

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: Goodness-of-fit Tests and Function Estimators for Receiver Operating Characteristic (ROC) Curves: Inference from Perpendicular Distances

Author(s): R. Bradley Patterson

Document No.: 240957

Date Received: January 2013

Award Number: 2011-CD-BX-0124

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant report available electronically.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

GOODNESS-OF-FIT TESTS AND FUNCTION ESTIMATORS FOR
RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES:
INFERENCE FROM PERPENDICULAR DISTANCES

by

R. Bradley Patterson
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Statistical Science

Committee:

_____ Dr. John Miller, Dissertation Director
_____ Dr. Linda Davis, Committee Member
_____ Dr. Christopher Saunders, Committee Member
_____ Dr. Liansheng Tang, Committee Member
_____ Dr. William Rosenberger, Department Chair
_____ Dr. Lloyd J. Griffiths, Dean, The Volgenau School
of Engineering

Date: _____ Spring Semester 2012
George Mason University
Fairfax, VA

Goodness-of-fit Tests and Function Estimators for Receiver Operating Characteristic (ROC)
Curves: Inference from Perpendicular Distances

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at George Mason University

By

R. Bradley Patterson
Master of Arts
University of California, Santa Barbara, 2004
Bachelor of Science
Stanford University, 2000

Director: Dr. John Miller, Professor
Department of Statistics

Spring Semester 2012
George Mason University
Fairfax, VA

Copyright © 2012 by R. Bradley Patterson
All Rights Reserved

Dedication

In memory of Robert M. Patterson (1946-2010)

Acknowledgments

I would like to thank my advisor, Professor Miller, for accepting me as a student and teaching me about research in statistics for over three years. I am grateful for his instantaneous insights into my work. His willingness to devote time helping and guiding me could not have been greater. To Professor Davis, I owe infinite thanks for advising me on writing this dissertation. Her role as practically a co-advisor this year was so generous and helpful. I am also indebted to Professor Saunders for excellent research opportunities and inspiration. In addition, I thank Professor Tang for many valuable discussions about ROC curves. I appreciate the assistance and advice provided by Professor Rosenberger throughout my time in graduate school. Professor Gantz also graciously involved me in additional research activity and offered support regularly. I must emphasize my gratitude to Liz Quigley for patiently and readily helping me with so much in the department. Additionally, I am grateful to Carl Barrelet for his friendship, which began as we both started our studies in science. Finally, I want to thank Lauren and Meredy Patterson for their continuous love and assurance.

This project was supported by Award No. 2011-CD-BX-0124, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect those of the Department of Justice.

Table of Contents

	Page
List of Tables	viii
List of Figures	x
Abstract	xii
1 Introduction	1
1.1 Overview of ROC curves	1
1.1.1 Classification methods, scores, and error rates	2
1.1.2 Underlying classes and distribution functions	5
1.1.3 Examples of ROC curves	10
1.1.4 Assessing performance with ROC curves	13
1.1.5 Two important mathematical properties of ROC curves	15
1.1.6 ROC curves as CDFs	17
1.2 Goodness-of-fit tests in general	18
1.3 Parameter estimation	18
1.4 Overview of this dissertation	19
1.4.1 Goodness-of-fit statistic for binormal ROC	19
1.4.2 Estimators for binormal ROC	21
1.4.3 Outline of this dissertation	22
2 Literature review	24
2.1 History of the ROC curve	24
2.2 Binormal model	25
2.3 Estimation of binormal ROC functions	27
2.3.1 Maximum likelihood approaches	27
2.3.2 Minimum distance estimation	29
2.3.3 Regression	31
2.3.4 Recent developments	36
2.4 Goodness-of-fit of binormal ROC functions	37
2.4.1 General background on goodness-of-fit statistics	37
2.4.2 Specific goodness-of-fit statistics for estimated binormal ROC functions	41

3	Development of goodness-of-fit tests for receiver operating characteristic curves	43
3.1	Cramér-von Mises family applied to ROC curves	43
3.1.1	Adaptation of Cramér-von Mises for ROC	44
3.1.2	Characterization of Cramér-von Mises statistics for ROC curves	48
3.2	Empirical ROC-based goodness-of-fit statistics	52
3.2.1	Motivation	52
3.2.2	Framework	53
3.2.3	Implementation	59
3.2.4	Preview of the ROC-PD statistic's distribution	65
3.2.5	Extensions	65
4	Characterization of goodness-of-fit tests	67
4.1	Outline of simulations	67
4.2	Detailed steps of simulations	70
4.2.1	Generation of simulation data (step one)	70
4.2.2	Formulation of hypotheses (step two)	71
4.2.3	Calculation of goodness-of-fit (step three)	72
4.2.4	Calculation of p-values by parametric bootstrap (step four)	72
4.3	List of simulations	73
4.3.1	Null distributions	73
4.3.2	Alternative distributions	74
4.4	Results	77
4.4.1	Presentation of results	77
4.4.2	Null distributions	81
4.4.3	Alternative distributions	83
4.5	Consistency of test	86
4.6	Application	95
4.6.1	ROC curve for species of Iris	95
5	Estimation of binormal ROC functions	98
5.1	New estimation methods	98
5.1.1	Minimum perpendicular distance (ROC-MPD)	98
5.1.2	Total least squares (ROC-TLS)	99
5.1.3	Minimum Cramér-von Mises distance (ROC-MCVM)	101
5.2	Design of simulations	102
5.2.1	Outline of simulations	102
5.2.2	Additional estimators	102

5.2.3	Measures of error	103
5.3	Estimates of binormal ROC functions in null cases	104
5.3.1	Perpendicular distance squared	104
5.3.2	Mean squared error and bias of parameter and AUC estimations	106
5.4	Estimates of binormal ROC functions in alternative cases	108
5.4.1	Biexponential	108
5.4.2	Normal-gamma	110
5.5	Bootstrap confidence regions	113
5.5.1	Development	114
5.5.2	Coverage probabilities	117
5.6	Application	119
5.6.1	ROC curve for audiology example	119
6	Conclusion	122
6.1	Goodness-of-fit statistic for binormal ROC	123
6.2	Estimators for binormal ROC	124
6.3	Parametric studentized bootstrap confidence regions	125
6.4	Future research	125
A	Results of simulations for goodness-of-fit tests	127
A.1	Binormal	127
A.2	Biexponential	176
A.3	Normal-gamma	189
A.4	Normal-mixture	202
A.5	Normal-Cauchy	215
B	Results of simulations for estimation methods	223
B.1	Binormal	224
B.2	Biexponential	250
B.3	Normal-gamma	254
C	A chi-squared goodness-of-fit test for continuous ROC curves	258
D	Computing environment	261
D.1	Software	261
D.2	Hardware	261
D.3	Times	261
	Bibliography	263

List of Tables

Table	Page
1.1 Possible outcomes for a single threshold on scores.	3
4.1 Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a null hypothesis.	73
4.2 Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a biexponential alternative.	74
4.3 Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a normal-gamma alternative.	75
4.4 Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a normal-mixture alternative.	76
4.5 Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a normal-Cauchy alternative.	77
5.1 Root mean summed perpendicular squared errors for $\rho = 1$ and $\delta = 0.5$	105
5.2 Root mean summed perpendicular squared errors for $\rho = 1$ and $\delta = 1.5$	105
5.3 Root mean summed perpendicular squared errors for $\rho = 0.5$ and $\delta = 0.5$	106
5.4 Root mean squared error (RMSE) and bias of estimations of binormal parameters for $\rho = 1$ and $\delta = 1$	107
5.5 Root mean squared error (RMSE) and bias of AUC estimations for $\rho = 1$ and $\delta = 1$	107
5.6 Root mean global squared errors for binormal functions fit to 5000 samples from biexponential with $\lambda_1 = 1$ and $\lambda_2 = 2$	109
5.7 Root mean global squared errors for binormal functions fit to 5000 samples from biexponential with $\lambda_1 = 1$ and $\lambda_2 = 15$	109
5.8 Root mean squared error (RMSE) and bias of AUC estimations from binormal functions fit to 5000 samples from biexponential with $\lambda_1 = 1$ and $\lambda_2 = 2$	110
5.9 Root mean squared error (RMSE) and bias of AUC estimations from binormal functions fit to 5000 samples from biexponential with $\lambda_1 = 1$ and $\lambda_2 = 15$	111

5.10	Root mean global squared errors for binormal functions fit to 5000 samples from normal-gamma ROC with parameters $\mu_1 = 0, \sigma_1 = 1, \alpha_2 = 2, \beta_2 = 0.5, \theta_2 = 5$	112
5.11	Root mean global squared errors for binormal functions fit to 5000 samples from normal-gamma ROC with parameters $\mu_1 = 0, \sigma_1 = 3, \alpha_2 = 3, \beta_2 = 2, \theta_2 = 3$	112
5.12	MSE and bias of AUC from binormal functions fit to 5000 samples from normal-gamma ROC with parameters $\mu_1 = 0, \sigma_1 = 1, \alpha_2 = 2, \beta_2 = 0.5, \theta_2 = 5$	113
5.13	MSE and bias of AUC from binormal functions fit to 5000 samples from normal-gamma ROC with parameters $\mu_1 = 0, \sigma_1 = 3, \alpha_2 = 3, \beta_2 = 2, \theta_2 = 3$	114
5.14	Coverage probabilities for 95% confidence regions of ρ and δ calculated over 1000 samples each with $n_1 = 100$ negative observations and $n_2 = 100$ positive observations.	118
5.15	Parameters estimates and ROC-PD goodness-of-fit tests for audiology data.	120

List of Figures

Figure	Page	
1.1	An ROC curve plots the true positive rate versus the false positive rate for all thresholds on scores. The step function is the empirical ROC curve for a sample of size 100 from the negative and positive distributions. The tables in the figure give the results at single values of the threshold t	4
1.2	Demonstration of the form of the binormal ROC curve for different separations between the negative and positive distributions. In this example, the negative class follow a normal distribution with mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 1$ while the positive class follows a normal distribution with variable mean μ_2 and constant standard deviation $\sigma_2 = 1$	10
1.3	Example of ROC curve when negative class follows a normal distribution and positive class follows a reflected gamma distribution.	11
1.4	Example of ROC curve when negative class follows a normal distribution and positive class follows a mixture of normals.	11
1.5	Example of ROC curve when negative class follows a normal distribution and positive class follows a Cauchy distribution.	12
3.1	Examples of the a and b coordinates on empirical ROC curves. In the bottom two plots, the pairs (a_4, b_4) and (a_5, b_5) refer to the point at the end.	46
3.2	Simulated distributions of the ROC-CVM statistic for different values of the binormal parameter δ under a simple null hypothesis. The parameter ρ has a constant value of one. One thousand iterations ran with sample sizes each 100.	49
3.3	Empirical ROC and ODC curves for the same data. The middle panel in the first row shows the ROC curve's horizontal segments, which appear as vertical segments in the right panel of the second row for the ODC.	54
3.4	Example of distance $D(p^*)$ between binormal parametric ROC curve and empirical ROC curve. The perpendicular line $h(p; p^*, \theta)$ intersects the point $(p^*, \text{ROC}_\theta(p^*))$ on the parametric ROC curve ROC_θ	58

3.5	Ratio of ROC-PD statistic for different values of the number (d) of points to the ROC-PD statistic for $d = 10^5$. The ratios for the first 1,000 simulations appear. The ratios approach one, which suggests the value of the P_1^2 statistic converges as d becomes large.	63
3.6	Box plot of all 10,000 errors in P_1^2 statistic for different choices of d relative to $d = 10^5$ value. For $d = 1000$, the relative error is below 0.1%.	64
3.7	Simulated distributions of the ROC-PD statistic for different values of the binormal parameter δ under a simple null hypothesis. The parameter ρ has a constant value of one. One thousand iterations ran with sample sizes each 100.	65
4.1	Design of simulations for goodness-of-fit statistics.	69
4.2	Goodness-of-fit results for binormal ROCs. $\rho = 1$, $\delta = 0.5$, $n_1 = 50$, $n_2 = 50$.	79
4.3	Goodness-of-fit results for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 100$, and $n_2 = 100$.	80
4.4	Distributions of scores (sepal widths measured in centimeters) for two classes composed of species of Iris.	95
4.5	ROC curves for application with Iris data.	96
5.1	Design of calculation of parametric studentized bootstrap confidence regions.	116
5.2	ROC function estimates for audiology data discussed by Pepe (2004) and Stover et al. (1996).	119
5.3	Simultaneous confidence regions from parametric studentized bootstrap for binormal ROC parameters.	121

Abstract

GOODNESS-OF-FIT TESTS AND FUNCTION ESTIMATORS FOR RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES: INFERENCE FROM PERPENDICULAR DISTANCES

R. Bradley Patterson, PhD

George Mason University, 2012

Dissertation Director: Dr. John Miller

Statistical inference with the receiver operating characteristic (ROC) curve, a tool to assess performance of classification methods, has traditionally been based on differences between empirical and parametric curves in the vertical direction. New and unique in this dissertation is the use of differences in a perpendicular direction for goodness-of-fit tests, function estimators, and confidence regions for the ROC curve.

Working along directions perpendicular to parametric binormal ROC curves, we designed a goodness-of-fit test similar to existing statistics based on the empirical distribution function (EDF) for a single random variable. We initially demonstrate the poor performance of the direct application of the Cramér-von Mises family of goodness-of-fit statistics, which operate in only the vertical direction, to ROC curves. Through large simulations, our new test exhibits uniformity of p-values under the null hypothesis, and we prove its consistency.

Additional simulations with alternatives also show the goodness-of-fit test's power in three general cases. First, the test can reject the binormal model estimated with a fully parametric approach and small samples. Second, the test can reject the binormal model estimated with either a semi-parametric or a fully parametric approach and large samples. Third, the test can reject the binormal model with small samples if one of the underlying distributions follows a truncated, mixture, or heavy-tailed distribution.

The original function estimators developed here minimize differences in perpendicular directions between empirical and parametric ROC curves. One function estimator performs a total least squares fit that follows simply from principal component analysis and runs quickly. This short calculation time permitted the implementation of a parametric studentized bootstrap confidence region, a computationally intensive technique with smaller error than simpler bootstrap methods. Large simulations suggest the new estimator has the smallest integrated squared error among semi-parametric estimators for ROC curves that are not too steep. The simulations also demonstrate coverage probabilities of the new confidence regions to be close to nominal levels.

Three extensions of this work are possible now. First, different weight functions in the new goodness-of-fit statistic might improve the test's efficiency and raise its power. Second, the total-least-squares estimator could handle covariates through regression. Third, other parametric forms, such as the biexponential and bigamma, of the ROC curve share some of the same fundamental geometry of the binormal and could benefit from adaptation of the new statistics here.

Chapter 1: Introduction

This work focuses on measuring the goodness-of-fit and estimating functions of parametric receiver operating characteristic (ROC) curves. The ROC curve assesses the performance of classification methods used to identify observations by type. For instance, the ROC curve may indicate the performance of (1) a medical diagnostic test which classifies patients as healthy or sick, (2) a forensic DNA test which classifies genetic samples as matching or not, (3) a software filter which classifies emails as spam or ham, or (4) an astrophysics technique which classifies galaxies as elliptical or spiral.

To produce an ROC curve, a sample of observations with known classes must be available. Still, a finite sample results in an ROC plot which appears stepped like a staircase, possibly making it difficult to use. Often, the true ROC function may be a continuous curve which remains unknown. Statistical methods lead to smooth, parametric estimates of the true ROC function from the jagged empirical one.

To measure the quality of smooth, parametric ROC function estimates, we have developed a unique statistical goodness-of-fit test. Inspired by that development, we have also created two original estimators of parametric ROC functions.

A summary of our research appears in this chapter after general introductions to ROC curves, goodness-of-fit tests, and parameter estimation methods.

1.1 Overview of ROC curves

The ROC curve measures the performance of classification methods. ROC curves appeared in signal detection theory in the 1950s as Swets et al. (2000) and Krzanowski and Hand (2009) discuss. Today, a variety of fields use ROC curves. In medicine, ROC curves have widespread adoption, particularly for evaluating diagnostic tests (Zhou et al., 2002; Pepe, 2004; Lasko et al., 2005; Zou et al., 2007).

The general introduction by Fawcett (2006) covers their utility in data mining and machine learning. More of the development and history of the ROC curve comes in the subsequent chapter.

Before establishing terminology to discuss the ROC curve in detail, we highlight the tool's benefits in assessing the performance of classification methods. The ROC curve depicts a classification method's capability to distinguish classes and shows all attainable error rates across thresholds. Furthermore, the ROC curve's standard axes of true positive and false positive rates facilitate comparisons of various classification methods.

1.1.1 Classification methods, scores, and error rates

This brief background on ROC curves begins with a review of the outcomes for a classification method. Suppose that we have a classification method that evaluates observations to determine to which one of two classes they belong. For consistency with established terminology of ROC curves, label one class the *negative* and the other the *positive*. For a given observation, the classification method returns a number that we call a score. Over multiple observations, the classification method generates two distributions of scores, one for the negative class and one for the positive class. Imagine that high scores suggest the positive class and that low scores suggest the negative. We wish to measure the method's performance in discriminating between observations from the two classes.

Next, for the purposes of testing this performance, assume that we have observations whose true sources we know. We apply the classification method to this test data to obtain scores for all observations. Then we choose a threshold t on the scores. If the score for a given observation is greater than t , we suppose the observation arose from the positive class. Otherwise, we favor the negative class as the source of the observation. If we support the wrong hypothesis, we commit an error. When we favor the negative class when in truth an observation comes from the positive class, we call the error a *false negative*. Similarly, when we favor the positive class when in truth an observation comes from the negative class, we call the error a *false positive*. Table 1.1 lists the possible outcomes for a specific threshold t on the scores.

With multiple observations, we can compute the rates of these errors. We will emphasize the false positive rate and true positive rate, which define the axes of the ROC plot. The observed *false positive*

Table 1.1: Possible outcomes for a single threshold on scores.

		True class:	
		positive	negative
Predicted class:	positive	true positives	false positives
	negative	false negatives	true negatives

rate refers to the fraction of times that the classification method *incorrectly* assigns observations from the negative class to the positive class. That is,

$$\text{false positive rate} = \frac{\text{number of false positives}}{\text{number of true negatives} + \text{number of false positives}}.$$

Similarly, the *false negative rate* refers to the fraction of times that the classification method *incorrectly* assigns observations from the positive class to the negative class. So we can express the observed false negative rate as:

$$\text{false negative rate} = \frac{\text{number of false negatives}}{\text{number of true positives} + \text{number of false negatives}}.$$

The *true positive rate*, used on the ROC plot, is the complement of the false negative rate:

$$\text{true positive rate} = 1 - \text{false negative rate}.$$

The observed true positive rate refers to the fraction of times that the classification method *correctly* assigns observations from the positive class to the positive class.

In the preceding, a single threshold on scores yielded one set of rates, but we could vary the threshold to find all possible rates. Figure 1.1 demonstrates the concept of constructing the ROC curve from all possible thresholds. First we sort the scores from greatest to least and then consider all possible values of the threshold t . For each value of t , we find the values of the false positive and

true positive rates as before. The set of pairs of false positive and true positive rates at each value of t produce the ROC curve. Thus, the ROC curve shows the true positive rate versus the false positive rate for all possible values of the threshold t . With the adoption of higher scores for the positive

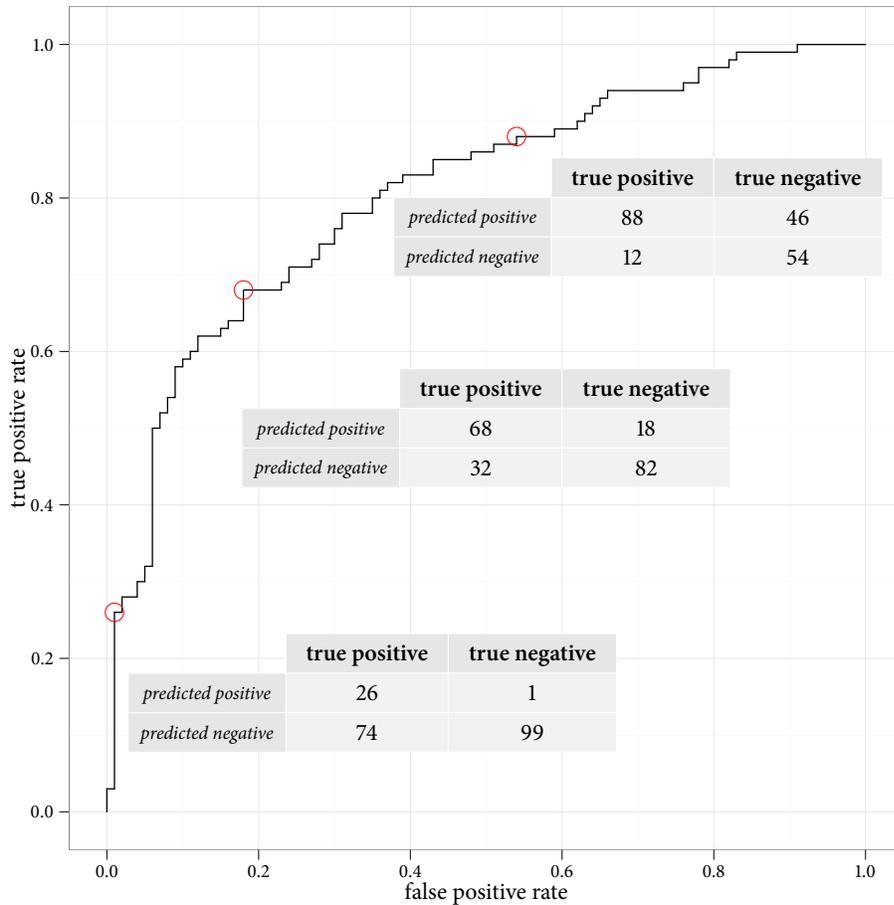


Figure 1.1: An ROC curve plots the true positive rate versus the false positive rate for all thresholds on scores. The step function is the empirical ROC curve for a sample of size 100 from the negative and positive distributions. The tables in the figure give the results at single values of the threshold t .

class, the value of the threshold t is greatest at the beginning of the curve. The ROC curve rises vertically for runs of true positives and crosses horizontally for runs of false positives. Below, we will see that only the relative ordering of scores from the positive and negative classes determines the

ROC curve. Because it depends on only the ranks of the scores, the ROC curve allows comparisons of classification methods that generate scores on different scales.

1.1.2 Underlying classes and distribution functions

In this section we introduce formal notation and definitions for ROC curves. We just gave an overview of the ROC curve, where we actually focused on an *empirical ROC curve*, which results from observed samples. Often, the empirical ROC curve may arise from some *true ROC curve*, which we may not actually know. However, we may believe that the true ROC curve belongs to a certain family of functions and has specific parameter values. Then we may seek an *estimated ROC curve*, which approximates the true ROC curve.

The true ROC curve relates to the true distributions of the negative and positive classes as follows. Let $F(s)$ and $G(s)$ be the cumulative distribution functions (CDFs) of the negative and positive classes over scores s . The corresponding probability density functions (PDFs) we denote by $f(s)$ and $g(s)$. We let U_{i_1} with $i_1 = 1, \dots, n_1$ be a random sample from the negative class and V_{i_2} with $i_2 = 1, \dots, n_2$ a random sample from the positive class. ROC curves depict the separation of the distributions $F(s)$ and $G(s)$ of scores from the two classes.

The ROC curve arises from equations for the horizontal and vertical coordinates parameterized by t , the threshold on scores s . Letting $x(t)$ denote the function for the horizontal coordinate and $y(t)$ the vertical, the equations are:

$$x(t) = 1 - F(t) \tag{1.1}$$

$$y(t) = 1 - G(t). \tag{1.2}$$

We classify observations with scores above t as positive and those with scores below t as negative. Then $x(t)$ equals the false positive rate and $y(t)$ the true positive rate.

Solving (1.1) for t , we have $t = F^{-1}(1 - x)$, where $F^{-1}(\cdot)$ is the inverse of the function $F(\cdot)$. (Throughout this work, we assume that $F^{-1}(\cdot)$ exists.) We may then substitute this result into (1.2)

to obtain an expression for $y(x)$:

$$y(x) = 1 - G[F^{-1}(1 - x)]. \quad (1.3)$$

In all that ensues, we will let the variable p denote the false positive rate and express the ROC curve as the following function:

$$\text{ROC}(p) = 1 - G[F^{-1}(1 - p)]. \quad (1.4)$$

Empirical and parametric ROC curves

We can generate an empirical ROC curve, which is a jagged step-function that estimates the actual, but unknown curve, using the true positive and false positive rates from test data. The expression for the empirical ROC curve involves both an empirical distribution function (EDF) and an empirical quantile function, which we first define. The EDF is an estimator of the true CDF. For a sample of size n_1 from the CDF $F(\cdot)$, Shao (2003) defines the EDF $\widehat{F}_{n_1}(\cdot)$ as:

$$\widehat{F}_{n_1}(s) = \frac{1}{n_1} \sum_{i_1=1}^{n_1} I_{(-\infty, s]}(U_{i_1}). \quad (1.5)$$

Csörgö (1983) defines the empirical quantile function, denoted by $\widehat{F}_{n_1}^{-1}(\cdot)$, as:

$$\begin{aligned} \widehat{F}_{n_1}^{-1}(q) &= \inf\{s : \widehat{F}_{n_1}(s) \geq q\} \\ &= U_{(i_1)} \text{ if } \frac{i_1 - 1}{n_1} < q \leq \frac{i_1}{n_1}, \quad i_1 = 1, \dots, n_1 \end{aligned} \quad (1.6)$$

where $U_{(i_1)}$ is the i_1^{th} order statistic. As Csörgö (1983, p. 2) writes, $\widehat{F}_{n_1}^{-1}(q)$ “is the left continuous inverse of the right continuous” \widehat{F}_{n_1} .

To define the empirical ROC curve, suppose that n_1 observations constitute the sample from the negative class with CDF $F(\cdot)$ and that n_2 observations make up the sample from the positive class

with CDF $G(\cdot)$. In terms of the EDF $\widehat{G}_{n_2}(\cdot)$ of the positive distribution and the empirical quantile function $\widehat{F}_{n_1}^{-1}(\cdot)$ of the negative distribution, we may write the empirical ROC as:

$$\widehat{\text{ROC}}(p) = 1 - \widehat{G}_{n_2} \left[\widehat{F}_{n_1}^{-1}(1 - p) \right]. \quad (1.7)$$

As discussed in the literature review, the empirical ROC is a pointwise consistent estimator of the true ROC (Hsieh and Turnbull, 1996).

If $F(\cdot)$ and $G(\cdot)$ both belong to parametric families, then the general form of the parametric ROC curve is:

$$\text{ROC}_{\theta}(p) = 1 - G_{\theta_2} \left[F_{\theta_1}^{-1}(1 - p) \right], \quad (1.8)$$

where θ_1 and θ_2 are the parameters for the negative and positive distributions, respectively.

When formulating a parametric model for the ROC curve, we could begin by considering the parameters for the underlying populations, or we could directly parameterize the ROC curve itself. Krzanowski and Hand (2009, p. 53) note that “postulating specific probability models for populations ... is generally termed a ... [fully] parametric approach to ROC curve estimation.” In contrast, “assuming that some unknown transformation converts both populations to a specific form ... is often referred to either as a ... [semiparametric] or a [parametric distribution-free] approach” (Krzanowski and Hand, 2009, p. 53). The semiparametric approach may be more sensible than the fully parametric one:

In my opinion, the idea of modeling ... [scores] in order to estimate the ROC curve is somewhat unnatural. The basic rationale for the ROC curve is to quantify the relationship between ... [the two distributions of scores] in a distribution-free manner. Thus, by definition, the ROC curve is invariant to monotone increasing transformations of the ... [scores]. However, ... methods that model the ... [scores] in order to estimate the ROC are not invariant to such data transformations. They are not distribution free in the sense that the ROC curve relies on the distributional forms for both [distributions

of scores], not just on their relationship or separation. There is a sort of philosophical rift between the ROC estimator produced by the approach ... [of modeling scores] and the basic framework of ROC analysis. Pepe (2004, p. 114)

Our focus for the ROC curve will be on semiparametric models, and we will explain the invariance property of ROC curves in more detail in the next section. We discuss the common semiparametric binormal ROC model below.

Binormal model

The binormal model for ROC curves assumes that under some unspecified, monotonic increasing transformation, the negative and positive classes follow normal distributions (Pepe, 2004). Let the negative class be normal with mean μ_1 and variance σ_1^2 and the positive class normal with mean μ_2 and variance σ_2^2 under this transformation. Representing the standard normal CDF by Φ , we write the CDFs of the negative and positive classes as:

$$\begin{aligned} F(t) &= \Phi\left(\frac{t - \mu_1}{\sigma_1}\right) \\ G(t) &= \Phi\left(\frac{t - \mu_2}{\sigma_2}\right). \end{aligned} \tag{1.9}$$

To arrive at a simple expression for $\text{ROC}_\theta(p)$, we take the inverse transformations for both of the preceding equations:

$$\begin{aligned} \Phi^{-1}[F(t)] &= \frac{t}{\sigma_1} - \frac{\mu_1}{\sigma_1} \\ \Phi^{-1}[G(t)] &= \frac{t}{\sigma_2} - \frac{\mu_2}{\sigma_2}. \end{aligned} \tag{1.10}$$

Rearranging the first equation leads to

$$t = \mu_1 + \sigma_1 \Phi^{-1}[F(t)], \tag{1.11}$$

which we can then substitute into the second:

$$\begin{aligned}\Phi^{-1}[G(t)] &= \frac{\mu_1 + \sigma_1 \Phi^{-1}[F(t)]}{\sigma_2} - \frac{\mu_2}{\sigma_2} \\ &= \frac{\mu_1 - \mu_2}{\sigma_2} + \frac{\sigma_1}{\sigma_2} \Phi^{-1}[F(t)].\end{aligned}\tag{1.12}$$

Introducing

$$\delta = \frac{\mu_2 - \mu_1}{\sigma_2} \quad \text{and} \quad \rho = \frac{\sigma_1}{\sigma_2},\tag{1.13}$$

we have

$$\Phi^{-1}[G(t)] = \rho \Phi^{-1}[F(t)] - \delta.\tag{1.14}$$

We may transform the above back to

$$G(t) = \Phi\{\rho \Phi^{-1}[F(t)] - \delta\},\tag{1.15}$$

which leads to

$$\begin{aligned}\text{ROC}_\theta(p) &= 1 - G[F^{-1}(1-p)] \\ &= 1 - \Phi\{\rho \Phi^{-1}[F(F^{-1}(1-p))] - \delta\} \\ &= 1 - \Phi[\rho \Phi^{-1}(1-p) - \delta] \\ &= 1 - \Phi[-\rho \Phi^{-1}(p) - \delta] \\ &= \Phi[\rho \Phi^{-1}(p) + \delta],\end{aligned}\tag{1.16}$$

where we used the facts that $\Phi^{-1}(1-p) = -\Phi^{-1}(p)$ and that $\Phi(z) = 1 - \Phi(-z)$. Thus, we see that the binormal ROC model has two parameters, δ and ρ , which one may estimate directly. We will discuss

our approach to estimating these parameters at the end of this chapter and in Chapter 5.

1.1.3 Examples of ROC curves

To give a sense of the form of ROC curves, we plot a few examples in this section. Figure 1.2 shows the movement of the ROC curve from the 45° line (which indicates performance no better than random guessing) to the upper left corner as the separation between the negative and positive distributions increases. For the examples in figure 1.2, the negative class follows a normal distribution with fixed mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 1$ while the positive class follows a normal distribution with variable mean μ_2 and constant standard deviation $\sigma_2 = 1$. In figures 1.3, 1.4, and 1.5 we present three different cases for the underlying distributions and the resulting ROC curves. We consider the case of a normal and a gamma, a normal and a mixture of normals, and a normal and a Cauchy for the negative and positive populations. We drew random samples of 50 observations from each class to produce empirical ROC curves in figures 1.3, 1.4, and 1.5.

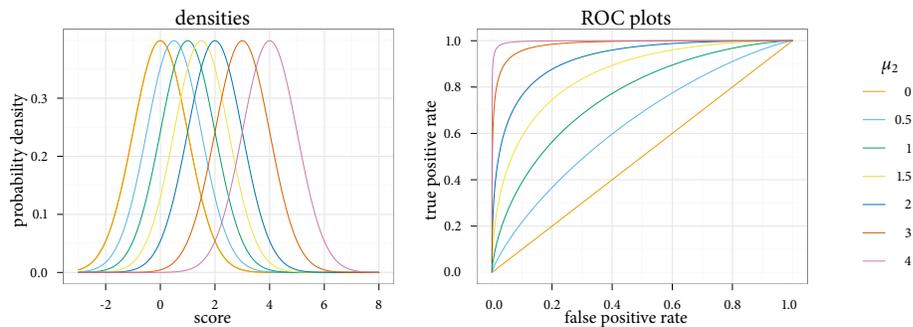


Figure 1.2: Demonstration of the form of the binormal ROC curve for different separations between the negative and positive distributions. In this example, the negative class follow a normal distribution with mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 1$ while the positive class follows a normal distribution with variable mean μ_2 and constant standard deviation $\sigma_2 = 1$.

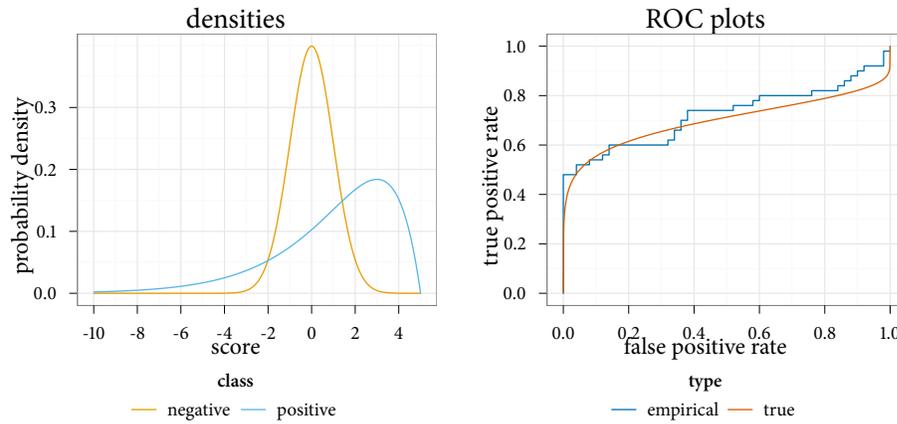


Figure 1.3: Example of ROC curve when negative class follows a normal distribution and positive class follows a reflected gamma distribution.

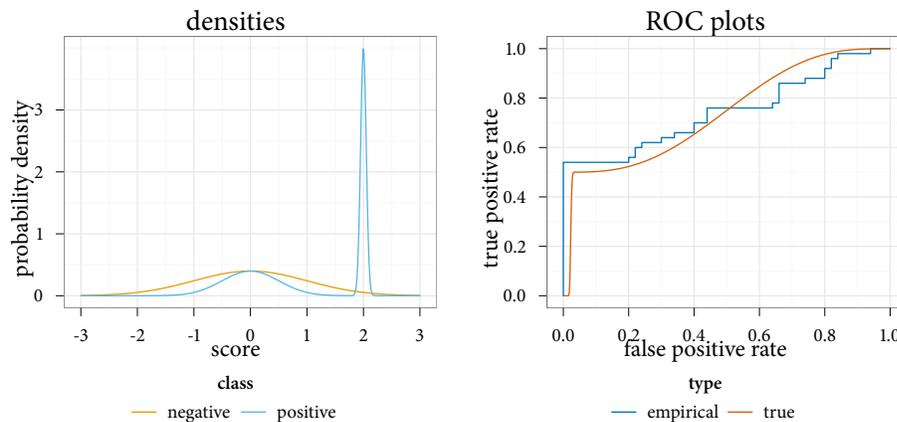


Figure 1.4: Example of ROC curve when negative class follows a normal distribution and positive class follows a mixture of normals.

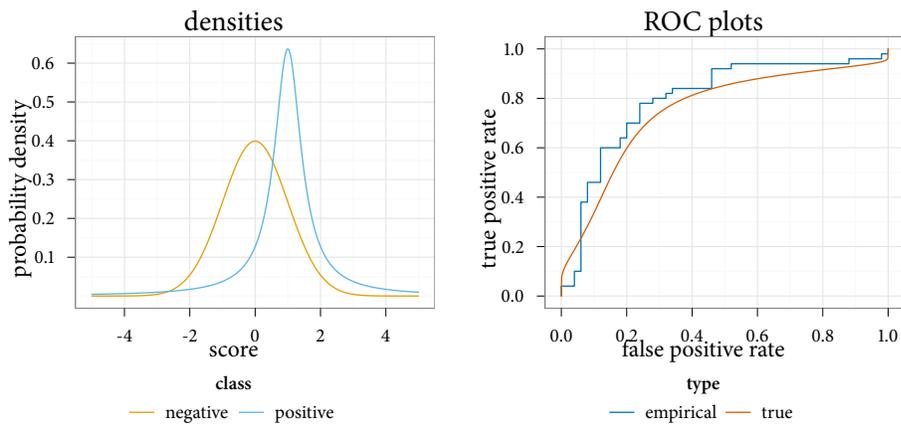


Figure 1.5: Example of ROC curve when negative class follows a normal distribution and positive class follows a Cauchy distribution.

1.1.4 Assessing performance with ROC curves

The ROC curve offers many benefits for evaluating the performance of classification methods.

Completeness and comparability

For assessing the effectiveness of classification methods, ROC curves offer the specific advantages of completeness and comparability. ROC curves present an entire picture of possible error rates achievable with a method. The ROC curve depicts this full range of error rates by plotting the true positive rate against the false positive rate at each possible threshold on the scores. So the ROC curve is a comprehensive tool that overcomes reliance on nominal thresholds or arbitrary ones based on scores. Instead, it allows more optimal possibilities based on the error rates.

By using standardized values of true positive and false positive rates, ROC curves make measures of performance comparable for distinct classification methods. ROC curves are independent of the scale or calibration of scores generated by a given method. The curves depend on only the relative order of scores from the positive and negative distributions. The ROC curve thus depicts a classification method's inherent capability of separating true positives and true negatives. Curves that approach the upper-left corner more closely indicate more separation between the positive and negative distributions.

Equal error rate (EER)

Several aspects of ROC curves provide further insight into the performance of classification methods. A typical value for measuring performance of methods in forensics and biometrics is the equal error rate (EER). The error rate achieved at the threshold where the false positive rate matches the false negative rate defines the EER. On the ROC plot, the EER occurs where a line running from the point (0,1) to (1,0) intersects the curve. The EER often does not match the error rates at nominal cutoffs.

Area under the curve (AUC)

Another important quantity, the area underneath the curve (AUC), defined as

$$\text{AUC} = \int \text{ROC}(p) dp, \quad (1.17)$$

gives two valuable indications of the performance of a classification method. First, the AUC equals the probability that a randomly selected observation V from the positive class would have a higher score than a randomly selected observation U from the negative (i.e., $\Pr[V > U]$). The AUC estimated from the empirical ROC relates to the Mann-Whitney U statistic (Bamber, 1975). Second, the AUC gives the mean true positive rate averaged uniformly across the false positive rate (Krzanowski and Hand, 2009). This average true positive rate is clearly distinct from the average error rate calculated as the mean of the false positive and false negative rates at a single threshold. By taking the average true positive rate over all possible false positive rates, we obtain a broader average measure of performance with the AUC. Hanley and McNeil (1982) share more on the AUC's interpretation.

Likelihood ratio

The likelihood ratio equals the probability density of scores from the positive class divided by the probability density of scores from the negative class at a given score. Not by coincidence, the slope of the ROC curve equals the likelihood ratio. As discussed in the next chapter, the development of ROC methodology proceeded directly from statistical decision theory. Swets (1973, p. 995) remarks that the “unification of several decision rules ... [by the likelihood ratio] established the generality of the ROC analysis.” To demonstrate that the ROC curve's derivative at a given pair of coordinates equals the likelihood ratio at the corresponding score, we note that if x and y are both parametrized in terms of t , then the derivative is:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}. \quad (1.18)$$

For the denominator, we have

$$\frac{dx}{dt} = \frac{d}{dt}(1 - F(t)) = -f(t), \quad (1.19)$$

and for the numerator, we have

$$\frac{dy}{dt} = \frac{d}{dt}(1 - G(t)) = -g(t). \quad (1.20)$$

We used the fact that the derivatives of the CDFs $F(t)$ and $G(t)$ are the PDFs $f(t)$ and $g(t)$. Therefore, the derivative of the ROC curve is

$$\frac{dy}{dx} = \frac{g(t)}{f(t)}, \quad (1.21)$$

which is the ratio of the probability density of the distribution of positive scores at the value t to the probability density of the distribution of negative scores at the value t . This corresponds to the likelihood ratio. Furthermore, this derivative presents a scale-free indication of the likelihood ratios obtainable with a classification method. The more steeply an ROC curve initially rises, the greater the possible likelihood ratios. Likewise, the more quickly the ROC curve flattens out to the value one, the smaller the possible likelihood ratios.

1.1.5 Two important mathematical properties of ROC curves

We next discuss two mathematical properties of ROC curves on which our development of a goodness-of-fit statistic for parametric models depends.

Invariant under strictly increasing transformations

First, the ROC curve is invariant under strictly increasing transformations of the scores (Pepe, 2004; Krzanowski and Hand, 2009). To demonstrate this property, let \mathcal{S} denote the set of scores with $\mathcal{S} \subset \mathbb{R}$

and $h(\cdot)$ a strictly increasing function. From Mattuck (1999), this definition means $h(a) < h(b)$ for all $a, b \in \mathcal{S}$ for which $a < b$. Therefore, for the random variables U and V from respectively the negative and positive class, $\Pr[U \leq t] = \Pr[h(U) \leq h(t)]$ and $\Pr[V \leq t] = \Pr[h(V) \leq h(t)]$. Points on the ROC curve for the transformed scores satisfy

$$x^*(t) = 1 - \Pr[h(U) \leq h(t)] = 1 - \Pr[U \leq t] = x(t) \quad (1.22)$$

$$y^*(t) = 1 - \Pr[h(V) \leq h(t)] = 1 - \Pr[V \leq t] = y(t).$$

Thus the ROC curve for the transformed scores has the same shape as the original ROC curve.

As mentioned in the previous section, this property of invariance under strictly increasing transformations of the scores motivates modeling the ROC curve directly. We briefly illustrate an undesirable result of modeling the two distributions of scores themselves. Assume that we treat the two as normal distributions and estimate separate means and variances for each. These estimates lead to an induced form for the ROC curve given by 1.16. However, applying a strictly increasing transformation that increases higher scores more than lower ones and then fitting the two distributions again would not in general result in the same ROC curve. For instance, such a transformation might magnify the standard deviation of the positive distribution more than the negative, and as a consequence, the parameter ρ in the binormal model would be smaller. Pepe (2004, p. 124) prefers “rank-based methods on the philosophical grounds that the ROC curve is invariant to monotone increasing data transformations and therefore so should its estimator.” By modeling the ROC curve itself, we ensure that the form depends on only the ranks of the scores from the distributions.

Monotonically increasing

The second mathematical property of ROC curves that we present is that they are monotonically increasing (Pepe, 2004; Krzanowski and Hand, 2009). We show this property by starting with (1.8) and noting that $F^{-1}(\cdot)$ is strictly increasing and $G(\cdot)$ is monotonically increasing (Casella and Berger, 2002). Consider two false positive values p_1 and p_2 such that $p_1 < p_2$. Then $1 - p_2 < 1 - p_1$, and $F^{-1}(1 - p_1) > F^{-1}(1 - p_2)$. Because $G(\cdot)$ is monotonically increasing, $G[F^{-1}(1 - p_1)] \geq G[F^{-1}(1 -$

$p_2)$], which implies

$$\text{ROC}(p_1) = 1 - G[F^{-1}(1 - p_1)] \leq 1 - G[F^{-1}(1 - p_2)] = \text{ROC}(p_2). \quad (1.23)$$

Thus, for false positive values p_1 and p_2 such that $p_1 < p_2$, $\text{ROC}(p_1) \leq \text{ROC}(p_2)$, and so the ROC curve is monotonically increasing with the false positive rate p .

1.1.6 ROC curves as CDFs

Recognizing the similarity between ROC curves and CDFs motivated our choice of EDF statistics. To show that the ROC curve is a CDF, we start with the idea of a placement value. As Pepe and Cai (2004) and Pepe and Longton (2005) discuss, in terms of a random observation V from the positive class and the CDF $F(\cdot)$ of the negative class, the term *placement value* refers to $1 - F(V)$. So the placement value equals the fraction of the negative distribution with a score higher than that of V . Here the negative class serves as the reference population. The ROC curve is then the CDF of the placement values (Pepe and Cai, 2004; Pepe and Longton, 2005; Pepe et al., 2009). Assuming continuous CDFs $F(\cdot)$ and $G(\cdot)$, we have in our notation

$$\begin{aligned} \Pr[1 - F(V) \leq p] &= \Pr[F(V) \geq 1 - p] = \Pr[V \geq F^{-1}(1 - p)] \\ &= 1 - \Pr[V \leq F^{-1}(1 - p)] = 1 - G[F^{-1}(1 - p)] \\ &= \text{ROC}(p). \end{aligned} \quad (1.24)$$

Treating the ROC curve as a CDF, we consider in detail adaptations of EDF-based goodness-of-fit tests in Chapter 3.

1.2 Goodness-of-fit tests in general

This short introduction to goodness-of-fit tests follows D’Agostino and Stephens (1986). The null hypothesis for a goodness-of-fit test is that a random variable has a given CDF $F(s)$. The null hypothesis may be simple or composite. A simple null hypothesis asserts that the data come from $F(s)$ with the parameters completely specified. Under a composite null hypothesis, we specify the family of distributions but must estimate the values of the parameters. The alternative hypothesis is that the data do not arise from the distribution specified in the null.

One of the more familiar goodness-of-fit tests is Pearson’s χ^2 , which Dorfman and Alf (1969) applied to their binormal model of the ROC curve for discrete scores. However, the focus of this work is the class of goodness-of-fit tests based on the EDF. Stephens (1986, p. 110) writes that “EDF statistics are usually much more powerful than the Pearson chi-square statistic.” In general, EDF-based goodness-of-fit tests measure the distance between the EDF $\widehat{F}_n(s)$ and the hypothesized CDF $F(s)$. We will consider in particular the following family of Cramér-von Mises statistics based on the EDF:

$$M_n^2 = n \int_{-\infty}^{\infty} [\widehat{F}_n(s) - F(s)]^2 \psi(s) dF(s), \quad (1.25)$$

where $\psi(s)$ is a weighting function. Setting $\psi(s) = 1$ gives the Cramér-von Mises statistic, and putting $\psi(s) = 1/\{F(s) \times [1 - F(s)]\}$ results in the Anderson-Darling statistic. While we investigate the adaptation of both of these statistics, we also explore other forms for $\psi(s)$. More details about these statistics appear in the literature review.

1.3 Parameter estimation

In addition to providing measures of goodness-of-fit, the distances discussed in the previous section also may serve as criteria for estimating parameters of a CDF. We refer to estimates that minimize such distances as minimum distance estimators (MDEs). In the review of the literature, we discuss existing MDEs for the binormal ROC curve. They do not include the type of weight given by the Cramér-von Mises statistics for the differences between the empirical and parameterized CDFs. We

also cover several other techniques of estimating the ROC curve's parameters such as maximum likelihood estimation and regression. Chapter 5 presents new semi-parametric estimators for binormal ROC curves.

1.4 Overview of this dissertation

We have broadly covered fundamentals of ROC curves, the binormal model, goodness-of-fit tests, and function estimation. Now we highlight the novel parts of our research on measuring the goodness-of-fit and estimating the functions of parametric binormal ROC curves.

1.4.1 Goodness-of-fit statistic for binormal ROC

A goodness-of-fit statistic for binormal ROC curves is the first new subject of this dissertation.

New developments and results

The new goodness-of-fit statistic uniquely measures a lack of fit between a parametric binormal and an empirical ROC curve in the perpendicular direction. To emphasize the perpendicular aspect, we refer to the goodness-of-fit test as the ROC-PD (perpendicular-direction) statistic. Measuring the perpendicular distance offers at least two advantages. While computing the vertical or horizontal distance between an empirical and parametric ROC curve along steep or shallow stretches can be unstable, computing the distance in the perpendicular direction provides stability. Furthermore, measuring the perpendicular distance gives balanced weight to the two types of errors for the negative and positive classes that underlie the ROC curve.

Simulations suggest the new ROC-PD statistic may be the most powerful test of the binormal model against various alternatives. The ROC-PD exhibits capability to reject a null hypothesis in two broad cases. First, if one of the underlying classes has a distinct bimodal distribution or has heavy tails, the ROC-PD statistic may be able to reject the null hypothesis of a binormal model for moderate sample sizes. Second, as the sample sizes of the two classes near 1000, the ROC-PD statistic may be able to reject the null hypothesis of a binormal model even if the true alternative ROC function

appears by eye to be in the binormal family. The inability to reject the binormal model for smaller sample sizes occurs mainly due to the variability of the empirical ROC function.

The ROC-PD statistic is a consistent statistical test of the goodness-of-fit of a parametric binormal ROC function to an empirical ROC function. The basis of the ROC-PD statistic is the empirical ROC function, which itself is a consistent estimator of the true ROC function. This property relates to the consistency of the EDF for the true CDF. As the ROC function is also a CDF, we initially investigated adapting the Cramer-von Mises and Anderson-Darling goodness-of-fit statistics, which are consistent tests based on the difference between an EDF and CDF. However, we found that Cramer-von Mises and Anderson-Darling statistics do not lead to reliable tests of the goodness-of-fit in the case of ROC curves. Still, by basing the ROC-PD statistic on the empirical ROC function, we developed a consistent test.

Finally, simulations aided in further characterizing the ROC-PD statistic. As desired in a goodness-of-fit test, the ROC-PD statistic appears to have a uniform distribution of p-values under a null hypothesis. The null distribution of the ROC-PD statistic itself depends on the specific values of the binormal parameters. Yet in all simulations, the distribution of the ROC-PD statistic remains one-sided and right-skewed.

Key steps

In developing the new ROC-PD statistic and obtaining results about its behavior, we completed several undertakings.

To guarantee the feasibility of computing the perpendicular distance, we established some basic properties of empirical and parametric binormal ROC functions. First, perpendicular lines from a parametric binormal ROC have unique intersections with the empirical ROC over the open interval of false positive rates from zero to one. Second, the slope of the binormal ROC changes from less than to greater than one or vice versa at most twice. This property ensures that the curve is relatively smooth and that computation of line integrals along the curve can be fairly stable.

Calculating an integrated perpendicular distance between the empirical and parametric binormal ROC functions requires considerable computation. We identified ordinary differential equations

(ODEs) that numerical ODE solvers can use to give accurate lengths along the arc of the binormal ROC curve. In addition, we implemented a fast search to find where perpendicular lines from the parametric binormal ROC intersect the empirical ROC.

After building a computational framework for the ROC-PD statistic, we conducted extensive computer simulations. The definition of the ROC-PD statistic does not lead to a closed form expression for its distribution. So the simulations enabled our study of the ROC-PD statistic's distribution for various parameter values and under null and alternative hypotheses.

Because the ROC-PD statistic does not have a closed form expression for its null distribution, we designed a parametric bootstrap to approximate p-values for the ROC-PD statistic.

We proved the consistency of the ROC-PD goodness-of-fit statistic.

1.4.2 Estimators for binormal ROC

The second new subject of this dissertation is a pair of estimators for the binormal ROC function.

