS REJECTED

25.0000	27.0000	19.0000	22.0000	17.0000	110.0000	0.2200
0.0500	0.0568	0.0424	0.0513	0.0418		
0.0000	2.0000	6.0000	6.0000	5.0000	19.0000	
0.0000	0.0000	0.0000	0.0000	3.0000	3.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

COLUMN TOTALS AND PERCENTAGES

25.0000	29.0000	25.0000	28.0000	25.0000	132.0000
0.0500	0.0580	0.0500	0.0560	0.0500	

METHOD 2

FREQUENCY DISTRIBUTION OF UNDETECTED LOSSES

SIZE OF UNDETECTED LOSS	FREQUENCY OF OCCURRENCE
-------------------------	-------------------------

0.0	25.0
-----	------

0.0	27.0
-----	------

0.0	19.0
-----	------

0.0	22.0
-----	------

0.0	17.0
-----	------

0.0	390.0
-----	-------

AVERAGE UNDETECTED LOSS = 0.0000

VARIANCE OF UNDETECTED LOSSES = 0.0000

STANDARD DEVIATION OF UNDETECTED LOSSES = 0.0000

RESULTS FOR METHOD 3: ROWS REPRESENT RESTARTS  
 COLUMNS REPRESENT PERIODS WHERE HO WAS REJECTED

25.0000	15.0000	25.0000	18.0000	9.0000	92.0000	0.1840
0.0500	0.0316	0.0543	0.0414	0.0216		
0.0000	2.0000	2.0000	3.0000	7.0000	14.0000	
0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

COLUMN TOTALS AND PERCENTAGES

25.0000	17.0000	27.0000	21.0000	17.0000	107.0000
0.0500	0.0340	0.0540	0.0420	0.0340	

METHOD 3

FREQUENCY DISTRIBUTION OF UNDETECTED LOSSES

SIZE OF UNDETECTED LOSS	FREQUENCY OF OCCURRENCE
-------------------------	-------------------------

0.0	25.0
-----	------

0.0	15.0
-----	------

0.0	25.0
-----	------

0.0	18.0
-----	------

0.0	9.0
-----	-----

0.0	408.0
-----	-------

AVERAGE UNDETECTED LOSS = 0.0000

VARIANCE OF UNDETECTED LOSSES = 0.0000

STANDARD DEVIATION OF UNDETECTED LOSSES = 0.0000

RESULTS FOR METHOD 4: ROWS REPRESENT RESTARTS  
 COLUMNS REPRESENT PERIODS WHERE HO WAS REJECTED

16.0000	22.0000	18.0000	16.0000	14.0000	86.0000	0.1720
0.0320	0.0455	0.0390	0.0360	0.0327		
0.0000	0.0000	4.0000	5.0000	5.0000	14.0000	
0.0000	0.0000	0.0000	0.0000	2.0000	2.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

COLUMN TOTALS AND PERCENTAGES

16.0000	22.0000	22.0000	21.0000	21.0000	102.0000	
0.0320	0.0440	0.0440	0.0420	0.0420		

METHOD 4

FREQUENCY DISTRIBUTION OF UNDETECTED LOSSES

SIZE OF UNDETECTED LOSS	FREQUENCY OF OCCURRENCE
0.0	16.0
0.0	22.0
0.0	18.0
0.0	16.0
0.0	14.0
0.0	414.0

AVERAGE UNDETECTED LOSS = 0.0000

VARIANCE OF UNDETECTED LOSSES = 0.0000

STANDARD DEVIATION OF UNDETECTED LOSSES = 0.0000

THE RESULTS FROM 500 ITERATIONS FOR THE DIVERSION STRATEGY  
 1.6000 1.6000 1.6000 1.6000 1.6000

RESULTS FOR METHOD 1: ROWS REPRESENT RESTARTS  
 COLUMNS REPRESENT PERIODS WHERE HO WAS REJECTED

309.0000	166.0000	24.0000	0.0000	0.0000	499.0000	0.9980
0.6180	0.8691	0.9600	0.0000	0.0000		
0.0000	147.0000	130.0000	136.0000	54.0000	467.0000	
0.0000	0.0000	22.0000	79.0000	113.0000	214.0000	
0.0000	0.0000	0.0000	2.0000	16.0000	18.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

COLUMN TOTALS AND PERCENTAGES

309.0000	313.0000	176.0000	217.0000	183.0000	1198.0000	
0.6180	0.6260	0.3520	0.4340	0.3660		

