Measures of Predictive Accuracy
Study Note for MAS-II

Anthony T. Salis, FCAS, MAAA, CSPA, CPCU

1 Introduction

Statistical models may be evaluated using a variety of measures of predictive accuracy. The purpose of this study note is to clarify some aspects of the calculation of some of these measures as tested on the Modern Actuarial Statistics II (MAS-II) exam. The study note does not introduce anything that has not been covered by the text references in the content outline but rather clarifies details of the calculations that are implied but not outrightly stated in the texts. The study note assumes that the reader is already familiar with Chapter 8 of the James et al. text and Chapter 7 for the GLM text and the use of these measures in model evaluation.

This study note will cover the following measures:

- Gini Index (for Node Purity)
- Entropy
- Lift
- Gini Index (for Lift)
- Confusion Matrix Ratios
- Area Under the Receiver Operator Characteristic Curve (AUROC)

Updated: November 6, 2025

2 Measures of Node Purity

Section 8.1.2 of the James et al. text describes two measures of node purity for classification trees: the Gini index and entropy. This section will cover clarifications on these calculations as well as how they can be used to measure the quality of a split.

2.1 Gini Index (for Node Purity)

The Gini index for node purity is separate and independent from the Gini index for lift which is discussed in Section 3.2 of this study note. The Gini index for node purity is only used in classification trees, whereas the Gini index for lift can measure the segmentation power of any type of predictive model.

The formula for the Gini index for node purity and an example calculation are given in Section 2.3 of this study note.

2.2 Entropy

The calculation of entropy includes a logarithm of $\hat{p}_{mk}$ (the proportion of training observations in the *m*th region that are from the *k*th class), which may leave the base of the logarithm ambiguous. The choice of logarithm only changes the calculation by a constant multiplicative factor, but it is necessary to have a single standard in the exam context. The normal interpretation of the logarithm as used in James et al. is that it refers to the natural logarithm. This is corroborated by the comparison of entropy to deviance (James et al. p. 353). Other texts that discuss entropy in the context of a decision tree may use a different base, but the natural logarithm will be assumed in the entropy calculations for the MAS-II exam:

$$D_m = -\sum_{k=1}^{K} \hat{p}_{mk} \log(\hat{p}_{mk})$$

It should also be noted that in the case where $\hat{p}_{mk} = 0$ for some class *k*, the absence of class *k* is an indication of greater purity of the node, and class *k* does not contribute to the entropy sum. The formula for entropy is still valid despite the mathematical error in taking the logarithm of 0 since:

$$\lim_{p \to 0^+} p \cdot \log(p) = 0$$

2.3 Split Quality

Section 8.1.2 of the James et al. text uses Gini index and entropy to evaluate the quality of a split by comparing these measures on a node before and after making a split. This requires defining these measures in a way in which they can capture the average node purity across multiple terminal nodes. The derivation in this section will focus on the Gini index, although the same steps can show the derivation for entropy.

The Gini index is based on the proportion of observations in each class and is defined for a particular region *m*. However, it can be rewritten as a weighted average of the “impurity” of each class, where the “impurity” of a class is the proportion of observations in the node that do not belong to that class $(1 - \hat{p}_{mk})$:

$$G_m = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$$

$$\hat{p}_{mk} = \frac{n_{mk}}{n_m}$$

$$G_m = \frac{1}{n_m} \sum_{k=1}^{K} n_{mk}(1 - \hat{p}_{mk})$$

where n_m is the number of observations in the *m*th region and n_{mk} is the number of observations in the *m*th region that belong to the *k*th class.