New developments and results

Two original estimators for the binormal ROC function minimize distance in the perpendicular direction between the empirical and parametric ROC functions. The first estimator uses the ROC curve's original scale of true positives versus false positives. The second estimator applies probit transformations to the ROC curve's axes to arrive at a linear expression for the binormal ROC function. By finding the line with the minimum perpendicular distance to the observations, we effectively perform total least squares regression. This estimator, which we refer to as the ROC-TLS (total least squares) estimator, leads to function estimates with the smallest mean integrated squared error for curves that are not too steep.

Another novel development consists of parametric, studentized bootstrap confidence intervals for the individual binormal ROC parameters and simultaneous confidence regions for the joint distribution of the parameters. We implemented the confidence intervals for the ROC-TLS estimator, but other estimators could use the technique as well. In simulations, the 95% studentized bootstrap

simultaneous confidence regions appear to have coverage probabilities very close to the nominal levels for a range of binormal ROC curves.

Key steps

As with the ROC-PD statistic, the implementation of the two binormal ROC function estimators required computational techniques. Numerical nonlinear optimization functions from the statistical software R calculated estimates for the estimator based on the original ROC scale. Applying principal component analysis for the ROC-TLS estimator, we used linear algebra functions in the computer language C to produce fast estimates. The fast calculations permit studentized bootstrap confidence intervals to be computed in reasonable time (less than one minute on a modern computer). Large-scale computer simulations helped to characterize the behavior of the two new estimators as well as four existing estimators.

1.4.3 Outline of this dissertation

The organization of this dissertation proceeds as follows. Chapter 2 offers a history of the ROC curve and the binormal model and reviews the literature on measuring the goodness-of-fit and estimating parametric functions of ROC curves.

Chapter 3 covers the adaptation of the Cramér-von Mises family of statistics for ROC curves and introduces our new goodness-of-fit statistic in depth. There the reader will find the technical details of the statistic's definition in terms of the perpendicular distance between an empirical and continuous binormal ROC curve. We also discuss the practicalities of implementing the statistic's computation.

Chapter 4 contains a characterization of the goodness-of-fit statistic. Large simulations demonstrate the ROC-PD statistic's null distribution for different binormal parameters. Additional simulations indicate the test's power against distinct alternatives over various sample sizes. Chapter 4 also includes the lemmae and theorem with proofs that establish the consistency of the ROC-PD test statistic. Lastly, an application of the goodness-of-fit statistic appears at the end of Chapter 4.

Chapter 5 presents two new estimators for the binormal ROC function. Substantial simulations

compare the errors of the new estimators to several existing ones. Chapter 5 also introduces a new parametric studentized bootstrap for obtaining confidence intervals and regions for binormal ROC curves. At the end of Chapter 5, we apply the parametric studentized bootstrap to find confidence regions for binormal ROC curves that assess the performance of a diagnostic test for hearing loss.

The concluding Chapter 6 summarizes this dissertation's research and proposes topics for further exploration in the future.

Appendix A includes figures that depict results from the simulations for the goodness-of-fit statistic. Appendix B contains tables that list the errors of the function estimators across various simulations. The remaining appendices discuss specific topics separate from the main discourse and describe the computing environment and software.

Chapter 2: Literature review

This chapter covers the development of ROC curves and the binormal model as well as goodness-of-fit tests based on the empirical distribution function (EDF). The history of the ROC curve appears first with background on the binormal model following. For the binormal model, this account focuses on its applicability and existing techniques of estimating the function. We present general details of EDF-based goodness-of-fit statistics, particularly the Cramér-von Mises and Anderson-Darling tests. Then we review one goodness-of-fit test of the binormal ROC model in the case of discrete data and two in the case of continuous data.

2.1 History of the ROC curve

The ROC curve relates back to statistical decision theory. In the 1950s, it appeared in the field of signal detection theory, which used statistical decision theory as a foundation. Green and Swets (1966, p. 1) discuss the origins of signal detection theory, which included “an almost direct translation of statistical decision theory.” The field’s fundamental papers by Peterson et al. (1954) and van Meter and Middleton (1954) incorporated work in mathematical statistics by Neyman and Pearson (1933); Wald (1947, 1950); and Grenander (1950) according to Green and Swets (1966).

Subsequently, the ROC methodology spread to psychology, and the book by Swets and Pickett (1982) popularized the methodology in the evaluation of diagnostic systems, especially in medical fields such as radiology. More discussion about the history of the ROC curve in detection theory and psychology appears in Swets (1973), which also describes the origin of the term ROC:

It is because the ROC is a comparison of two operating characteristics that I use the term “relative” operating characteristic, according to a suggestion by Birdsall. Originally, serving to confuse the ROC and OC (operating characteristic) in the detection context,

the R stood for “receiver.” That terminology stemmed from the broader perspective of communications, which views detection as part of the reception process. Swets (1973, p. 995)

Today, the word “receiver” seems more common than “relative.”

2.2 Binormal model

Under the binormal model, an increasing transformation exists to convert the underlying distributions for the negative and positive classes to normals. According to Swets (1973), the assumption of two normal distributions in studying discrimination arose at least as early as 1927 in work on differentiating two psychological stimuli (Thurstone, 1927a,b).

In ROC methodology specifically, Green and Swets (1966) offer worthwhile discussion of the binormal model. The justification of Green and Swets (1966) for the assumption of the binormal model in the field of signal detection theory is both theoretical and practical. “Since we often think that sensory events are composed of a multitude of similar, smaller events, which are by and large independent, the central limit theorem might be invoked to justify the assumption of a Gaussian distribution of net effects” (Green and Swets (1966), p. 58). Practically, the Gaussian assumption provides more tractable mathematics, and the authors comment that, “Ultimately, the justification of any scientific assumption is pragmatic...” (Green and Swets, 1966, p. 58).

Green and Swets (1966) also consider using the binormal model with underlying populations which either do not share a common variance or come from distributions other than the normal. They write,

The justification for the Gaussian model with unequal variance is, we believe, not to be made on theoretical but rather on practical grounds. It is a simple, convenient way to summarize the empirical data with the addition of a single parameter [in the case of underlying exponentials]. Many practical questions about statistical processes can be answered from information about their first and second moments. The assumption

of a Gaussian distribution with unequal variance seems to be motivated by the same pragmatic goals.” (Green and Swets, 1966, p. 79)

Still, for data originating from exponential distributions, Green and Swets (1966) note two advantages of the biexponential model over the binormal. “First, the decision axis is monotonic with likelihood ratio” (Green and Swets, 1966, p. 79). Also, the biexponential “is a one-parameter distribution, and thus the ROC curve can be summarized by one parameter rather than by two” (Green and Swets, 1966, p. 79). Simulations to study fitting biexponential ROC data with a binormal model appear in Chapters 4 and 5.

Hanley (1996, p. 1575) reports that “[u]ntil the early 1980s, the parametric methods used to fit ROC curves and to derive summary indices of accuracy were based entirely on the binormal model.” In their 1982 book that popularized ROC methodology for diagnostic systems, particularly in medicine, Swets and Pickett (1982) support the binormal model. Swets and Pickett (1982, p. 30) discuss “a highly robust, empirical result, which is now substantiated in dozens of diverse applications; namely, that the empirical ROC is very similar in form to a theoretical ROC derived from normal probability distributions.” According to Hanley (1996, p. 1526), other investigations (Hanley, 1988; Swets, 1986) from around this time could not “find cases where the binormal model would seriously mislead” in analyzing ordinal data.

Motivated by the common assumption of normal distributions, possibly on the scale of a latent variable, in the fields of radiology, psychology, and epidemiology, Goddard and Hinberg (1990) conducted an empirical study of the binormal model with continuous data in 1990. The authors fit separate normals to output from seven different medical kits for diagnosing prostate cancer and compared the induced fully parametric ROC to the empirical one. They calculated the mean and standard deviation using both the original data and the logarithmic transform. Their results show disparities between the fully parametric ROC fit to the raw data and the empirical ROC. While the log-transformed data often produced a parametric ROC that looked closer to the empirical one, gross differences between the two curves still occurred occasionally. “. . . [T]here were cases where the log-transform based normal ROC curve poorly approximated the non-parametric curve. . . . We found

that the routine use of such a normal theory based approach for these data would have led to incorrect inferences” (Goddard and Hinberg, 1990, p. 335). This paper is an important example of potentially misleading results generated by fully parametric fits of binormal models.

Hanley (1996) focuses on continuous data and demonstrates that pairs of underlying distributions that are uniform and polynomial, both exponential, and both beta can appear binormal by eye after transformation. Thus, he argues that a semi-parametric approach, such as Metz’s LABROC algorithm detailed below, for fitting the binormal model is justifiable despite raw data that do not appear to come from two normals. He relied on visual inspection because at the time binormality could not “be judged using the significance level from a formal goodness-of-fit test; the assessment is necessarily subjective, based on the magnitude of the departures from binormality” (Hanley, 1996, p. 1577). We will show with our goodness-of-fit statistic that fully parametric fits of the binormal model can go astray, but often a semi-parametric approach will fit fine. For the pairs of distributions considered, Hanley (1996) reaches a reasonable conclusion:

Several recent papers dealing with quantitative test results have made statements to the effect that ‘because our data were not Gaussian, we could not use the binormal model. It is hoped that this paper dispels this misunderstanding about the use of the binormal model. Hanley (1996, p. 1583)

2.3 Estimation of binormal ROC functions

2.3.1 Maximum likelihood approaches

Perhaps the most prevalent algorithm for fitting semi-parametric binormal ROC models in the case of continuous scores comes from Metz et al. (1998), who modified the method proposed by Dorfman and Alf (1969) for discrete scores.

Dorfman and Alf (1968) derive results for maximum likelihood estimators (MLEs) of the binormal model in the case of “yes-no” responses with different “instructions.” The authors stated that, “An instruction usually consists of an a priori probability and/or payoff matrix” (Dorfman and Alf, 1968, p. 118). A payoff matrix gives the reward or penalty for each of the four outcomes in the two-class

classification problem. We may consider the instructions as producing different thresholds by which to classify observations. Each instruction then leads to a point on an ROC plot. A similar approach to estimating the parameters of a binormal ROC curve using maximum likelihood but with categorical ratings data appears in Dorfman and Alf (1969), on which we will base our exposition.

Dorfman and Alf (1969) develop maximum likelihood estimators for the parameters of the binormal ROC curve for categorical responses. In this case, the set of scores consists of discrete, ordered values, $\mathcal{S} = \{s_1, \dots, s_{k+1}\}$, where the s_i are the response ratings as denoted by Dorfman and Alf (1969). The thresholds come from the set of rankings, and the ROC plot consists of discrete points. Furthermore, the authors assume that these ratings arose from a latent variable, W . Following the authors' Axiom 3, we introduce cutoffs $\gamma_1, \dots, \gamma_k$ on W so that the observed response would be s_1 if $W < \gamma_1$, s_{i+1} if $\gamma_i < W < \gamma_{i+1}$ for $i < k$, and s_{k+1} if $W > \gamma_k$. Under the binormal model, the authors assume that on the scale of the latent variable, the distribution of the negative population is normal with mean μ_1 and variance σ_1^2 and the distribution of the positive population is normal with mean μ_2 and variance σ_2^2 . Assume that $\Phi\left(\frac{\gamma_0 - \mu_1}{\sigma_1}\right) = \Phi\left(\frac{\gamma_0 - \mu_2}{\sigma_2}\right) = 0$ and that $\Phi\left(\frac{\gamma_{k+1} - \mu_1}{\sigma_1}\right) = \Phi\left(\frac{\gamma_{k+1} - \mu_2}{\sigma_2}\right) = 1$. Then the probability for an observation from the negative class to have rating s_i is:

$$q_{1_i} = \Phi\left(\frac{\gamma_i - \mu_1}{\sigma_1}\right) - \Phi\left(\frac{\gamma_{i-1} - \mu_1}{\sigma_1}\right). \quad (2.1)$$

Similarly,

$$q_{2_i} = \Phi\left(\frac{\gamma_i - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{\gamma_{i-1} - \mu_2}{\sigma_2}\right) \quad (2.2)$$

is the probability for an observation from the positive class to have rating s_i . Introducing $\delta = (\mu_2 - \mu_1)/\sigma_2$ and $\rho = \sigma_1/\sigma_2$, we follow the authors' approach and let $\kappa_i = (\gamma_i - \mu_1)/\sigma_1$. Then the preceding probabilities become

$$q_{1_i} = \Phi(\kappa_i) - \Phi(\kappa_{i-1}) \quad (2.3)$$

and

$$q_{2_i} = \Phi(\rho\kappa_i - \delta) - \Phi(\rho\kappa_{i-1} - \delta), \quad (2.4)$$

which correspond to equations (1) and (2) in Dorfman and Alf (1969). With these definitions, the ROC curve follows the binormal model given by: $\text{ROC}(p) = \Phi[\delta + \rho\Phi^{-1}(p)]$.

To estimate the $k + 2$ parameters $(\rho, \delta, \kappa_1, \dots, \kappa_k)$, the authors use maximum likelihood estimation. The distribution of categorical responses for observations from each population is multinomial, and so the likelihood equation is:

$$L(\delta, \rho, \kappa_1, \dots, \kappa_k | \mathbf{x}, \mathbf{y}) = q_{1_i}^{m_{1_1}} q_{2_i}^{m_{1_2}} \dots q_{1_{k+1}}^{m_{1_{k+1}}} q_{2_i}^{m_{2_1}} q_{2_i}^{m_{2_2}} \dots q_{2_{k+1}}^{m_{2_{k+1}}}, \quad (2.5)$$

where m_{1_i} and m_{2_i} are respectively the numbers of observations from the negative and positive classes with rating s_i . Equation (3) in Dorfman and Alf (1969) gives the log of the likelihood with their slightly different notation. The authors propose finding the maximum of the likelihood equation by adapting the Newton-Raphson method.

The preceding MLE approach of Dorfman and Alf (1969) applies only to discrete scores. For continuous scores, Metz et al. (1998) adapt the method of Dorfman and Alf (1969) as follows. Metz et al. (1998) rank the observed data by score and then look for runs (one or more consecutive instances) of positives and negatives. They discretize the scores by considering the runs as ratings. With the discretized scores, they apply the original approach of Dorfman and Alf (1969). The algorithm of Metz et al. (1998) may currently be the most popular, and the authors distribute a software implementation called ROCKIT.

2.3.2 Minimum distance estimation

Minimum distance estimators (MDEs) for the parameters of the binormal ROC curve in the case of continuous scores appear in Hsieh and Turnbull (1996). They propose estimating the parameters of the binormal model by minimizing the squared difference between the empirical and parametrized ordinal dominance curve (ODC). The ODC curve plots the true negative rate versus the false negative

rate, or $F(t)$ versus $G(t)$, and results for the ODC curve apply to the ROC curve. Hsieh and Turnbull (1996) formulate the problem of estimating the parameters $\theta = (\delta, \rho)'$ in terms of the following empirical process:

$$\xi_{n_1 n_2}(\theta) = \widehat{F}_{n_1}[\widehat{G}_{n_2}^{-1}(p)] - \Phi[\delta + \rho\Phi^{-1}(p)] \quad (2.6)$$

They define the MDE $\hat{\theta}_{n_1 n_2} = (\hat{\delta}, \hat{\rho})'$ by

$$\|\xi_{n_1 n_2}(\hat{\theta}_{n_1 n_2})\| = \inf_{\theta} \|\xi_{n_1 n_2}(\theta)\|, \quad (2.7)$$

where they use the L_2 -distance

$$\|\xi_{n_1 n_2}(\theta)\| = \int_0^1 [\xi_{n_1 n_2}(\theta)]^2 dp. \quad (2.8)$$

While Hsieh and Turnbull (1996) give their L_2 distance in terms of the ODC, the equivalent of equation (2.8) for the ROC curve is:

$$\|\xi_{n_1 n_2}(\theta)\| = \int_0^1 [\widehat{\text{ROC}}(p) - \text{ROC}(p)]^2 dp. \quad (2.9)$$

This distance differs from both the Cramér-von Mises and perpendicular distances in our work.

For the empirical ROC curve Hsieh and Turnbull (1996) provide two important theorems, which then allow them to use the theory from Millar (1984) to obtain asymptotic distributions for their estimators and the goodness-of-fit statistic. First, the authors prove that the empirical ROC curve is strongly consistent for the true ROC curve. Second, the authors give a strong approximation for the ODC curve. This strong approximation enables them to prove convergence in L_2 of the empirical process

$$\xi_{n_1 n_2}(\theta) = \widehat{F}_{n_1}[\widehat{G}_{n_2}^{-1}(p)] - F[G^{-1}(p)] \quad (2.10)$$

to a simpler process. This result then permits them to apply Theorem 3.6 from Millar (1984) to obtain the asymptotic distributions of minimum L_2 distance estimators for the binormal parameters. Hsieh and Turnbull (1996) prove that these estimators are asymptotically normal and also derive their asymptotic covariance matrix. In addition, they use these results to find the asymptotic distribution for their goodness-of-fit measure for the binormal model (Corollary 4.2 in Hsieh and Turnbull (1996)).

While we noted the difference in the L_2 distance and the Cramér-von Mises distances above, we elaborate more on the distinction. Whereas the distance suggested by Hsieh and Turnbull (1996) is the L_2 -norm of the difference between the empirical and parametric model, the Cramér-von Mises and Anderson-Darling distances include a weight. Hsieh and Turnbull (1996) also mentioned that the above method provides a goodness-of-fit value for the binormal model, but it again omits a weight. Hsieh and Turnbull (1996, p. 39) remarked that “in the definition of the distance criterion, we could introduce a weight function. . . . Corresponding asymptotic results can be derived. . . . These ideas are the subject for future study, as are the computational problems for finding the MDE.”

2.3.3 Regression

Another approach to estimating the parameters of the semi-parametric ROC curve comes from regression. We trace the historical development of the regression modeling framework that led to generalized linear models (GLMs) for ROC curves, known as the the ROC-GLM approach, of which we then provide a practical example.

History

An approach to the direct regression modeling of ROC curves with covariate effects for continuous data appears in Pepe (1997). Earlier, methods of handling covariates either modeled the scores in the underlying populations or modeled a summary index (such as the AUC) as functions of covariates. The objective function for the estimators takes the general form of a weighted difference between the empirical and parametric ROC, which depends on both a baseline function and a vector of covariates. To solve the equations for the estimators, Pepe (1997) uses the Newton-Raphson method. As

for analytic results, Pepe (1997) showed that the estimators are weakly consistent (i.e., converge in probability), but omits their asymptotic distributions. To compute standard errors for the estimates, the author uses the bootstrap.

Pepe (1998) compares three techniques of fitting regression models to ROC curves with covariates for continuous response variables. The article is similar to Pepe (1997), but applies the methods to continuous data from an audiology study. The three methods were:

1. fitting the scores for underlying populations,
2. fitting summary indices from ROC curves for different combinations of covariates, and
3. fitting the ROC curve itself.

Among the advances to Pepe (1997), the use of a weight function $w(p) = 1 / \{ROC(p)[1 - ROC(p)]\}$ raises the possibility of more efficient estimators. As Pepe (1998, p. 128) explains, “This choice of weight function gives more weight to contributions defined at false-positive rates ... with smaller variance ... and thus is likely more efficient than a choice of $w(p) = 1$, for example.”

Pepe (2000) casts the original regression framework for ROC curves from Pepe (1997) in terms of GLMs. Pepe (2000, p. 353) observes that “each point on the ROC curve can be interpreted as a conditional expectation of a binary random variable.” Letting $S(U) \equiv 1 - F(U)$ denote the survival function of U , Pepe (2000) specifically presents the following equation:

$$\begin{aligned}
 \Pr [V \geq U | S(U) = p] &= \Pr [V \geq U | F(U) = 1 - p] \\
 &= \Pr [V \geq U | U = F^{-1}(1 - p)] \\
 &= \Pr [V \geq F^{-1}(1 - p)] \\
 &= 1 - \Pr [V \leq F^{-1}(1 - p)] \\
 &= 1 - G [F^{-1}(1 - p)] \\
 &= ROC(p),
 \end{aligned}$$

where p is a given false positive rate. We modified the original equation to give results in terms of CDFs. This result states that given a false positive rate of p , the value of the ROC curve is the

conditional probability that the score from a random positive observation is greater than or equal to the scores for the upper $100 \times p$ percent of the negative distribution.

ROC-GLM Motivated by this observation, Pepe (2000) introduces the following binary variable:

$$W_{i_2 i_1} = I(V_{i_2} \geq U_{i_1}), \quad (2.11)$$

for which the conditional expected value

$$\begin{aligned} E[W_{i_2 i_1} | F(U_{i_1}) = 1 - p_{i_1}] &= \Pr[V_{i_2} \geq U_{i_1} | F(U_{i_1}) = 1 - p_{i_1}] \\ &= \text{ROC}(p_{i_1}). \end{aligned} \quad (2.12)$$

This random variable could serve to estimate the coefficients in the following binary GLM:

$$\text{ROC}(p) = r^{-1} \left(\sum_k \beta_k h_k(p) \right), \quad (2.13)$$

where $r(\cdot)$ is the link function and the $h_k(\cdot)$ are baseline functions of p that serve as covariates. This binary GLM could replace the more general regression model in Pepe (1997). Because the actual false positive rates are unknown, Pepe (2000) suggests using an empirical parametric or nonparametric estimator $\widehat{F}(\cdot)$ (*not necessarily the EDF*) to estimate $\hat{p}_{i_1} = \widehat{F}^{-1}(U_{i_1})$ for $i_1 = 1, \dots, n_1$. This choice leads to an $n_2 \times n_1$ matrix of responses with elements $W_{i_2 i_1}$. The estimates for the parameters follow from the maximum likelihood method for GLMs. Pepe (2000) also derives the asymptotic distributions for the estimates found as roots to the likelihood equations.

Alonzo and Pepe (2002) follow up on Pepe (2000) by identifying a smaller set of binary variables to use in the GLM regression. Specifically, Alonzo and Pepe (2002) propose a finite set $\mathcal{P} = \{p_{i_1}\}$ of false positive rates at which to compute $W_{i_2 p_{i_1}} = 1 - I[V_{i_2} \leq F^{-1}(1 - p_{i_1})]$. This reduces the number of columns to facilitate computing. Again, the estimates $\widehat{F}^{-1}(1 - p_{i_1})$ must substitute for the true

values of $F^{-1}(1 - p_{i_1})$. The authors consider different sets for \mathcal{P} , but always with equally spaced false positive rates. They note the largest set is $\mathcal{P} = \{1/n_1, \dots, (n_1 - 1)/n_1\}$, but find that a set of 10 elements resulted in reasonably high efficiency for estimating parameters with 500 observations from each of the underlying populations in the case of the binormal with true parameters $(\delta, \rho) = (0.75, 0.9)$ and $(1.5, 0.85)$. Alonzo and Pepe (2002) use the bootstrap technique to estimate standard errors and forgo discussion of the asymptotic distribution of the estimators.

In the remainder of this review of the ROC-GLM approach, we detail a specific example of applying the ROC-GLM approach from Alonzo and Pepe (2002) to the binormal ROC model without additional covariates. Suppose that we have scores U_1, \dots, U_{n_1} from the negative class and V_1, \dots, V_{n_2} from the positive and that we wish to apply the ROC-GLM approach by fitting a GLM to the ROC curve. We start by introducing a binary variable, which we denote as $W_{i_2 p_{i_1}}$, based on the value of the i_2^{th} score, V_{i_2} , from the positive population and its relation to the CDF of the scores from the negative population. Converting the notation from Pepe (2004), we have:

$$\begin{aligned} W_{i_2 p_{i_1}} &= I[S(V_{i_2}) \leq p_{i_1}] = 1 - I[S(V_{i_2}) \geq p_{i_1}] \\ &= 1 - I[1 - F(V_{i_2}) \geq p_{i_1}] = 1 - I[V_{i_2} \leq F^{-1}(1 - p_{i_1})] \end{aligned} \quad (2.14)$$

where S is the survival function of the negative population, F is the CDF, and p_{i_1} belongs to a set $\mathcal{P} = \{p_{i_1}\}$ of false positive rates.

Noting that at a given value p_{i_1} ,

$$E(W_{i_2 p_{i_1}}) = 1 - \Pr[V_{i_2} \leq F^{-1}(1 - p_{i_1})] = 1 - G(F^{-1}(1 - p_{i_1})) = \text{ROC}(p_{i_1}), \quad (2.15)$$

we follow Alonzo and Pepe (2002) to estimate the ROC curve using the following GLM:

$$E(W_{i_2 p_{i_1}}) = \text{ROC}(p_{i_1}) = r^{-1} \left(\sum_k \beta_k h_k(p) \right). \quad (2.16)$$

The effective data used in fitting the GLM depends on the estimates $\widehat{F}^{-1}(1 - p_{i_1})$. For each value of $p_{i_1} \in \mathcal{P}$, the estimated values for the response come from:

$$\widehat{W}_{i_2 p_{i_1}} = 1 - I[V_{i_2} \leq \widehat{F}^{-1}(1 - p_{i_1})]. \quad (2.17)$$

We may view the responses as:

$$\widehat{W}_{i_2 p_{i_1}} = \begin{pmatrix} p_1 & p_2 & \dots & p_{n_1} \\ \widehat{W}_{11} & \widehat{W}_{12} & \dots & \widehat{W}_{1n_1} \\ \widehat{W}_{21} & \widehat{W}_{22} & \dots & \widehat{W}_{2n_1} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{W}_{n_21} & \widehat{W}_{n_22} & \dots & \widehat{W}_{n_2n_1} \end{pmatrix} \quad (2.18)$$

where the p_{i_1} indicate the corresponding false positive rates.

In the specific case of the binormal model, we may write the ROC curve as:

$$\text{ROC}(p) = \Phi(\delta + \rho\Phi^{-1}(p)). \quad (2.19)$$

Under the ROC-GLM setup, the link function $r(\cdot)$ is $\Phi^{-1}(\cdot)$ for the binormal model. The covariates are $h_1(p_{i_1}) = 1$ and $h_2(p_{i_1}) = \Phi^{-1}(p_{i_1})$. The likelihood equations for the binormal model under the ROC-GLM approach become:

$$\sum_{i_2=1}^{n_2} \sum_{p_{i_1} \in \mathcal{P}} \frac{\phi(\delta + \rho\Phi^{-1}(p_{i_1}))}{\text{ROC}(p_{i_1})(1 - \text{ROC}(p_{i_1}))} (\widehat{W}_{i_2 p_{i_1}} - \text{ROC}(p_{i_1})) = 0 \quad (2.20)$$

$$\sum_{i_2=1}^{n_2} \sum_{p_{i_1} \in \mathcal{P}} \frac{\phi(\delta + \rho\Phi^{-1}(p_{i_1}))}{\text{ROC}(p_{i_1})(1 - \text{ROC}(p_{i_1}))} \Phi^{-1}(p_{i_1}) (\widehat{W}_{i_2 p_{i_1}} - \text{ROC}(p_{i_1})) = 0.$$

The solutions for δ and ρ are the MLEs for the ROC-GLM approach with the binormal model.

Recent publications such as Pepe et al. (2009) and Gu and Pepe (2010) use the ROC-GLM approach discussed in Alonzo and Pepe (2002).

2.3.4 Recent developments

Cai and Pepe (2002) further develop the work in Pepe (1997, 2000) by permitting the baseline function $h(\cdot)$ to be nonparametric. Cai (2004) estimates the ROC curve by casting the regression model in terms of the placement values of the positive observations with respect to the negative.

Davidov and Nov (2009) develop estimators for the parameters of the binormal model using minimum-norm estimation. They equate the approach to solving the following weighted linear regression model:

$$\Phi^{-1} [\widehat{F}(p_j)] = \{ \delta + \rho \Phi^{-1} [\widehat{G}(p_j)] \} + \varepsilon \quad (2.21)$$

with weights w_j . This method minimizes the weighted vertical distance between a line with slope ρ and intercept δ and the probit transformations of coordinates from the empirical ROC curve. The authors prove that their estimators are strongly consistent for the true parameters and find their asymptotic distributions. They also comment on the work of Hsieh and Turnbull (1996)'s minimum L_2 distance problem: "It appears to be less tractable, and is left open (only the asymptotic properties of the estimators are derived, through the theory develop by Millar, 1984)."

Tang and Zhou (2011) introduce a least squares approach to estimating the semiparametric ROC curve. The technique takes the probit transformation of both the empirical ROC function $\widehat{ROC}(\cdot)$ and observed false positive rates p . A linear regression then provides the least-squares estimates of the parameters ρ and δ .

Other approaches of estimating the parameters of the ROC curve include the following. Zou and Hall (2000) offer estimates based on maximizing the rank-based likelihood. The two methods of Cai and Moskowitz (2004) use profile likelihood and pseudo-maximum likelihood. Gu et al. (2008) and Gu and Ghosal (2009) present a Bayesian approach, which they claim usually outperforms the MLE method of Metz et al. (1998), the method of Zou and Hall (2000), and the GLM method of Alonzo

and Pepe (2002).

2.4 Goodness-of-fit of binormal ROC functions

The previous section covered methods of estimating parametric binormal ROC curves. Here we change our focus to assessing the goodness of such fits. First, we review goodness-of-fit tests that measure the quality of estimated functions in general. Second, we describe developments of goodness-of-fit tests specifically for binormal ROC curves.

2.4.1 General background on goodness-of-fit statistics

This section includes a brief background on existing goodness-of-fit statistics for the distribution of a single random variable. In Chapter 3, we will share our modifications of these well-known goodness-of-fit tests for use with ROC curves.

We will explain these tests in terms of a continuous random variable U . Following the notation of section 1.1.2, suppose that U has a distribution function $F(s)$ for $s \in \mathcal{S} \subset \mathbb{R}$.

Chi-squared

In 1900, Pearson proposed the χ^2 goodness-of-fit statistic (Pearson, 1900). This exposition of the χ^2 goodness-of-fit statistic follows Moore (1986) and considers the case of a continuous random variable. To test the fit of a hypothesized CDF $F(s)$ given a random sample U_1, \dots, U_n , the statistic requires that we divide the domain of $F(s)$ into cells. Then we compare the observed and expected counts in each cell.

Suppose that there are K cells labeled C_j for $j = 1, \dots, K$. To calculate the expected number of observations in the j^{th} cell, we integrate the probability density function $f(s)$ over the interval of values in the cell. Let q_j denote the the probability that an observations lands in cell C_j defined by

the interval $(a_j, b_j]$ for $a_j, b_j \in \mathcal{S}$ and $a_j < b_j$. Then we have

$$q_j = \int_{C_j} f(s) ds = \int_{a_j}^{b_j} f(s) ds = F(b_j) - F(a_j). \quad (2.22)$$

The expected number of observations in cell C_j is then nq_j . Express the count of the observed sample values u_1, \dots, u_n that fall in the cell C_j by o_j . That is,

$$o_j = \sum_{i=1}^n I(a_j < u_i \leq b_j), \quad (2.23)$$

where $I(\cdot)$ is the indicator function. Then the Pearson χ^2 goodness-of-fit statistic is:

$$\chi^2 = \sum_{j=1}^K \frac{(o_j - nq_j)^2}{nq_j}. \quad (2.24)$$

Under the null hypothesis, this statistic has a χ^2 asymptotic distribution with $K-1$ degrees of freedom.

As Moore (1986, p. 91) comments, “Chi-squared tests are generally less powerful than EDF tests and special-purpose tests of fit.” Most of our attention goes to EDF tests, covered in the ensuing section. For more on the χ^2 statistic, Greenwood and Nikulin (1996) provides a thorough account.

Cramér-von Mises and Anderson-Darling statistics

History We share the history of the Cramér-von Mises family of statistics as covered by Durbin (1973) and Thas (2009). In 1928, Cramér introduced the L_2 difference between the empirical and parametric CDFs as a statistic (Cramér, 1928). Von Mises then added a general weight function to the L_2 difference (von Mises, 1931). By modifying the statistic discussed by von Mises (1931) to the general form

$$M_{n_1}^2 = n_1 \int_{-\infty}^{\infty} [\widehat{F}_{n_1}(s) - F(s)]^2 \psi(s) dF(s), \quad (2.25)$$

Smirnov (1936, 1937) made it distribution-free. Anderson and Darling (1952, 1954) investigated the case of $\psi(s) = 1/\{F(s)[1 - F(s)]\}$ thoroughly, and so today it bears their name.

Definitions and computing formulas The complete forms of the Cramér-von Mises and Anderson-Darlings statistics appear below. Starting with (2.25) and letting $\psi(s) = 1$ lead to the Cramér-von Mises statistic:

$$W_{n_1}^2 = n_1 \int_{-\infty}^{\infty} [\widehat{F}_{n_1}(s) - F(s)]^2 dF(s). \quad (2.26)$$

In the Anderson-Darling statistic, $\psi(s) = 1/\{F(s)[1 - F(s)]\}$,

$$A_{n_1}^2 = n_1 \int_{-\infty}^{\infty} [\widehat{F}_{n_1}(s) - F(s)]^2 \frac{1}{F(s)(1 - F(s))} dF(s). \quad (2.27)$$

Formulas for the Cramér-von Mises and Anderson-Darling statistics that facilitate computation also exist. Versions of these equations appeared in Stephens (1986). Let $U_{(i_1)}$ be the i_1^{th} order statistic of the sample. Then we have the following different forms:

$$W_{n_1}^2 = \sum_{i_1=1}^{n_1} \left[F(u_{(i_1)}) - \widehat{F}_{n_1}(u_{(i_1)}) + \frac{1}{2n_1} \right]^2 + \frac{1}{12n_1}, \quad (2.28)$$

$$W_{n_1}^2 = \sum_{i_1=1}^{n_1} \left[F(u_{(i_1)}) - \frac{2i_1 - 1}{2n_1} \right]^2 + \frac{1}{12n_1}, \quad (2.29)$$

$$A_{n_1}^2 = -n_1 - \frac{1}{n_1} \sum_{i_1=1}^{n_1} (2i_1 - 1) \{ \log F(u_{(i_1)}) + \log[1 - F(u_{(n_1+1-i_1)})] \}, \quad \text{and} \quad (2.30)$$

$$A_{n_1}^2 = -n_1 - \frac{1}{n_1} \sum_{i_1=1}^{n_1} \{ (2i_1 - 1) \log F(u_{(i_1)}) + (2n_1 + 1 - 2i_1) \log[1 - F(u_{(i_1)})] \}. \quad (2.31)$$

Distributions Below we consider the exact and asymptotic distributions of the Cramér-von Mises and Anderson-Darling statistics under simple and complex null hypotheses. The simple null hypothesis is that observations for a random sample of U arose from the continuous CDF $F(s)$ with specific parameters θ_1 . Under the simple null hypothesis, the statistics $W_{n_1}^2$ and $A_{n_1}^2$ have known asymptotic distributions with some values tabulated. Furthermore, the null distributions of $W_{n_1}^2$ and $A_{n_1}^2$ do not depend on the hypothesized CDF $F(\cdot)$ in the case of a simple null hypothesis (DasGupta, 2008; Thas, 2009). Anderson and Darling (1952) derived exact expressions for the asymptotic distributions of $W_{n_1}^2$ and $A_{n_1}^2$ as infinite sums. Thas (2009, p. 139) explains that “the CvM and AD statistics are asymptotically equivalent in distribution to the random variables

$$W = \sum_{j=1}^{\infty} \frac{1}{j^2 \pi^2} Z_j^2 \quad (2.32)$$

and

$$A = \sum_{j=1}^{\infty} \frac{1}{j(j+1)} Z_j^2 \quad (2.33)$$

where Z_1, Z_2, \dots are iid standard normal random variables. These represent infinite weighted sums of independent chi-squared random variates...

For finite samples, Thas (2009, p. 137) writes, “The exact distribution of the CvM statistic has received much attention in the statistical literature already for more than 50 years, and still the exact distribution is only tabulated for $n = 1, \dots, 7$.” Thas (2009, p. 138) continues, “Less is known about the exact null distribution of the Anderson-Darling statistic. Lewis (1961) gave the exact distribution when $n = 1$, but, to our knowledge at least, there are no exact results for larger sample sizes.”

If we use estimates of θ_1 in the CDF $F(s)$ under a composite null hypothesis, the asymptotic distributions of $W_{n_1}^2$ and $A_{n_1}^2$ do not in general follow the forms previously mentioned for the simple hypothesis. Thas (2009, p. 140) writes that these distributions may depend “on the hypothesized distribution $[F(\cdot)]$, as well as on the unknown” parameters θ_1 . He adds that as a result, “the distribution

cannot be tabulated, and percentage points and p -values must be approximated using the bootstrap” (Thas, 2009, p. 140). If $F(\cdot)$ belongs to a location-scale family, the statistics will however depend on only $F(\cdot)$ and not the parameters θ_1 . Furthermore, Thas (2009, p. 140) notes that “no simple analytic expression for the asymptotic distribution” functions exist for $W_{n_1}^2$ and $A_{n_1}^2$ under the composite null hypothesis.

2.4.2 Specific goodness-of-fit statistics for estimated binormal ROC functions

In addition to maximum likelihood estimators for the binormal ROC curve in the case of discrete responses, Dorfman and Alf (1969) also suggest a χ^2 goodness-of-fit test. We modify their notation to express the estimated probabilities for each category as:

$$\hat{q}_{1_i} = \frac{m_{1_i}}{n_1} \qquad \hat{q}_{2_i} = \frac{m_{2_i}}{n_2} \qquad (2.34)$$

for the negative and positive populations, respectively. Then we may write their χ^2 goodness-of-fit test as:

$$\chi^2 = \sum_{i=1}^{k+1} n_1 \frac{(\hat{q}_{1_i} - q_{1_i})^2}{q_{1_i}} + \sum_{i=1}^{k+1} n_2 \frac{(\hat{q}_{2_i} - q_{2_i})^2}{q_{2_i}} \qquad (2.35)$$

where the values of q_{1_i} and q_{2_i} come from equations 2.3 and 2.4 but with the MLEs substituted for the values of $\delta, \rho, \kappa_1, \dots, \kappa_k$.

Ideas for testing the goodness-of-fit of continuous ROC curves have appeared more recently. Hsieh and Turnbull (1996) propose using the minimum L_2 distance

$$\inf_{\theta} \|\xi_{n_1 n_2}(\theta)\|,$$

as a goodness-of-fit criterion for the assumption of the binormal model. Zou et al. (2003) as well as Zou et al. (2005) base their goodness-of-fit test around the AUC, estimated from the parametric model and from the nonparametric Mann-Whitney Wilcoxon U-statistic. For a better large-sample approximation, they advise a probit transformation, $W = \Phi^{-1}(A)$ of the AUC. Then for the AUC

estimated from the parametric model, they have $\hat{W}_P = \Phi^{-1}(\hat{A}_P)$, and for the nonparametric estimate of the AUC, they have $\hat{W}_N = \Phi^{-1}(\hat{A}_N)$. Noting that $\text{Var}(|\hat{W}_N - \hat{W}_P|) = \text{Var}(\hat{W}_N) - \text{Var}(\hat{W}_P)$, they define their goodness-of-fit statistic as:

$$\hat{D} = \frac{|\hat{W}_N - \hat{W}_P|}{\sqrt{\text{Var}(\hat{W}_N) - \text{Var}(\hat{W}_P)}}.$$

Assuming that the parametric model is correct, the asymptotic distribution of the statistic is the same as that of the absolute value of a standard normal random variable. Cai and Zheng (2007) developed procedures based on cumulative residuals.

Chapter 3: Development of goodness-of-fit tests for receiver operating characteristic curves

The first two chapters provided an overview and some history of receiver operating characteristic (ROC) curves and goodness-of-fit statistics based on empirical distribution functions (EDFs). Indeed, the ROC curve is itself a cumulative distribution function (CDF). So let us consider adapting EDF statistics, which measure the goodness-of-fit of a parametric CDF to the EDF, to assess the goodness-of-fit of a parametric ROC curve to the empirical one. Ultimately, the EDF statistics suffer when directly applied to ROC curves, but lead to successful alternatives.

3.1 Cramér-von Mises family applied to ROC curves

In section 2.4.1 we discussed the family of Cramér-von Mises statistics for a single, continuous random variable. This family measures distance between an EDF and a CDF. Here we develop a version of the Cramér-von Mises family of statistics for ROC curves.

Recall that $\text{ROC}(p)$ and $\widehat{\text{ROC}}(p)$ denote a parametric and empirical ROC curve, respectively. Let the function $\psi(p)$ provide a weighting. Then the common form for a goodness-of-fit statistic of the ROC curve based on Cramér-von Mises statistics is:

$$\begin{aligned} M_{\text{ROC}}^2 &= \int_0^1 [\widehat{\text{ROC}}(p) - \text{ROC}(p)]^2 \psi[\text{ROC}(p)] d\text{ROC}(p) \\ &= \int_0^1 [\widehat{\text{ROC}}(p) - \text{ROC}(p)]^2 \psi[\text{ROC}(p)] \text{ROC}'(p) dp. \end{aligned} \tag{3.1}$$

In the ensuing sections, we explore the use of weights corresponding to the Cramér-von Mises and Anderson-Darling statistics. We derive a computational form for the Cramér-von Mises statistic in the case of ROC curves and study its distribution.

3.1.1 Adaptation of Cramér-von Mises for ROC

Equations (2.29) and (2.30) from Chapter 2 give computational forms for the Cramér-von Mises and Anderson-Darling statistics in the case of a continuous CDF $F(\cdot)$. Below we arrive at a similar result for the ROC curve treated as a CDF.

Derivation

The empirical ROC curve has important distinctions from the EDF $\widehat{F}_{n_1}(\cdot)$ of a single continuous random variable. The EDF $\widehat{F}_{n_1}(\cdot)$ is a step function, and if we assume no ties among observations of the continuous variable, the consecutive heights of the steps all have increments equal to $1/n_1$. However, for the empirical ROC curve, the heights of the steps may vary. As a result, the integral in equation (3.1) has a different computational form.

We introduce notation and definitions to simplify the adaptation of the Cramér-von Mises family of statistics to ROC curves. Assume that the empirical ROC curve comes from samples of sizes n_1 and n_2 from the negative and positive classes. Let k be the number of unique, contiguous runs of positive observations in the sample data. (For intuition, k is the number of vertical jumps in the step function.) We use a_i and b_i to denote respectively horizontal and vertical coordinates of the empirical ROC with $i = 0, 1, \dots, k + 1$. Define the following:

$$\begin{aligned} a_0 &\equiv 0 \\ b_0 &\equiv 0 \\ a_{k+1} &\equiv 1 \\ b_k &\equiv 1 \\ b_{k+1} &\equiv 1. \end{aligned} \tag{3.2}$$

The values of a_1, \dots, a_k come from the unique, observed false positive rates (FPRs) as determined

by the empirical ROC. More specifically, let $\widehat{F}_{n_1}(\cdot)$ be the EDF of observations u_1, \dots, u_{n_1} from the negative class. Let $v_{(1)}, \dots, v_{(n_2)}$ be the ordered observations from the positive class. A subset $\{w_{(1)}, \dots, w_{(k)}\}$ of $\{v_{(1)}, \dots, v_{(n_2)}\}$ with $k \leq n_2$ may denote ordered values from contiguous runs of positive observations. Then,

$$a_i = 1 - \widehat{F}_{n_1}(w_{(k+1-i)}) \quad \text{for } i = 1, \dots, k. \quad (3.3)$$

The values of b_1, \dots, b_{k-1} come from the heights of the empirical ROC:

$$b_i = \widehat{\text{ROC}}(p) \text{ for } a_i \leq p < a_{i+1} \text{ with } i = 1, \dots, k. \quad (3.4)$$

Figure 3.1 shows examples of the coordinates on four empirical ROC curves, for each of which $k = 4$. The four plots in figure 3.1 demonstrate the four combinations of either vertical or horizontal segments at the beginning and end of an ROC curve.

We may then express the value in equation (3.1) as the sum of integrals over the horizontal stretches of the stepped empirical ROC curve:

$$M_{\text{ROC}}^2 = \sum_{i=0}^k \int_{a_i}^{a_{i+1}} [b_i - \text{ROC}(p)]^2 \psi[\text{ROC}(p)] \text{ROC}'(p) dp. \quad (3.5)$$

Let $r = \text{ROC}(p)$. Then $dr = \text{ROC}'(p) dp$. Also define $c_i \equiv \text{ROC}(a_i)$. The distance becomes:

$$\begin{aligned} M_{\text{ROC}}^2 &= \sum_{i=0}^k \int_{a_i}^{a_{i+1}} [b_i - \text{ROC}(p)]^2 \psi[\text{ROC}(p)] \text{ROC}'(p) dp \\ &= \sum_{i=0}^k \int_{c_i}^{c_{i+1}} (b_i - r)^2 \psi(r) dr \\ &= \sum_{i=0}^k \int_{c_i}^{c_{i+1}} [b_i^2 \psi(r) - 2b_i r \psi(r) + r^2 \psi(r)] dr. \end{aligned} \quad (3.6)$$

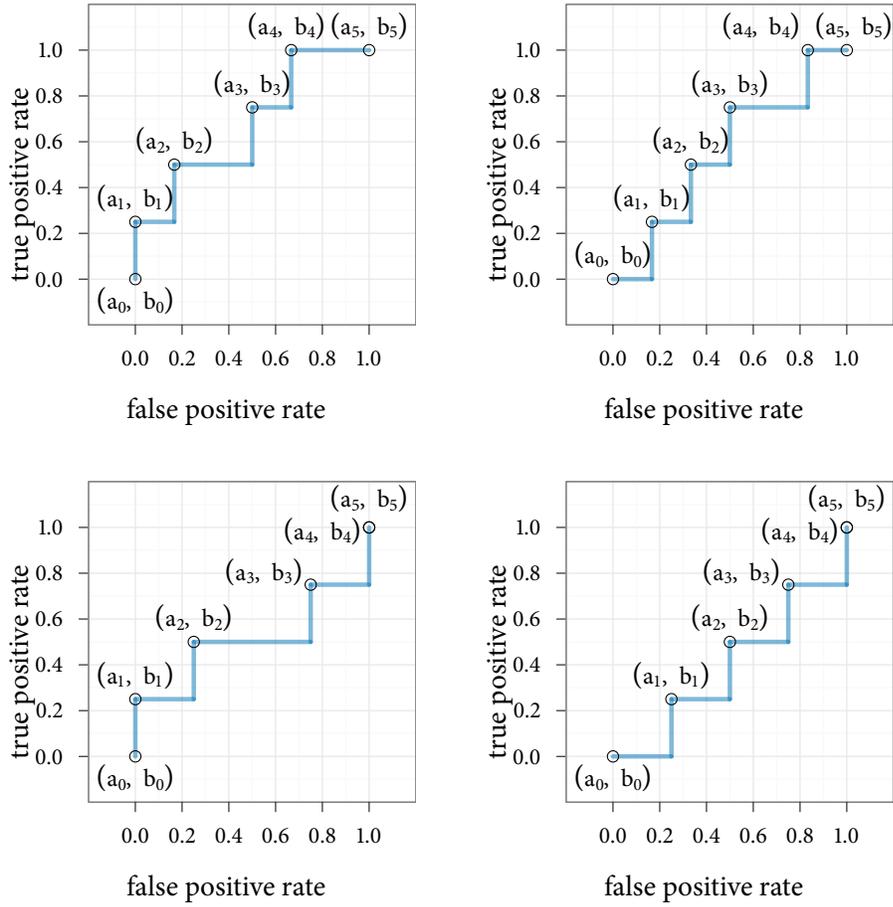


Figure 3.1: Examples of the a and b coordinates on empirical ROC curves. In the bottom two plots, the pairs (a_4, b_4) and (a_5, b_5) refer to the point at the end.

As in Durbin (1973), define

$$\begin{aligned} \phi_1(s) &= \int_0^s \psi(r) dr \\ \phi_2(s) &= \int_0^s r\psi(r) dr. \end{aligned} \tag{3.7}$$

Then

$$\begin{aligned}
M_{\text{ROC}}^2 &= \sum_{i=0}^k b_i^2 [\phi_1(c_{i+1}) - \phi_1(c_i)] - 2 \sum_{i=0}^k b_i [\phi_2(c_{i+1}) - \phi_2(c_i)] + \int_0^1 r^2 \psi(r) dr \\
&= \sum_{i=0}^k b_i^2 \phi_1(c_{i+1}) - \sum_{i=0}^k b_i^2 \phi_1(c_i) - 2 \sum_{i=0}^k b_i \phi_2(c_{i+1}) + 2 \sum_{i=0}^k b_i \phi_2(c_i) + \int_0^1 r^2 \psi(r) dr.
\end{aligned} \tag{3.8}$$

Now let $j = i + 1$ and express the preceding as

$$\begin{aligned}
M_{\text{ROC}}^2 &= \sum_{j=1}^{k+1} b_{j-1}^2 \phi_1(c_j) - \sum_{i=0}^k b_i^2 \phi_1(c_i) - 2 \sum_{j=1}^{k+1} b_{j-1} \phi_2(c_j) + 2 \sum_{i=0}^k b_i \phi_2(c_i) + \int_0^1 r^2 \psi(r) dr \\
&= b_k^2 \phi_1(c_{k+1}) - b_0^2 \phi_1(c_0) + \sum_{i=1}^k (b_{i-1}^2 - b_i^2) \phi_1(c_i) - 2b_k \phi_2(c_{k+1}) + 2b_0 \phi_2(c_0) \\
&\quad - 2 \sum_{i=1}^k (b_{i-1} - b_i) \phi_2(c_i) + \int_0^1 r^2 \psi(r) dr.
\end{aligned} \tag{3.9}$$

Recall that $b_0 = 0$, which leads to

$$\begin{aligned}
M_{\text{ROC}}^2 &= b_k^2 \phi_1(c_{k+1}) + \sum_{i=1}^k (b_{i-1}^2 - b_i^2) \phi_1(c_i) - 2b_k \phi_2(c_{k+1}) - 2 \sum_{i=1}^k (b_{i-1} - b_i) \phi_2(c_i) + \int_0^1 r^2 \psi(r) dr \\
&= \sum_{i=1}^k (b_{i-1} - b_i) [(b_{i-1} + b_i) \phi_1(c_i) - 2\phi_2(c_i)] + b_k^2 \phi_1(c_{k+1}) - 2b_k \phi_2(c_{k+1}) + \int_0^1 r^2 \psi(r) dr.
\end{aligned} \tag{3.10}$$

Because $b_k = 1$ and $c_{k+1} = \text{ROC}(a_{k+1}) = \text{ROC}(1) = 1$,

$$\begin{aligned}
b_k^2 \phi_1(c_{k+1}) - 2b_k \phi_2(c_{k+1}) + \int_0^1 r^2 \psi(r) dr &= \phi_1(1) - 2\phi_2(1) + \int_0^1 r^2 \psi(r) dr \\
&= \int_0^1 (1-r)^2 \psi(r) dr.
\end{aligned} \tag{3.11}$$

So we have

$$\begin{aligned}
M_{\text{ROC}}^2 &= \sum_{i=1}^k (b_{i-1} - b_i) [(b_{i-1} + b_i)\phi_1(c_i) - 2\phi_2(c_i)] + \int_0^1 (1-r)^2 \psi(r) dr \\
&= 2 \sum_{i=1}^k (b_i - b_{i-1}) \left[\phi_2(c_i) - \frac{b_{i-1} + b_i}{2} \phi_1(c_i) \right] + \int_0^1 (1-r)^2 \psi(r) dr.
\end{aligned} \tag{3.12}$$

Equation (3.12) is a general computational form of the family of Cramér-von Mises statistics for ROC curves. Different weight functions $\psi(\cdot)$ lead to particular statistics.