METHOD 1

FREQUENCY DISTRIBUTION OF UNDETECTED LOSSES

SIZE OF UNDETECTED LOSS	FREQUENCY OF OCCURRENCE
-------------------------	-------------------------

0.0	309.0
-----	-------

1.6	166.0
-----	-------

3.2	24.0
-----	------

4.8	0.0
-----	-----

6.4	0.0
-----	-----

8.0	1.0
-----	-----

AVERAGE UNDETECTED LOSS = 0.7008

VARIANCE OF UNDETECTED LOSSES = 0.9803

STANDARD DEVIATION OF UNDETECTED LOSSES = 0.9901



RESULTS FOR METHOD 2: ROWS REPRESENT RESTARTS  
 COLUMNS REPRESENT PERIODS WHERE HO WAS REJECTED

309.0000	133.0000	28.0000	11.0000	6.0000	487.0000	0.9740
0.6180	0.6963	0.4828	0.3667	0.3158		
0.0000	147.0000	81.0000	93.0000	41.0000	362.0000	
0.0000	0.0000	22.0000	61.0000	68.0000	151.0000	
0.0000	0.0000	0.0000	2.0000	14.0000	16.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

COLUMN TOTALS AND PERCENTAGES

309.0000	280.0000	131.0000	167.0000	129.0000	1016.0000	
0.6180	0.5600	0.2620	0.3340	0.2580		

METHOD 2

FREQUENCY DISTRIBUTION OF UNDETECTED LOSSES

SIZE OF UNDETECTED LOSS	FREQUENCY OF OCCURRENCE
-------------------------	-------------------------

0.0	309.0
-----	-------

1.6	133.0
-----	-------

3.2	28.0
-----	------

4.8	11.0
-----	------

6.4	6.0
-----	-----

8.0	13.0
-----	------

AVERAGE UNDETECTED LOSS = 0.9952

VARIANCE OF UNDETECTED LOSSES = 2.9322

STANDARD DEVIATION OF UNDETECTED LOSSES = 1.7124

RESULTS FOR METHOD 3: ROWS REPRESENT RESTARTS  
 COLUMNS REPRESENT PERIODS WHERE HO WAS REJECTED

309.0000	184.0000	7.0000	0.0000	0.0000	500.0000	1.0000
0.6180	0.9634	1.0000	0.0000	0.0000		
0.0000	147.0000	157.0000	127.0000	50.0000	481.0000	
0.0000	0.0000	22.0000	91.0000	137.0000	250.0000	
0.0000	0.0000	0.0000	2.0000	28.0000	30.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

COLUMN TOTALS AND PERCENTAGES

309.0000	331.0000	186.0000	220.0000	215.0000	1261.0000	
0.6180	0.6620	0.3720	0.4400	0.4300		

METHOD 3

FREQUENCY DISTRIBUTION OF UNDETECTED LOSSES

SIZE OF UNDETECTED LOSS	FREQUENCY OF OCCURRENCE
0.0	309.0
1.6	184.0
3.2	7.0
4.8	0.0
6.4	0.0
8.0	0.0

AVERAGE UNDETECTED LOSS = 0.6336

VARIANCE OF UNDETECTED LOSSES = 0.6854

STANDARD DEVIATION OF UNDETECTED LOSSES = 0.8279

RESULTS FOR METHOD 4: ROWS REPRESENT RESTARTS  
 COLUMNS REPRESENT PERIODS WHERE HO WAS REJECTED

253.0000	231.0000	14.0000	2.0000	0.0000	500.0000	1.0000
0.5060	0.9352	0.8750	1.0000	0.0000		
0.0000	85.0000	130.0000	180.0000	71.0000	466.0000	
0.0000	0.0000	6.0000	49.0000	115.0000	170.0000	
0.0000	0.0000	0.0000	0.0000	3.0000	3.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

COLUMN TOTALS AND PERCENTAGES

253.0000	316.0000	150.0000	231.0000	189.0000	1139.0000	
0.5060	0.6320	0.3000	0.4620	0.3780		

METHOD 4

FREQUENCY DISTRIBUTION OF UNDETECTED LOSSES

SIZE OF UNDETECTED LOSS	FREQUENCY OF OCCURRENCE
-------------------------	-------------------------

0.0	253.0
-----	-------

1.6	231.0
-----	-------

3.2	14.0
-----	------

4.8	2.0
-----	-----

6.4	0.0
-----	-----

8.0	0.0
-----	-----

AVERAGE UNDETECTED LOSS = 0.8480

VARIANCE OF UNDETECTED LOSSES = 0.8442

STANDARD DEVIATION OF UNDETECTED LOSSES = 0.9188

The set of summary statistics given in Table A.2 is taken from the more extensive output presented above. This summary information is output by the program, but is directed to a different printing unit (or output file) than the more extensive output.