When a node is split, the "impurity" of each class changes because the observations are split between two regions, creating more unique class-node combinations. If the split creates two new regions, m_1 and m_2, the Gini index of the original node after the split can be calculated as follows:

$$G'_m = \frac{1}{n_m}\left(\sum_{k=1}^{K} n_{m_1k}\left(1 - \hat{p}_{m_1k}\right) + \sum_{k=1}^{K} n_{m_2k}\left(1 - \hat{p}_{m_2k}\right)\right)$$

$$G'_m = \frac{1}{n_m}\left(n_{m_1} \cdot G_{m_1} + n_{m_2} \cdot G_{m_2}\right)$$

This shows that the Gini index of the original node after the split is equal to the weighted average of the Gini indices of the new terminal nodes where the weights are the number of observations in each terminal node.

The text also states that using the Gini index or entropy to evaluate the quality of a split is more sensitive to node purity than is the classification error rate (James et al. p. 336). This can be illustrated in the example in Table 1 from a tree predicting a trinary target. The majority class of the original node is Class A, and it has a classification error rate of 40%. After the split, not only is Class A still the majority class of both new terminal nodes, but the classification error rate is still 40% for both nodes. The split is valuable, not because it reduces the classification error rate, but rather because it increases node purity by better segmenting Class B from Class C. In this case, both the Gini index and entropy are lower in both new nodes than in the original node, so the average Gini index and entropy for the original node has decreased as well. The difference between the Gini indices or entropies for the original node before and after the split provides one measure of the quality of that split.

Table 1. Example Calculation of Gini Index and Entropy

Region	Observations by Class			Error Rate	Gini Index	Entropy
	A	B	C			
Original Node (before split)	9	3	3	0.40	0.56	0.95
New Node 1	6	3	1	0.40	0.54	0.90
New Node 2	3	0	2	0.40	0.48	0.67
Original Node (after split)				0.40	0.52	0.82
Difference				0.00	0.04	0.13

3 Measures of Model Lift

Section 7.2 of the GLM text describes multiple measures of model lift, which are ways of quantifying the "economic value" of a model. This study note will provide clarifications on the mechanics of the calculations for two of these measures: lift and the Gini index.

3.1 Lift

Within the context of measures of "model lift", the measure of "lift" is defined as "the vertical distance between the first and last quantiles" (GLM p. 77). There are several ways to measure this "vertical distance". In the insurance context (especially in predicting frequency, severity, pure premium, and loss ratio), the vertical distance is usually measured as the ratio of the actual target rates in the last and first quantiles. When predicting a binary target, lift can also be measured as the odds ratio of the last and first quantiles. In some contexts, the difference between the last and first quantiles can be used.

The ratio, odds ratio, and difference calculations are shown below, where L is the lift and Y_1 and Y_Q are the actual target rates in the first and last quantiles, respectively:

$$L_{ratio} = \frac{Y_Q}{Y_1}$$

$$L_{odds\ ratio} = \left(\frac{Y_Q}{1 - Y_Q}\right) \cdot \left(\frac{1 - Y_1}{Y_1}\right)$$

$$L_{difference} = Y_Q - Y_1$$

The MAS-II exam will not directly test the calculation of lift as a metric, but understanding how this "vertical distance" can be calculated can be useful in comparing competing models.

3.2 Gini Index (for Lift)

Section 7.2.4 of the GLM text describes the method to construct the points on the Lorenz curve but does not provide specific details on calculating the area between the Lorenz curve and the line of equality. There are several theoretical considerations for accurately representing the Gini index which are out of scope for this study note and the MAS-II exam. For the purposes of the exam, the Gini index will be calculated using the trapezoidal method on an empirical sample of data, which is illustrated in this section of the study note. In practice, this is often sufficient to make useful comparisons of lift between competing models.

In the insurance context, model targets are often represented as a ratio. Some of these common insurance targets are listed in Table 2. In building the Lorenz curve, the numerator and denominator of the target are separated, since the denominator will be used to calculate the *x*-coordinate, and the numerator will be used to calculate the *y*-coordinate.

Table 2. Common Insurance Targets

Target	Denominator	Numerator
Frequency	Exposure	Claim Count
Severity	Claim Count	Loss Amount
Pure Premium (Loss Cost)	Exposure	Loss Amount
Loss Ratio	Premium	Loss Amount

In this example, the model will be predicting pure premium. Table 3 displays the holdout dataset, which contains 10 observations that have been sorted by predicted loss cost.