3.1.2 Characterization of Cramér-von Mises statistics for ROC curves

With the adaptation of the family of Cramér-von Mises statistics to ROC curves in equation (3.12), we investigate their behavior.

The Cramér-von Mises statistic for ROC curves

Setting $\psi(\cdot) = 1$ in equation (3.12) leads to the Cramér-von Mises statistic for ROC curves, which we refer to as ROC-CVM. For $\psi(\cdot) = 1$, the functions $\phi_1(s)$ and $\phi_2(s)$ in equation (3.7) simplify to $\phi_1(s) = 1$ and $\phi_2(s) = s^2/2$. Also for $\psi(\cdot) = 1$, the integral in equation (3.12) reduces to a constant: $\int_0^1 (1-r)^2 \psi(r) dr = \int_0^1 (1-r)^2 dr = 1/3$. Inserting these results into equation (3.12), we arrive at the ROC-CVM statistic, denoted by W_{ROC}^2 :

$$W_{\text{ROC}}^2 = 2 \sum_{i=1}^k (b_i - b_{i-1}) \left[\frac{c_i^2}{2} - \frac{b_{i-1} + b_i}{2} c_i \right] + \frac{1}{3}. \tag{3.13}$$

Below we present examples of the ROC-CVM statistic's null distribution from simulations for the simple null hypothesis that the observed data come from a binormal ROC curve with the parameters specified.

We first explain the simulations conducted. Recall that for negative and positive classes with normal distributions $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, the binormal ROC curve has parameters $\rho = \sigma_1/\sigma_2$

and $\delta = (\mu_2 - \mu_1)/\sigma_2$. Fixing $\rho = 1$ and choosing δ from $\{0, 1, 2, 3, 4, 5\}$, we generated random samples each of size 100 from the negative and positive classes. Because of the ROC curve's invariance under monotonically increasing transformations of scores, we could assume $\mu_1 = 0$ and $\sigma_1 = 1$ for the negative class. Then using the definitions of δ and ρ , we determined μ_2 and σ_2 for sampling from the positive class with the equations

$$\sigma_2 = \frac{\sigma_1}{\delta} = \frac{1}{\delta} \text{ and} \tag{3.14}$$

$$\mu_2 = \delta\rho + \mu_1 = \delta\rho.$$

We then calculated the value of the ROC-CVM statistic for the randomly generated empirical ROC curve. Figure (3.2) shows histograms of the ROC-CVM statistic for 1,000 repetitions of the simulations for each value of δ .

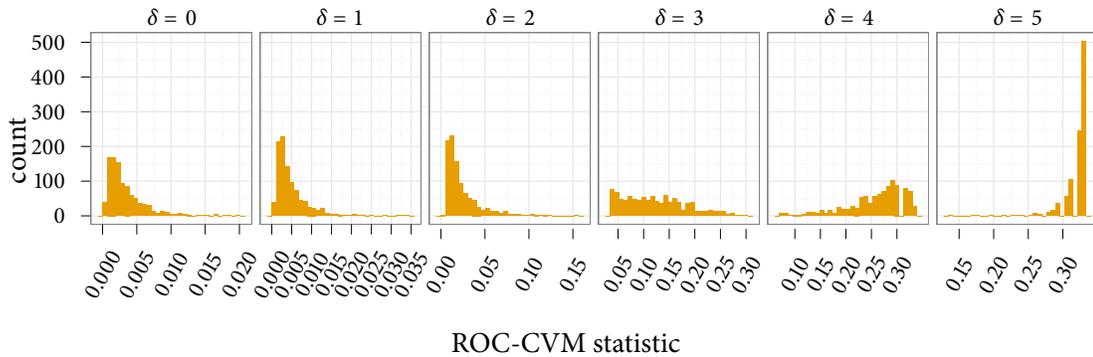


Figure 3.2: Simulated distributions of the ROC-CVM statistic for different values of the binormal parameter δ under a simple null hypothesis. The parameter ρ has a constant value of one. One thousand iterations ran with sample sizes each 100.

Two characteristics stand out in figure (3.2). First, even under the simple null hypothesis, the distribution of the ROC-CVM statistic depends on the parameters of the ROC curve. In contrast, recall from Chapter 2 that the Cramér-von Mises statistic for a single continuous random variable

has a distribution independent of the CDF's parameters under the simple null hypothesis. Second, the ROC-CVM statistic's distribution goes from right-skewed to left-skewed as δ increases from zero to five. Because the ROC-CVM statistic measures a distance between the empirical and parametric ROC curves, large values indicate departure from the null hypothesis. So ideally the distribution's large values would be extreme in the familiar sense of hypothesis testing. However, figure (3.2) demonstrates that for δ at least as small as four, the extreme values occur at smaller sizes of the statistic. This behavior seems undesirable in a goodness-of-fit statistic, and we consider the predominant terms in the sum in equation (3.12) for insight.

Take the extreme case of an empirical ROC with the area under the curve (AUC) equal to one. Then the value of k in equation (3.12) is one. The a and b coordinates are $(0, 0)$, $(0, 1)$, and $(1, 1)$. So the first part, $(b_i - b_{i-1})$, of the sum in equation (3.12) equals 0, and the ROC-CVM statistic equals $1/3$.

Next, consider the case of $k = 2$. From equation (3.12), we see that the ROC-CVM statistic depends entirely on the difference between terms b_1 and b_2 . In general, for small values of k , the ROC-CVM statistic depends heavily on the difference between the parametric and empirical ROC curves at small false positive rates.

Furthermore, the derivative of the ROC curve may exacerbate the dependence of the ROC-CVM statistic on the initial differences between the parametric and empirical ROC for small values of k . Suppose that k is small because the empirical ROC curve rises steeply and reaches the true positive rate of one at a small false positive rate. Then the derivative of the parametric ROC curve will tend to be large in this range. This large derivative will magnify the ROC-CVM statistic's dependence on gaps between the parametric and empirical ROC at small false positive rates.

So small values of k may destabilize the ROC-CVM statistic due to the dependence on a small number of differences between the empirical and parametric ROC and on a large derivative. Attempts to increase k by artificially dividing the increments of the empirical true positive rate into finer steps across the initial false positive rates showed little improvement.

The problem due to small values of k arises particularly for the ROC-CVM statistic, but not the original Cramér-von Mises statistic, because of ties. The original Cramér-von Mises statistic

applies to observations of a single random variable with a continuous CDF. In that case, ties do not occur among the values of the random variable, and the empirical CDF has steps of fixed increments. Viewing the empirical ROC as an empirical CDF, however, we see that effectively ties occur whenever observations from the positive class come in groups amid observations from the negative class. This issue seems unavoidable, and the resulting ties diminish the ROC-CVM's utility, especially for steep ROC curves, which occur with large values of the binormal parameter δ .

The Anderson-Darling statistic for ROC curves

For the Anderson-Darling statistic, the weight function $\psi[\text{ROC}(p)]$ would take the form

$$\psi[\text{ROC}(p)] = \frac{1}{\text{ROC}(p) [1 - \text{ROC}(p)]}. \quad (3.15)$$

As Anderson and Darling (1954) note, equation (3.12) does not lead to a simple solution because the integral $\int_0^1 (1-r)^2 \psi(r) dr$ does not exist for this form of $\psi(\cdot)$. Their method of arriving at the computational form in equation (2.30) for the CDF $F(\cdot)$ of a continuous random variable follows from expressing the statistic as

$$\begin{aligned} A_{n_1}^2 &= n_1 \int_{-\infty}^{\infty} [\widehat{F}_{n_1}(s) - F(s)]^2 \frac{1}{F(s)(1-F(s))} dF(s) \\ &= n_1 \int_{-\infty}^{u_{(1)}} \frac{F^2(s)}{F(s)(1-F(s))} dF(s) + n_1 \int_{u_{(1)}}^{u_{(2)}} \frac{[\widehat{F}_{n_1}(s) - F(s)]^2}{F(s)(1-F(s))} dF(s) \\ &\quad + \dots + n_1 \int_{u_{(n_1)}}^{\infty} \frac{[1-F(s)]^2}{F(s)(1-F(s))} dF(s), \end{aligned} \quad (3.16)$$

where $u_{(1)}, u_{(2)}, \dots, u_{(n_1)}$ are ordered observed values. In the case of the CDF of a continuous random variable with support $(-\infty, \infty)$, the first order statistic is greater than $-\infty$ and the last less than ∞ . The EDF $\widehat{F}_{n_1}(s)$ takes the value zero on $(-\infty, u_{(1)})$ and one on $(u_{(n_1)}, \infty)$, and so the authors could obtain the integrable first and last terms above.

However, this integral does not in general appear to have a closed form for the ROC curve as a CDF. The ROC curve as a CDF has support $[0, 1]$. Yet, the empirical $\widehat{\text{ROC}}(p)$ may well not equal zero at $p = 0$. Let $p_{(1)}, p_{(2)}, \dots$ be the observed false positive rates. Suppose that $p_{(1)} = 0$ and $\widehat{\text{ROC}}(p_{(1)}) \neq 0$. Then the first term of the sum in the statistic would be

$$\int_{p_{(1)}=0}^{p_{(2)}} [\widehat{\text{ROC}}(p) - \text{ROC}(p)]^2 \frac{1}{\text{ROC}(p)(1 - \text{ROC}(p))} dp, \quad (3.17)$$

which contains an undefined definite integral of the form

$$\int_0^c \frac{1}{z(1-z)} dz = [\log(z) - \log(z-1)]_{z=0}^{z=c}. \quad (3.18)$$

As a result, the statistic fails to have any solution when applied directly to ROC curves.

3.2 Empirical ROC-based goodness-of-fit statistics

Directly adapted to ROC curves, the Cramér-von Mises statistic depends strongly on the initial observed rates and the Anderson-Darling statistic has an unsuitable weight. However, the power of these statistics for ordinary CDFs suggests that another integrated distance measure between the empirical and parametric ROC curves might result in a beneficial goodness-of-fit statistic. In considering other possibilities, we sought to meet the objectives discussed below.

3.2.1 Motivation

We first wished to apply the goodness-of-fit statistic to the ROC curve itself (instead of the underlying classes) in order to work with semiparametric models. As reviewed in the first two chapters, the semiparametric approach to estimating ROC curves is invariant under monotonic increasing transformations and depends on only the ranks of underlying scores. Second, we aimed to base the statistic on the empirical ROC curve because of its consistency for the true ROC curve as proven by

Hsieh and Turnbull (1996). Balanced treatment of the observed negatives and positives became the third goal as the ROC curve depends on the relation between the two classes. The vertical distance between the empirical and parametric ROC curves alone does not fully capture this aspect.

With these three objectives, we focused on measuring the distance between the two curves in the direction perpendicular to the parametric one. The resulting line integrals in the next section will make this notion more precise.

3.2.2 Framework

Treatment of positives and negatives

As an empirical CDF, the empirical ROC curve technically is right-continuous. The function for the empirical ROC does not have the values represented by the vertical lines in the plotted empirical ROC. Instead, the plot should be a step function.

While the ROC curve presents error rates in terms of true positives and false positives, another approach conveys error rates in terms of true negatives and false negatives. The plot of true negatives versus false negatives is known as the ordinal dominance curve (ODC). For more details about the ODC, please refer to Bamber (1975) as well as Hsieh and Turnbull (1996). Here we emphasize that the ROC and ODC curves offer similar information. However, the horizontal segments of the empirical ROC curve appear as vertical segments of the empirical ODC curve, and vice versa. Figure 3.3 shows an example of this interchange between the empirical ROC and ODC curves for the same set of data.

By using a perpendicular distance between the parametric and empirical ROC curves, we can include both the horizontal and vertical segments. Then we give no preference to true positives and false positives over true negatives and false negatives.

Perpendicular distance Below we lay out the details of measuring distance between a parametric and an empirical ROC curve along perpendicular lines. To start, we need to specify the lines that run perpendicular to a given parametric ROC curve, which we will denote by $\text{ROC}_\theta(\cdot)$. Let $p^* \in (0, 1)$ denote a certain false positive rate. Then $(p^*, \text{ROC}_\theta(p^*))$ is a point on the parametric ROC curve, and we seek the expression for the line that is perpendicular to the ROC curve at that point.

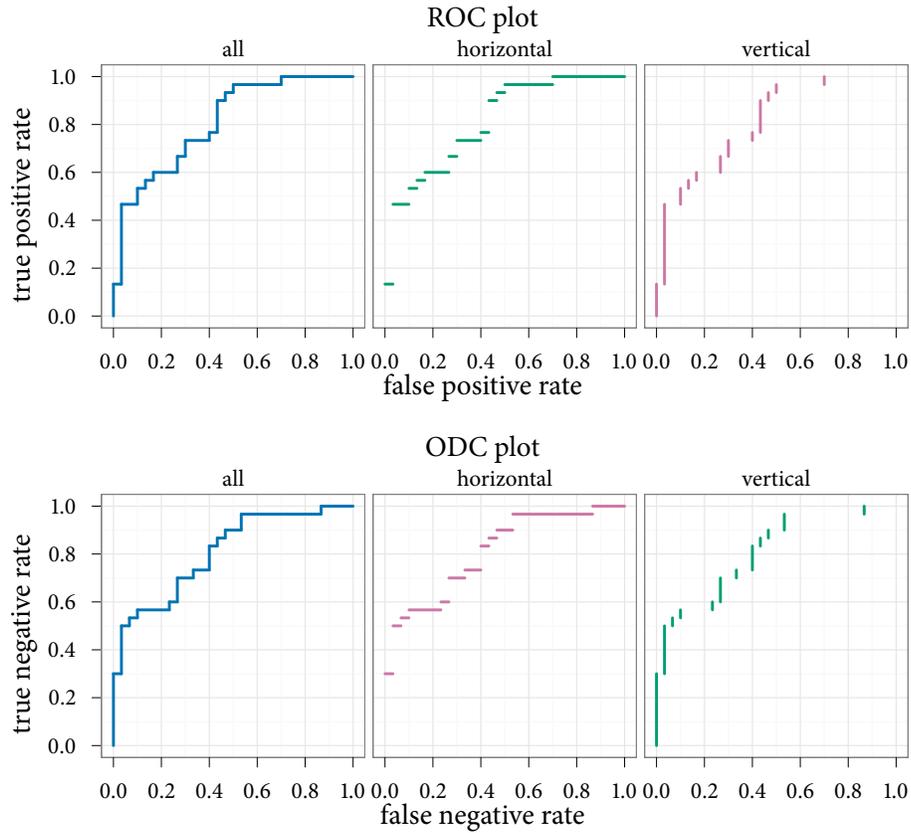


Figure 3.3: Empirical ROC and ODC curves for the same data. The middle panel in the first row shows the ROC curve's horizontal segments, which appear as vertical segments in the right panel of the second row for the ODC.

Writing the derivative $\frac{d\text{ROC}_\theta(p)}{dp}$ of the ROC curve at the false positive rate p^* as $\text{ROC}'_\theta(p^*)$, we introduce the following slope and intersection:

$$\beta^* = -\frac{1}{\text{ROC}'_\theta(p^*)} \tag{3.19}$$

$$\alpha^* = \text{ROC}_\theta(p^*) - \beta^* p^*.$$

The function

$$h(p; p^*, \theta) = \alpha^* + \beta^* p \tag{3.20}$$

then defines the line that runs perpendicular to the ROC curve at the point $(p^*, \text{ROC}_\theta(p^*))$.

For the distance to be well defined, perpendicular lines from the parametric curve must have unique intersections with the empirical curve. In the upcoming proof, we show that lines perpendicular to the binormal $\text{ROC}_\theta(p)$ always have negative slopes. Then we demonstrate that a line between any two points on the empirical ROC has a slope in the interval $[0, \infty]$. Because the slope of the perpendicular line can never equal the slope of a line between two points on the empirical ROC, the lines must not be equal. Therefore, any two points on the empirical ROC curve do not fall on the same perpendicular line from the parametric ROC curve. As a result, a given perpendicular line $h(p; p^*, \theta) = \alpha^* + \beta^* p$ for $p^* \in (0, 1)$ intersects the empirical ROC only once. We intentionally use an open interval because for $p^* = 0$ and $p^* = 1$, the perpendicular line may be horizontal or vertical and then have infinitely many points of intersection with a horizontal or vertical segment of the empirical ROC.

Proposition 3.2.1. *For the binormal ROC and false positive rate $p \in (0, 1)$, perpendicular lines from the parametric $\text{ROC}_\theta(p)$ have unique intersections with the empirical $\widehat{\text{ROC}}(p)$.*

Proof. Let $\text{ROC}_\theta(p)$ denote the binormal ROC curve as a function of the false positive rate:

$$\text{ROC}_\theta(p) = \Phi[\delta + \rho\Phi^{-1}(p)]. \quad (3.21)$$

The slope of the ROC curve is equal to the first derivative:

$$\begin{aligned} \text{ROC}'(p) &= \phi[\delta + \rho\Phi^{-1}(p)] \times \frac{d}{dp} \rho\Phi^{-1}(p) \\ &= \frac{\rho\phi[\delta + \rho\Phi^{-1}(p)]}{\phi[\Phi^{-1}(p)]} \\ &= \rho \frac{e^{-\frac{1}{2}[\delta + \rho\Phi^{-1}(p)]^2}}{e^{-\frac{1}{2}[\Phi^{-1}(p)]^2}}. \end{aligned} \quad (3.22)$$

The exponential function is always greater than zero, and hence the fraction is always greater than

zero. Recall that $\rho = \sigma_1/\sigma_2$ and that by definition a standard deviation is greater than zero. The product of two components that are always greater than zero must itself be greater than zero. So the slope of the binormal ROC curve is always positive for p in the interval $(0, 1)$.

The line perpendicular to the ROC curve at the point $(p^*, \text{ROC}(p^*))$ for $p^* \in (0, 1)$ has a slope β^* equal to $-1/\text{ROC}'(p^*)$. Because the denominator of β^* is always greater than zero and the numerator always less than zero, the slope β^* must be negative.

Now consider a point $(p_1, \widehat{\text{ROC}}(p_1))$ where a perpendicular line with slope β_1 from the binormal ROC intersects the empirical. From the preceding paragraph, the slope β_1 must be negative. Take any other point $(p_2, \widehat{\text{ROC}}(p_2))$ on the empirical ROC. Then there are four possibilities for the line between the points $(p_1, \widehat{\text{ROC}}(p_1))$ and $(p_2, \widehat{\text{ROC}}(p_2))$.

1. $p_1 = p_2$ and $\widehat{\text{ROC}}(p_1) \neq \widehat{\text{ROC}}(p_2)$: In this case, the slope of the line between the points is infinite.
2. $p_1 \neq p_2$ and $\widehat{\text{ROC}}(p_1) = \widehat{\text{ROC}}(p_2)$: In this case, the slope of the line between the points is zero.
3. $p_1 < p_2$ and $\widehat{\text{ROC}}(p_1) \neq \widehat{\text{ROC}}(p_2)$: Because the empirical ROC curve is monotonically increasing, $\widehat{\text{ROC}}(p_1) \leq \widehat{\text{ROC}}(p_2)$. In this case, the slope of the line between the points is positive.
4. $p_1 > p_2$ and $\widehat{\text{ROC}}(p_1) \neq \widehat{\text{ROC}}(p_2)$: Because the empirical ROC curve is monotonically increasing, $\widehat{\text{ROC}}(p_1) \geq \widehat{\text{ROC}}(p_2)$. In this case, the slope of the line between the points must also be positive.

In any case, the slope of the line between two points on the empirical ROC falls in the interval $[0, \infty]$.

The slope of any perpendicular line from the binormal ROC on the interval $p \in (0, 1)$ is negative. The slope of a line connecting any two points on the empirical ROC is non-negative. The slopes can never be equal, and so the two lines must be different. Hence, a perpendicular line from the parametric ROC that intersects a point on the empirical ROC does not intersect any other point on the empirical ROC.

Therefore, a perpendicular line from a binormal ROC at a false positive rate p on the interval $(0, 1)$ has a unique intersection with the empirical ROC. \square

Having defined the perpendicular lines from the parametric ROC curve to the empirical ROC, we next address distance. Let the distance along the perpendicular line $h(p; p^*, \theta)$ from $\text{ROC}_\theta(p^*)$ to the point of intersection on the empirical ROC curve be:

$$D(p^*) = \text{distance along } h(p; p^*, \theta) \text{ between } \text{ROC}_\theta(p^*) \text{ and } \widehat{\text{ROC}}(\cdot). \quad (3.23)$$

Specifying the distance's starting point depends on a given value of p^* , but finding its end point requires a search for the intersection along the empirical ROC. We share our algorithm for this simple search later in the chapter. Figure 3.4 depicts an example of the distance $D(p^*)$.

Line integrals The statistics introduced here follow from line integrals of $D(p^*)$ along the smooth parametric ROC curve. We also allow for a weighting function $\psi(p^*)$. For the line integrals of the distance and the distance squared, we denote the statistics by \tilde{P} and \tilde{P}^2 , respectively. A subscript on these terms will indicate the version of the weighting function used. Letting $d\ell$ denote the differential along the arc of the ROC curve, we write the first two statistics as

$$\begin{aligned} \tilde{P}_i^2 &= \int_{\text{ROC}} D^2(p^*) \psi_i(p^*) d\ell \\ \tilde{P}_i &= \int_{\text{ROC}} D(p^*) \psi_i(p^*) d\ell. \end{aligned} \quad (3.24)$$

We define the first weighting function $\psi_1(\cdot) \equiv 1$. Other possible weighting functions appear toward the chapter's end.

To relate these expressions to perhaps more familiar line integrals, we return to the notion of the ROC function as a curve with coordinates $x(t)$ and $y(t)$ parameterized by the threshold variable t .

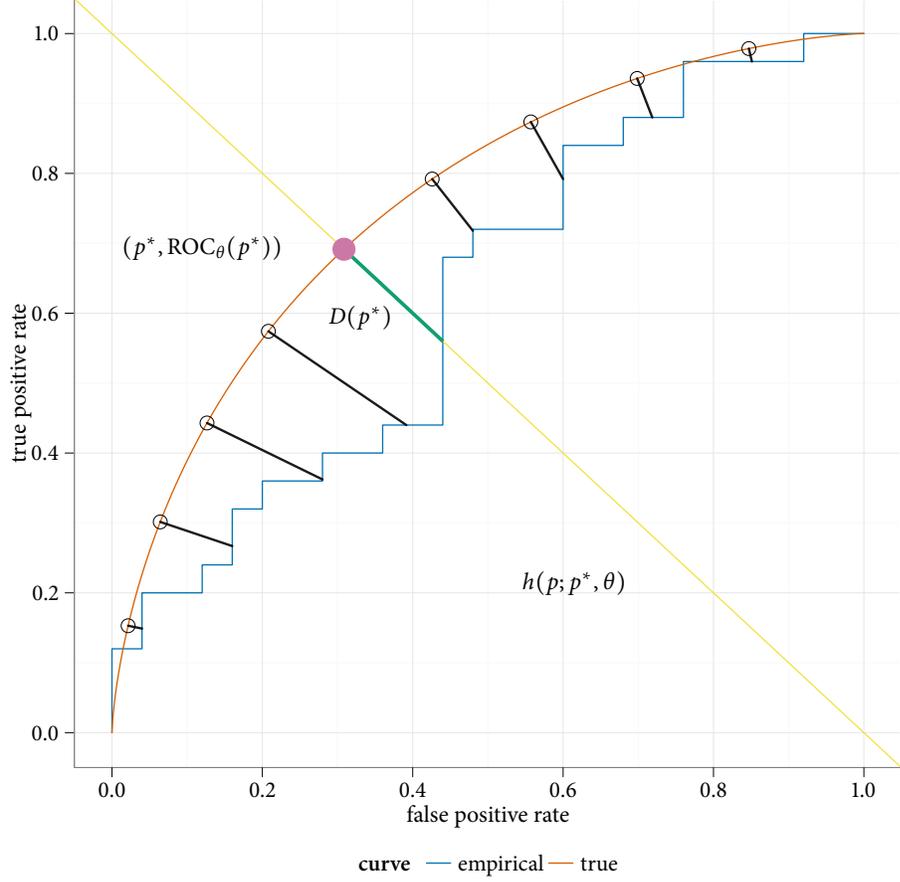


Figure 3.4: Example of distance $D(p^*)$ between binormal parametric ROC curve and empirical ROC curve. The perpendicular line $h(p; p^*, \theta)$ intersects the point $(p^*, \text{ROC}_\theta(p^*))$ on the parametric ROC curve ROC_θ .

Then the differential $d\ell$ follows from

$$\begin{aligned}
 d\ell &= \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt \\
 &= \sqrt{\left(\frac{d[1-F(t)]}{dt}\right)^2 + \left(\frac{d[1-G(t)]}{dt}\right)^2} dt \\
 &= \sqrt{\left(\frac{df(t)}{dt}\right)^2 + \left(\frac{dg(t)}{dt}\right)^2} dt
 \end{aligned} \tag{3.25}$$

where we used the expressions for the x and y coordinates in terms of distribution functions as in Chapter 1. Furthermore, we may view the functions of p^* in equation (3.24) as functions of x and y with the connections $x = p^*$ and $y = \text{ROC}(p^*)$. Thus, the line integrals in (3.24) relate to the more usual terms of x , y , and t .

With the concept of the \tilde{P} statistics established, observe that the integrals in equation (3.24) still do not have simple analytic solutions. However, a summation may approximate the line integral. Consider dividing the ROC curve into segments of equal lengths by spreading d points equidistantly apart along the arc. If we let the horizontal coordinate of the i^{th} point be p_i , then the first segment runs from $(0, 0)$ to $(p_1, \text{ROC}(p_1))$, the second from $(p_1, \text{ROC}(p_1))$ to $(p_2, \text{ROC}(p_2))$, and so on. The last segment runs from $(p_d, \text{ROC}(p_d))$ to $(1, 1)$. Let $\Delta\ell$ denote the small, fixed distance between consecutive points along the arc of the ROC curve. Expressing the total length of the ROC curve as ℓ , we have $\Delta\ell = \ell/(d + 1)$. The following summations give approximate values for the statistics.

$$\begin{aligned}\tilde{P}_1^2 &\approx P_1^2 = \sum_{i=1}^d D^2(p_i)\psi_1(p_i)\Delta\ell \\ \tilde{P}_1 &\approx P_1 = \sum_{i=1}^d D(p_i)\psi_1(p_i)\Delta\ell\end{aligned}\tag{3.26}$$

We refer to statistics of the form P_i^2 and P_i as the family of ROC-PD (ROC-perpendicular distance) statistics. The software implementation to compute such summations comes next.

3.2.3 Implementation

Computation of ROC-PD-based statistics

To start the computation of the sums in equation (3.26), we first calculate the total length ℓ of the arc of the parametric ROC curve. We evaluate the integral

$$\ell = \int_{\text{ROC}} d\ell = \int_0^1 \sqrt{1 + [\text{ROC}'(p)]^2} dp\tag{3.27}$$

numerically to determine the length. Then for a given number d of points at which to split the arc into segments, the distance by which to separate the starting points for the perpendicular lines is $\Delta\ell = \ell/(d + 1)$.

After obtaining $\Delta\ell$, we still need a technique to measure out distances along the arc to determine accurately the points $(p_i, \text{ROC}(p_i))$. Letting ℓ_i denote the distance along the arc from $(0, 0)$ to $(p_i, \text{ROC}(p_i))$, we have $\ell_i = i\Delta\ell$. With the value of ℓ_i known, we seek the corresponding coordinates $(p_i, \text{ROC}(p_i))$. We make use of the following ordinary differential equation (ODE):

$$\frac{d\ell}{dp} = \sqrt{1 + [\text{ROC}'(p)]^2} \quad (3.28)$$

which comes from differentiating equation (3.27). This result allows for the calculation of distances along the horizontal axis that correspond to distances of length $\Delta\ell$ along the arc. For example, given p_i and ℓ_i , the change in the horizontal coordinate to move a distance $\Delta\ell$ along the arc follows from finding p_{i+1} such that

$$\int_{p_i}^{p_{i+1}} \sqrt{1 + [\text{ROC}'(p)]^2} dp = \int_{\ell_i}^{\ell_i + \Delta\ell} d\ell. \quad (3.29)$$

Then $\Delta p_{i+1} = p_{i+1} - p_i$ is the change in the horizontal coordinate to move a distance $\Delta\ell$ along the arc. Given an initial condition of p_i and values for ℓ_i and $\Delta\ell$, numerical ODE solvers can compute the end point of p_{i+1} .

Where the ROC curve rises steeply, the derivative $\text{ROC}'(p)$ may change rapidly with p and become numerically unstable. To overcome this issue, we switch to integrating over the binormal ROC curve's inverse, expressed as a function of the true positive rate, when the slope rises above one. Then the steep derivative becomes a shallow one, and the numerical solution becomes more stable. As the upcoming proposition proves, the slope of the binormal ROC curve can change from less than to greater than one or vice versa at most twice for false positive rate $p \in (0, 1)$. This limited number of changes in the slope restricts the number of regions over which to select either equation (3.29) or its counterpart in terms of the binormal ROC curve's inverse to no more than three.

Proposition 3.2.2. *The slope of the binormal ROC curve equals the value one at most twice for values of $p \in (0, 1)$.*

Proof. Let $\text{ROC}(p)$ denote the ROC curve as a function of the false positive rate:

$$\text{ROC}(p) = \Phi[\delta + \rho\Phi^{-1}(p)]. \quad (3.30)$$

The slope of the ROC curve equals the first derivative, which from equation (3.22) is:

$$\text{ROC}'(p) = \frac{\rho e^{-\frac{1}{2}[\delta + \rho\Phi^{-1}(p)]^2}}{e^{-\frac{1}{2}[\Phi^{-1}(p)]^2}}. \quad (3.31)$$

Solving the following equation gives the values of the false positive rate p at which the slope equals one:

$$1 = \frac{\rho e^{-\frac{1}{2}[\delta + \rho\Phi^{-1}(p)]^2}}{e^{-\frac{1}{2}[\Phi^{-1}(p)]^2}}. \quad (3.32)$$

Multiplying both sides by $e^{-\frac{1}{2}[\Phi^{-1}(p)]^2}$ leads to

$$e^{-\frac{1}{2}[\Phi^{-1}(p)]^2} = \rho e^{-\frac{1}{2}[\delta + \rho\Phi^{-1}(p)]^2}, \quad (3.33)$$

for which we consider the cases of $\rho = 1$ and $\rho \neq 1$ separately.

If $\rho = 1$, we can take the log of both sides and proceed to the solution:

$$\begin{aligned} [\Phi^{-1}(p)]^2 &= \delta^2 + 2\delta\Phi^{-1}(p) + [\Phi^{-1}(p)]^2 \\ \Phi^{-1}(p) &= -\frac{\delta^2}{2\delta} = -\frac{\delta}{2} \end{aligned} \quad (3.34)$$

$$p = \Phi\left(\frac{\delta}{2}\right) \quad \text{solution if } \rho = 1.$$

If $\rho \neq 1$, we have an additional term after taking the log:

$$\begin{aligned} [\Phi^{-1}(p)]^2 &= \delta^2 + 2\rho\delta\Phi^{-1}(p) + \rho^2 [\Phi^{-1}(p)]^2 - 2\log(\rho) \\ 0 &= (\rho^2 - 1) [\Phi^{-1}(p)]^2 + 2\rho\delta\Phi^{-1}(p) + \delta^2 - 2\log(\rho). \end{aligned} \quad (3.35)$$

The solutions for $\Phi^{-1}(p)$ are:

$$\Phi^{-1}(p) = \frac{-\delta\rho \pm \sqrt{\delta^2 - 2\log(\rho) + 2\rho^2 \log(\rho)}}{\rho^2 - 1}, \quad (3.36)$$

which give the following two values for p

$$p = \Phi \left[\frac{-\delta\rho \pm \sqrt{\delta^2 - 2\log(\rho) + 2\rho^2 \log(\rho)}}{\rho^2 - 1} \right]. \quad (3.37)$$

Thus, if $\rho = 1$, the slope of the binormal ROC curve equals one at a single value of p , and if $\rho \neq 1$, it does so at two values of p . The slope of the binormal ROC therefore equals one at most twice. \square

Number of points The number, d , of points along the parametric ROC curve at which to evaluate the perpendicular distance does not have a strict definition, but larger values should result in more exact calculations of the ROC-PD statistic. We studied the convergence of and relative error in the values of the ROC-PD statistic for different values of d via simulation. Starting with a binormal ROC curve with parameters $\rho = 1$ and $\delta = 1$, we generate 10,000 empirical ROC curves from samples each of size 30 from the two classes. For each value of d , we compute the ROC-PD statistics under a simple null hypothesis and use $P_{1,d}^2$ to denote the values.

First, we investigate the convergence of $P_{1,d}^2$ values by taking their ratio to $P_{1,10^5}^2$ values. (We implicitly treat the results for $d = 10^5$ as close approximations to the exact values.) Figure 3.5 shows the ratio for the first 1,000 simulations versus the number of points used to calculate the ROC-PD statistic. Although an error of roughly 1 % may occur for $d = 100$, once we use a value of d around

1,000, the error becomes small. The convergence of the ratio as depicted in the plot suggests that the ROC-PD statistic converges as $d \rightarrow \infty$.

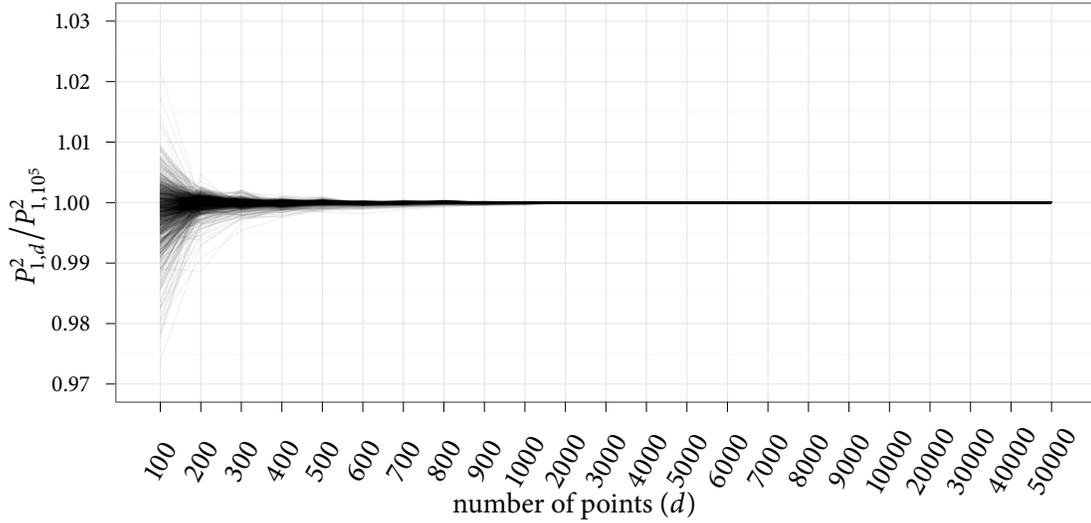


Figure 3.5: Ratio of ROC-PD statistic for different values of the number (d) of points to the ROC-PD statistic for $d = 10^5$. The ratios for the first 1,000 simulations appear. The ratios approach one, which suggests the value of the P_1^2 statistic converges as d becomes large.

Second, we consider the relative error in the ROC-PD statistic for different values of d . In particular, we study the error in the ROC-PD statistic for a given value d relative to the ROC-PD statistic for $d = 10^5$, which follows from:

$$\text{relative error in } P_{1,d}^2 = \frac{|P_{1,d}^2 - P_{1,10^5}^2|}{P_{1,10^5}^2}. \quad (3.38)$$

Figure 3.6 shows box plots of this relative error for a range of values of d . Based on these results, we choose to set $d = 1,000$ for the extensive simulations of subsequent chapters. With $d = 1,000$, the median relative error is well below 0.01% and the maximum relative error remains under 0.1%.

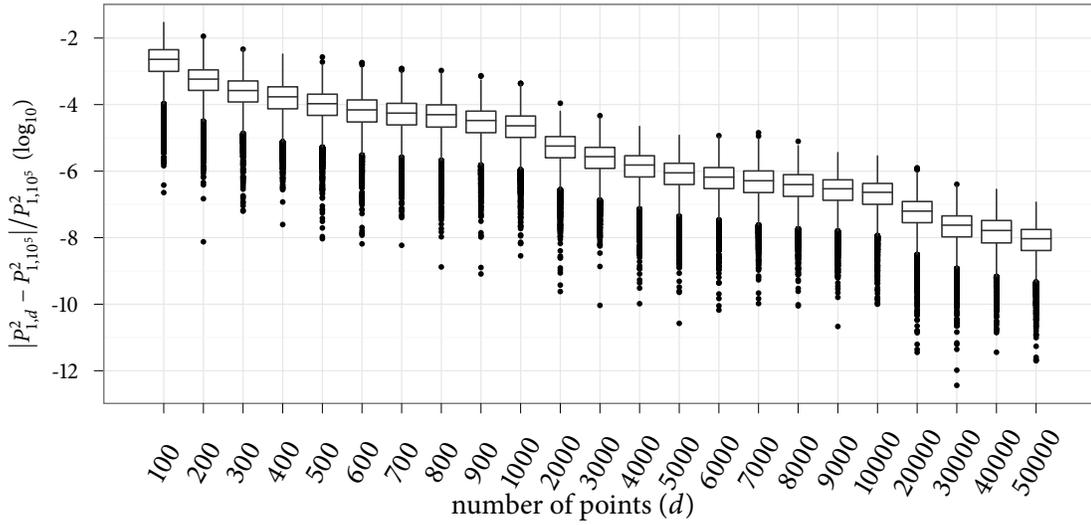


Figure 3.6: Box plot of all 10,000 errors in P_1^2 statistic for different choices of d relative to $d = 10^5$ value. For $d = 1000$, the relative error is below 0.1%.

Search for intersection with empirical ROC As previously noted, finding where the perpendicular line from a given point on the binormal parametric ROC curve intersects the empirical ROC curve requires a computational search. The empirical ROC curve consists of vertical and horizontal segments, which we may view as intervals on vertical and horizontal lines. If the perpendicular line intersects a vertical or horizontal line within such an interval, then the point of intersection is on the empirical ROC curve.

The maximum number of vertical and horizontal lines to check equals the total number of segments that constitute the empirical ROC curve. Our search, however, checks the vertical and horizontal lines by starting with the one along which the previous intersection occurred and then if necessary moving to the lines that contain the two segments next to the first. In this way, the search proceeds outward from the previous intersection along the empirical ROC curve. Because the binormal parametric ROC curve tends not to be rough, neighboring perpendicular lines tend to intersect the empirical ROC along the same or neighboring segments. So this approach reduces the number of segments through which we must search for intersections.

3.2.4 Preview of the ROC-PD statistic's distribution

Previously in this chapter, we conveyed our concern about the distribution of the ROC-CVM statistic. Figure 3.2 shows that the distribution under a simple null changes from right-skewed to left-skewed as the binormal parameter δ increases. By comparison, figure 3.7 gives the distribution of the ROC-PD statistic under a simple null with the same sample data. The ROC-PD statistic remains right-skewed

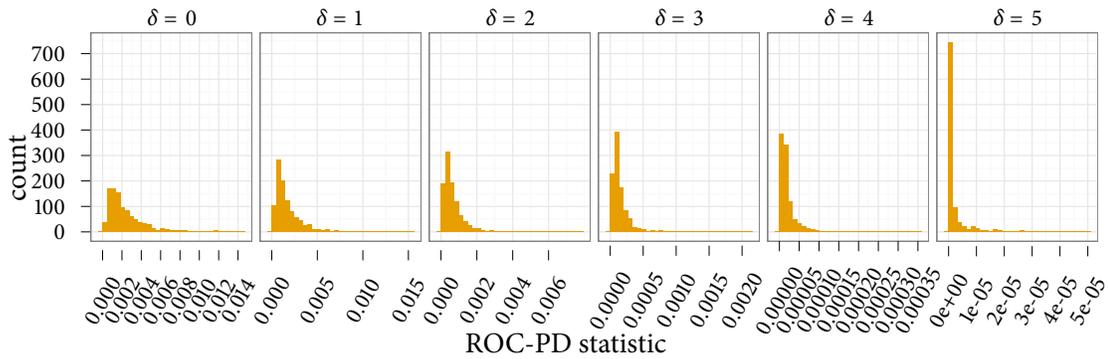


Figure 3.7: Simulated distributions of the ROC-PD statistic for different values of the binormal parameter δ under a simple null hypothesis. The parameter ρ has a constant value of one. One thousand iterations ran with sample sizes each 100.

as δ grows. This preview offers an indication that the ROC-PD statistic behaves as desired, but much more evidence follows in the next chapter.

3.2.5 Extensions

Weights

Earlier, we mentioned the use of weight functions other than $\psi_1(\cdot) = 1$ in the ROC-PD statistic. Two choices which both would de-emphasize regions of the binormal ROC curve with more variability appear in this section. The weights follow from the same motivation as the weight of the Anderson-Darling statistic for a single continuous random variable. Specifically, the weight functions seek to

lessen the contribution from distances where the binormal ROC varies most by dividing by the function's variability.

Pepe (2004, p. 100) approximates the distribution of the empirical ROC at a false positive rate p by a normal distribution with mean $\text{ROC}(p)$ and variance

$$\text{var}[\widehat{\text{ROC}}(p)] = \frac{\text{ROC}(p)[1 - \text{ROC}(p)]}{n_2} + \left(\frac{g(t)}{f(t)}\right)^2 \frac{p(1-p)}{n_1} \quad (3.39)$$

where $t = F^{-1}(1 - p)$. Equation (3.39) approximates the variance in the vertical direction. Given our focus on treating the horizontal direction fairly, we also include its variance. Letting the letter o indicate a true positive rate, the following approximates the variance in the horizontal direction

$$\text{var}[\widehat{\text{ROC}}^{-1}(o)] = \frac{\text{ROC}^{-1}(o)[1 - \text{ROC}^{-1}(o)]}{n_1} + \left(\frac{f(t)}{g(t)}\right)^2 \frac{o(1-o)}{n_2}, \quad (3.40)$$

which Pepe (2004, p. 101) derives. Using equations (3.39) and (3.40), we present the second weight function

$$\psi_2(p) = \frac{1}{\sqrt{\{\text{var}[\widehat{\text{ROC}}(p)]\}^2 + \{\text{var}[\widehat{\text{ROC}}^{-1}(o)]\}^2}}. \quad (3.41)$$

The third weight function also decreases the contribution along the middle of the ROC curve's arc length, but with a simpler form.

$$\psi_3(p_i) = \frac{1}{\ell_i(\ell - \ell_i)}. \quad (3.42)$$

Note that ℓ_i is the distance along the arc from the curve's origin and $(\ell - \ell_i)$ that from the end.

These additional weight functions likely provide more power to the ROC-PD goodness-of-fit statistic. However, their dependence on d , the number of points chosen along the ROC curve's arc, requires more study. So they remain the subject of future research.

Chapter 4: Characterization of goodness-of-fit tests

This chapter presents results from extensive computer simulations that characterize several goodness-of-fit statistics applied to binormal models of ROC curves. A description of the simulations follows in the first section. The second section contains specific results and discussion. Many more results appear in the figures of Appendix A.

4.1 Outline of simulations

As discussed in Chapter 3, the exact distributions of our new goodness-of-fit statistics appear analytically intractable. Therefore, we conducted various simulations to study the behavior of the statistics. Below we describe generating sample data, formulating hypotheses, approximating distributions of goodness-of-fit statistics, and finding p-values. We cover the statistics under both simple and composite null hypotheses as well as alternative hypotheses.

The following list may serve as a guide to the general steps of a single simulation. In referring to a single simulation, we mean a set of samples all from the same distributions with the same parameters and sizes, the relevant goodness-of-fit statistics, and their p-values. The subsequent sections will provide more details for each step.

1. Create simulation data set of samples.
 - a. Decide on distributions and parameters for negative and positive classes.
 - b. Choose sample sizes n_1 and n_2 .
 - c. Draw $M = 1000$ random samples for each class.
2. Formulate hypotheses for sample data.
 - a. For a simple null hypothesis, define the distributions and parameters.

- b. For a composite null hypothesis, estimate the parameters using fully parametric maximum likelihood estimates (FPMLEs) and ROCKIT's semi-parametric approach.
3. For each sample in the simulation data under each possible hypothesis, calculate the four following goodness-of-fit statistics:
 - a. ROC-PD,
 - b. ROC-CVM,
 - c. ROC- χ^2 , and
 - d. AUC-GOF.
4. For each goodness-of-fit statistic computed for each sample in the simulation data under each null hypothesis, conduct parametric bootstraps to approximate p-values.
 - a. Starting with the model specified by the null hypothesis, generate parametric bootstrap samples.
 - b. Measure the goodness-of-fit statistic for each bootstrap sample.
 - c. Calculate a p-value by using the original sample's goodness-of-fit statistic and the distribution of the bootstrapped samples' goodness-of-fit statistics.

In total, a single simulation has either eight or twelve (depending on whether or not a simple null hypothesis is appropriate) goodness-of-fit statistics and p-values for each sample. Figure 4.1 depicts the steps of our simulation for calculating p-values for one goodness-of-fit statistic under one type of hypothesis. We next discuss these steps in more detail.

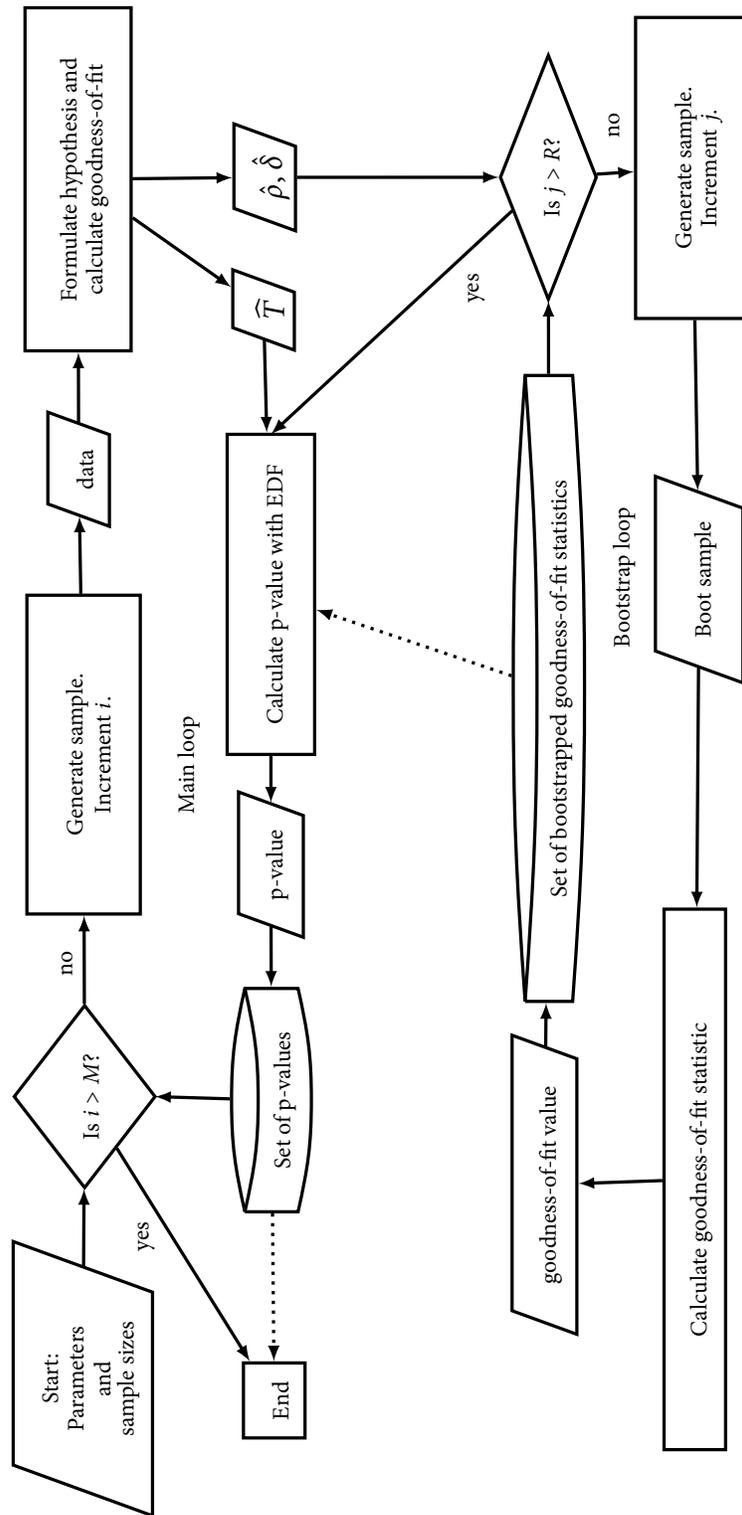


Figure 4.1: Design of simulations for goodness-of-fit statistics.

4.2 Detailed steps of simulations

4.2.1 Generation of simulation data (step one)

Each simulation began with the generation of sample data. We specified the distributions and their parameters for the negative and positive classes. Then we drew $M = 1000$ random samples, each of which included n_1 observations randomly generated from the negative class and n_2 observations randomly generated from the positive class. The resulting samples were stored in a file so that later steps used the same simulation data.

All of our null hypotheses consist of a binormal ROC model with parameters ρ and δ either given or estimated. In generating data for a specific null hypothesis, we used the assumption of the binormal model that under some monotonic increasing transformation, the underlying classes have normal distributions. (Let $\mathcal{N}(\mu, \sigma^2)$ denote a normal distribution with mean μ and variance σ^2 .) We could then set the distribution of the negative class to be $\mathcal{N}(0, 1)$ and that of the positive class to be $\mathcal{N}(\delta/\rho, 1/\rho^2)$ to obtain a binormal ROC with parameters ρ and δ .

For alternative distributions, we considered the following four cases:

1. An exponential distribution with mean λ_1 for the negative class and an exponential distribution with mean λ_2 for the positive class. We refer to the resulting ROC curve as a *biexponential ROC*.
2. A $\mathcal{N}(0, 1)$ distribution for the negative class and a left-running gamma distribution for the positive class. By “left-running gamma distribution,” we mean a gamma distribution reversed on the horizontal axis so that the support is $(-\infty, 0)$. The gamma distribution requires shape and rate parameters α and β . We also moved the upper bound of the support away from zero by permitting a position parameter that was an added constant. We refer to the resulting ROC curve as a *normal-reflected gamma ROC*.
3. A $\mathcal{N}(0, 1)$ distribution for the negative class and a mixture of normals for the positive class. For the mixture, we allowed for two normal distributions, $\mathcal{N}(\mu_2, \sigma_2^2)$ and $\mathcal{N}(\mu_3, \sigma_3^2)$, and a mixture coefficient. We refer to the resulting ROC curve as a *normal-mixture ROC*.

4. A $\mathcal{N}(0, 1)$ distribution for the negative class and a Cauchy distribution with location and scale parameters for the positive class. We refer to the resulting ROC curve as a *normal-Cauchy ROC*.

Tables 4.1, 4.2, 4.3, 4.4, and 4.5 later in this chapter list all of the parameters for the simulations run as well as the page numbers for the results.

4.2.2 Formulation of hypotheses (step two)

We examined the goodness-of-fit statistics under both simple and composite null hypotheses of binormal ROC curves. Under a simple null hypothesis, we used fixed values for the parameters ρ and δ . Under a composite null hypothesis, we first estimated the parameters.

Several methods of estimation of ρ and δ exist for the binormal model. A study of some of the methods forms part of the next chapter. Here we considered two estimation methods that have widespread use.