These summary statistics can be plotted, as in Figures 5.1-5.18, and used to compare false alarm rates, first-time detection probabilities, and expected undetected losses from the four testing procedures.

In this example, the same variance-covariance matrix was studied under two diversion patterns. These covariance/diversion pattern combinations are identified in column 1 of Table A.2. The content of the other columns is self explanatory.

Looking at the first-time detection probabilities in Column 6, when the goal quantity is 0, procedure 2 has the highest "false alarm rate" under the first diversion pattern. However, under the second diversion pattern, both procedures 2 and 3 show higher false alarm rates than the others. When the goal quantities are 8 or 12, the first-time detection probabilities are near 1 under both diversion patterns, and a better comparison of the procedures can be made by looking at the average undetected losses in Column 4.

When the goal quantity is 8, under either diversion pattern, procedure 2 is clearly the least powerful (i.e., results in considerably larger average undetected losses than the other procedures). It is also evident that under diversion pattern 2, larger undetected losses could be expected than under diversion pattern 1. This is consistent with the idea that diverting an amount proportional to  $\sqrt{\text{Var}(ID)}$  each period is an optimal diversion pattern.

In general, the results in Table A.2 are consistent with the results presented in Section 5.

A complete listing of the computer simulation code follows, starting on page A.28.

TABLE A.2. Summary Statistics for First-Time Detections

COVAR MATRIX	GOAL QUANTITY	TESTING PROCEDURE	AVERAGE UNDETECTED LOSS	STD.DEV. OF UNDETECTED LOSSES	FIRST-TIME DETECTION PROBABILITY
1	0.000000	1	0.000000	0.000000	0.150000
1	0.000000	2	0.000000	0.000000	0.220000
1	0.000000	3	0.000000	0.000000	0.184000
1	0.000000	4	0.000000	0.000000	0.172000
1	8.000000	1	0.760800	0.990009	0.998000
1	8.000000	2	0.995200	1.712380	0.974000
1	8.000000	3	0.633600	0.827870	1.000000
1	8.000000	4	0.848000	0.918800	1.000000
1	12.000000	1	0.235200	0.714270	1.000000
1	12.000000	2	0.235200	0.714270	1.000000
1	12.000000	3	0.235200	0.714270	1.000000
1	12.000000	4	0.379200	0.876260	1.000000
2	0.000000	1	0.000000	0.000000	0.148000
2	0.000000	2	0.000000	0.000000	0.196000
2	0.000000	3	0.000000	0.000000	0.200000
2	0.000000	4	0.000000	0.000000	0.152000
2	8.000000	1	1.092320	1.055000	1.000000
2	8.000000	2	1.688480	2.070780	0.956000
2	8.000000	3	0.961120	0.917250	1.000000
2	8.000000	4	1.198400	1.010910	1.000000
2	12.000000	1	0.586800	0.896790	1.000000
2	12.000000	2	0.668640	1.083910	1.000000
2	12.000000	3	0.568800	0.853040	1.000000
2	12.000000	4	0.766800	0.926700	1.000000



```

PROGRAM IDSIM
C
C
REAL*8 V(10,10),CH(10,10),ID(10),CID,X(10),VC(2,2),S2(10),TN
REAL*8 MU(10),RHO(10,10),DF,ZSCORE,CHI2,T,DABS,DSQRT,TMP,TCR
REAL*8 RR(4,10,12),T1,T2,RNOR,DET,ZSCO2,CTOT(4,2,11),FR(4,12)
REAL*8 ULOS(11),NULOS(4,11),AVULOS(4),VARLOS(4),SDLOS(4)
REAL*8 SLOS(4),SQLOS(4),CLOS,STRAT(10)
C
C FILE SIM.DAT CONTAINS INPUT DATA
C
C OPEN(UNIT=1,NAME='SIM.DAT',TYPE='OLD')
C
C PROGRAM WRITES EXTENSIVE SIMULATION OUTPUT TO SIM.RPT
C
C OPEN(UNIT=2,NAME='SIM.RPT',TYPE='NEW')
C
C PROGRAM WRITES SUMMARY STATISTICS TO LOSS.RPT
C INCLUDING EXPECTED UNDETECTED LOSSES AND
C FIRST-TIME DETECTION PROBABILITIES
C
C OPEN(UNIT=3,NAME='LOSS.RPT',TYPE='NEW')
C I1=13333
C I2=27579
C READ NDAT = OF DATA SETS TO BE INPUT
C
C READ(1,1111) NDAT
1111 FORMAT(I3)
DC 9999 IDAT=1,NDAT
C
C READ N = # OF MATERIAL BALANCE PERIODS
C NMU = # OF DIVERSION STRATEGY VECTORS TO BE SPECIFIED
C NIT = # OF ITERATIONS FOR EACH STRATEGY
C IVAR = 1 IF VARIANCES ARE TO BE GENERATED
C 2 IF VARIANCES ARE TO BE READ IN ADDITION TO CORRELATIONS
C 0 IF ONLY CORRELATIONS ARE TO BE USED AS READ
C DF = DEGREES OF FREEDOM FOR CHI SQUARE DISTRIBUTION
C USED TO GENERATE VAR( ID(I) ) = RHO(I,I) * CHISQ(DF)/DF
C NOTE: WHEN IVAR = 0 OR 2, DF CAN BE 0
C ZSCORE=UPPER 1-ALPHA/2 PERCENTILE POINT FROM THE N(0,1) DIST
C CHI2 = UPPER 1-ALPHA PERCENTILE OF THE CHI SQUARE
C DISTRIBUTION WITH 2 DEGREES OF FREEDOM
C ZSCO2 = UPPER 1-ALPHA/4 PERCENTILE OF THE N(0,1) DIST
C USED FOR "MULTIPLE COMPARISONS" TYPE TESTS
C
C READ(1,1) N,NMU,NIT,IVAR,DF,ZSCORE,CHI2,ZSCO2
1 FORMAT(4I4,4F10.0)
C
C READ THE LOWER TRIANGLE OF THE CORRELATION MATRIX
C
C DO 180 J=1,N
180 READ(1,2) (RHO(J,I),I=1,J)
2 FORMAT(10F10.0)
C
C PRINT THE LOWER TRIANGLE OF THE CORRELATION MATRIX