Table 3. Holdout Data for Gini Index

Policy ID	Predicted Loss Cost	Exposure	Loss
1	50	0.5	100
2	50	0.5	100
3	60	2.0	100
4	60	1.0	0
5	80	1.0	0
6	100	0.5	0
7	100	1.5	100
8	150	0.5	0
9	150	1.5	200
10	200	1.0	400

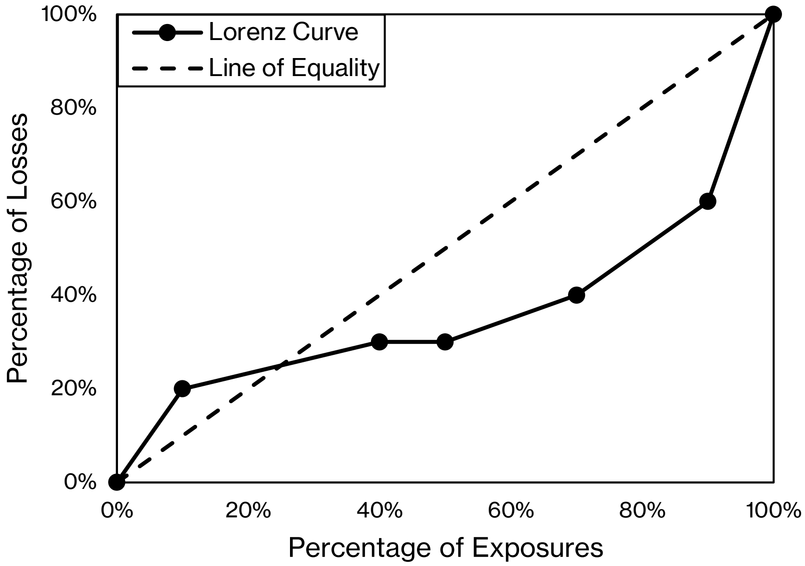
There are only 6 unique predicted loss costs, so the precise order in which the observations are sorted is ambiguous. Without any modification to the data structure, the calculation of cumulative exposures and losses would be different depending on the order in which observations with ties in the predicted loss cost are sorted. For this reason, the exposure and loss must first be aggregated by predicted loss cost and then the cumulative percentage of exposures and losses can be calculated. Table 4 has this aggregation and calculation of the cumulative percentage of exposures and losses.

Table 4. Plot Data for Lorenz Curve

Predicted Loss Cost	Exposure	Loss	Cumulative Percentage of Exposures (x_j)	Cumulative Percentage of Losses (y_j)
50	1.0	200	0.1	0.2
60	3.0	100	0.4	0.3
80	1.0	0	0.5	0.3
100	2.0	100	0.7	0.4
150	2.0	200	0.9	0.6
190	1.0	400	1.0	1.0
Total	10.0	1000		

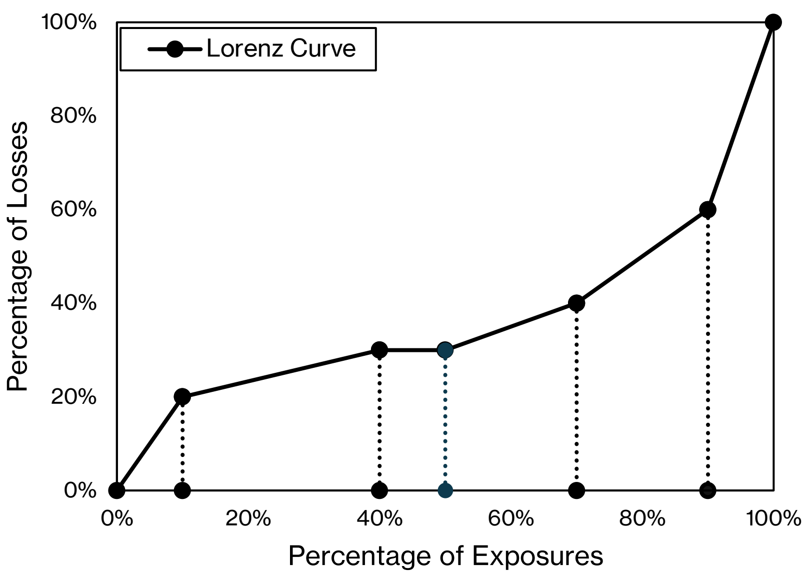
Adding the implied point at (0,0), the Lorenz curve can be plotted with the cumulative percentage of exposures on the *x*-axis and the cumulative percentage of losses on the *y*-axis as shown in Figure 1. The Lorenz curve connects each point with a straight line. This has an intuitive interpretation that each exposure within the unique predicted value contributes a constant amount of loss, which fits the assumption that there is no better way to sort the observations that contain the same predicted value.