The first estimation method we call fully parametric maximum likelihood estimation (FPMLE). This approach estimates the parameters μ_1 and σ_1 of the negative distribution and the parameters μ_2 and σ_2 of the positive distribution using familiar maximum likelihood methods. Let u_1, \dots, u_{n_1} be sample observations from the negative class and \bar{u} the sample mean. Then the MLEs are

$$\hat{\mu}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} u_i = \bar{u} \tag{4.1}$$

$$\hat{\sigma}_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (u_i - \bar{u})^2$$

with similar expressions for $\hat{\mu}_2$ and $\hat{\sigma}_2^2$ using v_1, \dots, v_{n_2} from the positive class. We then construct the FPMLEs of $\hat{\rho}$ and $\hat{\delta}$ from:

$$\hat{\rho} = \frac{\hat{\sigma}_1}{\hat{\sigma}_2} \tag{4.2}$$

$$\hat{\delta} = \frac{\hat{\mu}_2 - \hat{\mu}_1}{\hat{\sigma}_2}$$

The second estimation method, which is our preference, is to find semi-parametric estimates of ρ and δ for a composite null hypothesis. For these goodness-of-fit simulations, we used the software library `libroc` from ROCKIT to calculate the semi-parametric estimates (Metz et al., 1998).

4.2.3 Calculation of goodness-of-fit (step three)

Our primary goal in this chapter is to study the new ROC-PD statistic. We used the unweighted, squared version of the ROC-PD statistic as presented in Chapter 3 with $d = 1000$ points along the parametric ROC curve. Besides the ROC-PD statistic, we wished to demonstrate the behavior of other goodness-of-fit techniques for ROC curves. We included the adaptation of the Cramér-von Mises statistic to ROC curves, ROC-CVM, as formulated in the beginning of Chapter 3. However, *caveat lector*: the ROC-CVM test can misbehave and is not recommended. We also implemented a simple chi-squared goodness-of-fit test for continuous ROC curves, ROC- χ^2 , detailed in Appendix C. Finally, we added the goodness-of-fit statistic, AUC-GOF, based on the AUC as discussed in the literature review. For further information, please see Zou et al. (2003) as well as Zou et al. (2005).

4.2.4 Calculation of p-values by parametric bootstrap (step four)

Computation of approximate p-values for the goodness-of-fit statistics with a parametric bootstrap proceeded as follows. Assume after step three that for one of the pairs of samples and one of the hypotheses, we have estimates $\hat{\rho}$ and $\hat{\delta}$ for the binormal parameters and a value P_0 of the goodness-of-fit statistic. Then we generate $R = 1000$ bootstrap samples. For each sample, we draw n_1 negative observations from the $\mathcal{N}(0, 1)$ distribution and n_2 positive observations from $\mathcal{N}(\hat{\delta}/\hat{\rho}, 1/\hat{\rho}^2)$ to obtain an empirical ROC from the theoretical ROC with parameters $\hat{\rho}$ and $\hat{\delta}$. We apply the goodness-of-fit test to each of the bootstrap samples and collect all of the bootstrap goodness-of-fit values, P_i^* with $i = 1, \dots, R$, into an EDF. The percentile of the original measure P_0 of the goodness-of-fit statistic in the EDF of bootstrap measures leads to a p-value. The ROC-PD, ROC-CVM, and ROC- χ^2 statistics all used right-tailed p-values while the AUC-GOF statistic used two-tailed p-values. For a given sample, the hypothesis does not depend on the goodness-of-fit statistic, and so each goodness-of-fit

test worked on the same randomly generated bootstrap samples.

4.3 List of simulations

We conducted all simulations with $M = 1000$ samples from each class, $R = 1000$ bootstrap samples for approximating p-values, and $d = 1000$ points along the parametric ROC curve for the ROC-PD statistic.

4.3.1 Null distributions

For given ρ and δ , we drew observations for the negative class from the $\mathcal{N}(0, 1)$ distribution and observations for the positive class from the $\mathcal{N}(\delta/\rho, 1/\rho^2)$ distribution. Table 4.1 lists the parameters and references the page numbers in the appendix for all simulations of null distributions.

Table 4.1: Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a null hypothesis.

n_1	n_2	$\rho = 1$	$\rho = 1$	$\rho = 1$	$\rho = 1$	$\rho = 1$	$\rho = 0.5$	$\rho = 0.5$	$\rho = 0.75$
		$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$	$\delta = 0.5$	$\delta = 1.0$	$\delta = 0.5$
30	30	128	134	140	146	152	158	164	170
50	50	129	135	141	147	153	159	165	171
100	100	130	136	142	148	154	160	166	172
1000	1000	131	137	143	149	155	161	167	173
100	30	132	138	144	150	156	162	168	174
100	50	133	139	145	151	157	163	169	175

4.3.2 Alternative distributions

Biexponential

Observations for the negative class came from an exponential distribution with mean λ_1 , and observations for the positive class came from an exponential distribution with mean λ_2 . As a reference, the density of an exponential distribution with mean λ is:

$$g(x; \lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, \quad 0 \leq x < \infty, \quad \lambda > 0. \quad (4.3)$$

The parameters used in the simulations appear in table 4.2 along with references to the page numbers in the appendix of complete results.

Table 4.2: Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a biexponential alternative.

n_1	n_2	$\lambda_1 = 1, \lambda_2 = 2$	$\lambda_1 = 1, \lambda_2 = 15$
30	30	177	183
50	50	178	184
100	100	179	185
1000	1000	180	186
100	30	181	187
100	50	182	188

Normal-reversed gamma

We drew observations for the negative class from the $\mathcal{N}(0, 1)$ distribution. For the positive class, we drew from a left-skewed gamma distribution with a variable upper bound. Denoting the shape parameter as α , the rate parameter as β , and the upper-bound parameter as θ , we have the following

expression for the positive class's density:

$$g(x; \alpha, \beta, \theta) = \beta^\alpha \frac{1}{\Gamma(\alpha)} (\theta - x)^{\alpha-1} e^{-\beta(\theta-x)}, \quad -\infty \leq x < \theta, \quad \alpha, \beta > 0. \quad (4.4)$$

Table 4.3 gives all of the simulation parameters and references to the page numbers in the appendix of complete results.

Table 4.3: Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a normal-gamma alternative.

n_1	n_2	$\alpha = 2, \beta = 0.5, \theta = 5$	$\alpha = 3, \beta = 2, \theta = 3$
30	30	190	196
50	50	191	197
100	100	192	198
1000	1000	193	199
100	30	194	200
100	50	195	201

Normal-mixture

We started with observations for the negative class drawn from the $\mathcal{N}(0, 1)$ distribution. For the positive class, we created a mixture of two normals with mean μ_2 and variance σ_2^2 for the first, mean μ_3 and variance σ_3^2 for the second, and mixture coefficient α . With $\phi(\cdot)$ denoting the density of a standard normal, we may express the density used for the positive class as:

$$g(x; \mu_2, \sigma_2, \mu_3, \sigma_3, \alpha) = (1 - \alpha) \frac{1}{\sigma_2} \phi\left(\frac{x - \mu_2}{\sigma_2}\right) + \alpha \frac{1}{\sigma_3} \phi\left(\frac{x - \mu_3}{\sigma_3}\right). \quad (4.5)$$

The list of parameters for these simulations and references to the full results in the appendix are in table 4.4.

Table 4.4: Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a normal-mixture alternative.

n_1	n_2	$\mu_2 = 0, \sigma_2 = 0.5$	$\mu_2 = 10, \sigma_2 = 1$
		$\mu_3 = 2, \sigma_3 = 0.25, \alpha = 0.5$	$\mu_3 = 1.1, \sigma_3 = 0.1, \alpha = 0.5$
30	30	203	209
50	50	204	210
100	100	205	211
1000	1000	206	212
100	30	207	213
100	50	208	214

Normal-Cauchy

For the last set of simulations, we used a $\mathcal{N}(0, 1)$ distribution for the negative class, but a Cauchy distribution for the positive class. Writing the location parameter as θ and the scale parameter as σ , we have the following density for the positive class:

$$g(x; \theta, \sigma) = \frac{1}{\pi\sigma \left[1 + \left(\frac{x-\theta}{\sigma}\right)^2\right]} \quad (4.6)$$

Table 4.5 gives the simulation parameters and references the pages in the appendix with all results.

Table 4.5: Parameters and references to pages in appendix for simulations of goodness-of-fit tests under a normal-Cauchy alternative.

n_1	n_2	$\theta = 1, \sigma = 0.5$	$\theta = 5, \sigma = 1$
50	50	216	220
100	100	217	221
100	50	219	222

4.4 Results

The simulations constitute an empirical study of goodness-of-fit tests used with binormal ROC curves. Our primary interest lies in the new ROC-PD statistic, whose distribution does not have a closed form as previously mentioned. Below we explain our presentation of results, of which a few instances appear in this chapter. Appendix A contains our complete results. We also observe the behavior of the goodness-of-fit statistics and highlight possible alternatives against which the tests provide some power.

4.4.1 Presentation of results

Let us introduce the format of our results for the goodness-of-fit simulations. For each simulation, we created one figure with three sets of plots:

1. true, hypothesized, and empirical ROC curves;
2. distributions of the goodness-of-fit statistics under each hypothesis; and
3. p-values under each hypothesis.

Figures 4.2 and 4.3 show examples of each set of plots for the goodness-of-fit statistics under a null and an alternative hypothesis.

The first set of plots in figures 4.2 and 4.3 depicts all 1,000 instances of the true, empirical, and hypothesized ROC curves. The goodness-of-fit tests assessed the agreement between each of the

curves under the different hypotheses and the empirical one. Each test used the same empirical and estimated ROC curves. These plots offer a sense of the shape as well as the variability of the ROC curves in a simulation.

The second set of plots in figures 4.2 and 4.3 shows the distributions of the values of the goodness-of-fit statistics under each hypothesis. Each column corresponds to one goodness-of-fit test, and each row corresponds to one hypothesis. A goodness-of-fit test's general form under the different hypotheses should be evident in these plots.

The third set of plots in figures 4.2 and 4.3 draws the empirical distribution function (EDF) of the bootstrapped p-values for the goodness-of-fit statistics under each hypothesis. These plots reveal essential aspects of the goodness-of-fit tests. The p-values of a goodness-of-fit test under a null hypothesis should be uniformly distributed, and so the EDF should closely follow the forty-five degree line. Under alternatives, a goodness-of-fit test with some power of rejection will have more p-values distributed at small values, and so the EDF will rise more vertically in the beginning. Beside the EDF of the bootstrapped p-values, we list the p-values from three statistical tests to check uniformity.

The three statistical tests are

- B5** An exact binomial test that 5% of all bootstrapped p-values are less than or equal to 0.05. The one-sided alternative hypothesis is that more than 5% of all bootstrapped p-values are less than or equal to 0.05.
- B1** An exact binomial test that 1% of all bootstrapped p-values are less than or equal to 0.01. The one-sided alternative hypothesis is that more than 1% of all bootstrapped p-values are less than or equal to 0.01.
- KS** A Kolmogorov-Smirnov test of uniformity. The null hypothesis is that the bootstrapped p-values have a uniform distribution on (0,1). The one-sided alternative hypothesis is that the true distribution of the bootstrapped p-values is greater than the uniform.

We expect the tests to fail to reject the null hypotheses for the null distributions of the ROC goodness-of-fit statistics.

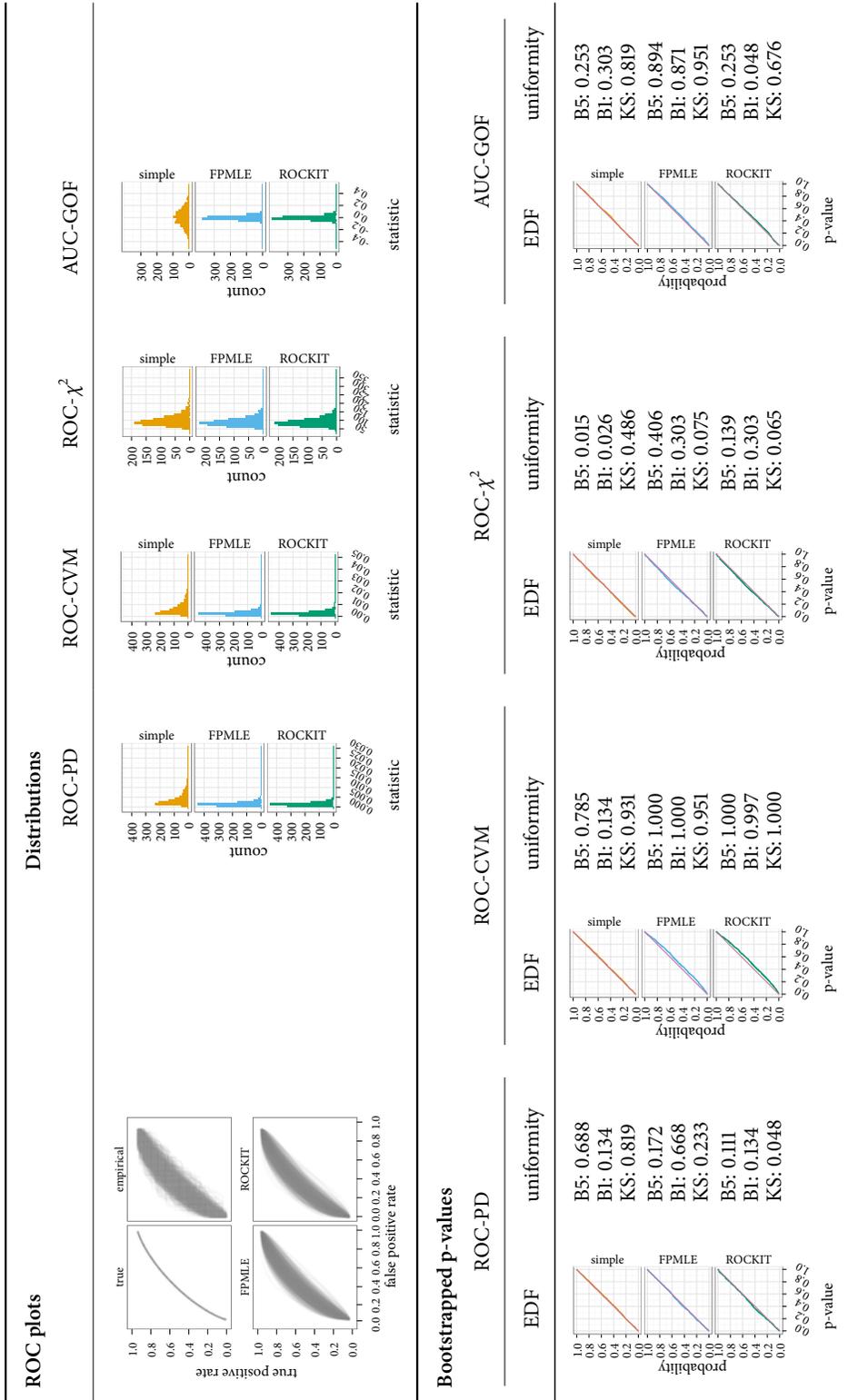


Figure 4.2: Goodness-of-fit results for binormal ROCs. $\rho = 1, \delta = 0.5, n_1 = 50, n_2 = 50$.

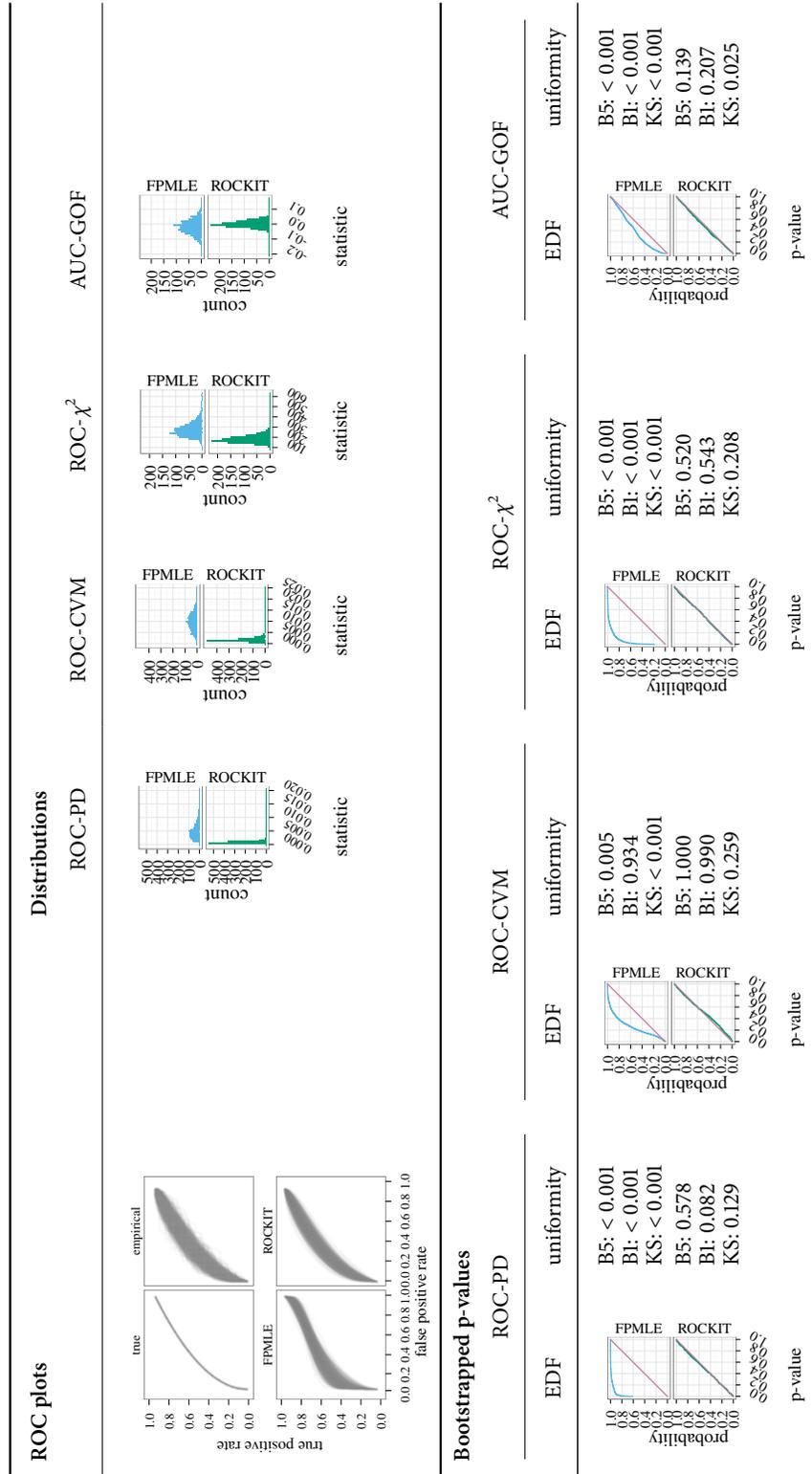


Figure 4.3: Goodness-of-fit results for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 100$, and $n_2 = 100$.

4.4.2 Null distributions

ROC-PD

Focusing mainly on the ROC-PD goodness-of-fit statistic, we check that the behavior under the null hypotheses satisfies basic properties of a useful goodness-of-fit statistical test. The principal results from the simulations conducted are:

1. The p-values of the ROC-PD statistic under a null hypothesis tend to follow a uniform distribution, particularly for p-values at or below the 0.1 level.
2. The ROC-PD statistic behaves as a one-sided test.
3. The distribution of the ROC-PD statistic depends on the type of hypothesis.
4. The distribution of the ROC-PD statistic depends on the values of binormal ROC parameters as well.

Below follows more discussion about these results.

The distributions of p-values from the ROC-PD statistic under a null hypothesis deserve close scrutiny. Across the simulations, the p-values appear uniform for p-values at or below the 0.1 level, which typically is the region of interest. The EDF plots of p-values do not depict anything of concern. The binomial tests of the 0.01 and 0.05 levels seldom reject, indicating that we do not detect a lack of uniformity at small p-values. Of all the checks of uniformity, the Komolgorov-Smirnov test is the only one to reject the null hypothesis occasionally, and those rejections seem to occur due to non-uniformity at larger p-values, as evident in the EDF plots.

Next, the ROC-PD statistic appears in the simulations to act as a one-sided test reliably under various null hypotheses. All of the distributions of the ROC-PD statistic in these simulations are right-skewed. Larger values of the ROC-PD statistic indicate more lack of fit. So the ROC-PD goodness-of-fit statistic leads to a one-sided, right-tailed test.

A clear distinction exists between the null distributions under a simple hypothesis and those under either composite hypothesis. The value of the ROC-PD goodness-of-fit statistic under the simple

null hypothesis is generally on a scale three times as large as that of the statistic under a composite null hypothesis. Under a composite null hypothesis, the estimates of ρ and δ may position the hypothesized parametric ROC curve fairly close to the empirical ROC curve. However, the fixed values of ρ and δ under a simple null hypothesis do not allow the same adjustment of proximity.

The distribution of the ROC-PD statistic under a null hypothesis is a function of the parameters of the binormal ROC curve. Comparing the distributions of the ROC-PD statistic from the simulations with $\rho = 1$ and $\delta = 0$ to the distributions of the ROC-PD statistic from the simulations with $\rho = 1$ and $\delta \neq 0$ indicates a dependence on the parameter δ . For $\rho = 1$, larger values of δ lead to larger values of the ROC-PD goodness-of-fit statistic.

ROC-CVM, ROC- χ^2 , and AUC-GOF

This section provides remarks on the three other goodness-of-fit statistics used in the simulations. While we already criticized the behavior of the ROC-CVM statistic in Chapter 3, we included it in the simulations to characterize it more completely. The p-values of the ROC-CVM statistic start to show hints of departure from uniformity once δ reaches one for $\rho = 1$. There are fewer small p-values and more large p-values than a uniform as indicted by the S-shape. This issue becomes more pronounced with larger delta. The ROC-CVM also gives fewer small p-values and more large p-values than expected from a uniform distribution for the asymmetric ROC curves with $\rho = 0.5$ and $\delta = 0.5$ as well as $\rho = 0.75$ and $\delta = 0.5$. The tests of uniformity do not necessarily reject these deviations because the one-sided alternatives are in the opposite direction (i.e., too many small p-values).

The chi-squared statistic shows occasional outliers. These may occur when an empirical ROC has a large value at a false positive rate of zero. Then the expected count from the parametric ROC curve, which has the value zero at a false positive rate of zero, may be appreciably smaller. As a result, the first term of the chi-squared statistic may contribute an unusually large value to the sum, which may lead to an appreciably higher value for the statistic. The chi-squared statistic also shows slightly more mid-range p-values than expected for $\rho = 1$ and $\delta = 1.5$.

The null distributions of the AUC-GOF test behave as expected across these simulations. The

distribution of the AUC-GOF statistic under a null hypothesis appears approximately normal.

4.4.3 Alternative distributions

Plots of some alternative ROC curves appear in figures 1.3, 1.4, and 1.5.

In Chapter 1 we mentioned that the binormal ROC curve can often fit ROC curves for populations which do not obviously follow normal distributions. The simulations below apply the goodness-of-fit tests to binormal ROC curves fit to empirical ROC curves generated for several alternative distributions of the underlying populations. We emphasize two points in sharing the results. First, a fully parametric approach to fitting a binormal ROC curve may fail while a semi-parametric approach may succeed. Second, ROC curves may arise where the binormal model does not apply and the ROC-PD statistic may detect the lack of fit. One important instance for which the binormal model may fail includes observations from a mixture distribution.

Biexponential

The results for studies of the goodness-of-fit of binormal models to data from two exponential populations appear on pages 177 through 188 in Appendix A. The perils of a fully parametric approach to estimating the parameters of a binormal model now become evident. The simulations suggest that the FPMLE models may fit the empirical ROC curves inadequately. Indeed, turning to the EDFs of p-values for the different goodness-of-fit tests, we see that the ROC-PD statistic readily rejects the FPMLE models. Of all the goodness-of-fit tests, the ROC-PD also seems to be the most powerful against this alternative. In these simulations, the AUC-GOF test does not appear to detect the FPMLE's lack of fit as well.

Still, using a semi-parametric fit, like those estimated by ROCKIT, yields curves that are difficult to discern from the empirical ones. Almost none of the goodness-of-fit tests offers much power against this alternative when fit semi-parametrically. The one exception may be the ROC-PD test applied with large sample sizes (e.g., over 1000).

Below we derive a straightforward analytic result to exemplify the failure of the FPMLE approach

of fitting binormal models to ROC curves from observations that arise from exponential distributions. We show that the AUC calculated from a binormal ROC estimated by the FPML method differs appreciably from the true AUC of a biexponential ROC as the means of the two underlying exponentials grow apart.

Suppose the negative observations come from an exponential distribution with mean λ_1 and the positive observations come from an exponential distribution with mean λ_2 . Again using the ROC curve's original definition,

$$\text{ROC}(p) = 1 - G[F^{-1}(1 - p)], \quad (4.7)$$

we may find the expression for the biexponential ROC curve to be

$$\text{ROC}(p) = p^{\lambda_1/\lambda_2}. \quad (4.8)$$

Integration gives the formula for the AUC of the biexponential ROC curve:

$$\text{AUC}_{\text{biexponential}} = \frac{1}{1 + \lambda_1/\lambda_2}. \quad (4.9)$$

The above expressions are for a true biexponential ROC curve.

Next, consider the theoretical results for a fully parametric binormal ROC curve fit to biexponential data. To start, because the mean equals the standard deviation for an exponential, we have for the mean and standard deviation of the two normals:

$$\mu_1 = \lambda_1, \quad \mu_2 = \lambda_2, \quad \sigma_1 = \lambda_1, \quad \text{and} \quad \sigma_2 = \lambda_2. \quad (4.10)$$

These result in the following values for the binormal parameters:

$$\begin{aligned}\rho &= \frac{\sigma_1}{\sigma_2} = \frac{\lambda_1}{\lambda_2} \\ \delta &= \frac{\mu_2 - \mu_1}{\sigma_2} = \frac{\lambda_2 - \lambda_1}{\lambda_2}.\end{aligned}\tag{4.11}$$

Let us compare the AUCs from the two approaches as we move the underlying exponentials far apart. Using equation (4.9), we have for the true AUC:

$$\lim_{(\lambda_1/\lambda_2) \rightarrow 0} \text{AUC}_{\text{biexponential}} = \lim_{(\lambda_1/\lambda_2) \rightarrow 0} \frac{1}{1 + (\lambda_1/\lambda_2)} \rightarrow 1.\tag{4.12}$$

For the FPML estimate of the binormal parameters, note that $\lim_{(\lambda_1/\lambda_2) \rightarrow 0} \rho = \lim_{(\lambda_1/\lambda_2) \rightarrow 0} \sigma_1/\sigma_2 = \lim_{(\lambda_1/\lambda_2) \rightarrow 0} (\lambda_1/\lambda_2) \rightarrow 0$. So the expression for the AUC from the binormal model becomes

$$\lim_{(\lambda_1/\lambda_2) \rightarrow 0} \text{AUC}_{\text{binormal}} = \lim_{(\lambda_1/\lambda_2) \rightarrow 0} \Phi\left(\frac{1 - \rho}{\sqrt{1 + \rho^2}}\right) \rightarrow \Phi(1).\tag{4.13}$$

Normal-gamma

Next we review the results of simulations in which the negative observations come from a normal distribution, but the positives come from a reflected (i.e., left-tailed) gamma distribution. In the fits on pages 190 through 201, the FPMLs look obviously poor. The ROC-PD statistic does as well as any of the goodness-of-fit tests. All of the tests struggle to discern a lack of fit for the semi-parametric approach when the sample sizes are below 100. The variability in the empirical ROC curves with the smaller samples sizes likely obscures the shapes of the true curves. Still, for sufficiently large sample sizes (i.e., over 100), the ROC-PD test rejects the binormal model for a normal-gamma ROC. Please see the results on page 199 for evidence.

Normal-mixture

In this section we consider again a normal distribution for the negative class but now a mixture of normals for the positive class. The mixture of normals may occur, for example, if two subpopulations of the positive class with different values of a covariate separately have normal distributions. Pages 204 through 214 depict the results of our simulations. The simulations suggest the following. First, the goodness-of-fit tests pick up the disparities between the binormal models and the empirical normal-mixture ROC curves for samples as small as 50. The binormal model seems incongruent with such mixtures even when fitted with a semi-parametric approach. Second, the EDFs of p-values indicate that the ROC-PD statistic has the most power against the alternatives in these simulations.

Normal-Cauchy

Suppose that the negative population is normally distributed while the positive population contains outliers. To simulate this aspect of the positive population, we used a Cauchy distribution. The fully parametric fits in the figures on pages 216 through 222 plainly disagree with the empirical ROC curves. However, the semi-parametric fits in the same figures might be harder to assess by eye. The EDFs of bootstrapped p-values reveal that the goodness-of-fit tests can distinguish the fitted binormal models from the empirical curves for appreciably overlapping underlying distributions (e.g., standard normal and Cauchy with $\theta = 1$ and $\sigma = 0.5$). Demonstrating the binormal model's lack of fit to much steeper normal-Cauchy ROC curves, like those generated in the case of a standard normal and a Cauchy with $\theta = 5$ and $\sigma = 1$, proves to be a lot harder.

4.5 Consistency of test

This section establishes the consistency of the ROC-PD goodness-of-fit test. For a consistent statistical test, the probability of rejection asymptotically goes to zero if the null hypothesis is true and approaches one under any alternative hypothesis.

Before the main proof, we establish definitions and assumptions used throughout this section, and we present a few lemmata. There are four different ROC curves that we need to define:

1. $R \equiv$ true, continuous ROC function (not necessarily binormal).
2. $\hat{R} \equiv$ empirical ROC.
3. $B \equiv$ best-fitting binormal to the true ROC curve, R , with respect to a metric Δ .
4. $\hat{B} \equiv$ binormal function estimated with \hat{R} using a continuous mapping. Let \hat{C} be an empirical ROC function and M be a continuous mapping from CDFs on $(0, 1)$ to the binormal family, \mathcal{B} , of CDFs. Then \hat{B} is such that

$$M(\hat{C}) = \underset{B \in \mathcal{B}}{\operatorname{argmin}} \Delta(B, \hat{C}). \quad (4.14)$$

We also make the same assumptions about sample sizes as Hsieh and Turnbull (1996, p. 28):

1. $n_1 = n_1(n_2)$.
2. $n_1/n_2 \rightarrow \lambda > 0$.

In addition, Hsieh and Turnbull (1996, p. 28) note that the slope, $f[G^{-1}(t)]/g[G^{-1}(t)]$, of the binormal ODC curve is bounded on any subinterval (o_1, o_2) of $(0,1)$, $0 < o_1 < o_2 < 1$. The slope, $g[F^{-1}(t)]/f[F^{-1}(t)]$, of the binormal ROC is bounded on any subinterval (p_1, p_2) of $(0,1)$, $0 < p_1 < p_2 < 1$ as well.

We proceed to the lemmata and then the proof.

Lemma 4.5.1. *The ROC-PD statistic \tilde{P}_1^2 calculated with respect to a reference ROC curve, R_C , that is continuous with a finite first derivative on $(0, 1)$ is a norm.*

Proof. Let R_C be a continuous ROC curve. Define the following:

$$\|\hat{E} - K\|_{1, R_C}^2 = \int_{R_C} D^2(p) d\ell = \int D^2(p) \sqrt{1 + [R'_C(p)]^2} dp \quad (4.15)$$

where $D^2(p)$ is the perpendicular distance squared between an empirical ROC curve, \hat{E} , and the point p on a continuous ROC, K . For a given value of p , $D(p)$ is the length of the hypotenuse of a right triangle. Let H and V be the lengths of the sides of this triangle in the horizontal and vertical directions. Then $D^2(p) = H^2(p) + V^2(p)$, and

$$\begin{aligned}\|\hat{E} - K\|_{\perp, R_C}^2 &= \int_{R_C} [H^2(p) + V^2(p)] \sqrt{1 + [R'_C(p)]^2} dp \\ &= \int_{R_C} H^2(p) \sqrt{1 + [R'_C(p)]^2} dp + \int_{R_C} V^2(p) \sqrt{1 + [R'_C(p)]^2} dp \\ &= \|H\|_{R_C}^2 + \|V\|_{R_C}^2.\end{aligned}\tag{4.16}$$

The norm $\|\cdot\|_{R_C}$ is simply the L_2 norm weighted by the positive, finite term $\sqrt{1 + [R'_C(p)]^2}$.

With the details above, we demonstrate that $\|\hat{E} - K\|_{\perp, R_C}^2$ meets the criteria of a norm.

1. If $\hat{E} - K = 0$, then

$$\begin{aligned}\|\hat{E} - K\|_{\perp, R_C} &= \left\{ \int_{R_C} D^2(p) \sqrt{1 + [R'_C(p)]^2} dp \right\}^{1/2} \\ &= \left\{ \int_{R_C} 0 \sqrt{1 + [R'_C(p)]^2} dp \right\}^{1/2} = 0.\end{aligned}\tag{4.17}$$

So $\|\hat{E} - K\|_{\perp, R_C} = 0$ if $\hat{E} - K = 0$.

2. Let $a \in \mathbb{R}$, then

$$\begin{aligned}\|a(\hat{E} - K)\|_{\perp, R_C} &= \left\{ \int_{R_C} a^2 D^2(p) \sqrt{1 + [R'_C(p)]^2} dp \right\}^{1/2} \\ &= |a| \left\{ \int_{R_C} D^2(p) \sqrt{1 + [R'_C(p)]^2} dp \right\}^{1/2} = |a| \|\hat{E} - K\|_{\perp, R_C}.\end{aligned}\tag{4.18}$$

Thus, $\|a(\hat{E} - K)\|_{\perp, R_C} = |a| \|\hat{E} - K\|_{\perp, R_C}$.

3. The last property of a norm is the triangle inequality. In the sum of two functions, $(\hat{E}_1 - K_1)$ and $(\hat{E}_2 - K_2)$, the horizontal and vertical components may add or subtract. In either case, for the squared norm of the sum, we have

$$\|(\hat{E}_1 - K_1) + (\hat{E}_2 - K_2)\|_{\perp, R_C}^2 \leq \|H_1 + H_2\|_{R_C}^2 + \|V_1 + V_2\|_{R_C}^2. \quad (4.19)$$

The sum of the individual squared norms is

$$\|\hat{E}_1 - K_1\|_{\perp, R_C}^2 + \|\hat{E}_2 - K_2\|_{\perp, R_C}^2 = \|H_1\|_{R_C}^2 + \|H_2\|_{R_C}^2 + \|V_1\|_{R_C}^2 + \|V_2\|_{R_C}^2. \quad (4.20)$$

Now, the weighted L_2 norms satisfy the triangle inequality. That is,

$$\begin{aligned} \|H_1 + H_2\|_{R_C}^2 &\leq \|H_1\|_{R_C}^2 + \|H_2\|_{R_C}^2 \text{ and} \\ \|V_1 + V_2\|_{R_C}^2 &\leq \|V_1\|_{R_C}^2 + \|V_2\|_{R_C}^2. \end{aligned} \quad (4.21)$$

Using these inequalities in equations (4.19) and (4.20), we have

$$\|(\hat{E}_1 - K_1) + (\hat{E}_2 - K_2)\|_{\perp, R_C}^2 \leq \|\hat{E}_1 - K_1\|_{\perp, R_C}^2 + \|\hat{E}_2 - K_2\|_{\perp, R_C}^2. \quad (4.22)$$

Therefore, $\|\hat{E} - K\|_{\perp, R_C}$ follows the triangle inequality.

Given that $\|\hat{E} - K\|_{\perp, R_C}$ fulfills criteria one through three, we conclude that it is a norm. \square

Lemma 4.5.2. *For a norm, the following reverse triangle inequality holds: $|\|f\| - \|g\|| \leq \|f - g\|$.*

Proof. Start with the triangle inequality, which holds for any norm:

$$\|f + g\| \leq \|f\| + \|g\|. \quad (4.23)$$

First substitute $f - g$ for f in equation (4.23):

$$\|(f - g) + g\| \leq \|(f - g)\| + \|g\|. \quad (4.24)$$

Reorganizing terms yields:

$$\begin{aligned} \|(f - g) + g\| - \|g\| &\leq \|(f - g)\|, \text{ or} \\ \|f\| - \|g\| &\leq \|f - g\|. \end{aligned} \quad (4.25)$$

Now start with g and replace it with $g - f$ in equation (4.23):

$$\|f + (g - f)\| \leq \|f\| + \|(g - f)\|. \quad (4.26)$$

Rearranging terms gives:

$$\begin{aligned} \|f + (g - f)\| - \|f\| &\leq \|(g - f)\|, \text{ or} \\ \|g\| - \|f\| &\leq \|g - f\|. \end{aligned} \quad (4.27)$$

Thus, we have that $|\|f\| - \|g\|| \leq \|f - g\|$. □

Lemma 4.5.3. *For the empirical ROC curve \hat{R} and true ROC curve R , the limit of the perpendicular norm with respect to the continuous ROC curve R_C with finite derivative tends to zero almost surely as $n_1, n_2 \rightarrow \infty$. That is,*

$$\lim_{n_2 \rightarrow \infty} \|(\hat{R}_N - R)\|_{\perp, R_C} \rightarrow 0 \text{ almost surely.} \quad (4.28)$$

Proof. Below we establish a bound on the length of the perpendicular line from the parametric ROC curve to the empirical ROC curve. The bound is in terms of the vertical and horizontal distances between the parametric ROC curve and the empirical ROC curve. Let us introduce the following four points on the ROC plot with horizontal x -axis and vertical y -axis:

1. (x_θ, y_θ) : the point of interest on the parametric ROC curve.
2. (x_e, y_e) : the point where the perpendicular line from the parametric ROC curve intersects the empirical ROC curve.
3. (x_h, y_h) : the point where the horizontal line $y = y_\theta$ from the parametric ROC curve intersects the empirical ROC curve.
4. (x_v, y_v) : the point where the vertical line $x = x_\theta$ from the parametric ROC curve intersects the empirical ROC curve.

Previously, we established that the slope of these perpendicular lines must be negative. As a result, there are two possible cases for the relationship between (x_r, y_r) and (x_e, y_e) .

- Case 1: $(x_\theta < x_e)$ and $(y_\theta > y_e)$
 - For the horizontal line, $y_h = y_\theta$, and since $y_\theta > y_e$, $y_h > y_e$. Because the empirical ROC is monotonically increasing, $x_h \geq x_e$ if $y_h > y_e$. So for the distance in the horizontal direction $\Delta x \leq x_h - x_\theta$.
 - For the vertical line, $x_v = x_\theta$, and since $x_\theta < x_e$, $x_v < x_e$. Because the empirical ROC is monotonically increasing, $y_v \leq y_e$ if $x_v < x_e$. So for the distance in the vertical direction $\Delta y \leq y_\theta - y_v$.
- Case 2: $(x_\theta > x_e)$ and $(y_\theta < y_e)$
 - For the horizontal line, $y_h = y_\theta$, and since $y_\theta < y_e$, $y_h < y_e$. Because the empirical ROC is monotonically increasing, $x_h \leq x_e$ if $y_h < y_e$. So for the distance in the horizontal direction $\Delta x \leq x_\theta - x_h$.
 - For the vertical line, $x_v = x_\theta$, and since $x_\theta > x_e$, $x_v > x_e$. Because the empirical ROC is monotonically increasing, $y_v \geq y_e$ if $x_v > x_e$. So for the distance in the vertical direction $\Delta y \leq y_v - y_\theta$.

In both cases, we have

$$H = \Delta x \leq |x_h - x_\theta| \tag{4.29}$$

$$V = \Delta y \leq |y_v - y_\theta|.$$

Then for the distance D in the perpendicular direction, we have

$$D^2 \leq (\Delta x)^2 + (\Delta y)^2. \tag{4.30}$$

In their Theorem 2.1, Hsieh and Turnbull state the pointwise strong consistency of the empirical ODC curve for the true ODC curve:

$$\sup_{0 \leq q \leq 1} |\hat{F}_{n_1}[\hat{G}_{n_2}^{-1}(t)] - F[G^{-1}(t)]| \xrightarrow{a.s.} 0 \text{ as } n_2 \rightarrow \infty. \tag{4.31}$$

Similarly, we have the pointwise strong consistency of the empirical ROC curve for the true ROC curve:

$$\sup_{0 \leq p \leq 1} |\hat{G}_{n_2}[\hat{F}_{n_1}^{-1}(t)] - G[F^{-1}(t)]| \xrightarrow{a.s.} 0 \text{ as } n_1 \rightarrow \infty. \tag{4.32}$$

Equation 4.32 indicates that the vertical distance between the empirical and true ROC curves goes to zero at every point. That is,

$$\sup |y_v - y_\theta| \rightarrow 0. \tag{4.33}$$

The distance in the horizontal direction is just the difference between the empirical and true ODC as covered earlier. Therefore, from equation 4.31, we have

$$\sup |x_h - x_\theta| \rightarrow 0. \tag{4.34}$$

Recalling that

$$\|\hat{R} - R\|_{\perp, R_C}^2 = \|H\|_{R_C}^2 + \|V\|_{R_C}^2, \tag{4.35}$$

we see that

$$\lim_{n_2 \rightarrow \infty} \|(\hat{R}_N - R)\|_{\perp, R_C} = \lim_{n_2 \rightarrow \infty} \|H\|_{R_C} + \lim_{n_2 \rightarrow \infty} \|V\|_{R_C} = 0. \quad (4.36)$$

□

The arguments in lemma 4.5.3 also lead to a limit for $\|(\hat{B}_N - B)\|_{\perp, R_C}$. Under this section's assumptions, $M(\cdot)$ is a continuous mapping. Given that $\hat{R} \xrightarrow{a.s.} R$ and the continuous mapping theorem, we have

$$\hat{B} = M(\hat{R}) \xrightarrow{a.s.} M(R) = B. \quad (4.37)$$

Bounds on the horizontal and vertical distances between \hat{B} and B exist like those in lemma 4.5.3. Then by similar arguments,

$$\lim_{n_2 \rightarrow \infty} \|(\hat{B}_N - B)\|_{\perp, R_C} = 0. \quad (4.38)$$

Theorem 4.5.4. *The ROC-PD statistic is a consistent test of the goodness-of-fit of a binormal model for the ROC curve. That is,*

$$\lim_{n_2 \rightarrow \infty} \|\hat{R} - \hat{B}\| = \|R - B\| = \begin{cases} 0 \text{ almost surely} & \text{if the true ROC curve is binormal} \\ c > 0 \text{ almost surely} & \text{if the true ROC curve is not binormal} \end{cases} \quad (4.39)$$

Proof. Consider:

$$\|\hat{R} - \hat{B}\| = \|(\hat{R} - R) + (B - \hat{B}) + (R - B)\|. \quad (4.40)$$

Let \hat{R}_N be a consistent estimator for R and \hat{B}_N a consistent estimator for B . Lemma 4.5.2 gives

$$\left| \|(\hat{R}_N - R) + (B - \hat{B}_N) + (R - B)\| - \|R - B\| \right| \leq \|(\hat{R}_N - R) + (B - \hat{B}_N) + (R - B) - (R - B)\| \quad (4.41)$$

$$\leq \|(\hat{R}_N - R) + (B - \hat{B}_N)\| \quad (4.42)$$

$$\leq \|\hat{R}_N - R\| + \|B - \hat{B}_N\|. \quad (4.43)$$

Now, given any ε , choose N such that

$$\|\hat{R}_N - R\| + \|B - \hat{B}_N\| \leq \varepsilon. \quad (4.44)$$

(Such an N exists if the two estimators are consistent.) Then, for such an N ,

$$\left| \|(\hat{R}_N - R) + (B - \hat{B}_N) + (R - B)\| - \|R - B\| \right| \leq \varepsilon, \quad (4.45)$$

or

$$\lim_{n_2 \rightarrow \infty} \|(\hat{R}_N - R) + (B - \hat{B}_N) + (R - B)\| = \|R - B\|. \quad (4.46)$$

If R belongs to the family of binormals, then $B = R$ and $\|R - B\| = 0$. Otherwise, $B \neq R$ and $\|R - B\| = c > 0$, where c is a positive constant. \square

4.6 Application

4.6.1 ROC curve for species of Iris

As an application of the ROC-PD goodness-of-fit statistic, we explore one ROC curve generated from the well-known Iris data reported by Anderson (1935) and analyzed by Fisher (1936). The data consist of physical aspects of flowers measured across three different species with 50 observations from each. The true species of each flower is one of: *setosa*, *versicolor*, or *virginica*.

Let us reduce the species to two classes by treating *virginica* as the negative class and combining the *setosa* and *versicolor* species into the positive class. We use one variable, a flower's sepal width measured in centimeters, as a score in a straightforward classification scheme. Observations from the negative class (*virginica*) have the distribution of scores shown on the left of figure 4.4, and observations from the positive class (not *virginica*) have the distribution of scores shown on the right of figure 4.4.

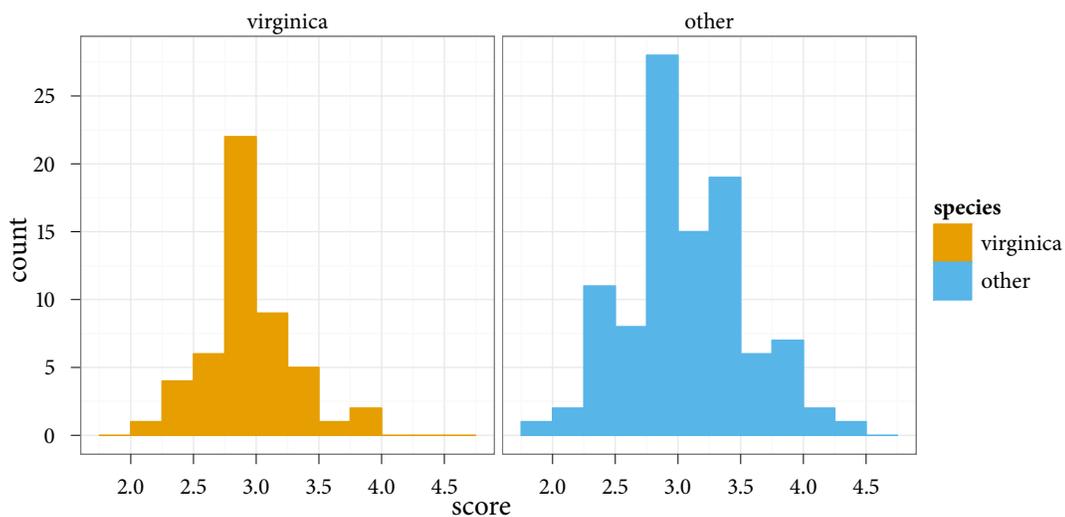


Figure 4.4: Distributions of scores (sepal widths measured in centimeters) for two classes composed of species of Iris.

To assess the performance of classifying the flowers as *virginica* or not using only the sepal width, we plot the empirical ROC curve as shown in figure 4.5. Now we wish to test the composite null hypothesis that the observed ROC curve arose in truth from a binormal ROC function. In testing this hypothesis, we use the semi-parametric estimates from ROCKIT and apply the new ROC-PD goodness-of-fit statistic. The fitted binormal ROC curve appears in figure 4.5 with the empirical. The parameter estimates are $\hat{\rho} = 0.677$ and $\hat{\delta} = 0.247$.

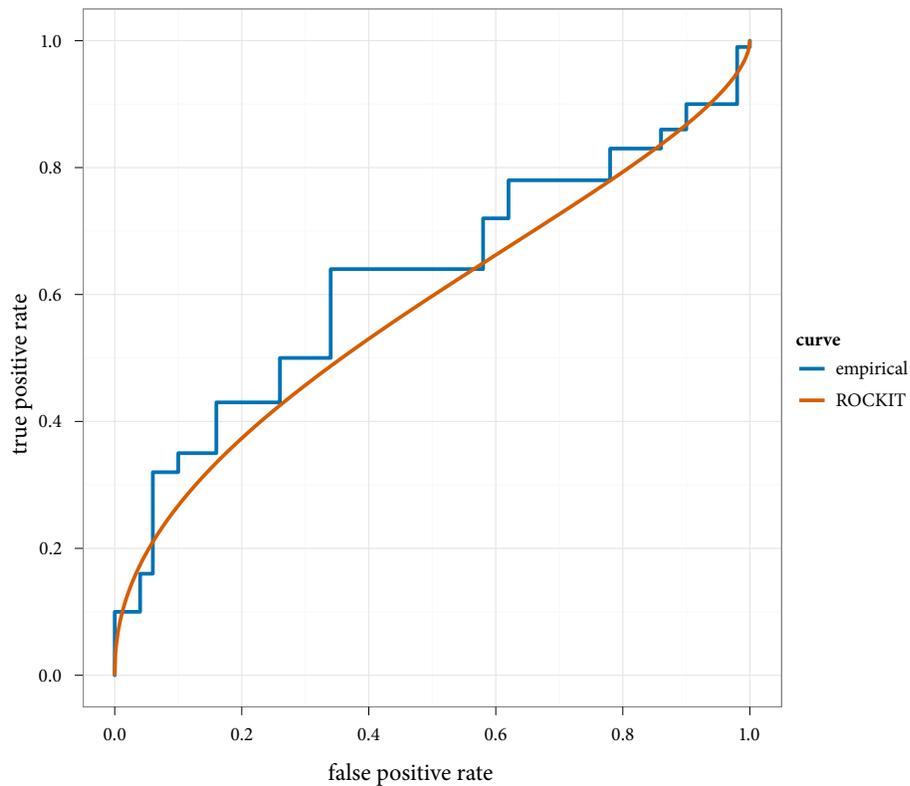


Figure 4.5: ROC curves for application with Iris data.

We compute the ROC-PD statistic and conduct a parametric bootstrap of the test statistic to estimate the p-value. The small p-value of approximately 0.018 suggests that we reject the null hypothesis that the true ROC curve belongs to the binormal family.

This application demonstrates the capability of the ROC-PD statistic to reject the binormal model when one of the underlying classes consists of a mixture of distributions. The sample sizes of 50 for the negative and 100 for the positive class reveal the statistic's potential usefulness in practical settings.

Chapter 5: Estimation of binormal ROC functions

Estimating binormal ROC functions constitutes the second part of this research. We develop three new sets of estimators for binormal ROC functions and explore their behavior through computer simulations. Our ultimate objective is to estimate the binormal ROC function itself, not just the parameters. So our focus is on global measures of error between the estimated and true ROC curves. Furthermore, we may use such global measures to study errors of binormal functions estimated from data that do not originate from underlying normals. For example, we look at the integrated squared error between an estimated *binormal* ROC function and a true *bixponential* ROC function to understand the flexibility of the binormal model.

In addition, we introduce a parametric bootstrap methodology for calculating separate and simultaneous confidence intervals for the parameters of the binormal ROC curve.

5.1 New estimation methods

5.1.1 Minimum perpendicular distance (ROC-MPD)

Recall that the new ROC-PD goodness-of-fit statistic is based on the distance along perpendicular lines from the parametric ROC curve to the empirical ROC curve. Denote the binormal ROC curve with parameters ρ and δ as $\text{ROC}_\theta(\cdot)$, where $\theta = (\rho, \delta)$, and the empirical ROC curve as $\widehat{\text{ROC}}(\cdot)$. Then let $D(p^*)$ be the distance from a single point $(p^*, \text{ROC}_\theta(p^*))$ on the parametric ROC curve to the point of intersection on $\widehat{\text{ROC}}(\cdot)$. That is,

$$D(p^*) = \text{perpendicular distance between } \text{ROC}_\theta(p^*) \text{ and } \widehat{\text{ROC}}(\cdot). \quad (5.1)$$

The ROC-PD goodness-of-fit statistic, introduced in equation (3.26) initially, is

$$P_1^2 = \sum_{i=1}^d D^2(p_i) \Delta \ell,$$

where the points p_i are false positive rates which divide the arc of the parametric ROC curve into equal segments.