```

```

C
WRITE(2,3)
3  FORMAT('1','THE LOWER TRIANGLE OF THE CORRELATION MATRIX'//)
   DO 181 J=1,N
181 WRITE(2,22) (RHO(J,I),I=1,J)
C
C   READ THE VAR(ID) QUANTITIES IF IVAR = 2
C
   IF (IVAR.NE.2) GO TO 6
   READ(1,2) (S2(I),I=1,N)
   WRITE(2,4) (S2(I),I=1,N)
4  FORMAT('1','THE VAR(ID) QUANTITIES FOR THE N PERIODS'//' ',10F13.4)
6  CONTINUE
C
C   CONSTRUCT UPPER TRIANGLE OF VARIANCE-COVARIANCE MATRIX FOR
C   THE N SUCCESSIVE ID'S
C
   DO 10 I=1,N
   DO 11 J=1,N
11  CH(I,J)=0.D0
   IF(IVAR.EQ.1) GO TO 5
   IF(IVAR.EQ.2) GO TO 7
   CH(I,I)=RHO(I,I)
   GO TO 10
7  CH(I,I)=S2(I)
   GO TO 10
5  K=DF
   DO 12 J=1,K
12  CH(I,I)=CH(I,I)+RHO(I,I)*(RNOR(I1,I2)**2)/DF
10  CONTINUE
   DO 13 I=1,N
   DO 13 J=I,N
13  CH(I,J)=DSQRT(CH(I,I)*CH(J,J))*RHO(J,I)
C
C   CONSTRUCT FULL VARIANCE-COVARIANCE MATRIX V
C
   DO 15 I=1,N
   DO 15 J=I,N
15  V(I,J)=CH(I,J)
   V(J,I)=CH(I,J)
C
C   PRINT V
C
   WRITE(2,21) N
21  FORMAT('1','THE VARIANCE-COVARIANCE MATRIX FOR THE',I3,
1' SUCCESSIVE ID QUANTITIES'//)
   DO 20 I=1,N
20  WRITE(2,22) (V(I,J),J=1,N)
22  FORMAT(' ',10F13.4)
C
C   PERFORM THE CHOLESKY FACTORIZATION OF THE VARIANCE-COVARIANCE MATRIX
C
   CALL CHLSKY(CH,N,N)
C
C