Figure 1. Gini Index Plot



There are many approaches to calculate the area between the Lorenz curve and the line of equality. A simple way is to calculate the area under the Lorenz curve and subtract it from the area under the line of equality. The shape formed under the line of equality is a triangle with a base and height both equal to 1, which has an area of 0.5. The shape under the Lorenz curve can be divided into a series of trapezoids as shown in Figure 2. Each trapezoid has base lengths of y_{j-1} and y_j and a height of $x_j - x_{j-1}$. (The first "trapezoid" in the series is actually a triangle, but it can be thought of as a special case of a trapezoid where one of the bases is 0.)

Figure 2. Lorenz Curve Trapezoids



The Gini index can then be calculated as twice the difference between 0.5 and the sum of the areas of these trapezoids:

$$G = 2\left(\frac{1}{2} - \sum_{j=1}^{J} \frac{1}{2}(x_j - x_{j-1})(y_{j-1} + y_j)\right)$$

$$G = 1 - \sum_{j=1}^{J} (x_j - x_{j-1})(y_{j-1} + y_j)$$

Using the second equation, Table 5 shows the calculation of each partial sum. (By definition, $x_0 = y_0 = 0$.) The Gini index is the difference between 1 and the sum of the partial sums, which is 0.27.

Table 5. Gini Index Calculation

j	x_j	y_j	Partial Sum
0	0.00	0.00	
1	0.10	0.20	0.02
2	0.40	0.30	0.15
3	0.50	0.30	0.06
4	0.70	0.40	0.14
5	0.90	0.60	0.20
6	1.00	1.00	0.16
Sum			0.73

In the original economics use case of the Gini index, the Lorenz curve can never cross the line of equality. This is because households are rank ordered by their income and then that same income is measured on the *y*-axis. However, in a modeling context, the rank ordering of observations comes from the model prediction and the values for the *y*-axis come from the actual experience in the data. Due to random variation in the target, it is possible that the Lorenz curve may be above the line of equality, which happens in this example and is shown in Figure 1. This is a similar effect as reversals appearing in the simple quantile plot. The area that under the Lorenz curve that is above the line of equality has a negative contribution to the Gini index, which effectively penalizes the model for this reversal.

4 Measures for Binary Models

Section 7.3 of the GLM text describes the confusion matrix and receiver operator characteristic (ROC) curve to describe the fit of the model. This study note will cover the additional material on the confusion matrix from Section 4.4.2 of the James et al. text as well as clarify the calculation of AUROC.