In the previous chapter, we used this statistic as a measure of distance between the empirical ROC and a parametric one. However, we may also develop a new method of estimating the parameters of the binormal ROC with it. Observe that the value of the statistic depends on the distances $D(p_i)$, which themselves depend on the parameterization of the binormal ROC curve. Starting with an empirical ROC curve and the family of binormal ROC curves, we can find estimates of the parameters $\theta = (\rho, \delta)$ by minimizing the goodness-of-fit statistic. Specifically, this new method of minimum distance estimation for the binormal ROC parameters yields the estimates $\hat{\theta} = (\hat{\rho}, \hat{\delta})$ defined as

$$\hat{\theta}_P = \underset{\theta=(\rho,\delta)}{\operatorname{argmin}} P_1^2.$$

We will refer to this method as the minimum perpendicular distance estimator, or the ROC-MPD estimator. Finding the estimates $\hat{\theta}_P = (\hat{\rho}, \hat{\delta})$ that minimize the goodness-of-fit statistic requires a numerical optimization routine.

5.1.2 Total least squares (ROC-TLS)

The preceding section introduced a new method of estimating the parameters of a binormal ROC curve by minimizing a sum of squared perpendicular distances between the parametric and empirical ROC curves. On the standard scale of ROC curves, with the true positive rate and false positive rate as axes, the binormal ROC curve is nonlinear. As mentioned above, the search for the parameter

estimates that minimize the perpendicular distance on this scale necessitates numerical optimization. However, a simple transformation leads to a linear function for the binormal ROC curve. Furthermore, minimizing perpendicular distances to a line follows simply from principal component analysis.

Let us consider the transformation first. As discussed in Chapter 1, the binormal ROC function may be expressed as

$$\text{ROC}(p) = \Phi[\rho\Phi^{-1}(p) + \delta].$$

Applying a probit transformation to both sides of the previous equation results in

$$\Phi^{-1}[\text{ROC}(p)] = \rho\Phi^{-1}(p) + \delta.$$

So under probit transformations of both the true positive rate $\text{ROC}(p)$ and false positive rate p , the binormal ROC curve is a line with slope ρ and intercept δ .

Next we will discuss a method to estimate the binormal parameters ρ and δ on this scale. Principal component analysis (PCA) provides a method to find a line that minimizes the orthogonal distance to the data points. Recall that in general the first principal component is a vector that runs in the direction of maximum variability among the data. The second principal component must be orthogonal to the first principal component. Now, in a plane, which has just two dimensions, any two orthogonal directions split the total sum of squares. We wish to determine the line with the smallest sum of squared orthogonal distances to the data points. Maximizing the sum of squared distances along the line must minimize the sum of orthogonal distances. The line that maximizes the sum of squared distances follows the vector that maximizes the variability, i.e., the first principal component. So identifying the first principal component gives the line that minimizes the sum of orthogonal distances squared.

Let (\hat{p}_i, \hat{R}_i) be coordinates along the empirical ROC curve. Consider \mathcal{X} to be the following data

matrix

$$\mathcal{X} = \begin{pmatrix} \Phi^{-1}(\hat{p}_1) & \Phi^{-1}(\hat{R}_1) \\ \Phi^{-1}(\hat{p}_2) & \Phi^{-1}(\hat{R}_2) \\ \vdots & \vdots \\ \Phi^{-1}(\hat{p}_n) & \Phi^{-1}(\hat{R}_n) \end{pmatrix}$$

and $\bar{\mathcal{X}}$ a matrix of the column means. Conducting a PCA with the centered data $(\mathcal{X} - \bar{\mathcal{X}})$ leads to two principal components, \hat{e}_1 and \hat{e}_2 . The vector \hat{e}_1 points along the direction of the line that minimizes the sum of the squared orthogonal distances from the centered data points to the line. This line has a slope of $\hat{\rho} = \hat{e}_{12}/\hat{e}_{11}$, where \hat{e}_{11} and \hat{e}_{12} are the first and second components of the vector \hat{e}_1 . With the column means of \mathcal{X} denoted by \bar{x}_1 and \bar{x}_2 , we calculate the intercept $\hat{\delta}$ by using $\hat{\delta} = \bar{x}_2 - \rho\bar{x}_1$.

Other names for this procedure include *orthogonal regression* and *total least squares regression*. We will call this estimator associated with total-least squares the ROC-TLS estimator.

5.1.3 Minimum Cramér-von Mises distance (ROC-MCVM)

Recall the definition of the Cramér-von Mises statistic for ROC curves, denoted as ROC-CVM, and initially given in equation (5.1.3) of Chapter 3:

$$W_{\text{ROC}}^2 = 2 \sum_{i=1}^k (b_i - b_{i-1}) \left[\frac{c_i^2}{2} - \frac{b_{i-1} + b_i}{2} c_i \right] + \frac{1}{3}. \quad (5.2)$$

Similar to the minimum ROC-MPD estimator, we can use this distance as an objective function for estimating the parameters of the binormal ROC function. We define the minimum Cramér-von Mises (MCVM) estimator for the parameters of the binormal as:

$$\hat{\theta}_{\text{CvM}} = \underset{\theta=(\rho,\delta)}{\operatorname{argmin}} W_{\text{ROC}}^2.$$

We refer to it as the ROC-MCVM estimator.

5.2 Design of simulations

5.2.1 Outline of simulations

The simulations for studying estimators of binormal ROC curves follow the simulations for examining the goodness-of-fit tests in Chapter 4. For details on the generation of sample data and the different distributions of the underlying negative and positive classes, the reader may refer to sections 4.1–4.3 in particular. Below, we will describe our application of several additional estimators for binormal ROC functions.

5.2.2 Additional estimators

Besides the three new methods of estimating the binormal parameters ρ and δ introduced above, the simulations include the following estimators.

Fully parametric maximum likelihood (FPML)

We discussed fully parametric maximum likelihood estimation (FPMLE) in section 4.2.2 and so will just briefly review the method here. This approach calculates the maximum likelihood estimates for the mean and standard deviation of the negative and positive classes separately. Functions of these four estimates then yield estimates of the two parameters of the binormal ROC. This method provides the only full parametric estimation of the binormal ROC curve's parameters. The remaining methods are all semi-parametric.

ROCKIT

Also applied in Chapter 4, the software library `libroc` from ROCKIT calculates semi-parametric estimates of the binormal ROC curve's parameters. For details, the reader may check the literature review or (Metz et al., 1998). The ROCKIT algorithm has one user-defined parameter, which is the maximum number of categories to create from continuous runs of negative or positive scores. We chose a value of 50 as recommended in the software library's documentation (<http://metz-roc.uchicago.edu/MetzROC/software/how-to-call-metz-roc-functions>).

Generalized linear model (ROC-GLM)

Section 2.3.3 reviews the use of generalized linear models (GLMs) to estimate the parameters of the binormal ROC curve. The implementation in our simulations follows the development in Alonzo and Pepe (2002) and Pepe (2004). The ROC-GLM technique relies on a choice of false positive rates at which to evaluate the height of the empirical ROC curve. With n_1 the number of observations from the negative class, we used the set $\mathcal{P} = \{1/n_1, \dots, (n_1 - 1)/n_1\}$ of false positive rates.

Semi-parametric least squares (ROC-SLS)

Another semi-parametric approach to estimating the binormal ROC curve's parameters received mention in Section 2.3.4 and appears in Tang and Zhou (2011). After taking a probit transform of points along the empirical ROC, this semi-parametric least squares (SLS) technique estimates ρ and δ by ordinary linear regression. The points used along the empirical ROC include the set $\mathcal{P} = \{1/n_1, \dots, (n_1 - 1)/n_1\}$ of false positive rates within the bounds of $(0.00001, 0.9999]$.

5.2.3 Measures of error

As mentioned at the start, the real goal of this chapter lies in estimating ROC curves with the binormal function. Therefore, global measures of an estimation method's errors are useful. We report the integrated squared error (ISE) between the true parametric ROC curve and the estimated binormal ROC curve in both the vertical and horizontal directions. In addition, we compute a summed squared perpendicular error between the curves. Unlike the ROC-PD statistic, which adds squared distances in the perpendicular direction from the parametric ROC curve to the empirical ROC curve, this summed perpendicular error captures distance between two parametric ROC curves. All tables list the root mean value of these three squared errors to portray the errors on the same scale as the functions themselves.

The global integrated and summed squared errors just discussed indicate performance even if the true parametric ROC from which we draw simulation samples is not in the binormal family. When in fact the true parametric ROC belongs to the binormal family, we can also calculate the root mean

squared error (RMSE) and bias of the different estimates of ρ and δ .

As another indication of error, the area under the curve (AUC) may apply. As discussed in Chapter 1, the empirical AUC is equivalent to the Mann-Whitney U statistic for estimating the probability that a random observation from the positive class has a higher score than a random observation from the negative class. More details are available in Bamber (1975). Here we choose to report the RMSE and bias of the AUC because it is perhaps the most common single measure of performance derived from the ROC curve. In addition to errors for the AUC from the estimated binormal ROC curves, we also give errors for the empirical AUC.

5.3 Estimates of binormal ROC functions in null cases

This section presents results for estimates of binormal ROC functions fit to data actually generated from a binormal ROC. Complete tables for all of the simulations appear in Appendix B. Below we have selected a few scenarios to demonstrate the behavior of the function estimates for different ρ and δ as well as different sample sizes.

Overall, the ROC-TLS method produces the smallest global errors among the semi-parametric methods for ROC curves that roughly fall in the middle of the ROC plot. For steeper ROC curves with the true parameter $\delta \gtrsim 1.5$ and smaller sample sizes, other methods may do slightly better. In this case of estimating binormal ROC functions with observations drawn from binormal ROC curves, the FPML estimates usually have smaller errors than the semi-parametric methods.

5.3.1 Perpendicular distance squared

Listed in tables 5.1, 5.2, and 5.3 are the root mean values of the summed squared perpendicular distance between the true ROC function and the estimated ROC function. Three combinations of ρ and δ offer examples of the estimators' behaviors for different binormal ROC curves. Simulations with different sample sizes also help to characterize the estimators.

Table 5.1: Root mean summed perpendicular squared errors for $\rho = 1$ and $\delta = 0.5$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	sample sizes		
	$n_1 = 30, n_2 = 30$	$n_1 = 100, n_2 = 100$	$n_1 = 100, n_2 = 50$
ROC-CVM	7.77 ^g	4.10 ^g	5.09 ^g
ROC-MPD	7.56 ^e	4.03 ^f	4.99 ^f
ROC-TLS	7.03 ^b	3.75 ^b	4.65 ^a
FPML	7.02 ^a	3.75 ^a	4.65 ^b
ROC-GLM	7.58 ^f	3.92 ^e	4.85 ^e
ROC-SLS	7.37 ^c	3.82 ^c	4.72 ^c
ROCKIT	7.41 ^d	3.82 ^d	4.75 ^d

Table 5.2: Root mean summed perpendicular squared errors for $\rho = 1$ and $\delta = 1.5$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	sample sizes		
	$n_1 = 30, n_2 = 30$	$n_1 = 100, n_2 = 100$	$n_1 = 100, n_2 = 50$
ROC-CVM	5.71 ^g	3.25 ^g	3.93 ^g
ROC-MPD	5.38 ^c	2.89 ^f	3.58 ^f
ROC-TLS	5.42 ^e	2.76 ^b	3.42 ^b
FPML	4.88 ^a	2.64 ^a	3.27 ^a
ROC-GLM	5.40 ^d	2.84 ^c	3.52 ^d
ROC-SLS	5.59 ^f	2.87 ^e	3.56 ^e
ROCKIT	5.29 ^b	2.85 ^d	3.48 ^c

Table 5.3: Root mean summed perpendicular squared errors for $\rho = 0.5$ and $\delta = 0.5$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	sample sizes		
	$n_1 = 30, n_2 = 30$	$n_1 = 100, n_2 = 100$	$n_1 = 100, n_2 = 50$
ROC-CVM	7.70 ^e	4.02 ^e	5.48 ^e
ROC-MPD	7.74 ^f	4.11 ^g	5.59 ^g
ROC-TLS	7.41 ^b	3.90 ^b	5.28 ^b
FPML	6.93 ^a	3.73 ^a	5.03 ^a
ROC-GLM	7.82 ^g	4.04 ^f	5.49 ^f
ROC-SLS	7.68 ^d	3.99 ^d	5.42 ^d
ROCKIT	7.56 ^c	3.90 ^c	5.33 ^c

The following remarks summarize some of the estimators' behavior in these simulations. Of all the semi-parametric approaches, the ROC-TLS estimator has the smallest global perpendicular squared error when the ROC curve is not too steep (i.e., for $\delta \lesssim 1.5$). The ROC-MPD estimator tends to exhibit a larger error than the semi-parametric methods besides the ROC-CVM estimator, which does most poorly.

For all of the methods, the errors decrease as the sample sizes increase. Augmenting the number of observations from the negative class, which may be easier to include in practice, also seems to improve the errors.

5.3.2 Mean squared error and bias of parameter and AUC estimations

Because these simulations contain sample data drawn randomly from binormal ROC functions with known parameters, the MSE and bias of the parameters are available. Table 5.4 gives results from simulations in which $\rho = 1$, $\delta = 1$, and the number of observations from the negative and positive classes are each 100.

Table 5.4: Root mean squared error (RMSE) and bias of estimations of binormal parameters for $\rho = 1$ and $\delta = 1$ over 5000 samples each with $n_1 = 100$ and $n_2 = 100$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	ρ		δ	
	RMSE	bias	RMSE	bias
ROC-CVM	15.89 ^g	1.29 ^f	20.06 ^g	2.22 ^g
ROC-MPD	14.41 ^f	0.96 ^e	17.51 ^f	1.61 ^e
ROC-TLS	11.97 ^b	0.74 ^d	16.36 ^c	0.94 ^b
FPML	10.21 ^a	0.47 ^b	16.17 ^a	1.71 ^f
ROC-GLM	12.54 ^d	-1.93 ^g	16.62 ^e	1.35 ^d
ROC-SLS	12.64 ^e	0.15 ^a	16.41 ^d	-1.20 ^c
ROCKIT	12.28 ^c	0.71 ^c	16.36 ^b	0.61 ^a

Table 5.5: Root mean squared error (RMSE) and bias of AUC estimations for $\rho = 1$ and $\delta = 1$ over 5000 samples each with $n_1 = 100$ and $n_2 = 100$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	RMSE	bias
ROC-CVM	3.25 ^a	0.06 ^e
ROC-MPD	3.47 ^g	-0.03 ^c
ROC-TLS	3.26 ^b	0.02 ^a
FPML	3.41 ^f	-0.05 ^d
ROC-GLM	3.28 ^d	0.10 ^g
ROC-SLS	3.51 ^h	0.03 ^b
ROCKIT	3.28 ^c	-0.09 ^f
empirical	3.33 ^e	0.11 ^h

A few general behaviors of the estimators stand out in these tables. The trend of the MSEs for

the parameter and AUC estimates tends to follow the results for the squared errors of the function estimates in the previous section. The MSEs of the ROC-TLS estimates tend to be among the smallest of the semi-parametric methods, especially for $\delta < 1.5$. For all methods, the variability dominates the MSE of the estimates of ρ , δ , and the AUC.

5.4 Estimates of binormal ROC functions in alternative cases

The simulations in this section use data generated from parametric ROC curves that do not belong to the binormal family. Of interest are two highlights. First, the FPML method goes awry, but the semi-parametric methods often fit a binormal ROC function close to the true parametric ROC. Whether the true ROC curve includes an exponential or a gamma distribution, the semi-parametric estimates do not deviate unreasonably. Second, among the semi-parametric methods, the ROC-TLS estimator typically results in the smallest global errors for ROC curves that lie toward the middle of the plot as in the previous section with observations from a true binormal ROC curve.

5.4.1 Biexponential

From the previous chapter, we saw that while the FPML method fails to fit the biexponential, the semi-parametric method of ROCKIT can produce a rather close fit. The results in this section cover additional semi-parametric fits of the binormal ROC function to samples from biexponential ROC curves.

Function errors

Tables 5.6 and 5.7 give the global squared errors in the perpendicular, vertical, and horizontal directions. Here the FPML method produces appreciably larger function errors than all of the semi-parametric methods. While the semi-parametric methods have similar function errors, an emerging trend may be deducible. The ROC-TLS method has some of the smallest function errors for ROC curves which have more gradual curvatures. For instance, the function errors of the ROC-TLS method are smallest of all estimators for the biexponential with $\lambda_1 = 1$, $\lambda_2 = 2$. Still, as the ROC curve

becomes flatter either vertically or horizontally, like in the case of the biexponential with $\lambda_1 = 1$ and $\lambda_2 = 15$, other semi-parametric methods may do better.

Table 5.6: Root mean global squared errors for binormal functions fit to 5000 samples from biexponential with $\lambda_1 = 1$ and $\lambda_2 = 2$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	$n_1 = 30, n_2 = 30$			$n_1 = 100, n_2 = 100$		
	perpendicular	vertical	horizontal	perpendicular	vertical	horizontal
ROC-CVM	7.55 ^f	8.68 ^f	9.36 ^f	3.94 ^e	4.61 ^e	4.98 ^f
ROC-MPD	7.43 ^d	8.61 ^e	9.20 ^d	3.94 ^f	4.65 ^f	4.97 ^e
ROC-TLS	6.95 ^a	8.08 ^a	8.62 ^a	3.70 ^a	4.36 ^a	4.67 ^a
FPML	9.29 ^g	10.67 ^g	10.73 ^g	7.21 ^g	8.55 ^g	8.62 ^g
ROC-GLM	7.46 ^e	8.60 ^d	9.25 ^e	3.84 ^d	4.52 ^d	4.85 ^d
ROC-SLS	7.22 ^c	8.39 ^c	8.95 ^b	3.76 ^c	4.44 ^c	4.73 ^c
ROCKIT	7.22 ^b	8.33 ^b	8.97 ^c	3.74 ^b	4.39 ^b	4.72 ^b

Table 5.7: Root mean global squared errors for binormal functions fit to 5000 samples from biexponential with $\lambda_1 = 1$ and $\lambda_2 = 15$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	$n_1 = 30, n_2 = 30$			$n_1 = 100, n_2 = 100$		
	perpendicular	vertical	horizontal	perpendicular	vertical	horizontal
ROC-CVM	4.57 ^e	5.82 ^e	8.45 ^c	2.20 ^f	2.49 ^c	4.97 ^f
ROC-MPD	3.67 ^a	4.40 ^b	7.60 ^a	2.10 ^d	2.49 ^e	4.54 ^e
ROC-TLS	7.14 ^f	11.04 ^f	9.74 ^d	2.02 ^b	2.46 ^b	4.27 ^a
FPML	23.16 ^g	11.53 ^g	26.40 ^g	24.07 ^g	11.92 ^g	27.59 ^g
ROC-GLM	4.18 ^c	4.83 ^d	11.65 ^f	2.10 ^c	2.49 ^d	4.54 ^d
ROC-SLS	4.22 ^d	4.70 ^c	10.89 ^e	2.14 ^e	2.64 ^f	4.40 ^c
ROCKIT	3.95 ^b	4.21 ^a	7.92 ^b	2.00 ^a	2.28 ^a	4.30 ^b

Mean squared error and bias of AUC

In estimating the AUC, the semiparametric methods have very little bias (on the order of 10^{-3}) for all simulations with $\lambda_1 = 1$ and $\lambda_2 = 2$. The biases are appreciably larger (on the order of 10^{-2}) for the simulations with $\lambda_1 = 1$ and $\lambda_2 = 2$, in which case they also are all negative except for the empirical. The MSE of the AUC tends to be smaller for $\lambda_1 = 1$ and $\lambda_2 = 15$ than for $\lambda_1 = 1$ and $\lambda_2 = 2$. The variability of the AUC drives the MSE.

Table 5.8: Root mean squared error (RMSE) and bias of AUC estimations from binormal functions fit to 5000 samples from biexponential with $\lambda_1 = 1$ and $\lambda_2 = 2$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	$n_1 = 30, n_2 = 30$		$n_1 = 100, n_2 = 100$	
	RMSE	bias	RMSE	bias
ROC-CVM	7.03 ^g	-0.26 ^h	3.84 ^g	-0.11 ^e
ROC-MPD	6.84 ^c	0.23 ^g	3.61 ^b	0.00 ^a
ROC-TLS	7.20 ^h	-0.22 ^f	3.84 ^h	0.03 ^c
FPML	6.63 ^a	0.01 ^b	3.52 ^a	0.24 ^h
ROC-GLM	6.81 ^b	0.01 ^a	3.70 ^e	0.12 ^f
ROC-SLS	6.89 ^d	-0.09 ^e	3.68 ^d	0.18 ^g
ROCKIT	7.01 ^e	0.06 ^d	3.65 ^c	-0.11 ^d
empirical	7.01 ^f	0.02 ^c	3.77 ^f	-0.01 ^b

5.4.2 Normal-gamma

This section contains measures of error in fitting a binormal ROC function to ROC curves generated from negative observations from the $\mathcal{N}(0, 1)$ distribution and positive observations from a gamma distribution reflected about the vertical axis and translated along the horizontal axis. Please see section 4.3.2 for the precise definition.

Of importance here are two aspects of these distributions. First, the binormal model does not hold exactly because the positive class does not follow a normal distribution. Yet the semi-parametric binormal function estimates generally have small global errors.

Table 5.9: Root mean squared error (RMSE) and bias of AUC estimations from binormal functions fit to 5000 samples from biexponential with $\lambda_1 = 1$ and $\lambda_2 = 15$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	$n_1 = 30, n_2 = 30$		$n_1 = 100, n_2 = 100$	
	RMSE	bias	RMSE	bias
ROC-CVM	5.53 ^h	-2.21 ^h	4.58 ^h	-1.71 ^h
ROC-MPD	5.21 ^b	-1.85 ^b	4.52 ^d	-1.64 ^e
ROC-TLS	5.43 ^f	-2.09 ^f	4.54 ^f	-1.65 ^g
FPML	5.51 ^g	-2.12 ^g	4.49 ^c	-1.50 ^b
ROC-GLM	5.36 ^d	-2.06 ^e	4.52 ^e	-1.59 ^d
ROC-SLS	5.39 ^e	-2.02 ^d	4.47 ^b	-1.53 ^c
ROCKIT	5.31 ^c	-1.98 ^c	4.55 ^g	-1.64 ^f
empirical	3.34 ^a	-0.02 ^a	1.79 ^a	0.01 ^a

Second, one simulation uses a gamma distribution which leads to ROC curves that run through the middle of the plot area, but the other simulation draws from a gamma distribution which produces ROC curves not unlike binormal ROC curves with $1.5 \leq \delta \leq 2$. As we might expect, the ROC-TLS method often generates the smallest global errors for the first case, but other methods often have smaller global errors for the second. These simulations suggest that the ROC-TLS method does best in terms of global error for moderately steep ROC curves.

Function errors

The global squared errors in the perpendicular, vertical, and horizontal directions appear in tables 5.10 and 5.11. The FPML estimator does noticeably worse compared to the semi-parametric methods in the first table. Among the semi-parametric methods the integrated errors are similar, but the ROC-TLS estimator consistently has the smallest values. The errors are largest in the horizontal direction.

In table 5.11, the performance of ROC-TLS degrades in comparison to some of the other semi-parametric methods. A similar result occurred in the simulations with binormal ROCs with the parameter $\delta = 1.5$ and $\delta = 2$. Again, the slight degradation of the ROC-TLS method appears to come with ROC curves that closely follow the plot's boundaries. However, in this case the ROC-MPD

Table 5.10: Root mean global squared errors for binormal functions fit to 5000 samples from normal-gamma ROC with parameters $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 2$, $\beta_2 = 0.5$, $\theta_2 = 5$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	$n_1 = 30, n_2 = 30$			$n_1 = 100, n_2 = 100$		
	perpendicular	vertical	horizontal	perpendicular	vertical	horizontal
ROC-CVM	8.29 ^e	8.54 ^f	13.20 ^f	4.75 ^f	4.88 ^f	7.66 ^f
ROC-MPD	8.11 ^b	8.24 ^c	12.91 ^b	4.54 ^d	4.66 ^d	7.35 ^d
ROC-TLS	7.95 ^a	8.16 ^a	12.64 ^a	4.46 ^a	4.57 ^a	7.19 ^a
FPML	10.55 ^g	10.81 ^g	16.60 ^g	8.04 ^g	8.45 ^g	13.06 ^g
ROC-GLM	8.31 ^f	8.27 ^d	13.12 ^e	4.50 ^c	4.61 ^c	7.30 ^c
ROC-SLS	8.17 ^c	8.31 ^e	12.95 ^c	4.50 ^b	4.61 ^b	7.28 ^b
ROCKIT	8.18 ^d	8.19 ^b	12.98 ^d	4.60 ^e	4.72 ^e	7.45 ^e

estimator seems to have the best performance among the semi-parametric methods.

Table 5.11: Root mean global squared errors for binormal functions fit to 5000 samples from normal-gamma ROC with parameters $\mu_1 = 0$, $\sigma_1 = 3$, $\alpha_2 = 3$, $\beta_2 = 2$, $\theta_2 = 3$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	$n_1 = 30, n_2 = 30$			$n_1 = 100, n_2 = 100$		
	perpendicular	vertical	horizontal	perpendicular	vertical	horizontal
ROC-CVM	5.77 ^e	7.98 ^d	9.07 ^g	4.09 ^g	5.24 ^g	7.38 ^g
ROC-MPD	5.34 ^a	7.77 ^b	8.03 ^a	2.98 ^a	4.68 ^a	4.72 ^b
ROC-TLS	5.85 ^g	8.37 ^f	8.50 ^e	3.61 ^f	5.22 ^e	5.19 ^e
FPML	5.59 ^c	7.60 ^a	8.54 ^f	3.52 ^e	4.88 ^d	5.80 ^f
ROC-GLM	5.45 ^b	7.92 ^c	8.27 ^b	3.01 ^b	4.85 ^b	4.57 ^a
ROC-SLS	5.81 ^f	8.38 ^g	8.48 ^d	3.31 ^d	5.23 ^f	4.91 ^c
ROCKIT	5.67 ^d	8.22 ^e	8.35 ^c	3.28 ^c	4.85 ^c	4.93 ^d

Mean squared error and bias of AUC

Results for estimates of the AUC from simulations in which the positive class follows a gamma distribution appear in tables 5.12 and 5.13. In general, the RMSE of the empirical AUC is smallest. The bias of the empirical AUC also tends to be low. The difference between the results for the empirical

AUC and those from parametric binormal function estimates appears more appreciable in table 5.12. In that case, the gamma distribution of the positive class has parameters $\alpha_2 = 2$, $\beta_2 = 0.5$, and $\theta_2 = 5$, and the resulting ROC curve runs close to the center of the plot. The AUC exhibits more variability, which affects the parametric estimates. The ROC curve produced in the case of a gamma distribution with parameters $\alpha_2 = 3$, $\beta_2 = 2$, and $\theta_2 = 3$ for the positive class reveals less difference in AUC estimates.

Table 5.12: MSE and bias of AUC from binormal functions fit to 5000 samples from normal-gamma ROC with parameters $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 2$, $\beta_2 = 0.5$, $\theta_2 = 5$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	$n_1 = 30, n_2 = 30$		$n_1 = 100, n_2 = 100$	
	RMSE	bias	RMSE	bias
ROC-CVM	7.81 ^c	-0.99 ^b	4.80 ^b	-1.18 ^e
ROC-MPD	7.91 ^d	-1.48 ^g	4.92 ^e	-1.21 ^f
ROC-TLS	8.16 ^h	-1.36 ^f	4.85 ^c	-1.05 ^c
FPML	8.04 ^f	-1.05 ^c	4.89 ^d	-0.95 ^b
ROC-GLM	7.64 ^b	-1.21 ^e	5.05 ^h	-1.49 ^g
ROC-SLS	8.09 ^g	-1.20 ^d	5.02 ^g	-1.54 ^h
ROCKIT	8.04 ^e	-1.49 ^h	4.94 ^f	-1.05 ^d
empirical	7.40 ^a	-0.28 ^a	4.13 ^a	-0.07 ^a

5.5 Bootstrap confidence regions

If a binormal ROC model appears suitable, then a confidence region allows for additional inference about the binormal parameters ρ and δ . We present multivariate studentized confidence regions for the binormal parameters developed with a parametric bootstrap (Davison and Hinkley, 1997; Wasserman, 2005). Simulations suggest that their coverage probabilities closely approximate the nominal confidence levels.

Table 5.13: MSE and bias of AUC from binormal functions fit to 5000 samples from normal-gamma ROC with parameters $\mu_1 = 0$, $\sigma_1 = 3$, $\alpha_2 = 3$, $\beta_2 = 2$, $\theta_2 = 3$. All values are multiplied by 100. The letters in superscript indicate the rank within a column.

method	$n_1 = 30, n_2 = 30$		$n_1 = 100, n_2 = 100$	
	RMSE	bias	RMSE	bias
ROC-CVM	4.72 ^a	-0.16 ^b	2.83 ^c	0.10 ^e
ROC-MPD	4.86 ^e	-0.35 ^e	2.86 ^d	0.14 ^f
ROC-TLS	4.81 ^d	-0.40 ^f	2.92 ^f	0.00 ^a
FPML	4.80 ^c	0.06 ^a	2.95 ^h	0.01 ^b
ROC-GLM	5.02 ^g	-0.48 ^g	2.91 ^e	-0.14 ^h
ROC-SLS	5.16 ^h	-0.52 ^h	2.93 ^g	-0.14 ^g
ROCKIT	4.90 ^f	-0.22 ^d	2.70 ^b	0.05 ^c
empirical	4.73 ^b	-0.19 ^c	2.58 ^a	-0.06 ^d

5.5.1 Development

Let $\hat{\theta} = (\hat{\rho}, \hat{\delta})$ be the vector of the initial point estimates of the parameters ρ and δ for a binormal ROC curve produced with negative and positive samples of sizes n_1 and n_2 . Denote the number of bootstrap samples by R . Over $i = 1, \dots, R$, the i^{th} bootstrap sample consists of observations u_{ij}^* from the negative class with $j = 1, \dots, n_1$ and observations v_{ik}^* from the positive class with $k = 1, \dots, n_2$. These primary bootstrap samples, which carry an asterisk as a superscript, come from the $\mathcal{N}(0, 1)$ distribution for the negatives and the $\mathcal{N}(\hat{\delta}/\hat{\rho}, 1/\hat{\rho}^2)$ distribution for the positives. The bootstrap is a parametric one that uses the binormal ROC model with parameters $\hat{\rho}$ and $\hat{\delta}$.

The primary bootstrap samples lead to bootstrap estimates $\hat{\theta}_i^* = (\hat{\rho}_i^*, \hat{\delta}_i^*)$ as well as a bootstrap covariance estimate V^* (without a subscript). To arrive at each bootstrap estimate $\hat{\theta}_i^* = (\hat{\rho}_i^*, \hat{\delta}_i^*)$, we apply our method of parameter estimation to the observations u_{ij}^* and v_{ik}^* . The bootstrap covariance estimate V^* follows from

$$V^* = \frac{1}{R} \sum_{i=1}^R (\hat{\theta}_i^* - \bar{\theta}_i^*)' (\hat{\theta}_i^* - \bar{\theta}_i^*) \quad (5.3)$$

where $\bar{\hat{\theta}}_i^*$ is the mean of the $\hat{\theta}_i^*$. The quantity

$$Q = (\hat{\theta} - \theta)' V^{-1} (\hat{\theta} - \theta) \quad (5.4)$$

approximately follows a χ_2^2 distribution. So percentiles of the χ_2^2 distribution may serve to form an approximate confidence region:

$$\{\theta : (\hat{\theta} - \theta)' V^{-1} (\hat{\theta} - \theta) \leq \chi_{2,1-\alpha}^2\}. \quad (5.5)$$

A more theoretically accurate, but computationally intensive confidence region follows from approximating the quantiles by

$$Q_i^* = (\hat{\theta}_i^* - \hat{\theta})' V_i^{*-1} (\hat{\theta}_i^* - \hat{\theta}) \quad (5.6)$$

where V_i^* (with a subscript) is the bootstrap covariance estimate of the primary bootstrap estimate $\hat{\theta}_i^*$. To compute V_i^* , we must perform a secondary parametric bootstrap using $\hat{\theta}_i^* = (\hat{\rho}_i^*, \hat{\delta}_i^*)$ as the initial estimates. We draw another R samples, in each of which n_1 negative observations come from the $\mathcal{N}(0, 1)$ distribution and n_2 positive observations come from the $\mathcal{N}[\hat{\delta}_i^*/\hat{\rho}_i^*, (1/\hat{\rho}_i^*)^2]$ distribution. For each of the secondary bootstrap samples for the i^{th} primary bootstrap estimate, we obtain secondary bootstrap estimates $\hat{\theta}_{i\ell}^* = (\hat{\rho}_{i\ell}^*, \hat{\delta}_{i\ell}^*)$ with $\ell = (1, \dots, R)$, where the star as a superscript distinguishes the secondary from the primary bootstrap estimates. Now the bootstrap covariance estimate V_i^* of $\hat{\theta}_i^*$ follows from

$$V_i^* = \frac{1}{R} \sum_{\ell=1}^R (\hat{\theta}_{i\ell}^* - \hat{\theta}_i^*)' (\hat{\theta}_{i\ell}^* - \hat{\theta}_i^*). \quad (5.7)$$

Figure 5.1 charts the calculation of the parametric studentized bootstrap confidence region.

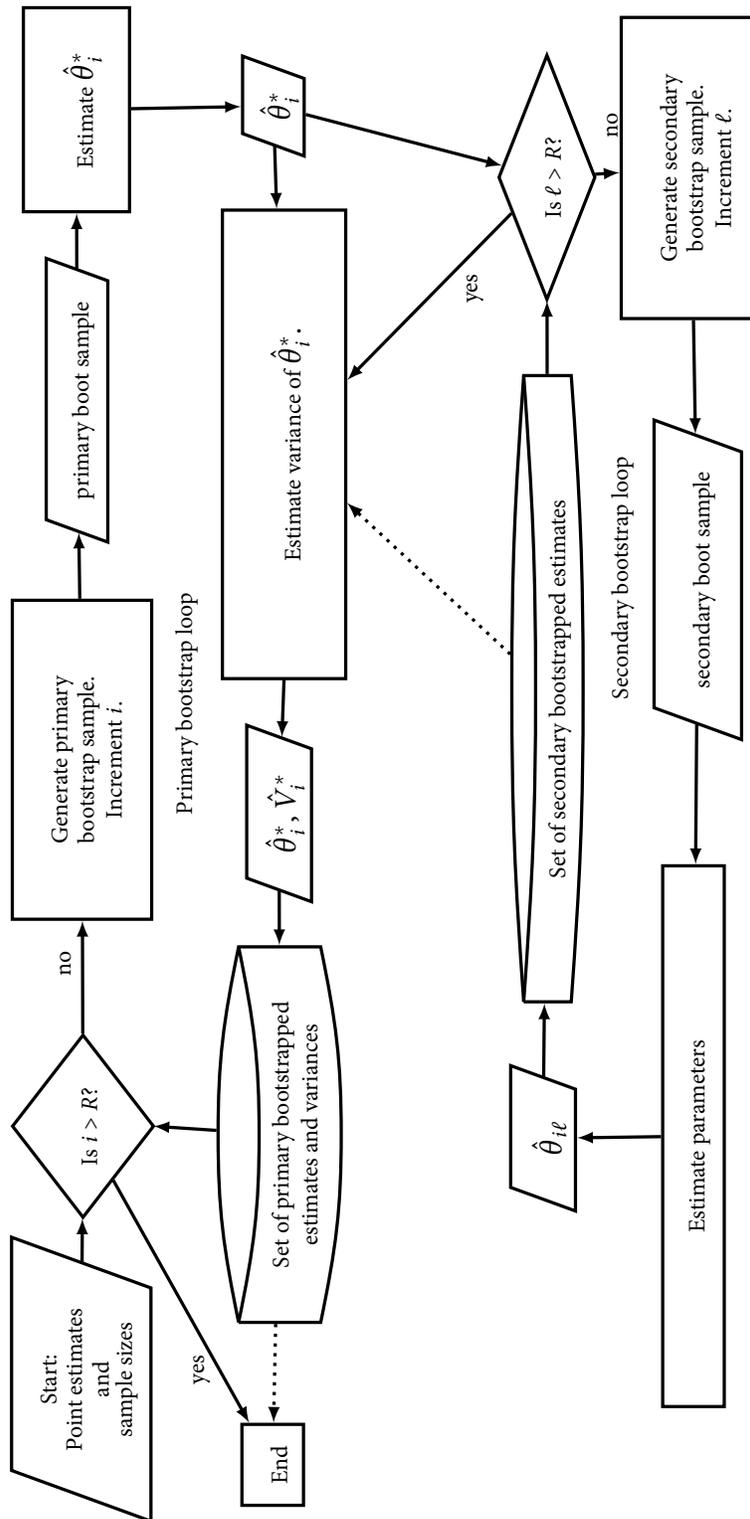


Figure 5.1: Design of calculation of parametric studentized bootstrap confidence regions.

5.5.2 Coverage probabilities

Computer simulations give insight into the the new parametric studentized bootstrap confidence intervals and regions for the binormal ROC. For a given pair of binormal parameters, we conducted five runs, which each consisted of 1000 simulations. In a single simulation, we drew samples each of size 100 from the negative and positive distributions and then calculated parametric studentized bootstrap confidence intervals and regions at the 95% level using $R = 999$ primary bootstrap samples.

Table 5.14 lists the coverage probabilities for the different confidence intervals and regions calculated in the simulations. The fourth and fifth columns give the coverage probabilities of the univariate parametric studentized bootstrap confidence interval. These simulations imply that the univariate confidence intervals for the binormal parameters ρ and δ may be slightly liberal.

The last two columns contain coverage probabilities for two forms of multivariate confidence regions. In the first, we checked whether the univariate confidence intervals for the parameters ρ and δ both included the true values. Such an approach neglects the covariance of the parameters and leads to coverage probabilities around 90%. In contrast, our new simultaneous confidence region has a coverage probability quite close to the nominal 95% level as indicted in the final column.

Table 5.14: Coverage probabilities for 95% confidence regions of ρ and δ calculated over 1000 samples each with $n_1 = 100$ negative observations and $n_2 = 100$ positive observations.

ρ	δ	run	ρ CI	δ CI	ρ and δ CI	simultaneous CR
1	0	1	93.40	94.50	88.50	95.20
		2	94.80	94.90	89.70	95.00
		3	93.20	94.30	87.70	95.30
		4	93.50	95.00	88.90	95.80
		5	92.80	95.20	88.20	95.10
		mean	93.54	94.78	88.60	95.28
1	0.5	1	93.60	94.70	88.90	95.10
		2	95.10	95.60	91.00	96.00
		3	94.00	94.00	88.40	95.50
		4	94.20	95.90	90.30	96.00
		5	92.90	94.90	88.10	94.10
		mean	93.96	95.02	89.34	95.34
1	1	1	92.60	95.40	89.10	95.10
		2	94.50	94.30	89.70	95.50
		3	94.80	94.20	89.90	95.10
		4	94.50	95.30	90.30	95.80
		5	93.90	93.90	88.50	94.40
		mean	94.06	94.62	89.50	95.18
1	1.5	1	91.50	93.10	86.90	94.40
		2	93.70	93.00	88.70	94.90
		3	95.30	93.00	89.70	94.90
		4	94.60	94.40	90.20	96.00
		5	93.20	93.10	88.20	94.30
		mean	93.66	93.32	88.74	94.90
0.5	0.5	1	94.30	95.10	90.00	94.60
		2	96.30	95.20	91.80	95.80
		3	95.90	94.70	91.00	95.70
		4	95.60	96.10	91.70	95.60
		5	94.50	95.10	90.40	93.80
		mean	95.32	95.24	90.98	95.10

5.6 Application

5.6.1 ROC curve for audiology example

To apply the new ROC-TLS estimator and parametric studentized bootstrap confidence region, we consider a set of audiology data which Pepe (2004) also uses. Stover et al. (1996) originally presents the data, which consist of signal-to-noise response ratios of distortion product otoacoustic emissions (DPOAE). As Pepe (2004) does, we use the data for a stimulus intensity of 65 dB and stimulus frequency of 2001 Hz. The negative signal-to-noise response ratio from the test serves as a score. We

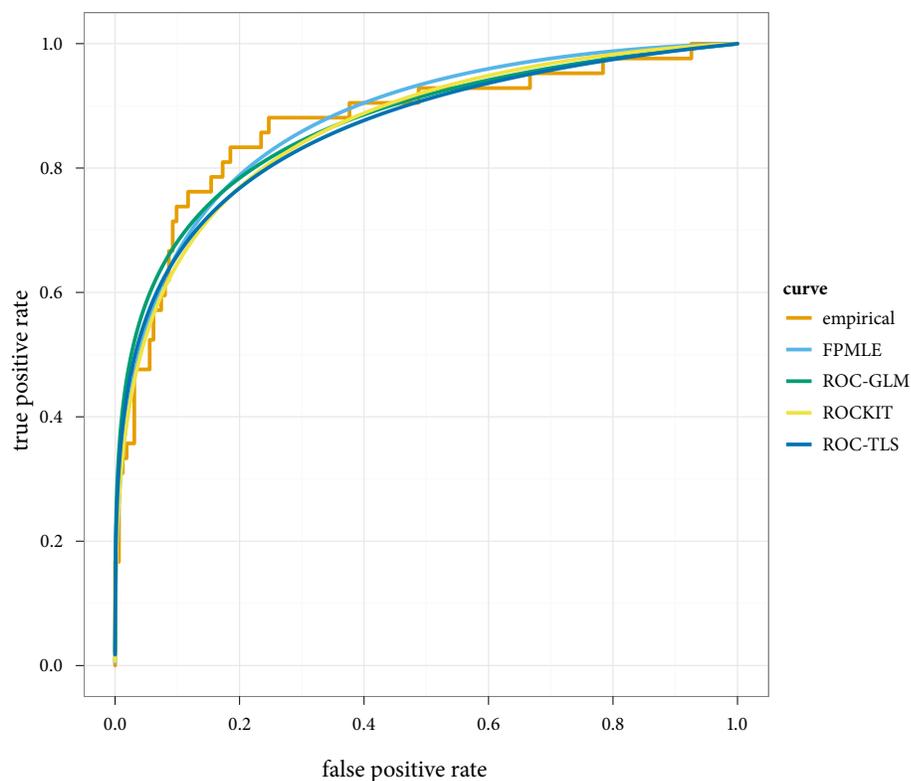


Figure 5.2: ROC function estimates for audiology data discussed by Pepe (2004) and Stover et al. (1996).

define the negative and positive classes based on whether an observation has hearing loss. The true class of an observation follows from a gold standard behavioral test. In this set, we have 103 observations from the negative class (normal hearing) and 107 observations from the positive class (impaired hearing).

We applied four different binormal ROC function estimators to the data. Figure 5.2 plots the empirical and parametric ROC curves. Table 5.15 gives the ROC-PD goodness-of-fit statistic and p-values estimated from a parametric bootstrap of the test statistic for each function estimate. None of the p-values suggests that we reject the fitted binormal ROC curves.

Table 5.15: Parameters estimates and ROC-PD goodness-of-fit tests for audiology data.

method	ρ	δ	ROC-PD	p-value
FPMLE	0.863	1.527	0.0005904	0.415
ROCKIT	0.819	1.425	0.0007469	0.241
ROC-GLM	0.714	1.386	0.0005723	0.299
ROC-TLS	0.726	1.342	0.0007996	0.198

We proceed to compute parametric studentized bootstrap confidence regions for ρ and δ with our implementation discussed in the previous section. Figure 5.3 depicts the confidence ellipses at 90%, 95%, and 99% levels.

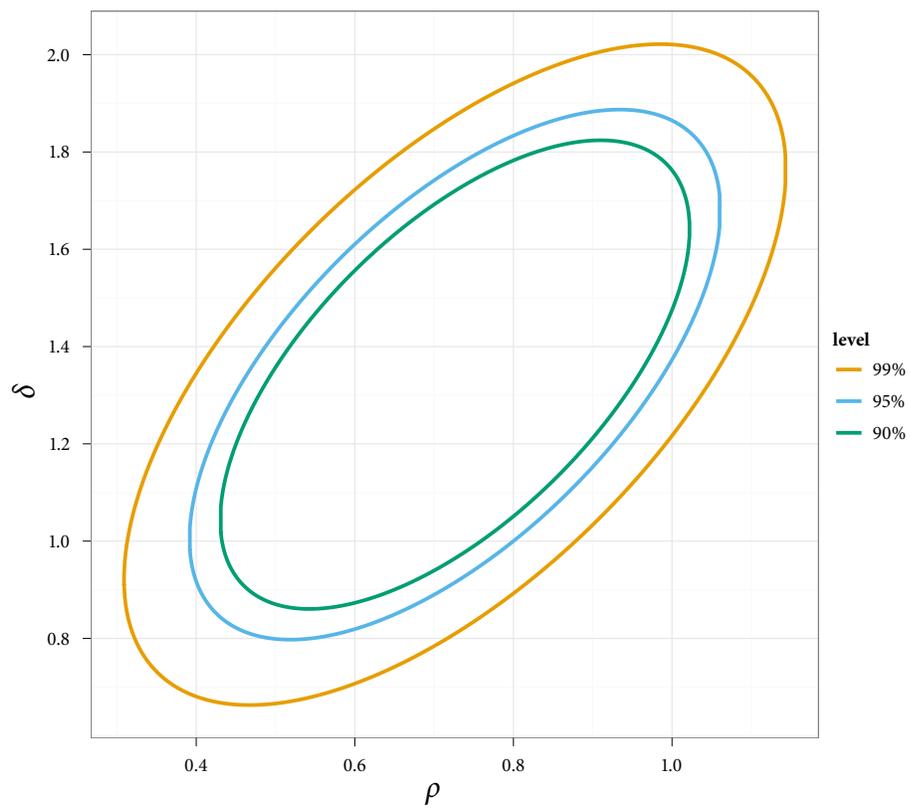


Figure 5.3: Simultaneous confidence regions from parametric studentized bootstrap for binormal ROC parameters.

Chapter 6: Conclusion

Overall, the novelty of this research results from considering perpendicular distances between empirical and parametric receiver operating characteristic (ROC) curves. This new approach has led to several new contributions to statistical inference with ROC curves, particularly for the parametric binormal model. First, we have developed a new goodness-of-fit test with power to reject the binormal model estimated from small samples with a fully parametric approach and from large samples with a semi-parametric approach. The test also has the capability to reject the binormal model estimated from small samples with a semi-parametric approach if one of the underlying distributions consists of a mixture or has heavy tails. Second, we have created two new estimators for binormal ROC functions. One of them appears to have smaller integrated squared errors than existing approaches for binormal ROC curves that are not too steep. Third, we designed a parametric studentized bootstrap confidence region for the binormal ROC curve's parameters with coverage probabilities near nominal levels. This bootstrap requires heavy computation, which runs quickly because the new estimator follows from a closed-form calculation.

Before detailing the contributions further, we wish to consider their context. In general, this research has sought to improve a technique for assessing the performance of classification methods. As new methods of classification arise and as new applications of current methods emerge, the ROC curve serves as a valuable tool in studying performance. The ROC curve can help to identify methods that discern classes better as well as variables that allow better discrimination between classes.

Our focus has been on parametric models of ROC curves. The advantages of a parametric over a non-parametric model of an ROC curve are twofold. First, the parametric model facilitates interpretation and use. For instance, reading error rates from or calculating a functional (such as an integral) of a parametric ROC curve tends to be easier. Second, tests based on parametric models tend to be more powerful than those based on nonparametric models. Furthermore, we may believe

that a smooth, parametric function exists to describe the ROC curve if the underlying classes have continuous distributions.

We have concentrated specifically on the parametric ROC function of the binormal model. The binormal model primarily assumes that there exists an unspecified, increasing transformation, under which the scores of the negative and positive classes follow normal distributions. While the two underlying normal distributions of the negative and positive classes have four parameters in total, the binormal model itself effectively has two parameters.

Usually, we do not know the true parameters of the binormal ROC function and instead estimate them. A given estimator results in an estimate of the binormal ROC function. Before we proceed to apply the function estimate, checking its fit to observed data is prudent.

Below we reiterate the key aspects of our contributions to performing inference with ROC curves.

6.1 Goodness-of-fit statistic for binormal ROC

We developed an original test of a binormal ROC function's goodness-of-fit to an empirical ROC function based on the perpendicular distance between the two curves. Before we realized the advantages of measuring distance in the perpendicular distance, we began with two interesting aspects of ROC curves. First, an ROC curve is a cumulative distribution function (CDF). Second, the empirical ROC curve is a consistent estimator for the true ROC curve.

Given these two results from the literature, we hoped that the Cramér-von Mises family of goodness-of-fit statistics, based on the empirical distribution function and estimated CDF for a single random variable, would apply to ROC curves. We derived the Cramér-von Mises statistic for ROC curves and showed that the Anderson-Darling statistic was not defined. In studying the Cramér-von Mises statistic for ROC curves under a simple null hypothesis, we learned that the distribution switches from right-skewed to left-skewed as the ROC curve becomes steeper. This change in skewness would not be well suited for a one-sided goodness-of-fit statistic. So we established the inadequacies of the Cramér-von Mises and Anderson-Darling statistics for ROC curves.

Our attention turned to measuring the perpendicular distance between the empirical and parametric ROC curves partly because large vertical gaps between the curves seemed to contribute to the Cramér-von Mises statistic's instability. The calculation of distance in the perpendicular direction results in more regular values. Of no less importance, measuring the perpendicular distance between the empirical and parametric ROC curves gives balanced weight to false negatives and false positives. Furthermore, because the new goodness-of-fit statistic remains based on the empirical ROC function, it results in a consistent test.

The new goodness-of-fit statistic, referred to as the ROC-PD (perpendicular distance) statistic, requires computation. Simulations characterize the distribution of the statistic and a parametric bootstrap produces p-values. The statistic itself has a right-skewed null distribution that depends on the binormal parameters. The null distributions of p-values all appear uniform.

Against alternative hypothesis, the ROC-PD statistic seems to have the most power of current goodness-of-fit tests for ROC curves. The ROC-PD statistic shows high power in at least two situations. First, the ROC-PD statistic may be able to reject the null hypothesis of a binormal model for moderate sample sizes if one of the underlying classes has a distinct bimodal distribution or heavy tails. Second, for sample sizes close to 1,000 or larger, the ROC-PD statistic may be able to reject the binormal model despite a close fit visually.

6.2 Estimators for binormal ROC

In addition to the goodness-of-fit statistic, we developed two new estimators of binormal ROC functions. The ROC-PD statistic measures distance between the empirical and parametric ROC curves. So a natural estimator of the parametric curve is the binormal ROC function that minimizes this perpendicular distance. The resulting ROC-MPD (minimum perpendicular distance) estimator depends on a numerical optimization routine to find the binormal ROC function that minimizes the ROC-PD goodness-of-fit statistic.