```

```

C PRINT T-TRANPOSE MATRIX
C
C WRITE(2,26)
26 FORMAT('1','THE TRANSPOSE OF THE CHOLESKY MATRIX')
DO 25 I=1,N
25 WRITE(2,22)(CH(J,I),J=1,I)
C
C BEGIN SIMULATION
C
C READ A DIVERSION PATTERN STRATEGY VECTOR
C
C READ(1,23)(STRAT(J),J=1,N)
DO 100 IMU=1,NMU
C
C READ A DIVERSION STRATEGY CONSTANT MULTIPLIER
C
C READ(1,23) CLOS
23 FORMAT(10F10.0)
C
C COMPUTE AMOUNT DIVERTED FROM EACH PERIOD, I.E., DIVERSION STRATEGY VECTOR
C
C DO 1820 I=1,N
1820 MU(I)=CLOS*STRAT(I)
C
C COMPUTE UNDETECTED LOSS VECTOR
C
C ULOS(1)=0.D0
DO 1801 I=1,N
1801 ULOS(I+1)=ULOS(I)+MU(I)
C
C INITIALIZE COUNTERS RR AND CTOT
C
C DO 24 K=1,4
DO 24 I=1,N
CTOT(K,1,1)=0.D0
CTOT(K,2,1)=0.D0
DC 24 J=1,N+2
24 RR(K,I,J)=0.D0
C
C CARRY OUT SPECIFIED OF ITERATIONS
C
C DO 99 IIT=1,NIT
C
C GENERATE ID(I), I = 1,...,N
C
C DO 30 I=1,N
30 X(I)=RNOR(I1,I2)
DO 31 I=1,N
ID(I)=MU(I)
DO 31 J=1,I
31 ID(I)=ID(I)+X(J)*CH(J,I)
C
C
C METHOD 1 BIVARIATE TEST OF ID AND CID
C

```

```

C
C THIS SECTION WILL PERFORM THE TESTS BY RESTARTING THE
C CUMULATIVE CALCULATIONS THE PERIOD AFTER HO IS REJECTED.
C
C INITIALIZE COUNTERS NRSTRT = OF RESTARTS
C AND NFIRST = PERIOD AT WHICH RESTART BEGINS
C
      NRSTRT=0
      NFIRST=1
45     NRSTRT=NRSTRT+1
C
C PERFORM Z-TEST FOR STARTING PERIOD (I.E., AT STARTING PERIOD CID=ID.)
C
      IF(DABS(ID(NFIRST)/DSQRT(V(NFIRST,NFIRST))).LT.ZSCORE)GO TO 50
      RR(1,NRSTRT,NFIRST)=RR(1,NRSTRT,NFIRST)+1.0
      NFIRST=NFIRST+1
      IF(NFIRST.GT.N) GO TO 70
      GO TO 45
50     IF(NFIRST.EQ.N)GO TO 70
C
C PERFORM BIVARIATE TESTS FOR ID AND CID WHEN CID INCLUDES TWO OR MORE ID'S
C
C CONSTRUCT VAR(ID,CID)
C
      DO 60 I=NFIRST+1,N
      VC(1,1)=V(I,I)
      VC(1,2)=0.00
      DO 61 J=NFIRST,I
61     VC(1,2)=VC(1,2)+V(J,I)
      VC(2,2)=0.00
      DO 62 J=NFIRST,I
      DO 62 K=NFIRST,I
62     VC(2,2)=VC(2,2)+V(J,K)
C
C INVERT VC
C
      DET=VC(1,1)*VC(2,2)-VC(1,2)**2
      TMP=VC(1,1)/DET
      VC(1,1)=VC(2,2)/DET
      VC(2,2)=TMP
      VC(1,2)=-VC(1,2)/DET
C
C COMPUTE CID
C
      CID=0.00
      DO 63 J=NFIRST,I
63     CID=CID-ID(J)
C
C COMPUTE CHI SQUARE TEST STATISTIC
C
      T1=ID(I)*VC(1,1)+CID*VC(1,2)
      T2=ID(I)*VC(1,2)+CID*VC(2,2)
      T=ID(I)*T1+CID*T2
      IF(T.LT.CHI2) GO TO 60
      RR(1,NRSTRT,I)=RR(1,NRSTRT,I)+1.00