4.1 Confusion Matrix Ratios

Tables 6 and 7 show some ratios (with their commonly used names) that can be derived from a confusion matrix (shown with its marginals):

Table 6. Confusion Matrix

		Predicted Class		
		Positive	**Negative**	**Total**
Actual (True) Class	**Positive**	TP	FN	P
	Negative	FP	TN	N
	Total	P*	N*	T

Table 7. Confusion Matrix Ratios

Calculation	**Name(s)**
(FN + FP) / T	• Error Rate
TP / P	• True Positive Rate • Sensitivity • Recall • Power
FN / P	• False Negative Rate • Type II Error
TN / N	• True Negative Rate • Specificity
FP / N	• False Positive Rate • Type I Error • 1 – Specificity
TP / P*	• Positive Predicted Value • Precision
FP / P*	• False Discovery Proportion
TN / N*	• Negative Predicted Value

4.2 AUROC

This section covers an example calculation of the area under the receiver operator characteristic curve (AUROC). The holdout dataset is shown in Table 8. It contains 15 observations with their predicted probabilities (of an outcome of 1) and actual outcome, sorted by predicted probability.

Table 8. Holdout Data for AUROC

Observation	Predicted Probability	Actual
1	0.10	0
2	0.10	0
3	0.10	0
4	0.10	0
5	0.20	0
6	0.20	0
7	0.20	1
8	0.30	1
9	0.30	0
10	0.40	0
11	0.40	1
12	0.50	0
13	0.60	1
14	0.70	1
15	0.80	0

Each point plotted on the receiver operator characteristic (ROC) curve is derived from the specificity and sensitivity associated with a discrimination threshold. As an example, Table 9 shows the construction of the confusion matrix for a discrimination threshold of 0.55. At this discrimination threshold, the specificity is 9/10 = 0.9 and the sensitivity is 2/5 = 0.4.

Table 9. Confusion Matrix for Table 8 with Discrimination Threshold of 0.55

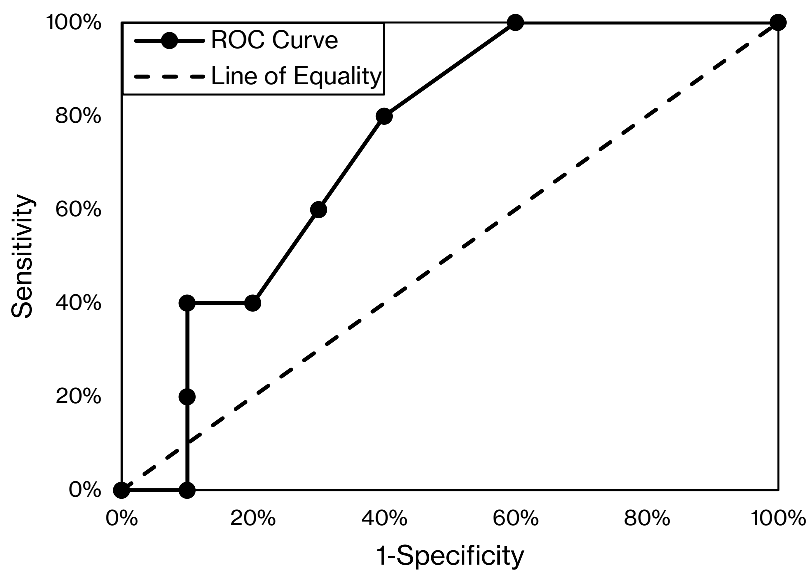
		Predicted Class		
		Positive	Negative	Total
Actual (True) Class	Positive	2	3	5
	Negative	1	9	10
	Total	3	12	15

Since there are 8 unique predicted probabilities, there will be 9 unique discriminations: 7 from discrimination thresholds lying between the unique predicted probabilities and 2 from a discrimination threshold above the highest predicted probability and a discrimination threshold below the lowest predicted probability. The 9 selected discrimination thresholds with their associated points on the ROC curve are shown in Table 10 and plotted in Figure 3.