The second estimator also minimizes a perpendicular distance, but after applying a straightforward transformation. By taking the probit of the binormal ROC function, we observe that the resulting function is a line with a slope and intercept equal to the parameters of the binormal ROC. After the probit transformation, principal component analysis (PCA) gives the line with the minimum perpendicular distance to points from the empirical ROC. We call this the ROC-TLS (total least squares) estimator. In simulations, the ROC-TLS function estimates have the smallest mean integrated squared error for curves that are not too steep. Also, the ROC-TLS estimator requires nothing more than linear algebra and works very quickly.

6.3 Parametric studentized bootstrap confidence regions

With a short computational time, the ROC-TLS estimator permits a new application of the parametric studentized bootstrap to obtain confidence intervals and regions for the binormal parameters. The studentized bootstrap requires secondary bootstrap samples to estimate the variances of the estimates from the original bootstrap samples. For example, if the desired number of bootstrap samples is $R = 1000$, then the studentized bootstrap uses $R^2 = 10^6$ samples in all. A compiled computer language on a modern desktop can process such a calculation to give confidence intervals and regions in under one minute using the ROC-TLS estimator for the binormal ROC. Additionally, the coverage probabilities closely agree with the nominal levels in simulations.

6.4 Future research

Several subjects came up in this research that would be interesting to pursue further. Investigation into the weight functions for ROC-PD goodness-of-fit statistic could start immediately. Weights that have inverse relationships with the ROC curve's variability, such as those proposed in Section 3.2.5, should improve the test's efficiency and increase its power. Also, extension of the ROC-TLS estimator to a general regression framework could contribute to the study of covariate effects on ROC curves.

Broadening the approaches developed here for binormal ROC curves to other parametric functions of the ROC curve would be intriguing as well. Comparing a goodness-of-fit statistic implemented for the parametric biexponential ROC to the ROC-PD statistic could lead to a way to select the better parametric form, particularly for small sample sizes. Additionally, the ROC-TLS estimator could apply to other parametric ROC curves with the right modifications. The changes could lead to new confidence intervals and regions for the relevant parameters by using the parametric studentized bootstrap techniques developed for the binormal parameters.

Appendix A: Results of simulations for goodness-of-fit tests

This appendix contains all of the results of the simulations to characterize the goodness-of-fit tests for binormal ROC curves.

A.1 Binormal

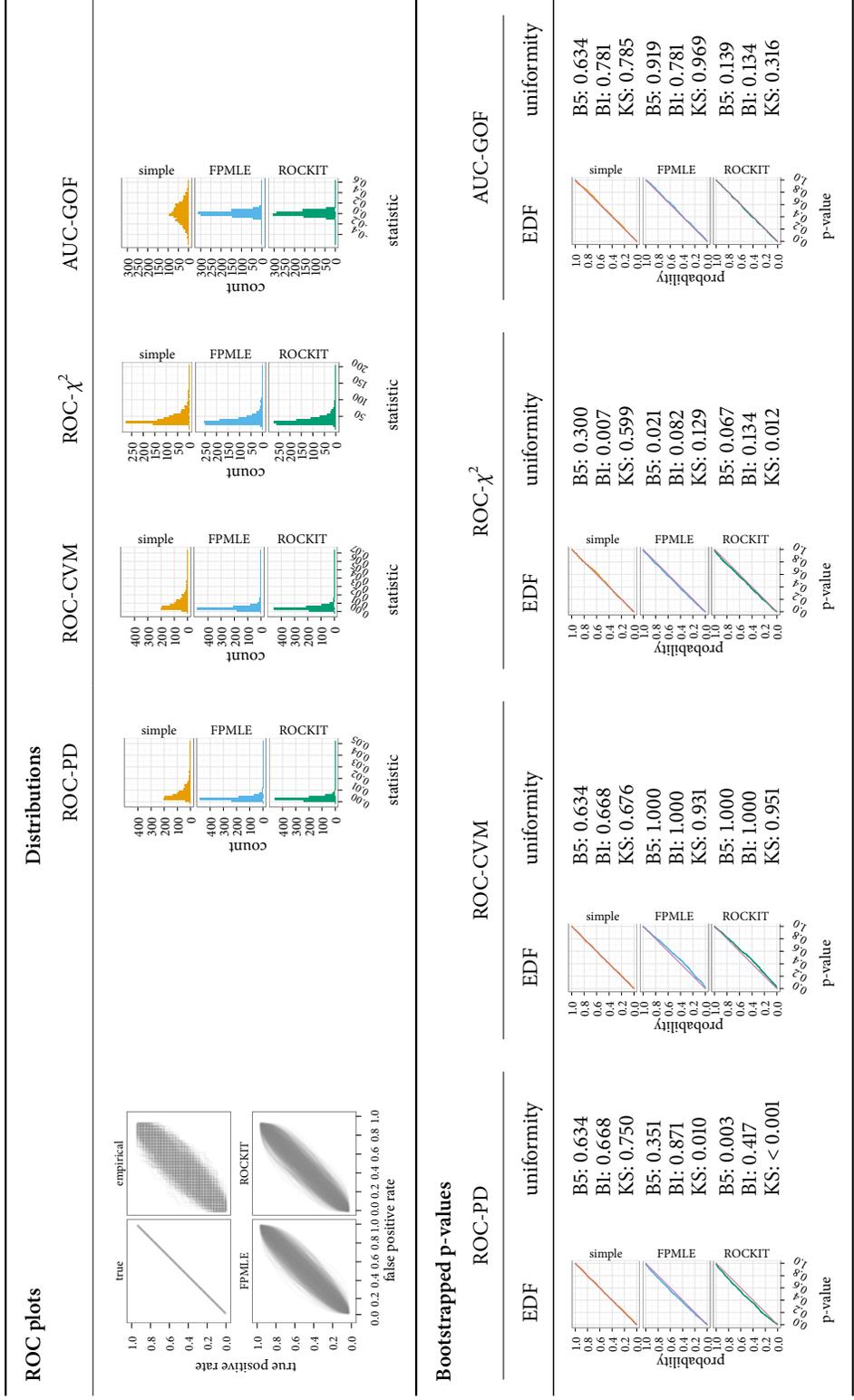


Figure A.1: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1, \delta = 0, n_1 = 30, n_2 = 30$.

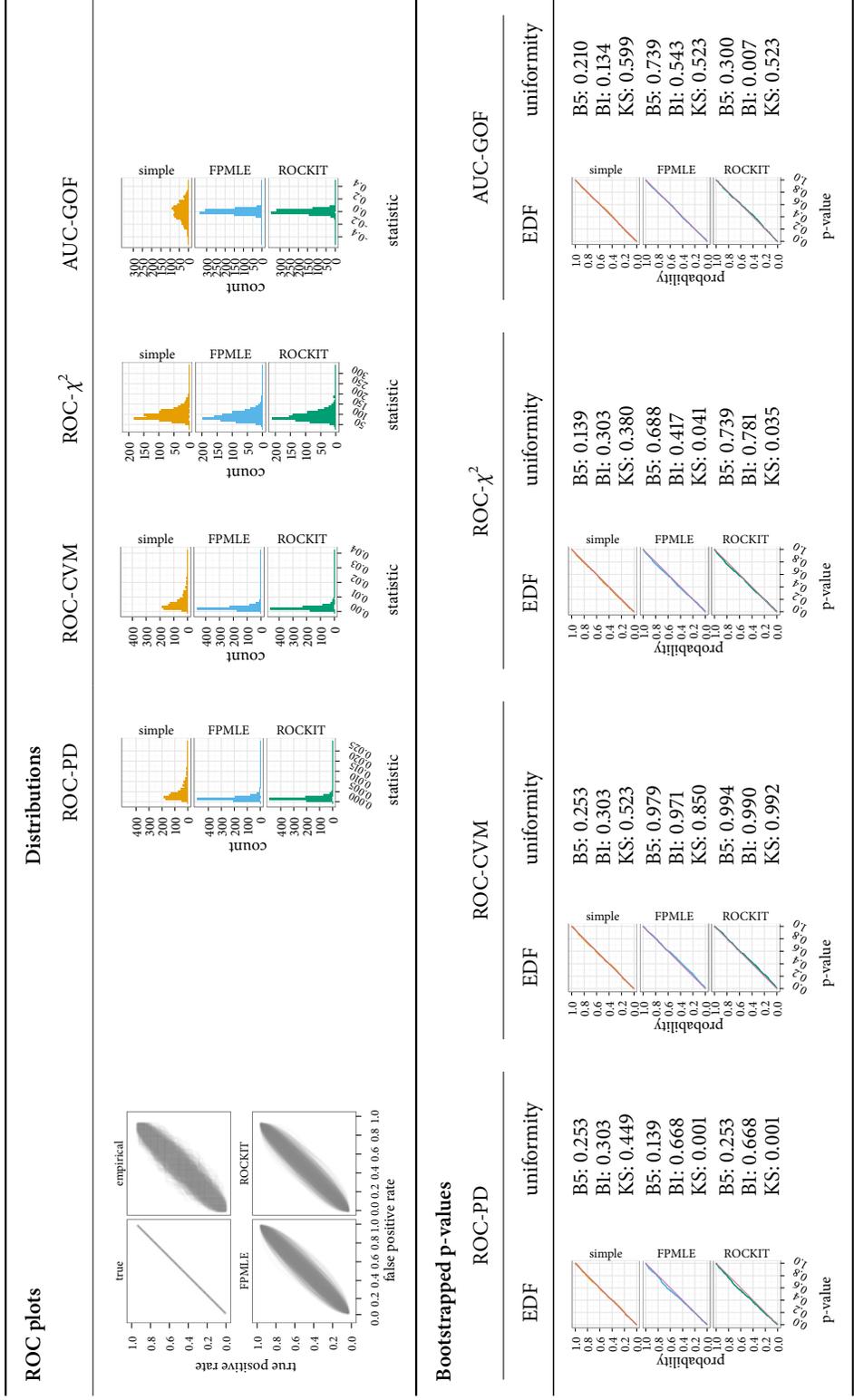


Figure A.2: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1, \delta = 0, n_1 = 50, n_2 = 50$.

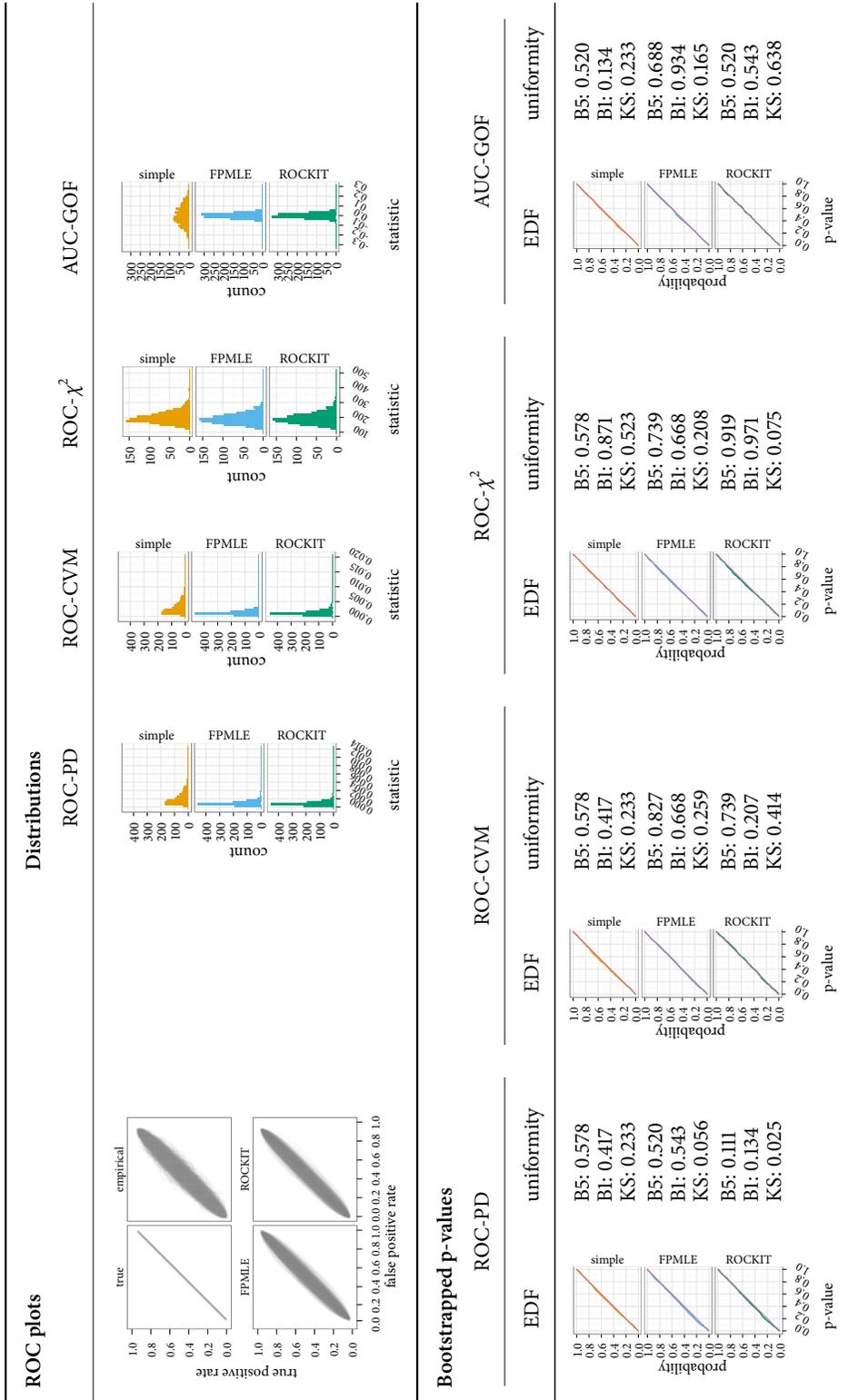


Figure A.3: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0$, $n_1 = 100$, $n_2 = 100$.

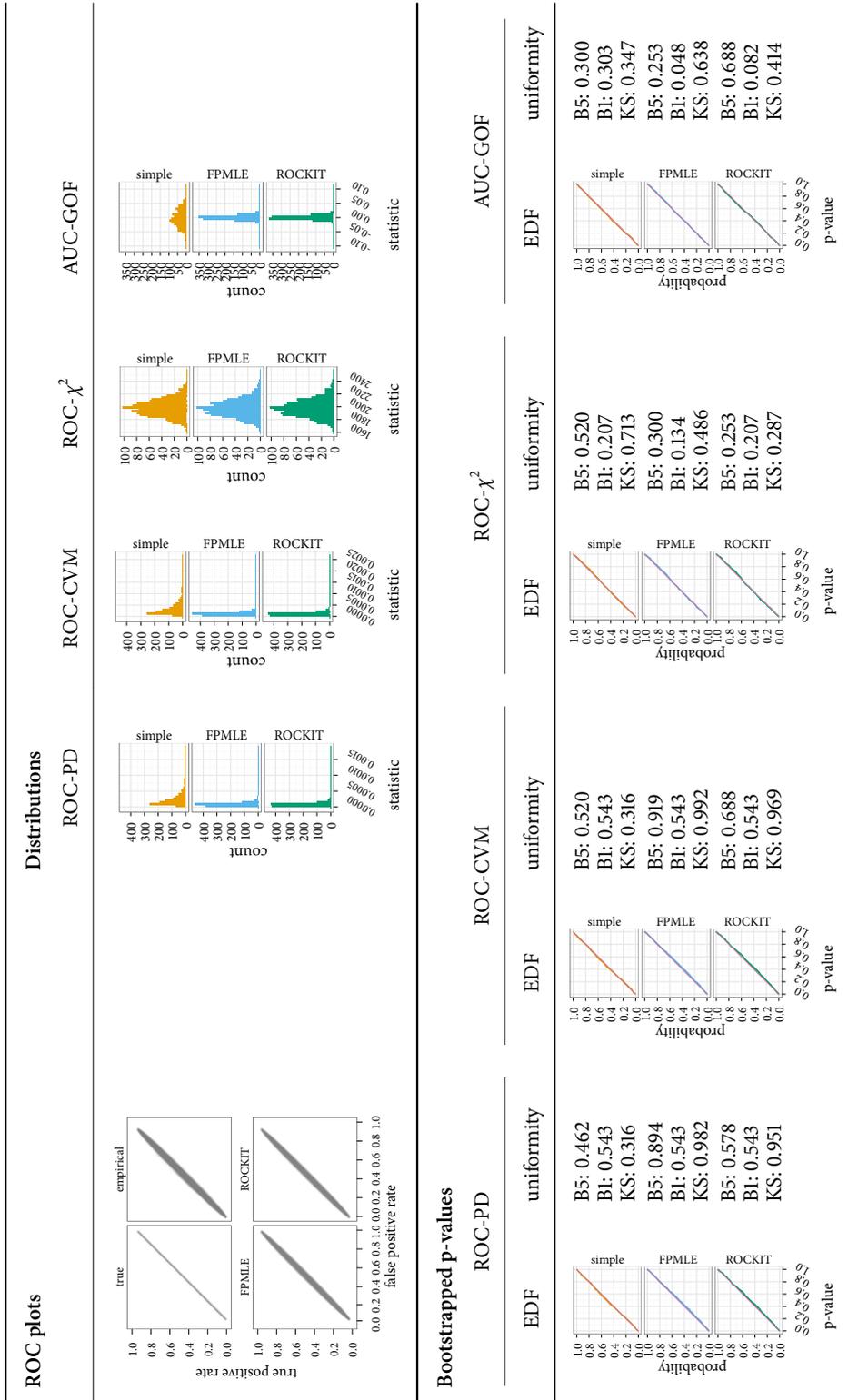


Figure A.4: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0$, $n_1 = 1000$, $n_2 = 1000$.

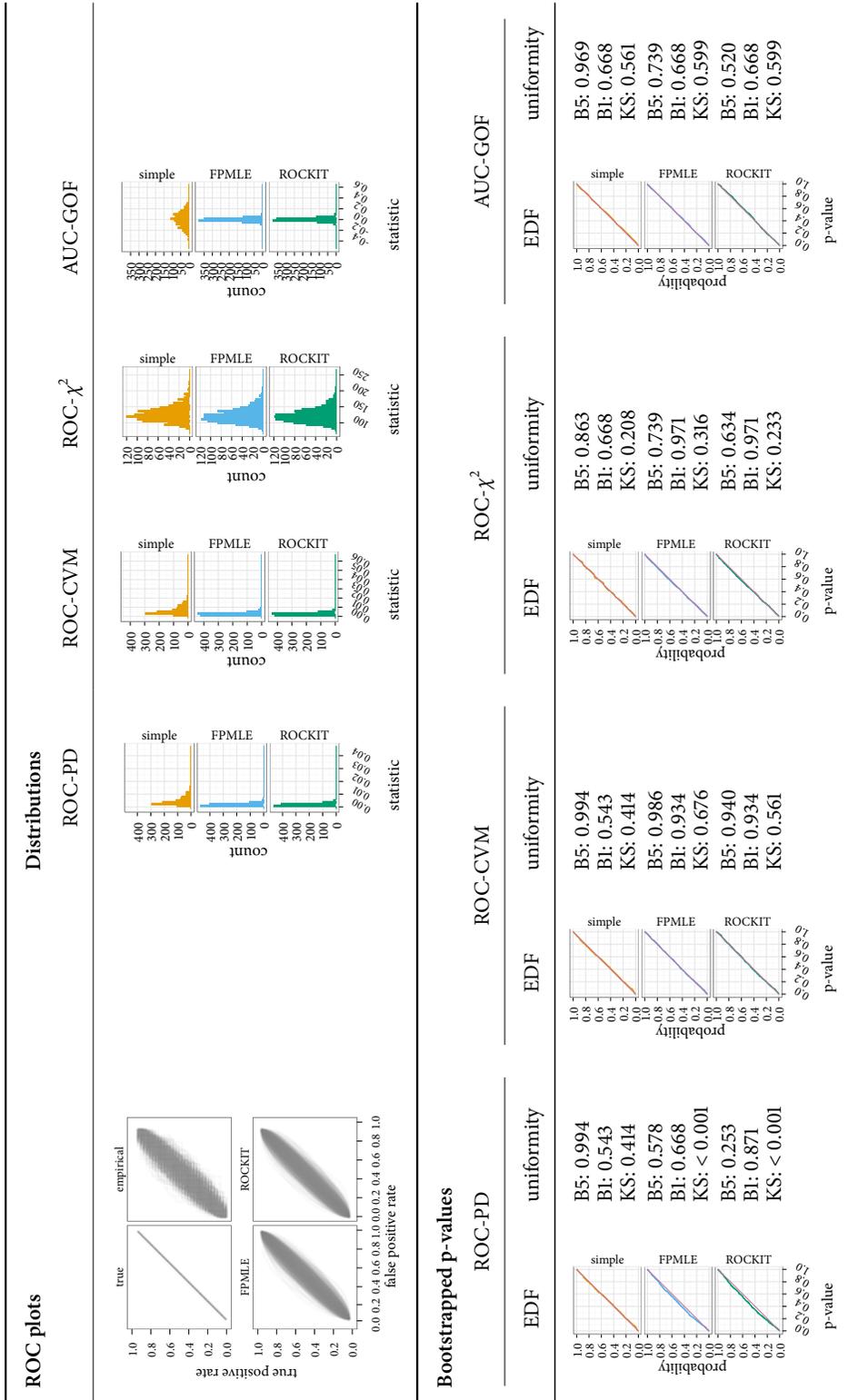


Figure A.5: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0$, $n_1 = 100$, $n_2 = 30$.

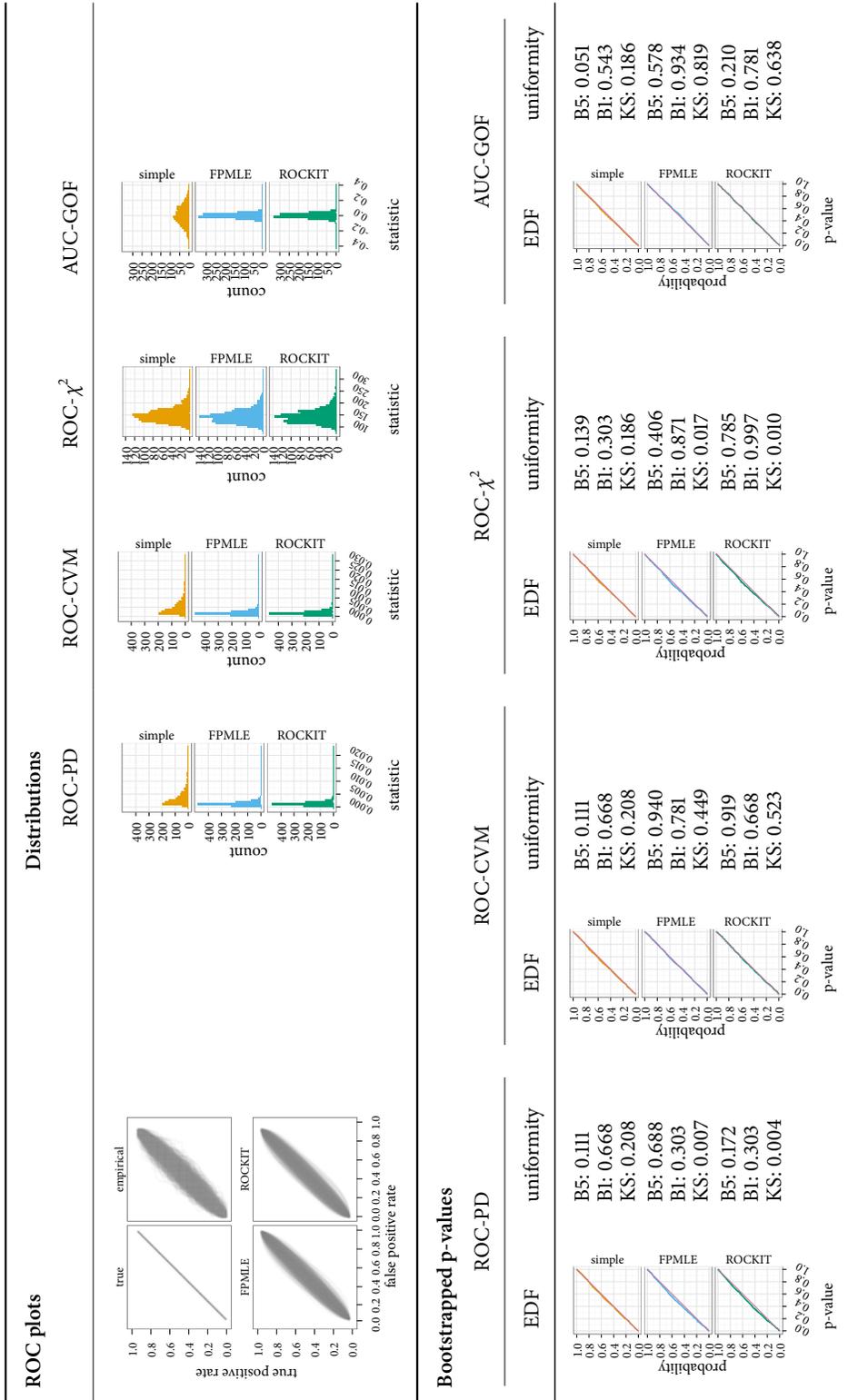


Figure A.6: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0$, $n_1 = 100$, $n_2 = 50$.

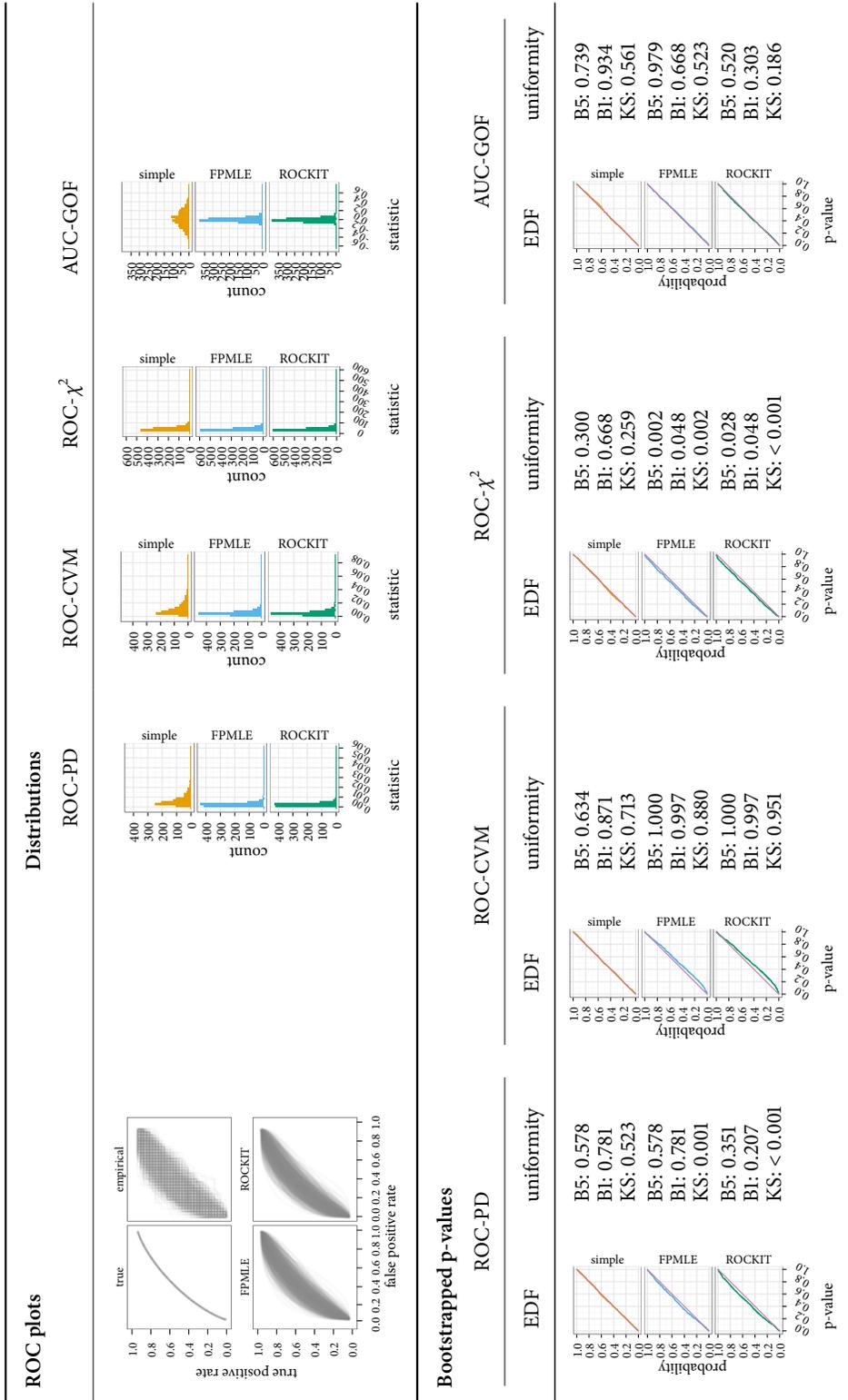


Figure A.7: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0.5$, $n_1 = 30$, $n_2 = 30$.

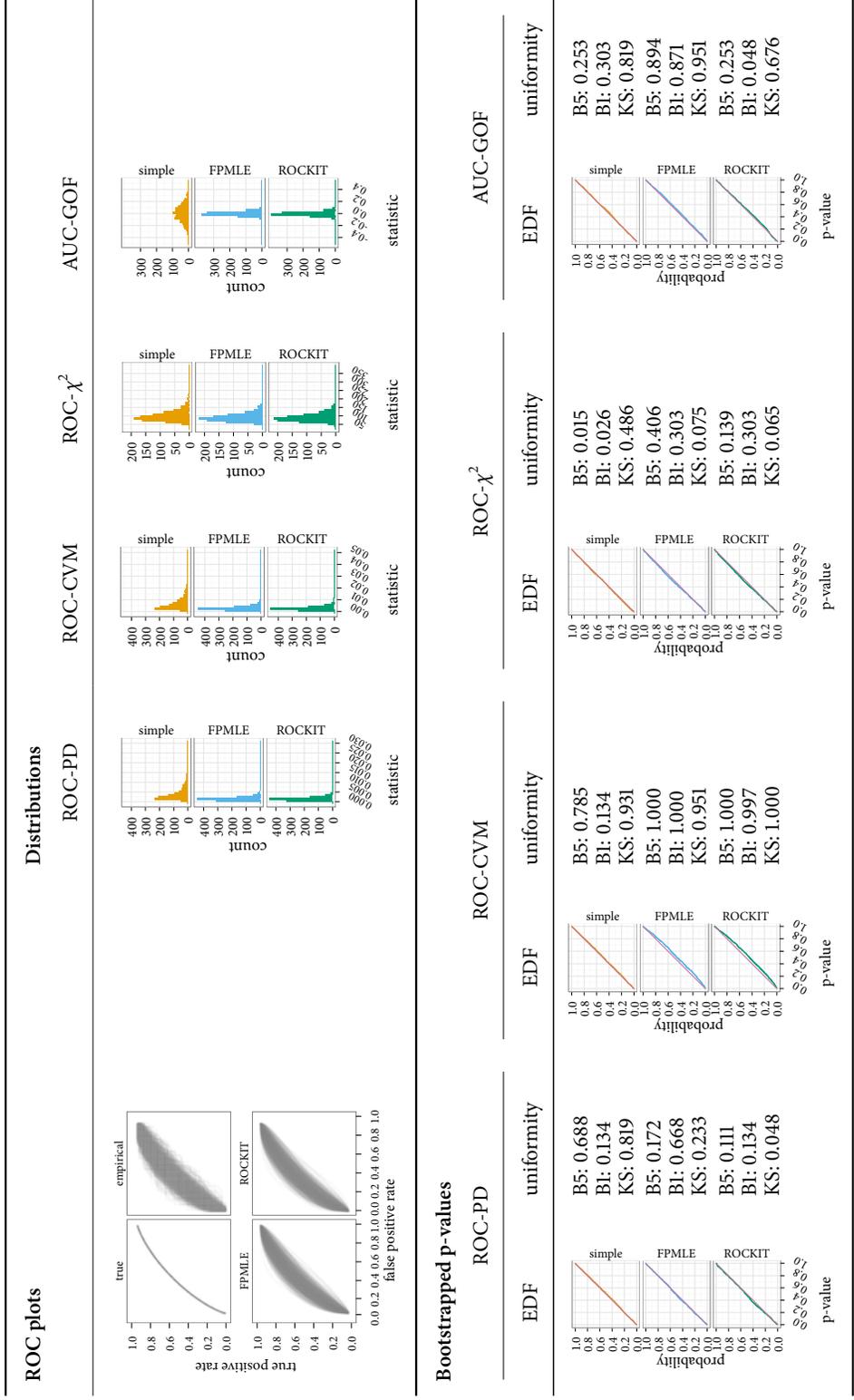


Figure A.8: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0.5$, $n_1 = 50$, $n_2 = 50$.

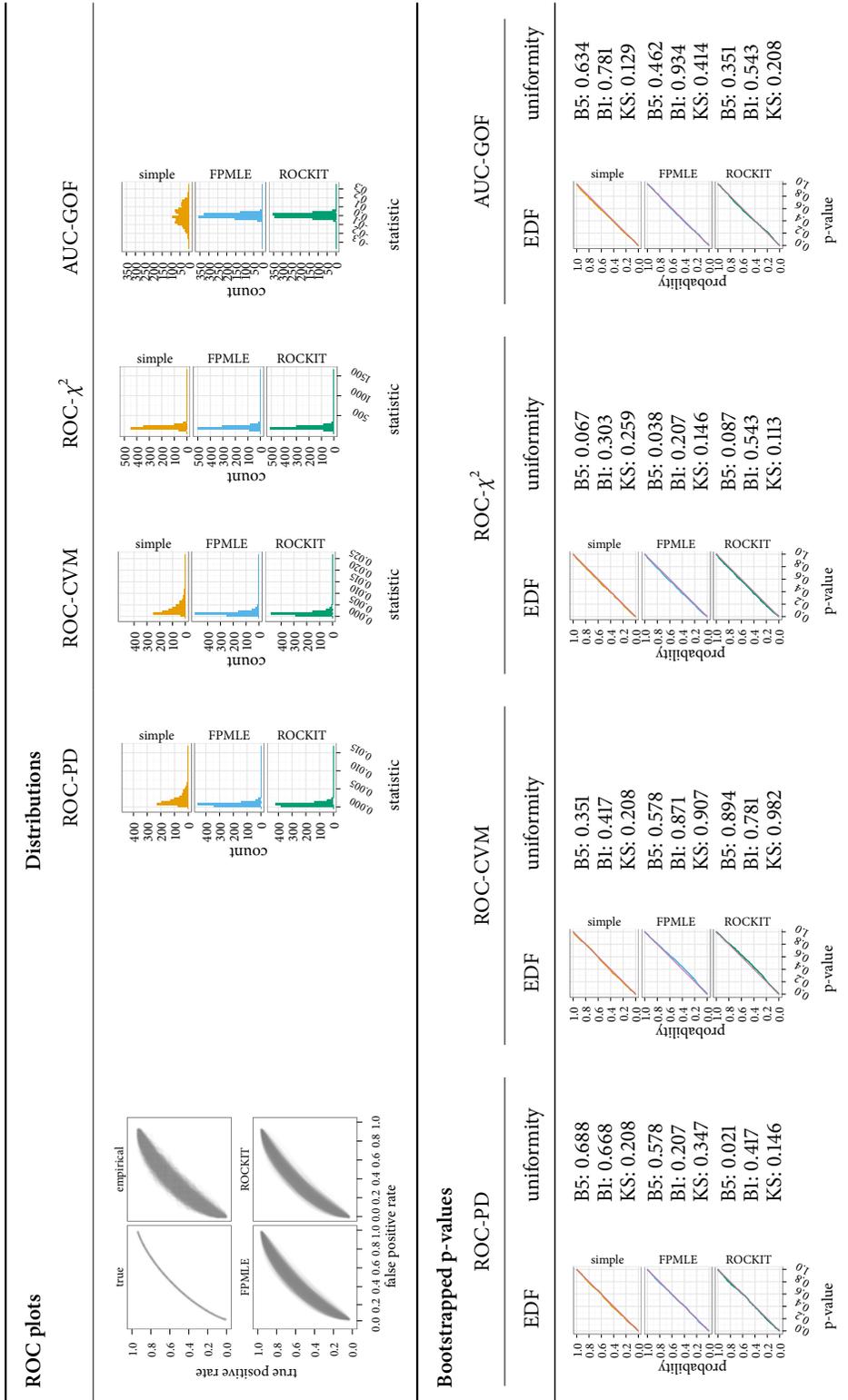


Figure A.9: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0.5$, $n_1 = 100$, $n_2 = 100$.

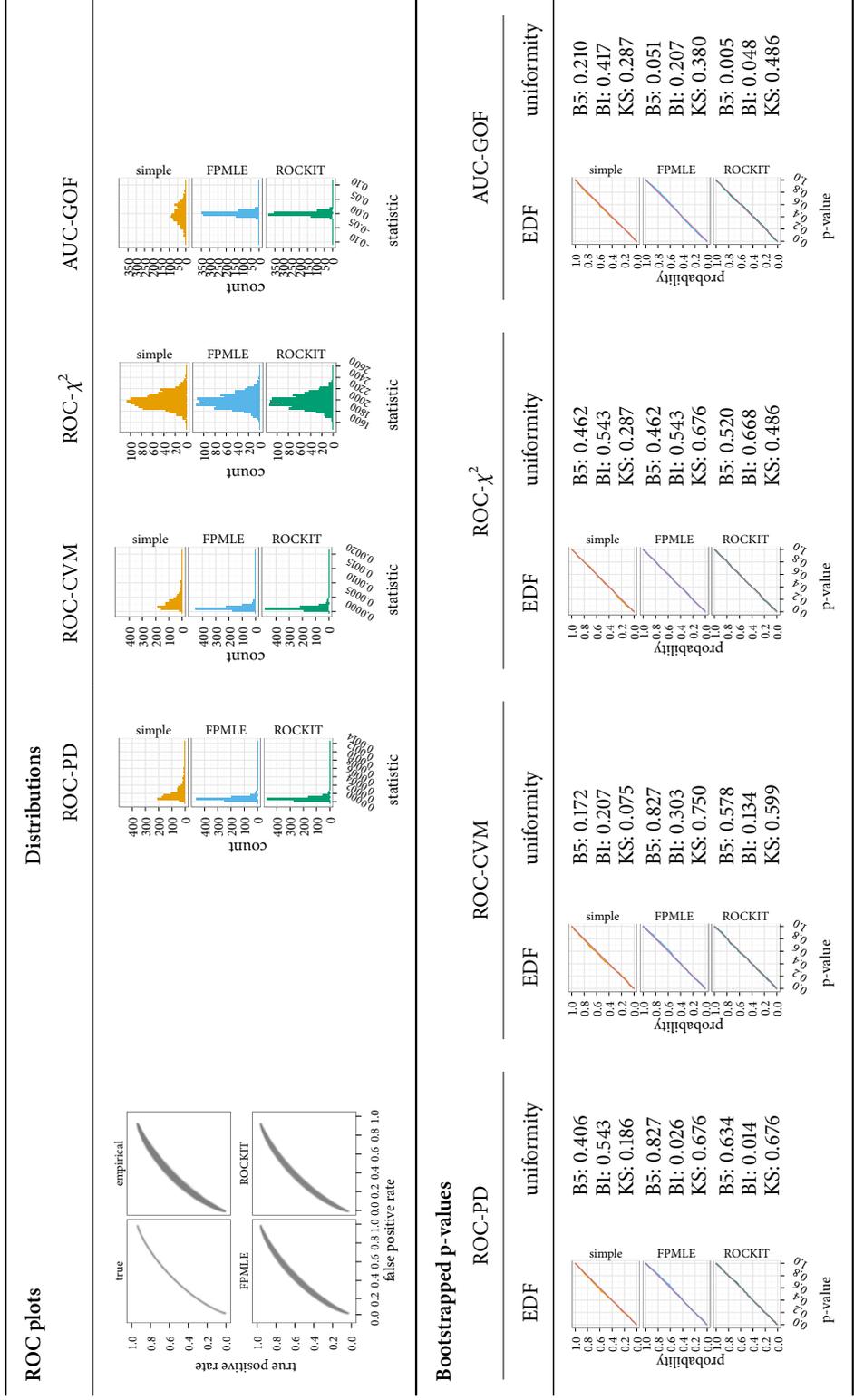


Figure A.10: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0.5$, $n_1 = 1000$, $n_2 = 1000$.

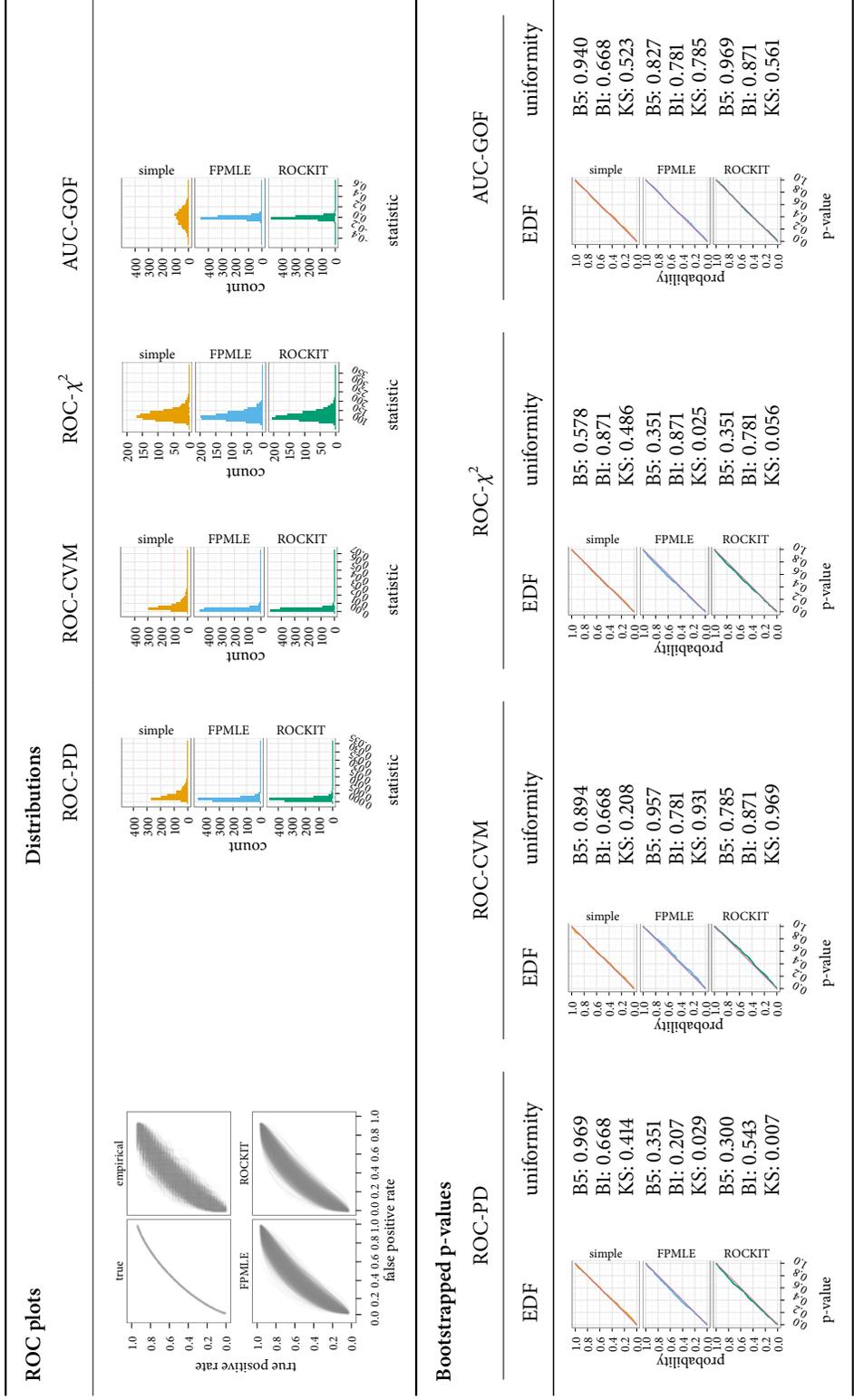


Figure A.11: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0.5$, $n_1 = 100$, $n_2 = 30$.

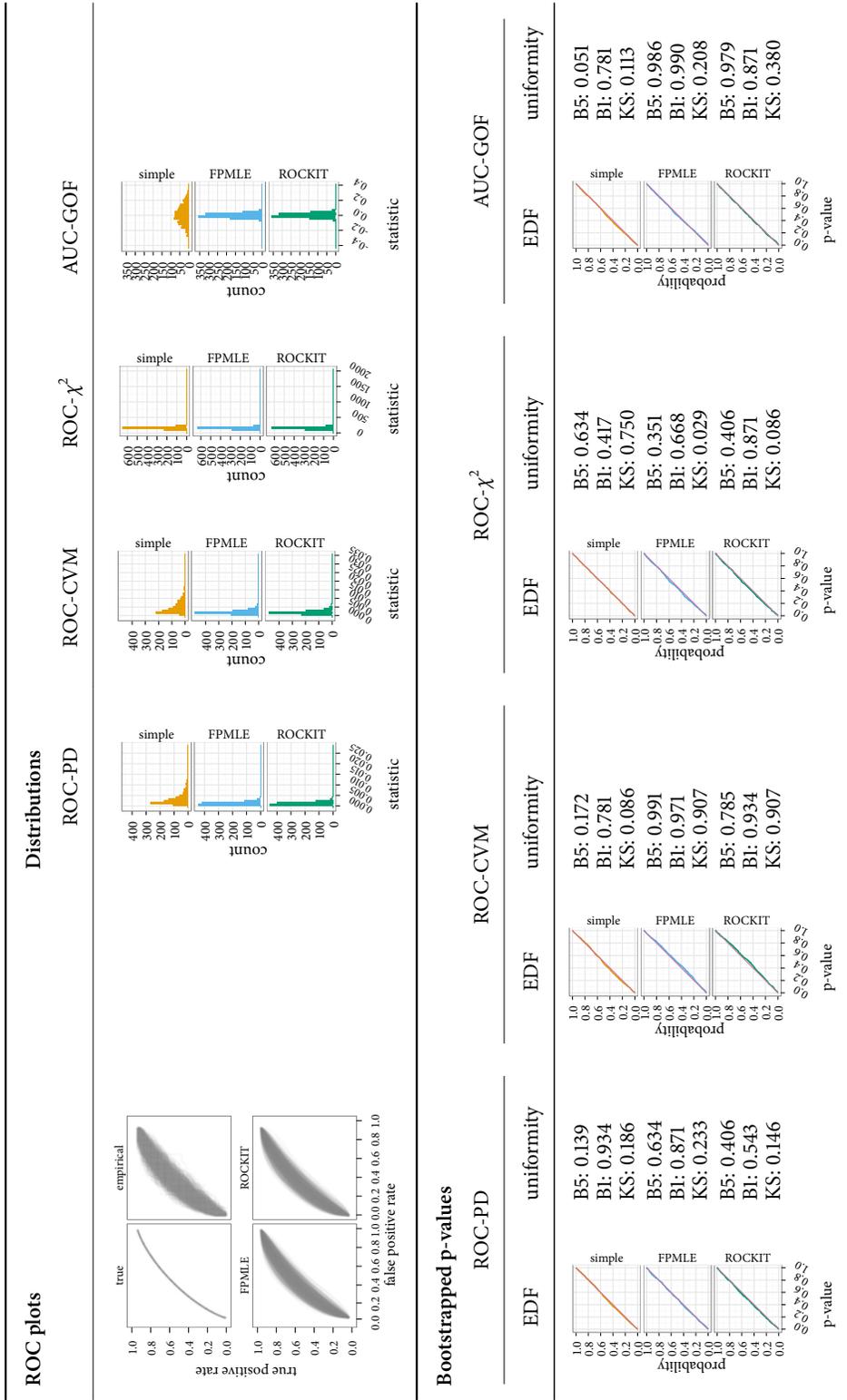


Figure A.12: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 0.5$, $n_1 = 100$, $n_2 = 50$.

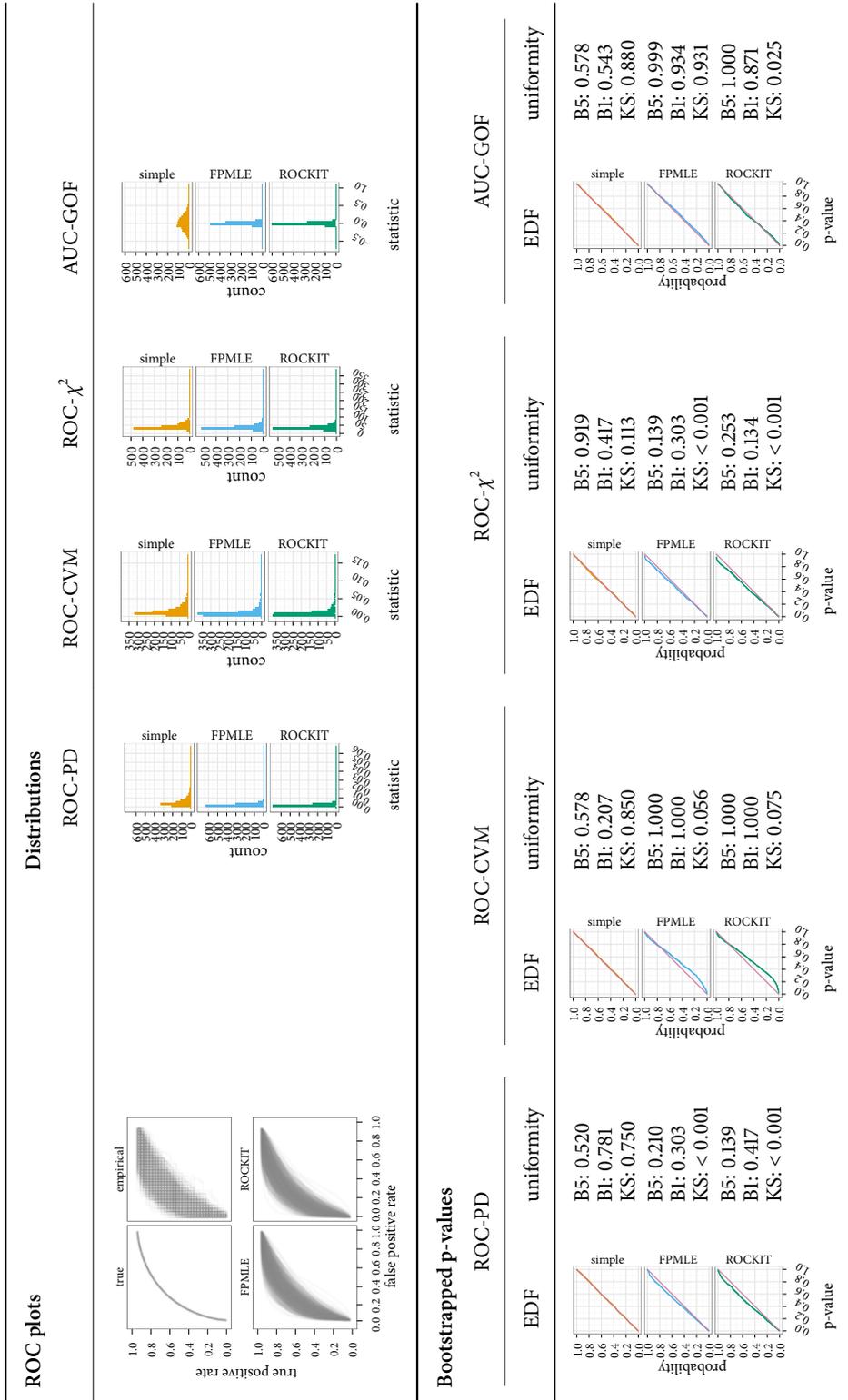


Figure A.13: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1$, $n_1 = 30$, $n_2 = 30$.