```

```

        NFIRST=I+1
        IF(NFIRST.GT.N) GO TO 70
        GO TO 45
60     CONTINUE
70     CONTINUE
C
C
C METHOD 2 TEST OF ID ONLY (I.E., THE CURRENT REQUIREMENT)
C
C
        NRJ=0
        DO 400 I=1,N
        IF(DABS(ID(I)/DSQRT(V(I,I))).LT.ZSCORE)GO TO 400
        NRJ=NRJ+1
        RR(2,NRJ,I)=RR(2,NRJ,I)+1.D0
400    CONTINUE
C
C
C METHGD 3 TEST OF CID ONLY
C
C
        NRSTRT=0
        NFIRST=1
450    NRSTRT=NRSTRT+1
        VC(2,2)=0.D0
        CID=0.D0
        DO 460 I=NFIRST,N
        CID=CID+ID(I)
        VC(2,2)=VC(2,2)+V(I,I)
        IF((I-NFIRST).EQ.0) GO TO 451
        DO 461 J=NFIRST,I-1
461    VC(2,2)=VC(2,2)+2*V(I,J)
451    IF(DABS(CID/DSQRT(VC(2,2))).LT.ZSCORE)GO TO 460
        RR(3,NRSTRT,I)=RR(3,NRSTRT,I)+1.D0
        NFIRST=I+1
        IF(NFIRST.GT.N)GO TO 470
        GO TO 450
460    CONTINUE
470    CONTINUE
C
C
C METHOD 4 "MULTIPLE COMPARISONS" USING ID VS LEID AND CID VS LECID
C TESTING EACH AT THE ALPHA/2 LEVEL (I.E., USING BONFERRONI'S INEQUALITY)
C
        NRSTRT=0
        NFIRST=1
650    NRSTRT=NRSTRT+1
        VC(2,2)=0.D0
        CID=0.D0
        DO 660 I=NFIRST,N
        CID=CID+ID(I)
        VC(2,2)=VC(2,2)+V(I,I)
        IF((I-NFIRST).EQ.0)GO TO 651
        DO 661 J=NFIRST,I-1
661    VC(2,2)=VC(2,2)+2*V(I,J)

```

```

651  IF(DABS(ID(I)/DSQRT(V(I,I))).LT.ZSCO2.AND.DABS(CID/DSQRT(V(2,2))).
1  LT.ZSCO2) GO TO 660
RR(4,NRSTRT,I)=RR(4,NRSTRT,I)+1.D0
NFIRST=I+1
IF(NFIRST.GT.N)GO TO 670
GO TO 650
660  CONTINUE
670  CONTINUE
C
99   CONTINUE
C
C   PRINT MU,NIT,RR, AND R
C
C   COMPUTE ROW TOTALS AND CONDITIONAL PROBABILITIES
C
      TNIT=NIT
      DO 177 I=1,4
      IN=NIT
      TCR=0.D0
      DO 178 K=1,N
      FR(I,K)=0.0
      RR(I,1,N+1)=RR(I,1,N+1)+RR(I,1,K)
      IF(TN.EQ.0.0)GO TO 178
      FR(I,K)=RR(I,1,K)/TN
      TCR=TCR+TN
173  TN=TN-RR(I,1,K)
177  RR(I,1,N+2)=RR(I,1,N+1)/TNIT
      DO 179 I=1,4
      DO 179 J=2,N
      DO 179 K=1,N
179  RR(I,J,N+1)=RR(I,J,N+1)+RR(I,J,K)
C
C   COMPUTE FREQUENCY DISTRIBUTION, MEAN, VARIANCE, STANDARD DEVIATION
C   FOR UNDETECTED LOSSES
C
      TMP=NIT
      DO 1802 I=1,4
      SLOS(I)=0.D0
      SQLOS(I)=0.D0
      DO 1803 J=1,N
      NULOS(I,J)=RR(I,1,J)
      SLOS(I)=SLOS(I)+NULOS(I,J)*ULOS(J)
1803  SQLOS(I)=SQLOS(I)+NULOS(I,J)*ULOS(J)**2
      NULOS(I,N+1)=TMP-RR(I,1,N+1)
      SLOS(I)=SLOS(I)+NULOS(I,N+1)*ULOS(N+1)
      SQLOS(I)=SQLOS(I)+NULOS(I,N+1)*ULOS(N+1)**2
      AVULOS(I)=SLOS(I)/TMP
      VARLOS(I)=(SQLOS(I)-(SLOS(I)**2)/TMP)/(TMP-1.D0)
      SDLOS(I)=DSQRT(VARLOS(I))
1802  CONTINUE
C
C   COMPUTE COLUMN TOTALS AND PERCENTS
C
      TMP=NIT
      DO 199 I=1,4