Table 10. Plot Data for ROC Curve

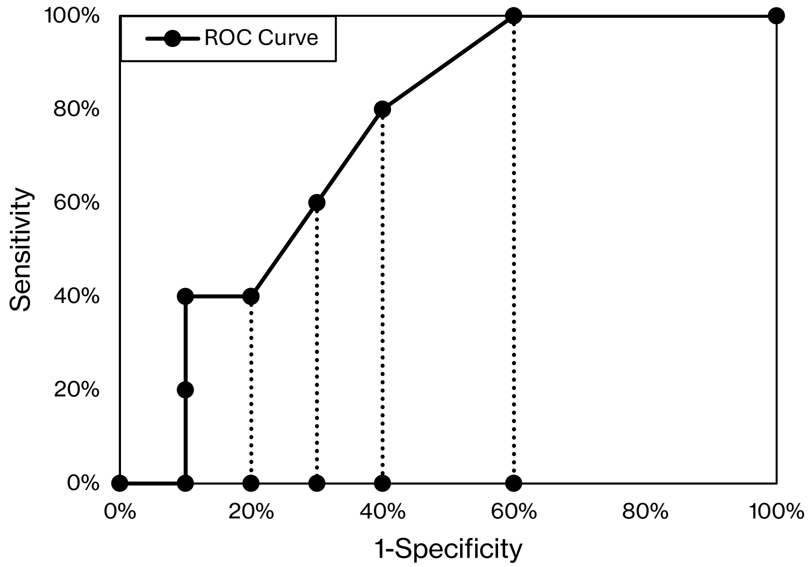
Discrimination Threshold	1-Specificity (x_j)	Sensitivity (y_j)
0.90	0.0	0.0
0.75	0.1	0.0
0.65	0.1	0.2
0.55	0.1	0.4
0.45	0.2	0.4
0.35	0.3	0.6
0.25	0.4	0.8
0.15	0.6	1.0
0.05	1.0	1.0

Figure 3. ROC Curve



AUROC is simply the area under the curve. Like the Gini index, it is calculated as a sum of the area of trapezoids, which are constructed in Figure 4. In this calculation, the line of equality is not used, so it has been removed from the graph.

Figure 4. ROC Curve Trapezoids



Each trapezoid has base lengths of y_{j-1} and y_j and a height of $x_j - x_{j-1}$. AUROC can then be calculated the sum of the areas of all these trapezoids:

$$AUROC = \sum_{j=1}^{J} \frac{1}{2}(x_j - x_{j-1})(y_{j-1} + y_j)$$

Table 11 shows the calculation of the trapezoidal areas. AUROC is the sum of the trapezoidal areas, which is 0.74.

Table 11. AUROC Calculation

j	x_j	y_j	**Trapezoidal Area**
0	0.0	0.0	
1	0.1	0.0	0.00
2	0.1	0.2	0.00
3	0.1	0.4	0.00
4	0.2	0.4	0.04
5	0.3	0.6	0.05
6	0.4	0.8	0.07
7	0.6	1.0	0.18
8	1.0	1.0	0.40
Sum			0.74

Another equivalent formulation of the table is to aggregate rows by unique predicted value (in descending order), which is shown in Table 12. The predicted probability of 1 is included to fill the index of 0.

Table 12. Alternative AUROC Calculation

i	**Predicted Probability**	**Number of Actual Negatives (n_i)**	**Number of Actual Positives (p_i)**	**Cumulative Actual Positives (c_i)**	**$n_i(c_{i-1} + c_i)$**
0	1.0	0	0	0	
1	0.8	1	0	0	0
2	0.7	0	1	1	0
3	0.6	0	1	2	0
4	0.5	1	0	2	4
5	0.4	1	1	3	5
6	0.3	1	1	4	7
7	0.2	2	1	5	18
8	0.1	4	0	5	40
Total		10	5		74

The number of negatives (n_i) corresponds to the difference in specificities ($x_j - x_{j-1}$) and the cumulative number of positives (c_i) corresponds to the sensitivities (y_j). The equation for AUROC can be rewritten as:

$$AUROC = \frac{1}{2NP}\sum_{i=1}^{I} n_i(c_{i-1} + c_i)$$

where *N* is the total number of observations with a target in the negative class and *P* is the total number of observations with a target in the positive class. Then AUROC = 74 / (2 • 10 • 5) = 0.74.