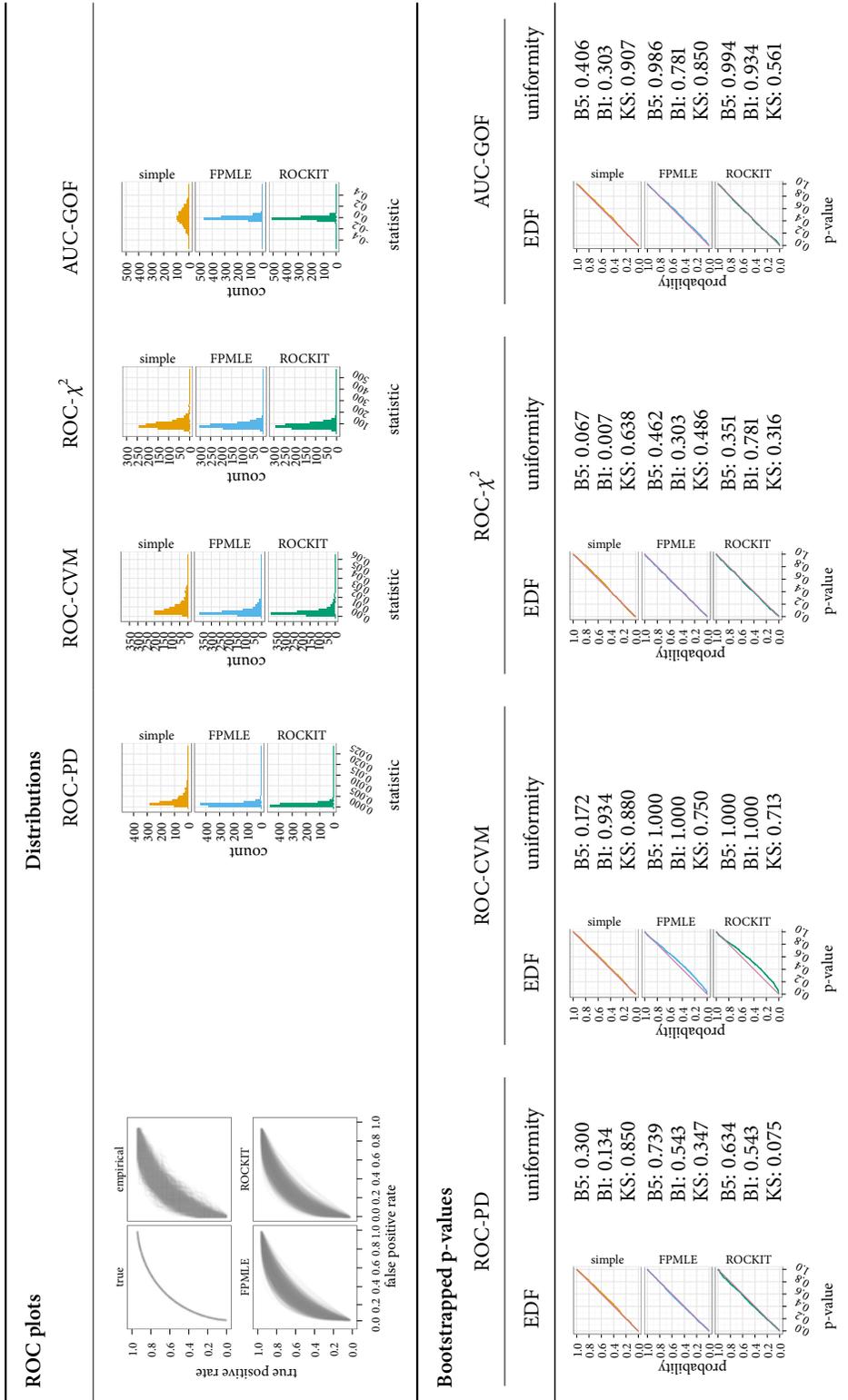


Figure A.14: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1$, $n_1 = 50$, $n_2 = 50$.

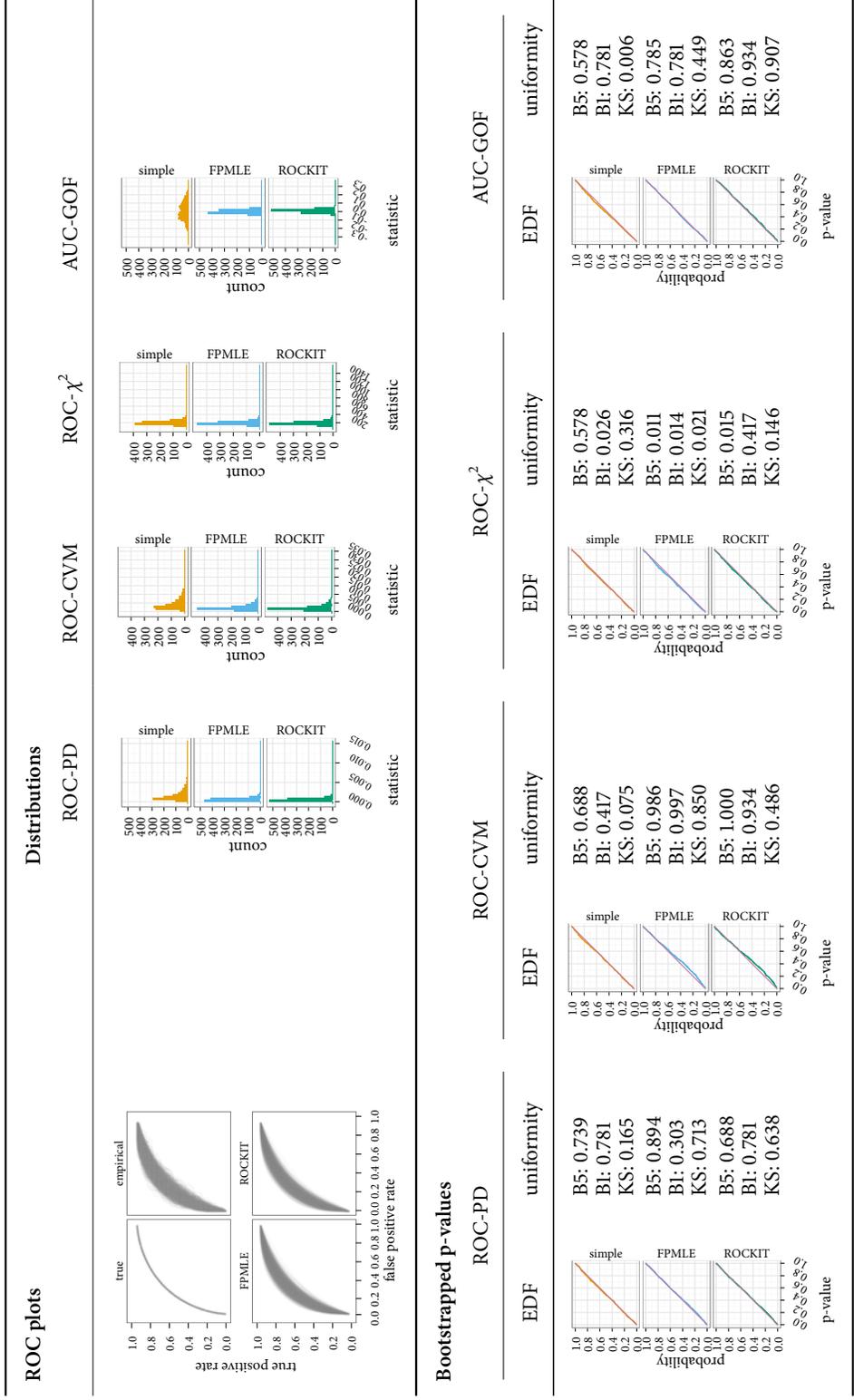


Figure A.15: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1, \delta = 1, n_1 = 100, n_2 = 100$.

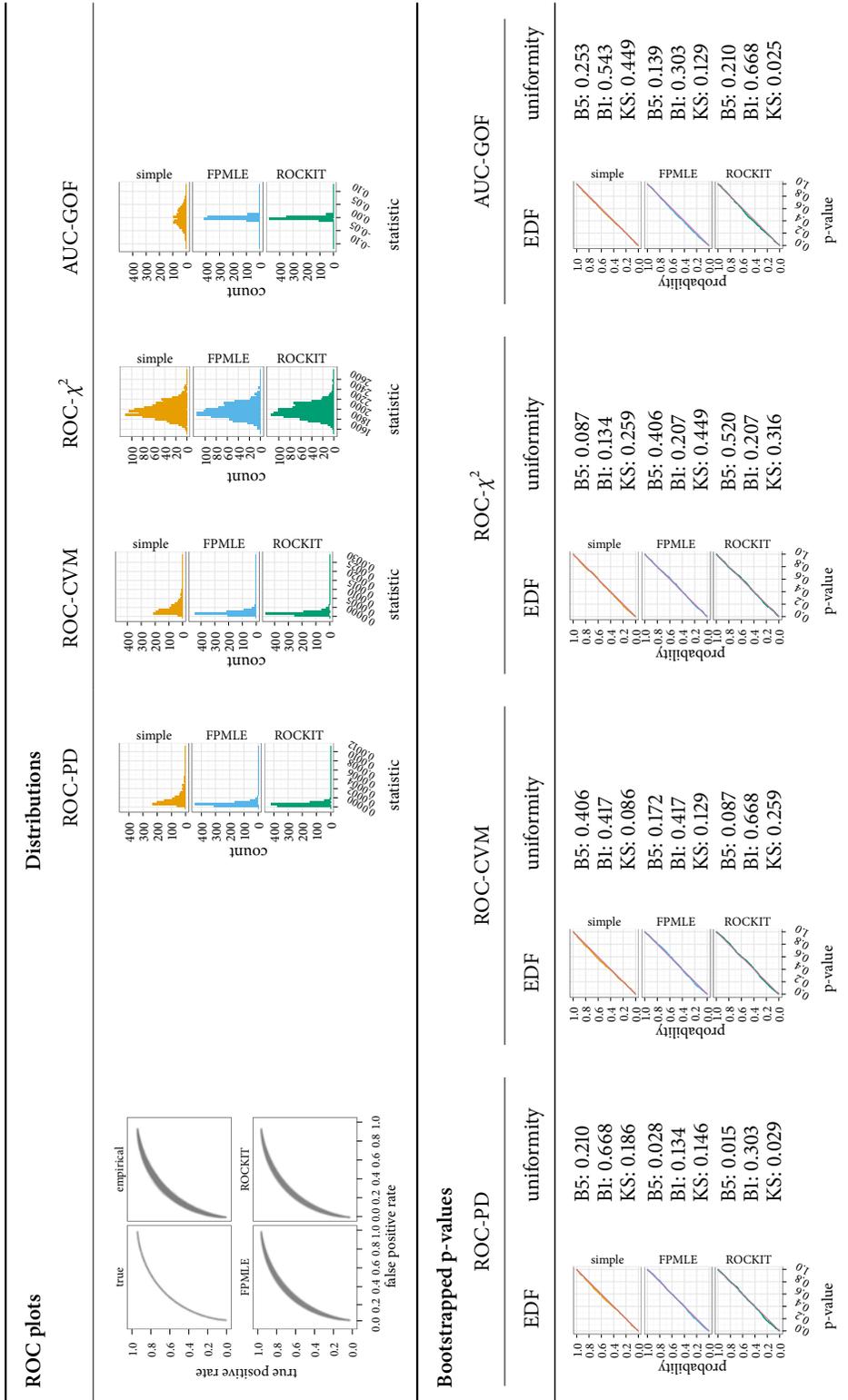


Figure A.16: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1$, $n_1 = 1000$, $n_2 = 1000$.

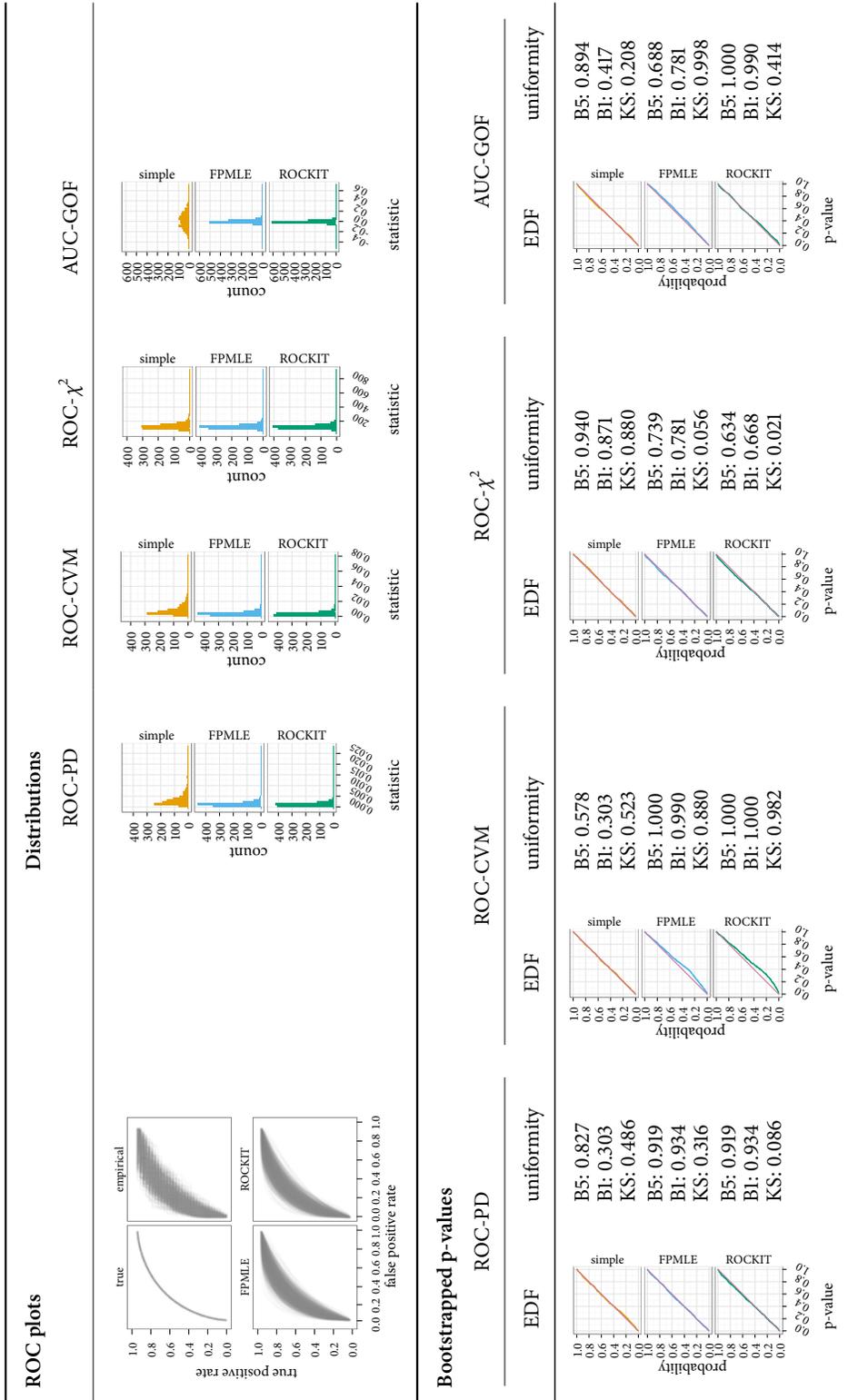


Figure A.17: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1$, $n_1 = 100$, $n_2 = 30$.

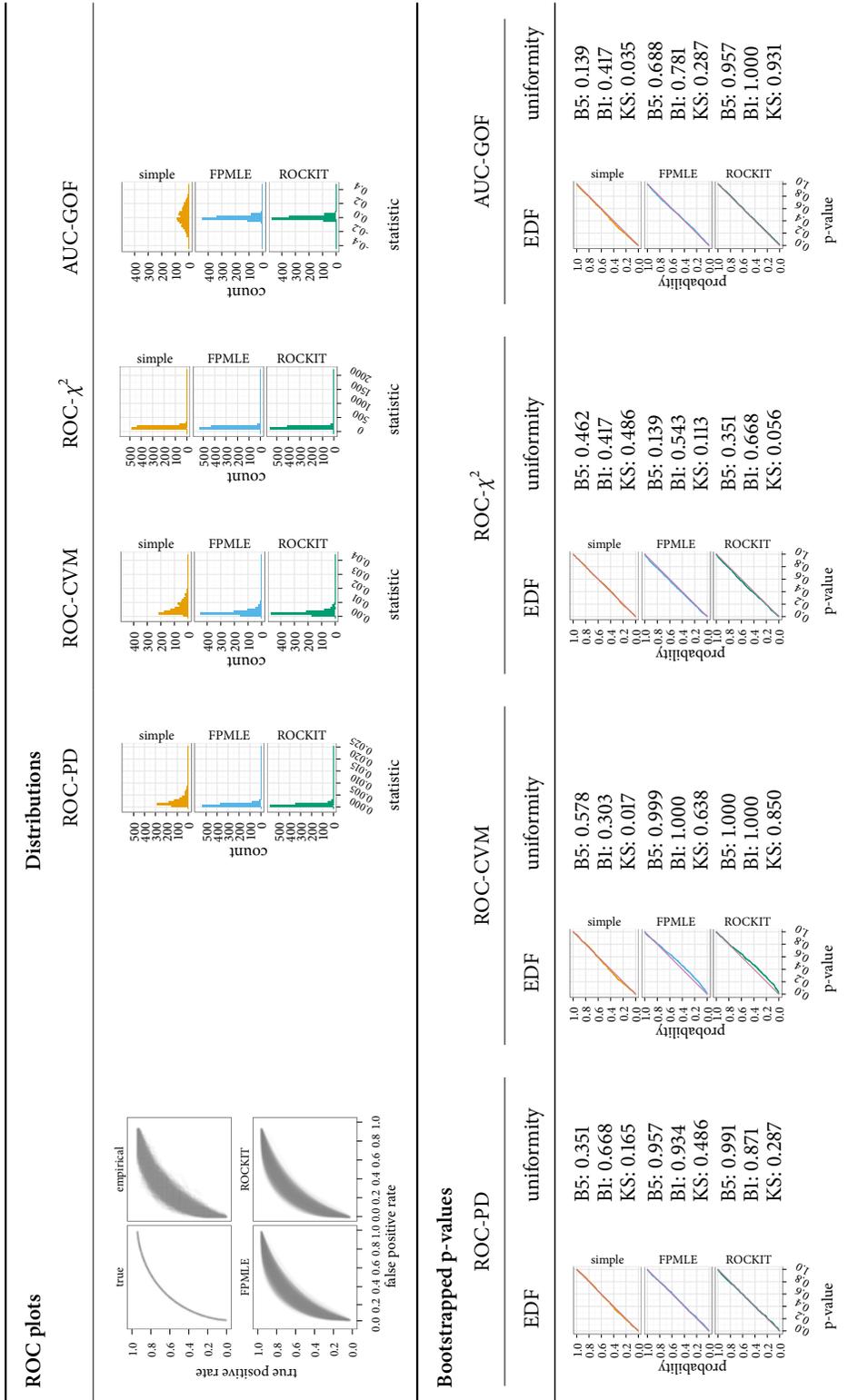


Figure A.18: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1$, $n_1 = 100$, $n_2 = 50$.

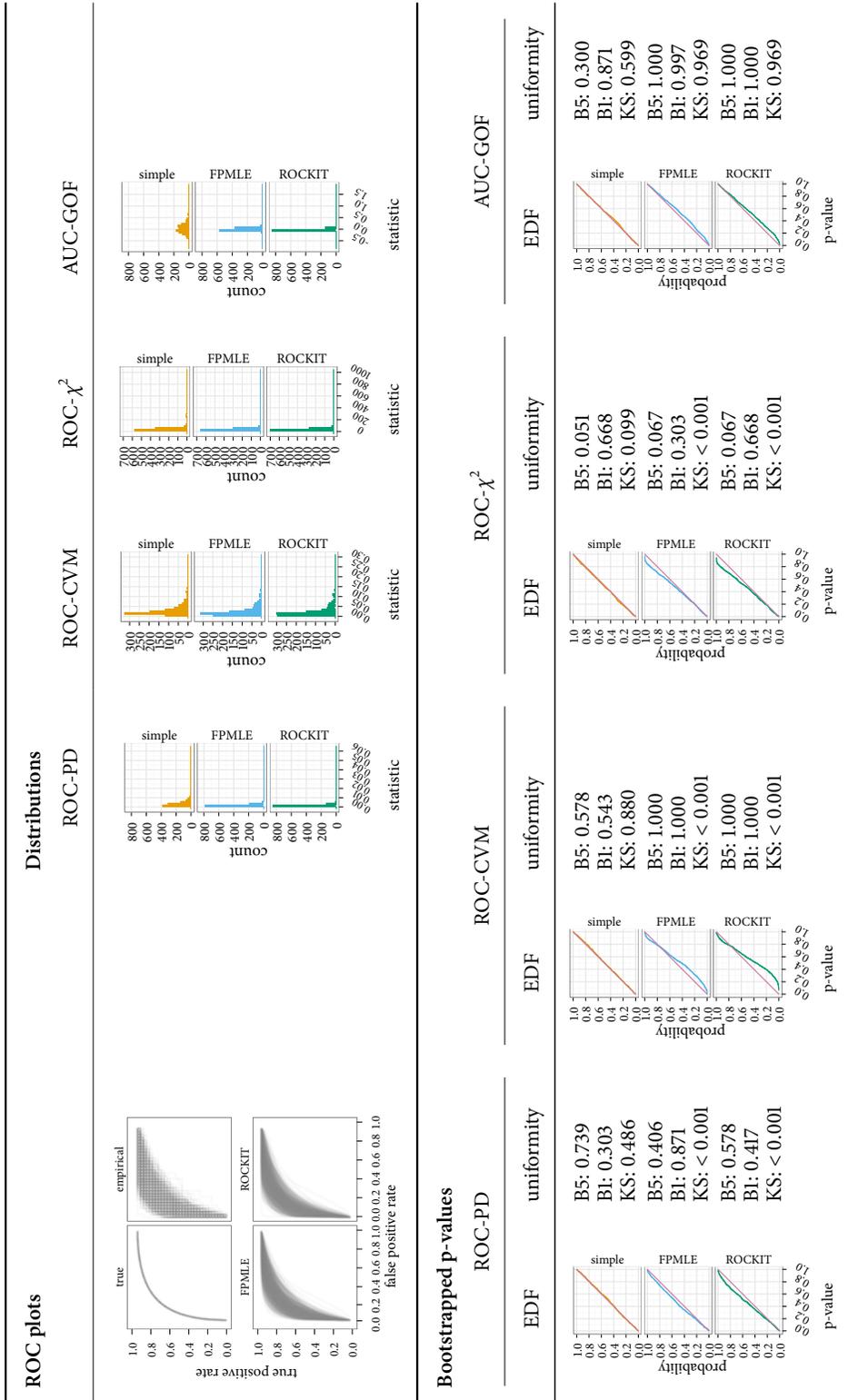


Figure A.19: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1.5$, $n_1 = 30$, $n_2 = 30$.

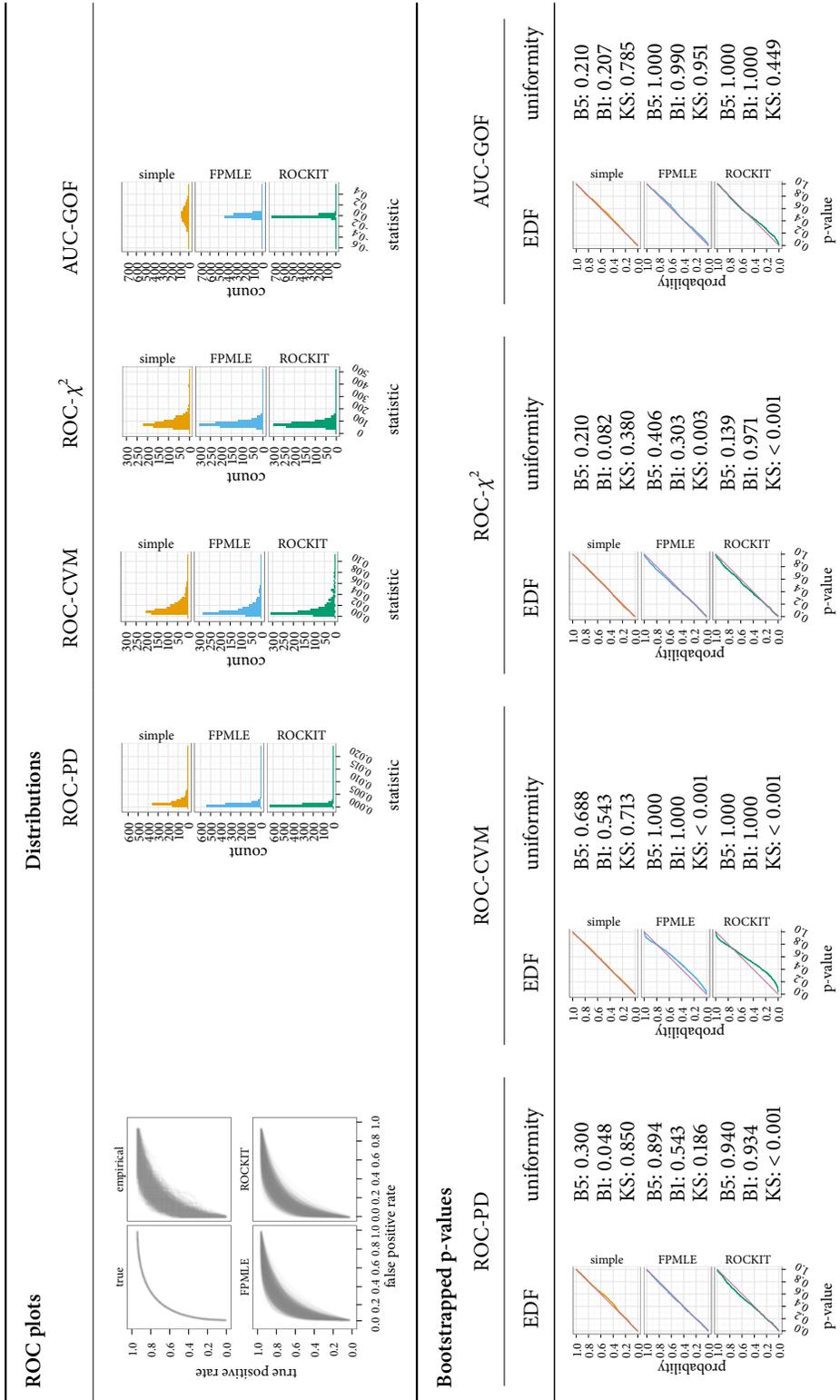


Figure A.20: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1.5$, $n_1 = 50$, $n_2 = 50$.

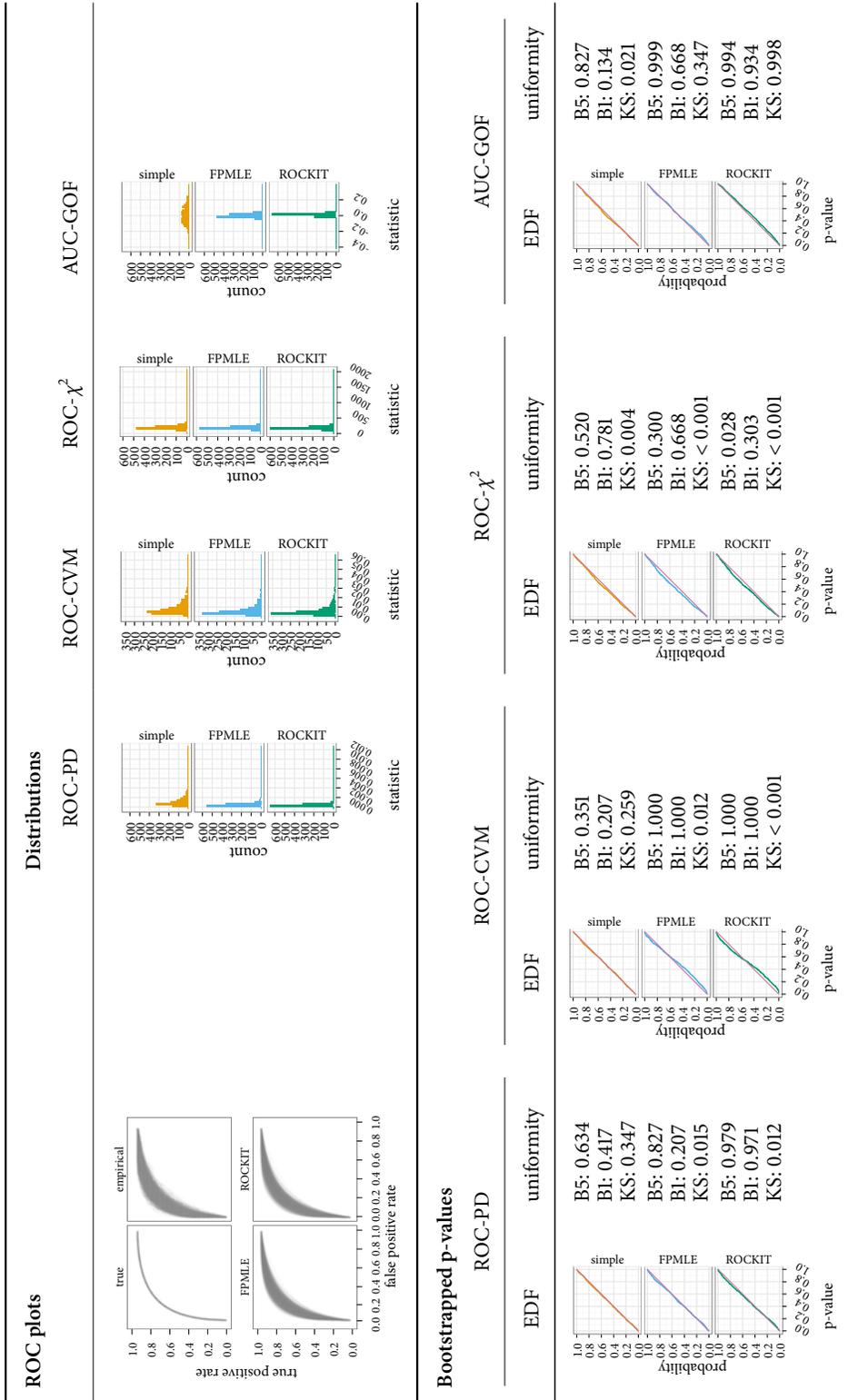


Figure A.21: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1.5$, $n_1 = 100$, $n_2 = 100$.

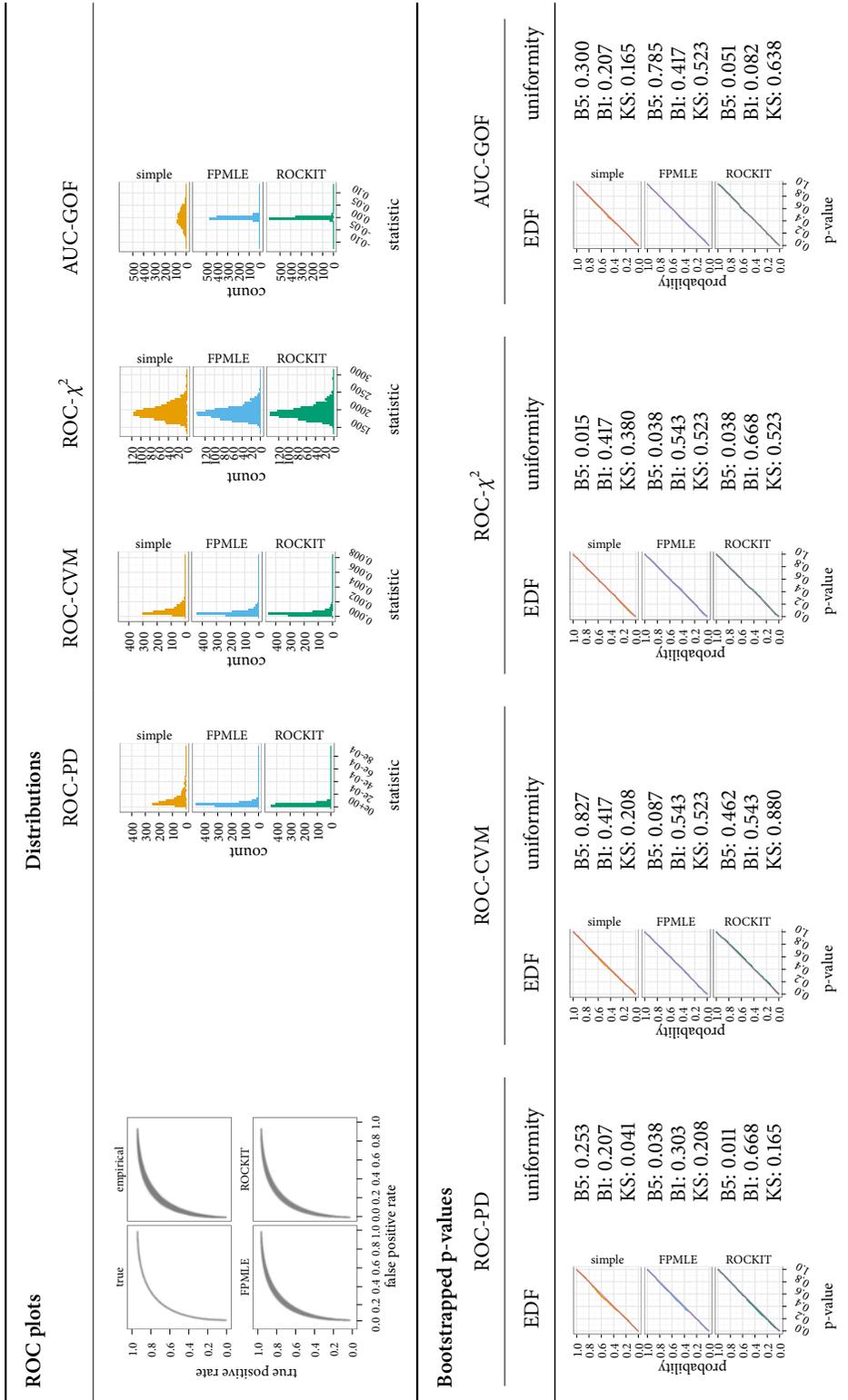


Figure A.22: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1.5$, $n_1 = 1000$, $n_2 = 1000$.

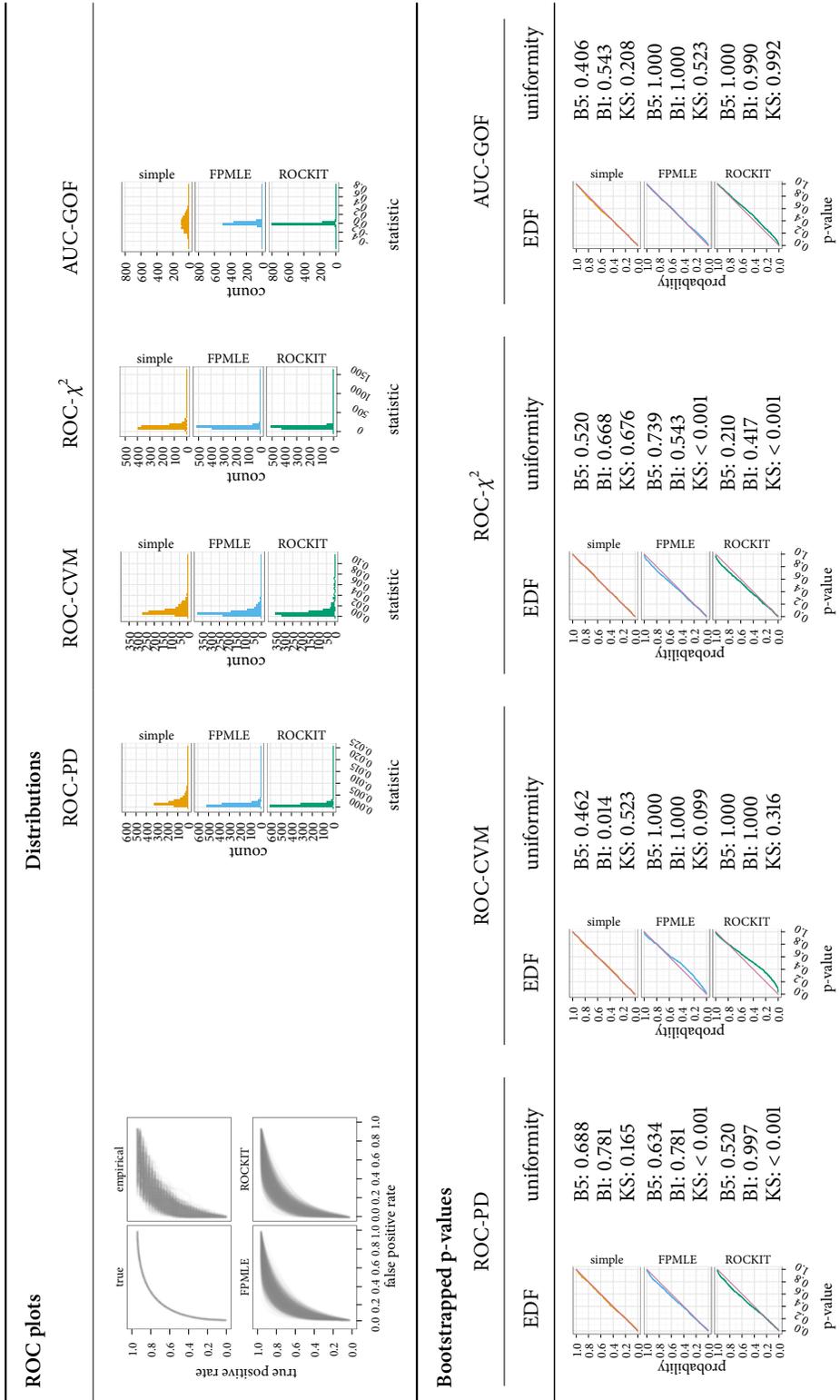


Figure A.23: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1.5$, $n_1 = 100$, $n_2 = 30$.

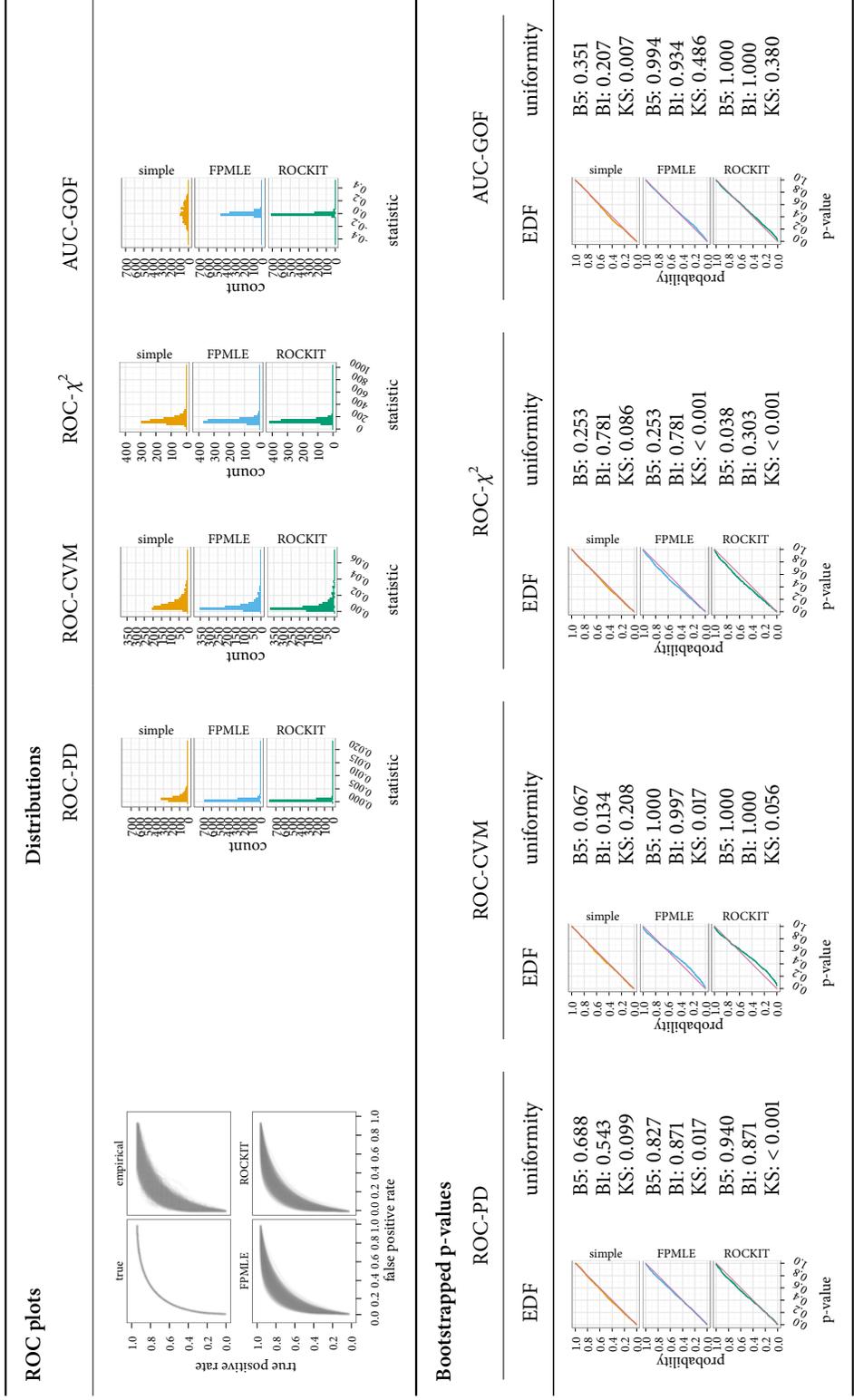


Figure A.24: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 1.5$, $n_1 = 100$, $n_2 = 50$.

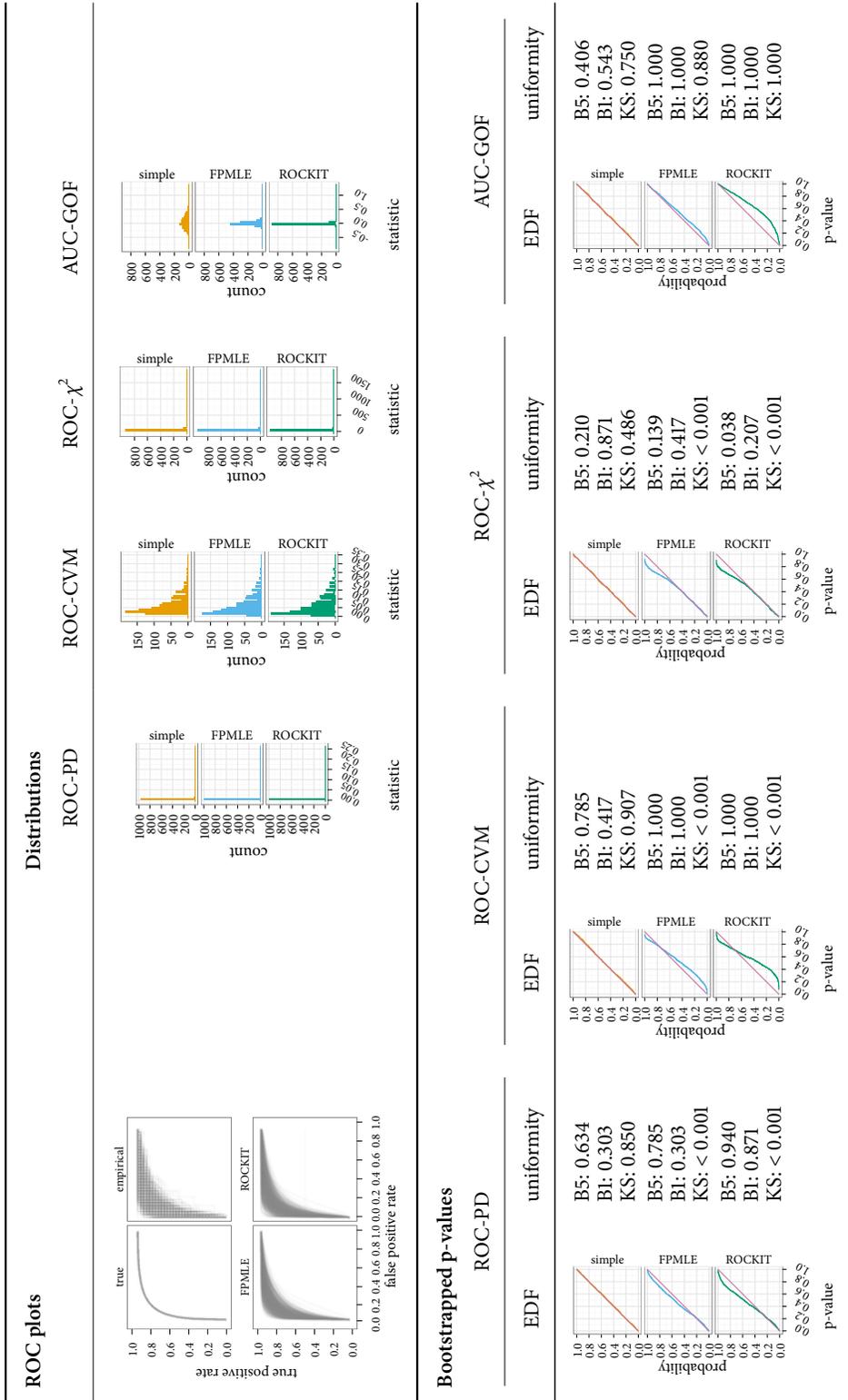


Figure A.25: Goodness-of-fit results for binormal ROCs. $\rho = 1$, $\delta = 2$, $n_1 = 30$, $n_2 = 30$.

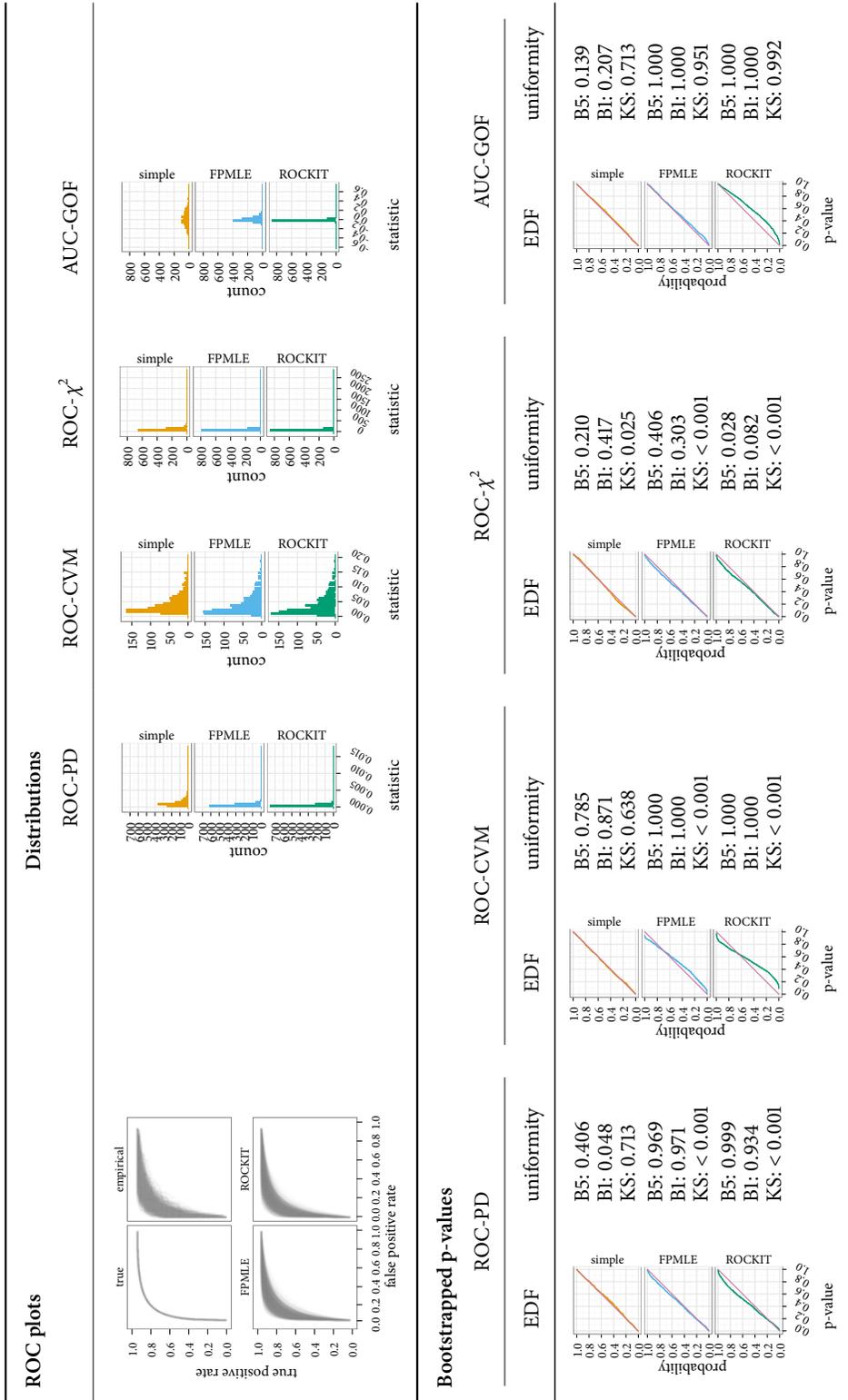


Figure A.26: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 2$, $n_1 = 50$, $n_2 = 50$.