```

```

DO 199 J=1,N
DO 198 K=1,N
198 CTOT(I,1,J)=CTOT(I,1,J)+RR(I,K,J)
199 CTOT(I,2,J)=CTOT(I,1,J)/TMP
C
C COMPUTE OVERALL TOTAL AND PERCENT
C
TMP= NIT
DO 298 I=1,4
CTOT(I,1,N+1)=0.D0
DO 299 J=1,N
299 CTOT(I,1,N+1)=CTOT(I,1,N+1)+CTOT(I,1,J)
298 CTOT(I,2,N+1)=CTOT(I,1,N+1)/TMP
WRITE(2,97) NIT,(MU(I),I=1,N)
97 FORMAT('1','THE RESULTS FROM ',I4,' ITERATIONS FOR THE DIVERSION
1 STRATEGY'/' ',10F12.4/)
DO 90 I=1,4
WRITE(2,96) I
96 FORMAT('1','RESULTS FOR METHOD ',I1,': ROWS REPRESENT RESTARTS' /
1 ' ',22X,'COLUMNS REPRESENT PERIODS WHERE HO WAS REJECTED'//)
DO 91 J=1,N
WRITE(2,95){RR(I,J,K),K=1,N+2)
IF(J.NE.1)GO TO 91
WRITE(2,95){FR(I,K),K=1,N)
91 CONTINUE
C
C WRITE COLUMN TOTALS AND PERCENTAGES
C
WRITE(2,195)
195 FORMAT('0','COLUMN TOTALS AND PERCENTAGES'//)
DO 191 J=1,2
191 WRITE(2,95){CTOT(I,J,K),K=1,N+1)
95 FORMAT(' ',/12F10.4)
C
C WRITE FREQUENCY DISTRIBUTION FOR UNDETECTED LOSSES
C
WRITE(2,1807)I
1807 FORMAT('1','METHOD',I2//'FREQUENCY DISTRIBUTION OF UNDETECTED
LOSSES',////11X,'SIZE OF',12X,'FREQUENCY',/10X,'UNDETECTED',
1 14X,'OF'/13X,'LOSS',13X,'OCCURRENCE'//)
DO 1804 J=1,N+1
1804 WRITE(2,1805) HLOS(J),NULOS(I,J)
1805 FORMAT('0',10X,F7.1,13X,F7.1//)
WRITE(2,1806) AVULOS(I),VARLOS(I),SDLOS(I)
1806 FORMAT('0',////5X,'AVERAGE UNDETECTED LOSS =',F16.4,
1 //5X,'VARIANCE OF UNDETECTED LOSSES =',F16.4//
1 5X,'STANDARD DEVIATION OF UNDETECTED LOSSES =',F16.4)
C
C WRITE OUT EXPECTED UNDETECTED LOSS, FIRST TIME PROBABILITY OF DETECTION,
C OVERALL PROB. OF DET. GIVEN RESTARTS BUT UNCORRECTED DIVERSIN STRATEGY.
C THESE QUANTITIES ARE WRITTEN OUT TO THE "PLOTING FILE" ON UNIT 3.
C
WRITE(3,1810) IDAT,CLOS,I,AVULOS(I),SDLOS(I),RR(I,1,N+2),CTOT(I,2,N+1)
1810 FORMAT(' ',I3,F16.5,I3,4F16.5)
90 CONTINUE

```

```

100    CONTINUE
9999  CONTINUE
      STOP
      END

C
C
      FUNCTION RNOR(I1,I2)
C
C  GENERATES A RANDOM DEViate FROM THE N(0,1) DISTRIBUTION
C
      REAL*8 DSQRT,DCOS,DLOG,PI,RNOR,U1,U2,X1,X2,SUM
      PI=3.141592654
      SUM=0.D0
      DO 10 I=1,8
      CALL RANF(I1,I2,U1)
      CALL RANF(I1,I2,U2)
      X1=DSQRT(-2.0*DLOG(U1))*DCOS(2.0*PI*U2)
      X2=DSQRT(-2.0*DLOG(U1))*DSIN(2.0*PI*U2)
10    SUM=SUM+X1+X2
      RNOR=SUM/4.0
      RETURN
      END

C
C
C
      SUBROUTINE CHLSKY(A,N,M)
C
C  PERFORMS CHOLESKY FACTORIZATION OF SYMMETRIC MATRIX A
C  A IS INPUT AS AN UPPER TRIANGULAR MATRIX AND IS RETURNED
C  WITH THE CHOLESKY MATRIX T IN THE UPPER TRIANGLE AND THE
C  GENERALIZED INVERSE OF T IN THE LOWER TRIANGLE, WHERE THE
C  DIAG ELEMENTS OF T-INVERSE ARE THE RECIPROCALLS OF THE
C  DIAG ELEMENTS OF T
C
      REAL*8 A(10,10),FAC,ZERO,DSQRT,DABS
      ZERO=0.E-8
      FAC=DSQRT(A(1,1))
100   A(1,I)=A(1,I)/FAC
      DO 800 I=2,N
      II=I-1
      DO 500 J=1,N
      FAC=0.D0
      DO 300 K=1,II
      IF(K.LT.J.AND.J.LT.I) GO TO 300
      IF(A(K,K).LE.ZERO) GO TO 300
      IF(K.EQ.J) FAC=FAC+A(K,I)/A(K,K)-A(K,I)*A(K,K)
      FAC=FAC+A(K,I)*A(K,J)
300   CONTINUE
      A(I,J)=A(I,J)-FAC
      IF(DABS(A(I,J)).LT.ZERO) A(I,J)=0.D0
      IF(J.NE.I) GO TO 500
      IF(DABS(A(I,I)).GT.ZERO) GO TO 500
      DO 400 K=I,N
400   A(I,K)=0.D0

```



```
GO TO 800
500 CONTINUE
FAC=DSQRT(A(I,I))
DO 700 K=1,N
700 A(I,K)=A(I,K)/FAC
800 CONTINUE
RETURN
END
```

DISTRIBUTION

No. of  
Copies

OFFSITE

A. A. Churm  
DOE Patent Division  
9800 South Cass Avenue  
Argonne, IL 60439

2 DOE Technical Information Center

(a) NRC

ONSITE

30 Pacific Northwest Laboratory  
W. M. Bowen (10)  
L. C. Davenport (10)  
D. L. Hall  
J. A. Mahaffey  
R. J. Sorenson  
Publishing Coordination (2)  
Technical Information (5)

---

(a) Sufficient copies for NRC to distribute according to Report Category RS.



NRC FORM 335 (7-77)		U.S. NUCLEAR REGULATORY COMMISSION <b>BIBLIOGRAPHIC DATA SHEET</b>		1. REPORT NUMBER (Assigned by DDC) NUREG/CR-2483 PNL-4085	
4. TITLE AND SUBTITLE (Add Volume No., if appropriate)  Evaluation of Simultaneous Testing Procedures for Nuclear Materials Control and Accounting				2. (Leave blank)	
7. AUTHOR(S)  W. M. Bowen				3. RECIPIENT'S ACCESSION NO.	
9. PERFORMING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)  Pacific Northwest Laboratory P. O. Box 999 Richland, WA 99352				5. DATE REPORT COMPLETED MONTH   YEAR December   1981	
12. SPONSORING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)  U. S. Nuclear Regulatory Commission Division of Facility Operations Office of Nuclear Regulatory Research Washington, DC 20555				6. (Leave blank)	
10. PROJECT/TASK/WORK UNIT NO.				8. (Leave blank)	
11. CONTRACT NO.  NRC FIN B0208				12. TYPE OF REPORT	
13. TYPE OF REPORT				PERIOD COVERED (Inclusive dates)	
15. SUPPLEMENTARY NOTES				14. (Leave blank)	
16. ABSTRACT (200 words or less)  This report presents the results of a comparative evaluation of four statistical testing procedures for use in the control and accounting of special nuclear materials. Of primary interest is a bivariate procedure that simultaneously tests ID and CID. Descriptions of the four testing procedures are presented with the necessary formulas and special considerations for their implementation. Results of a simulation study indicate the conditions under which each of the tests would provide superior protection against "trickle" diversions.					
17. KEY WORDS AND DOCUMENT ANALYSIS			17a. DESCRIPTORS		
17b. IDENTIFIERS/OPEN-ENDED TERMS					
18. AVAILABILITY STATEMENT  Unlimited			19. SECURITY CLASS (This report) Unclassified		21. NO. OF PAGES
			20. SECURITY CLASS (This page) Unclassified		22. PRICE \$