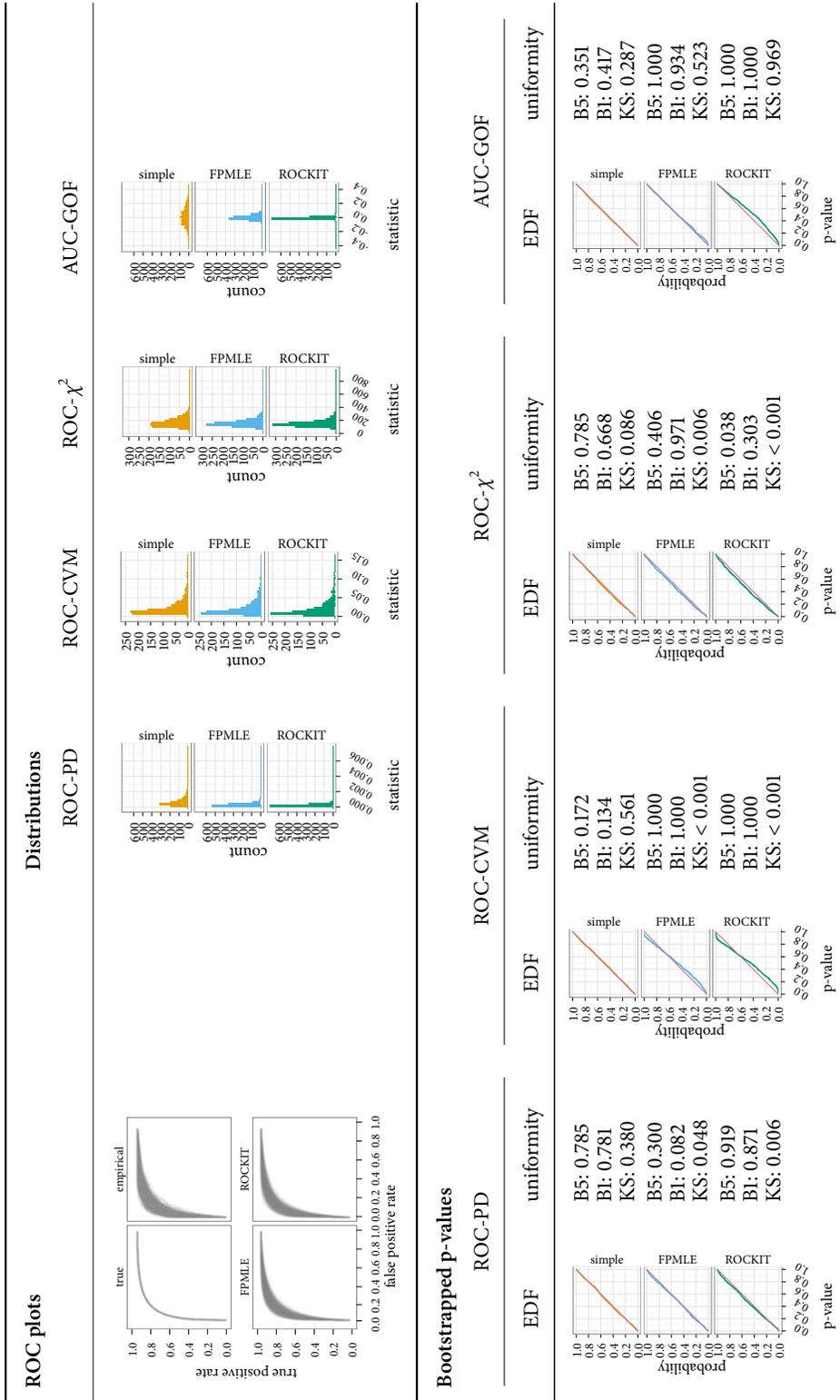


Figure A.27: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 2$, $n_1 = 100$, $n_2 = 100$.

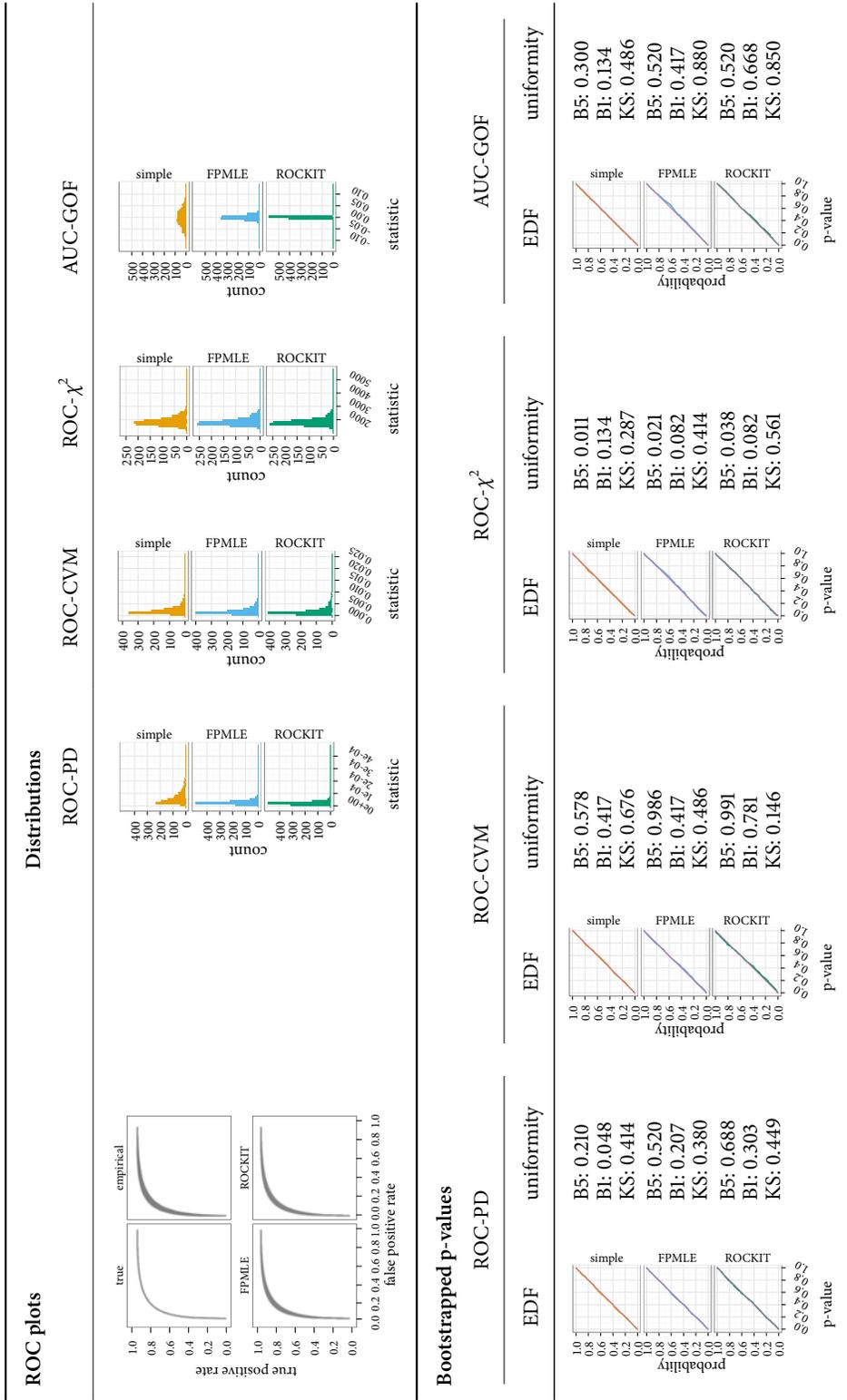


Figure A.28: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 2$, $n_1 = 1000$, $n_2 = 1000$.

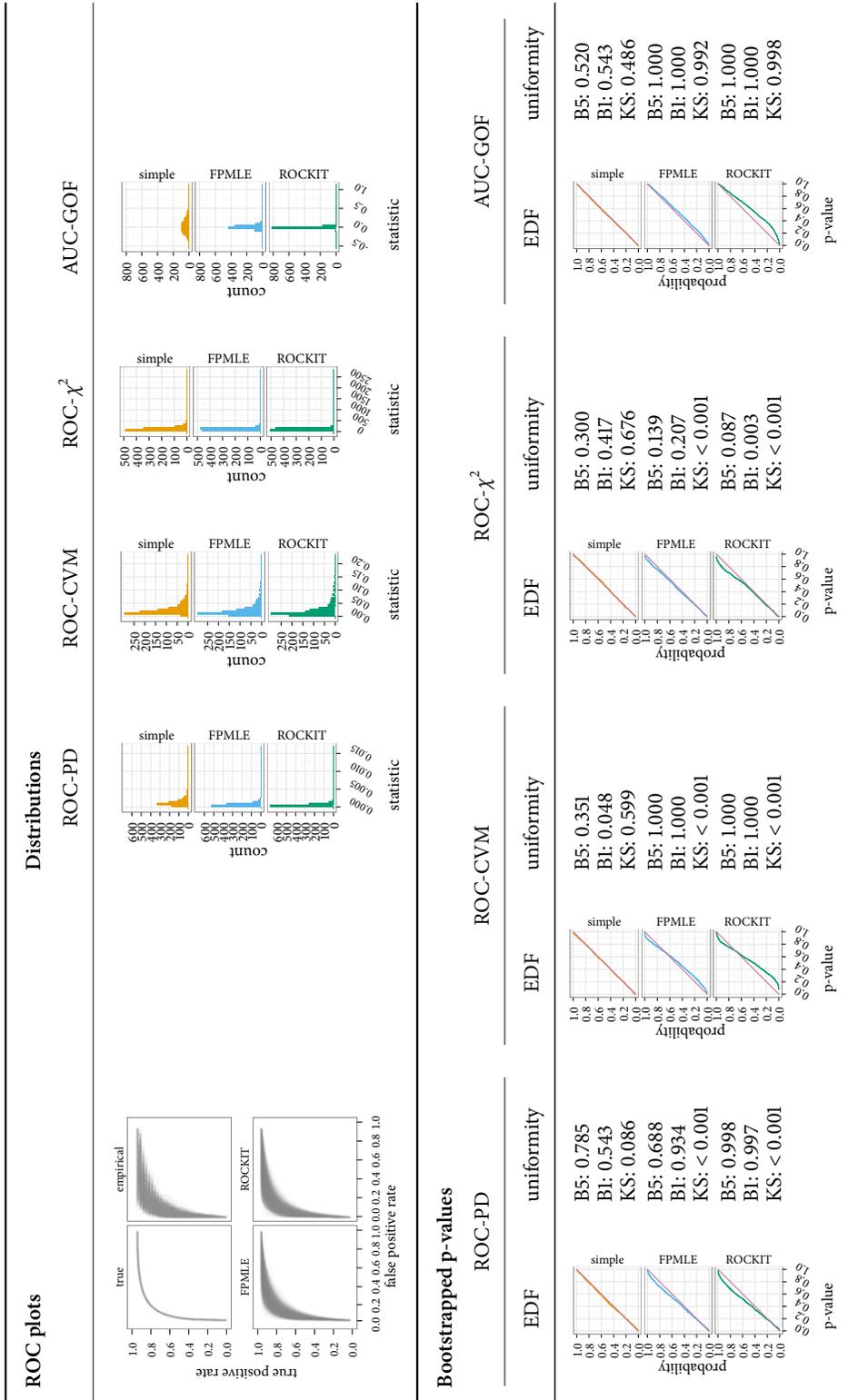


Figure A.29: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 1$, $\delta = 2$, $n_1 = 100$, $n_2 = 30$.

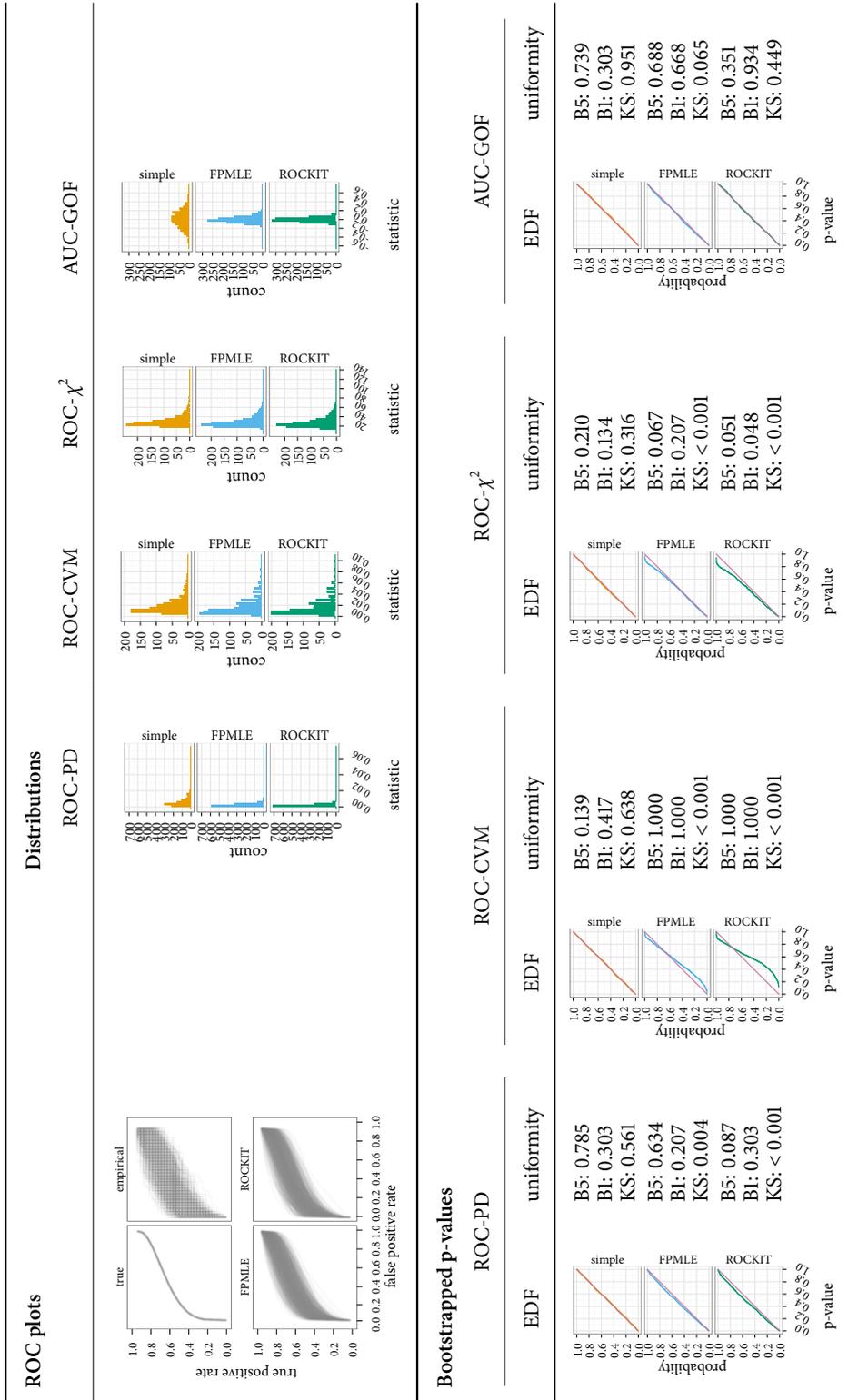


Figure A.31: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 0.5$, $n_1 = 30$, $n_2 = 30$.

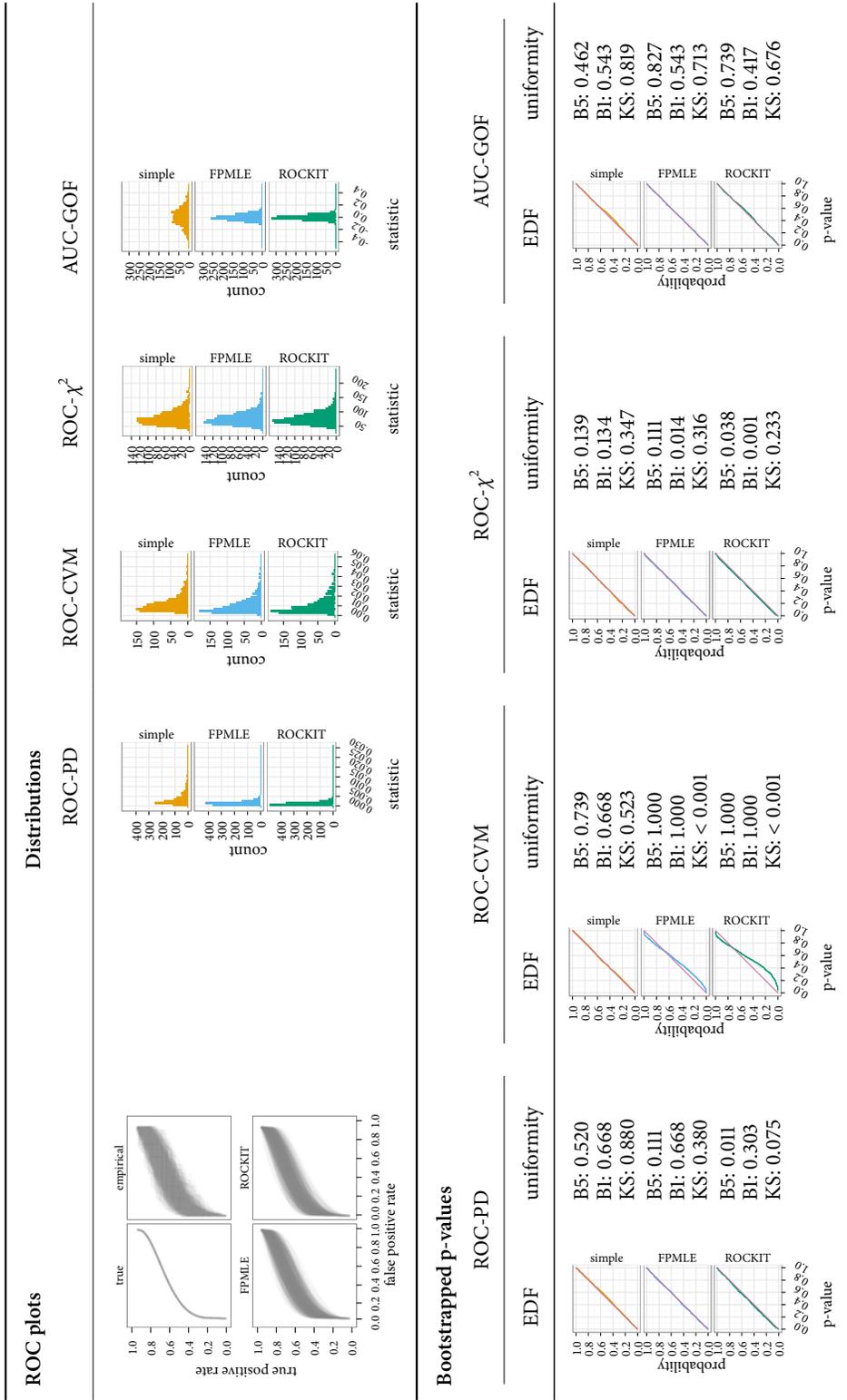


Figure A.32: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 0.5$, $n_1 = 50$, $n_2 = 50$.

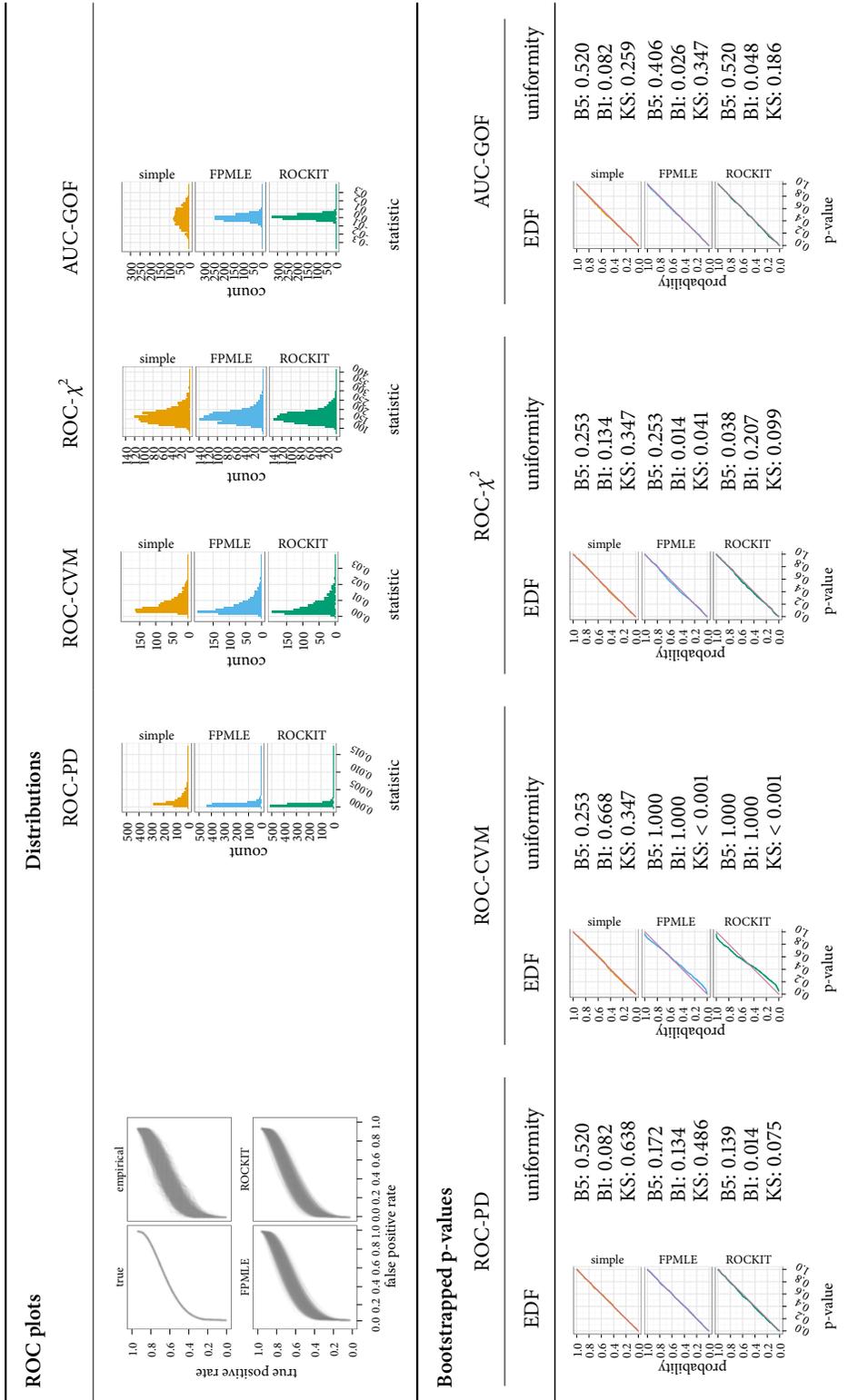


Figure A.33: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 0.5$, $n_1 = 100$, $n_2 = 100$.

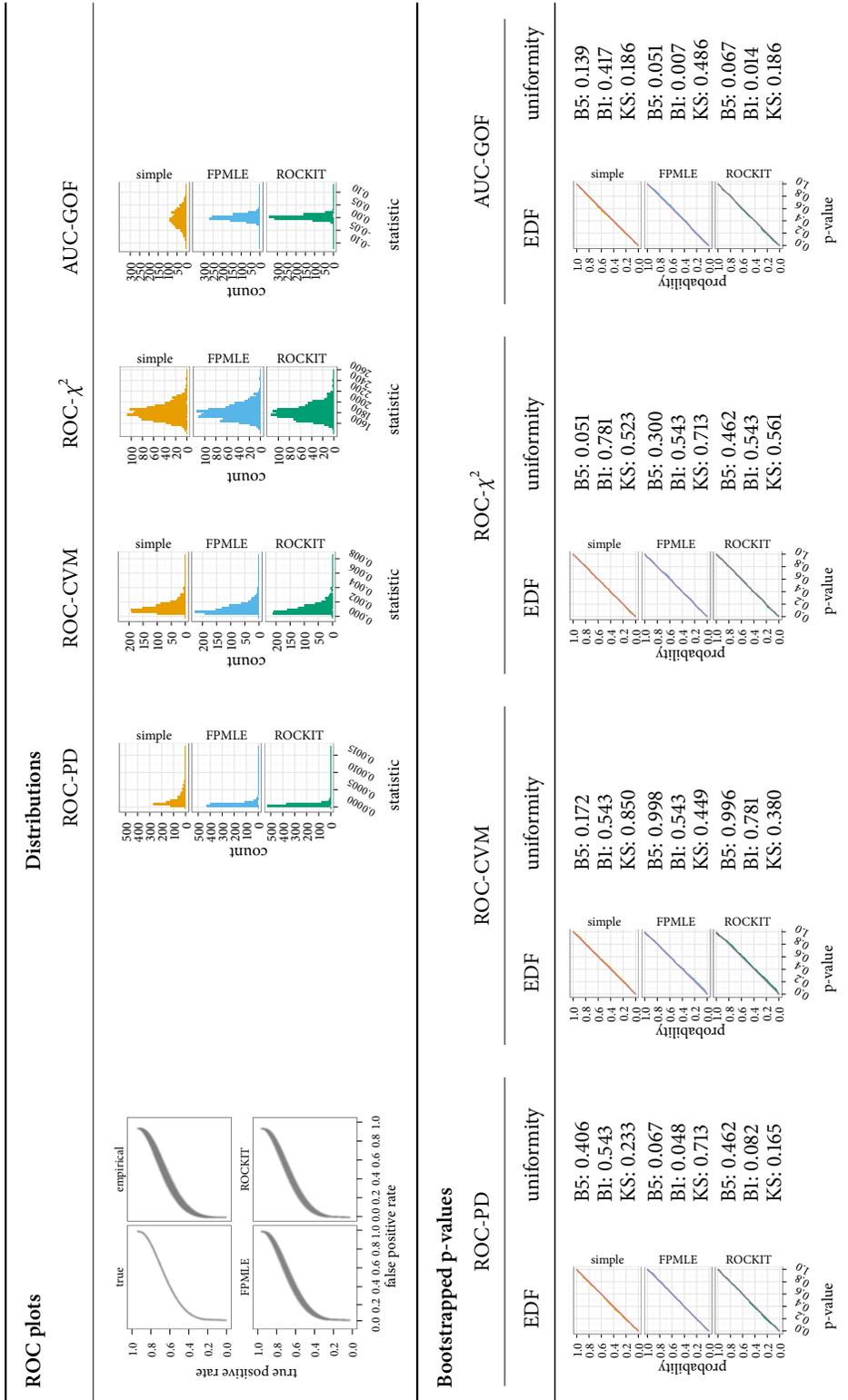


Figure A.34: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 0.5$, $n_1 = 1000$, $n_2 = 1000$.

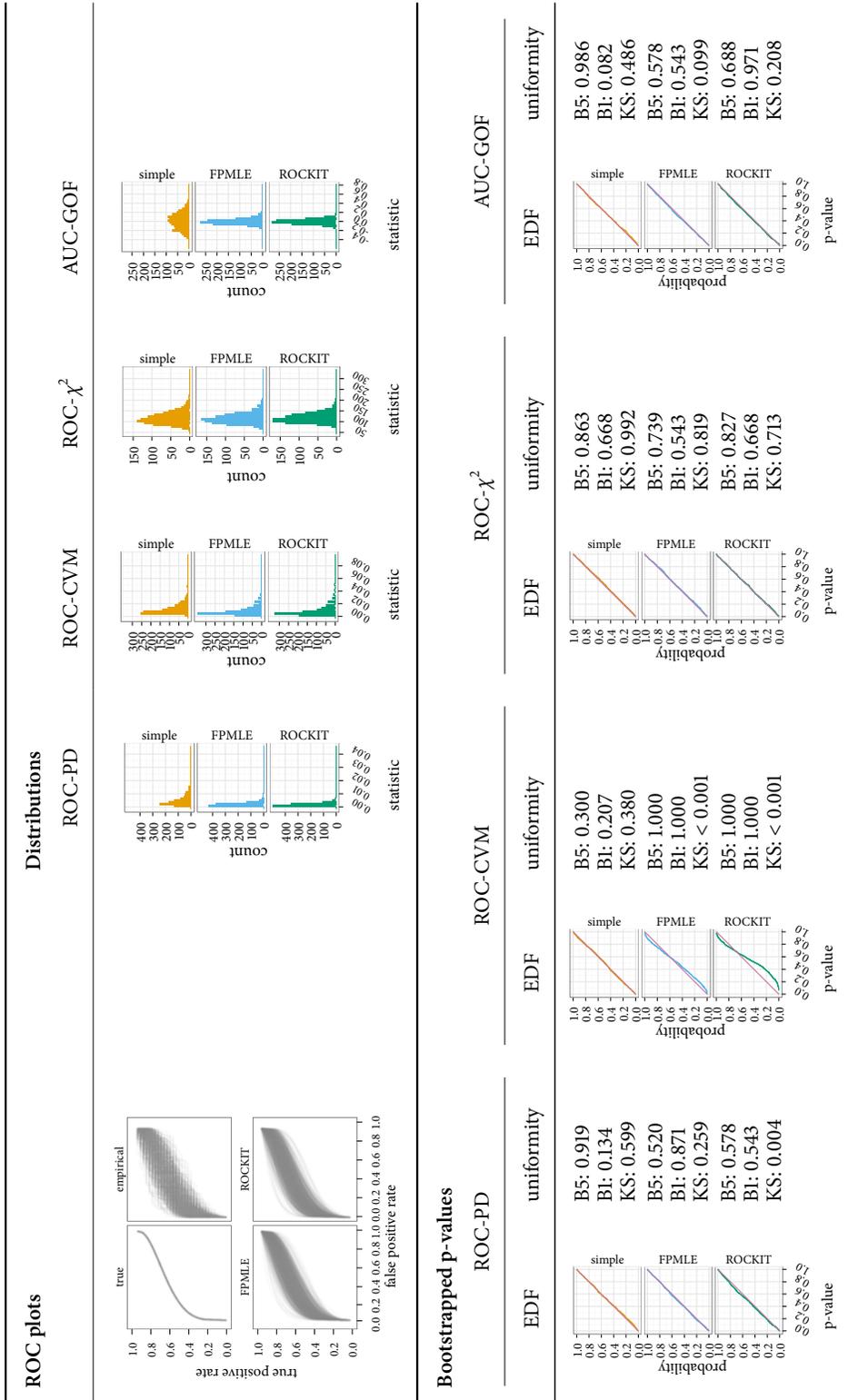


Figure A.35: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 0.5$, $n_1 = 100$, $n_2 = 30$.

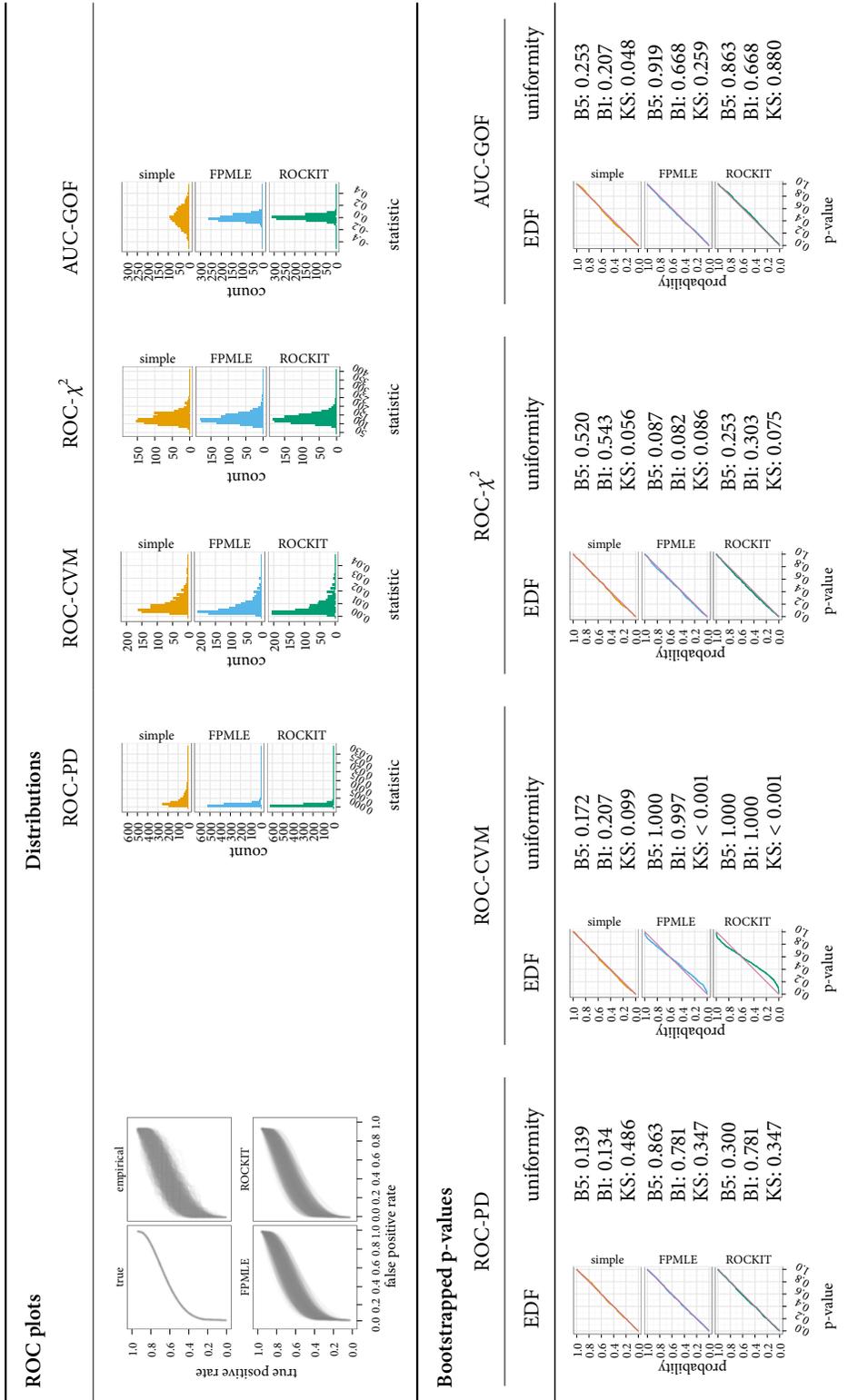


Figure A.36: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 0.5$, $n_1 = 100$, $n_2 = 50$.

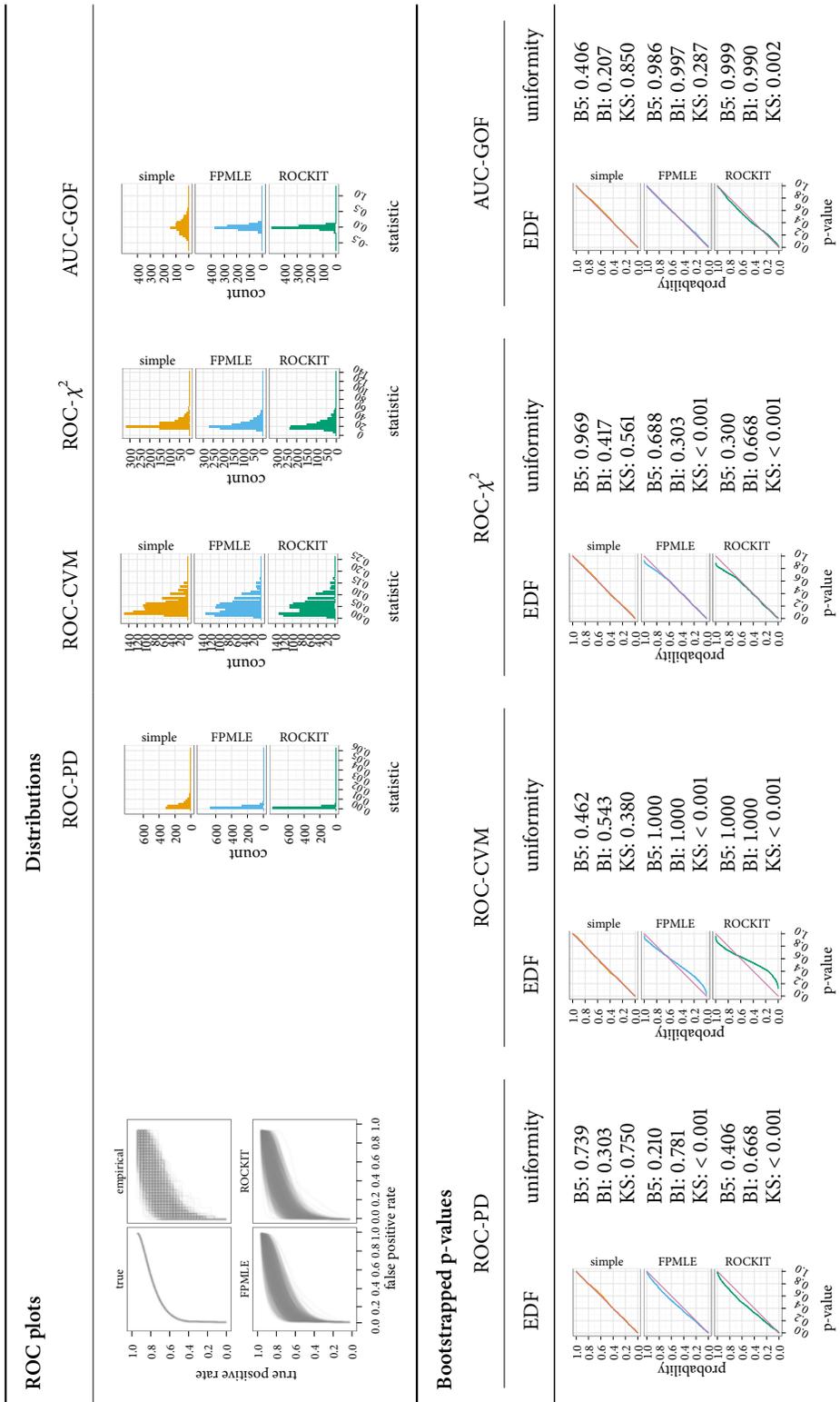


Figure A.37: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 1$, $n_1 = 30$, $n_2 = 30$.

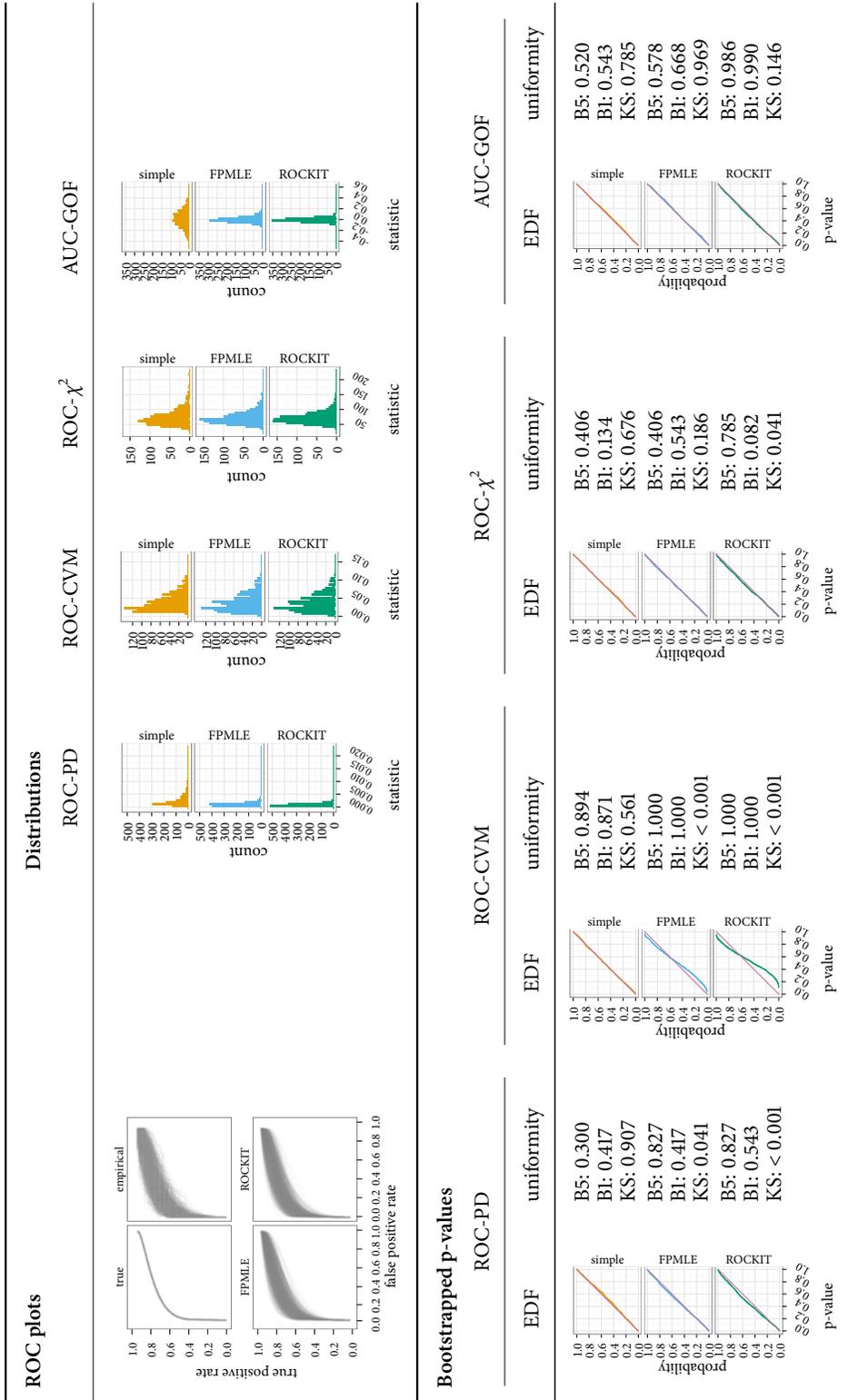


Figure A.38: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 1$, $n_1 = 50$, $n_2 = 50$.

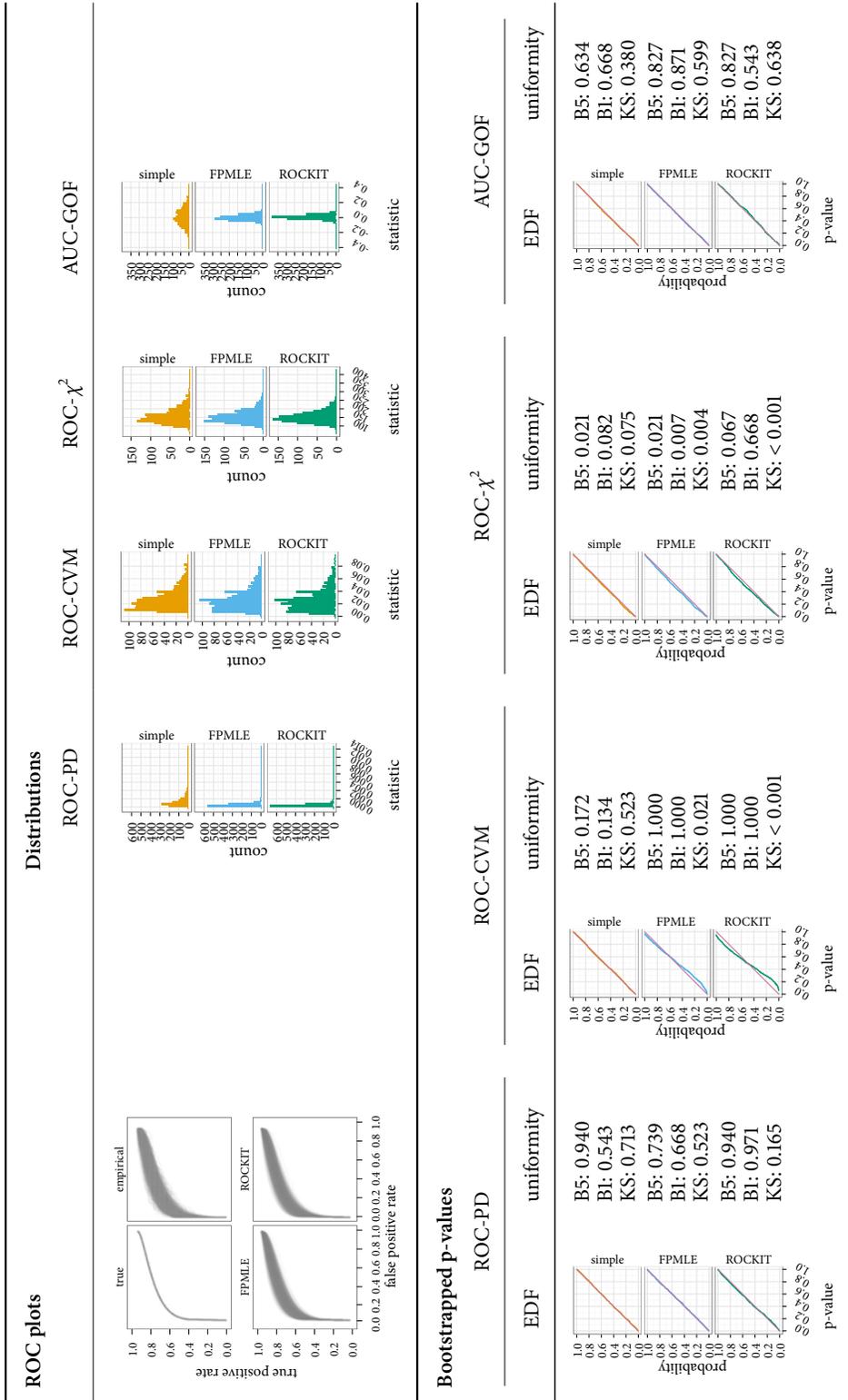


Figure A.39: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 1$, $n_1 = 100$, $n_2 = 100$.

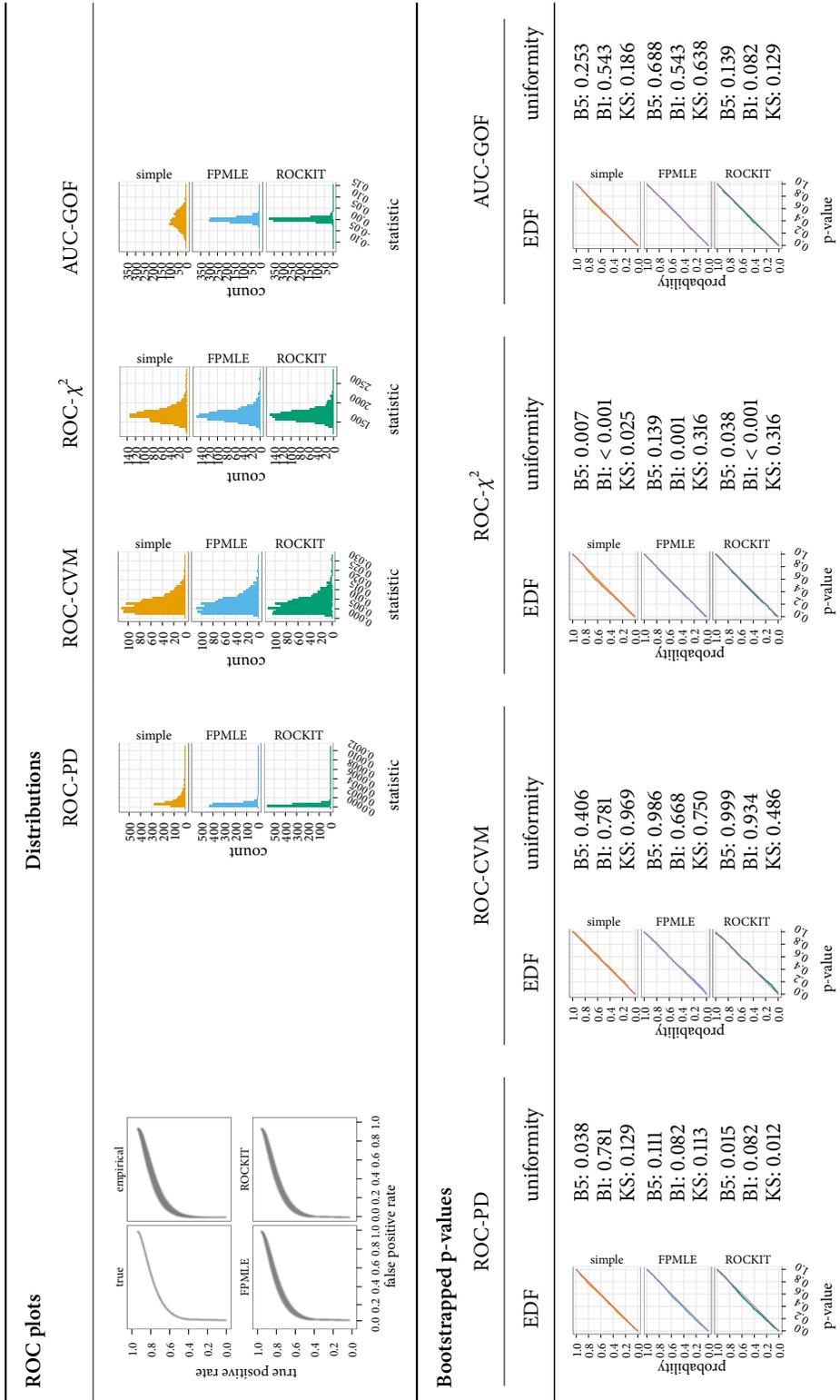


Figure A.40: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 1$, $n_1 = 1000$, $n_2 = 1000$.

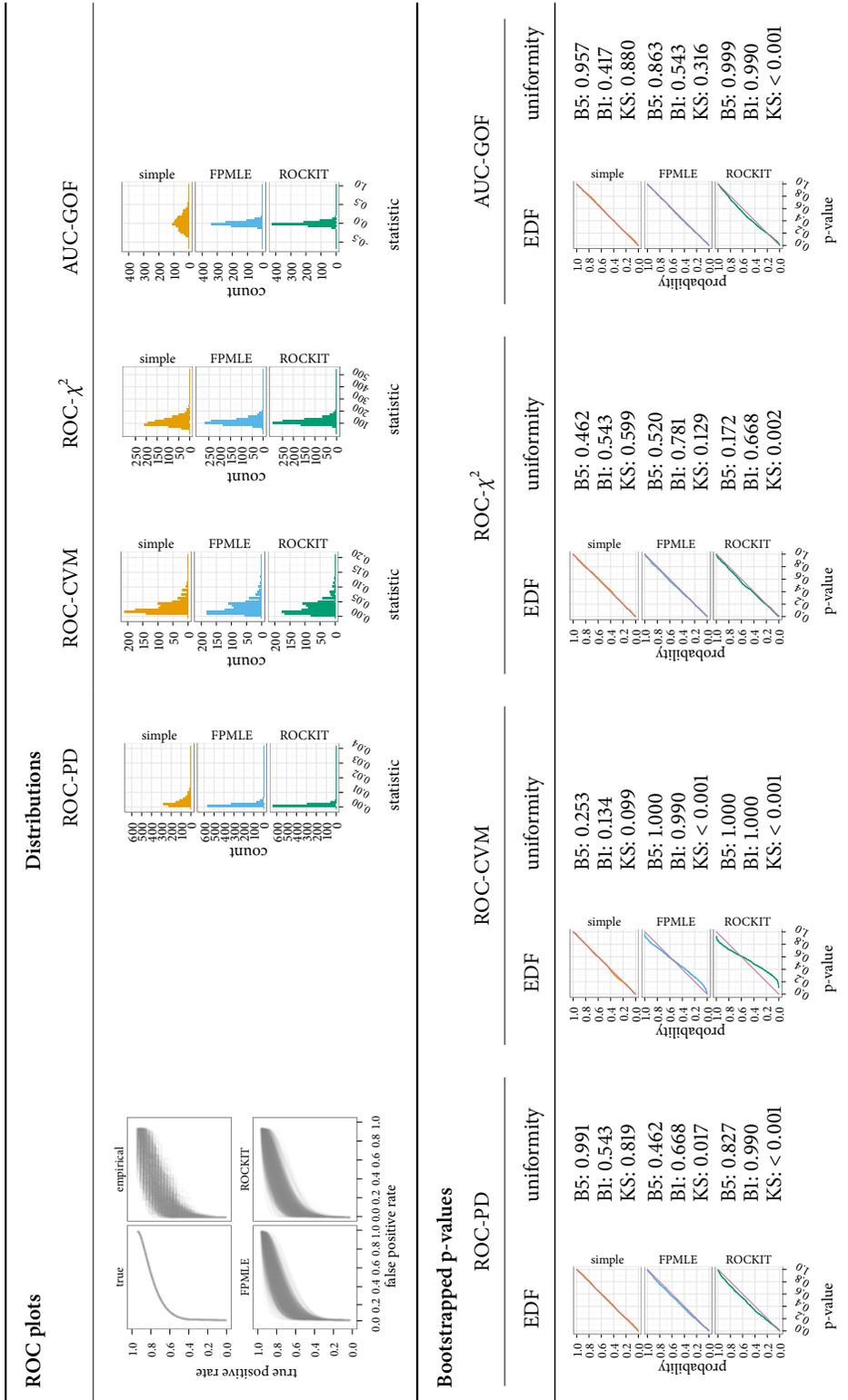


Figure A.41: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 1$, $n_1 = 100$, $n_2 = 30$.

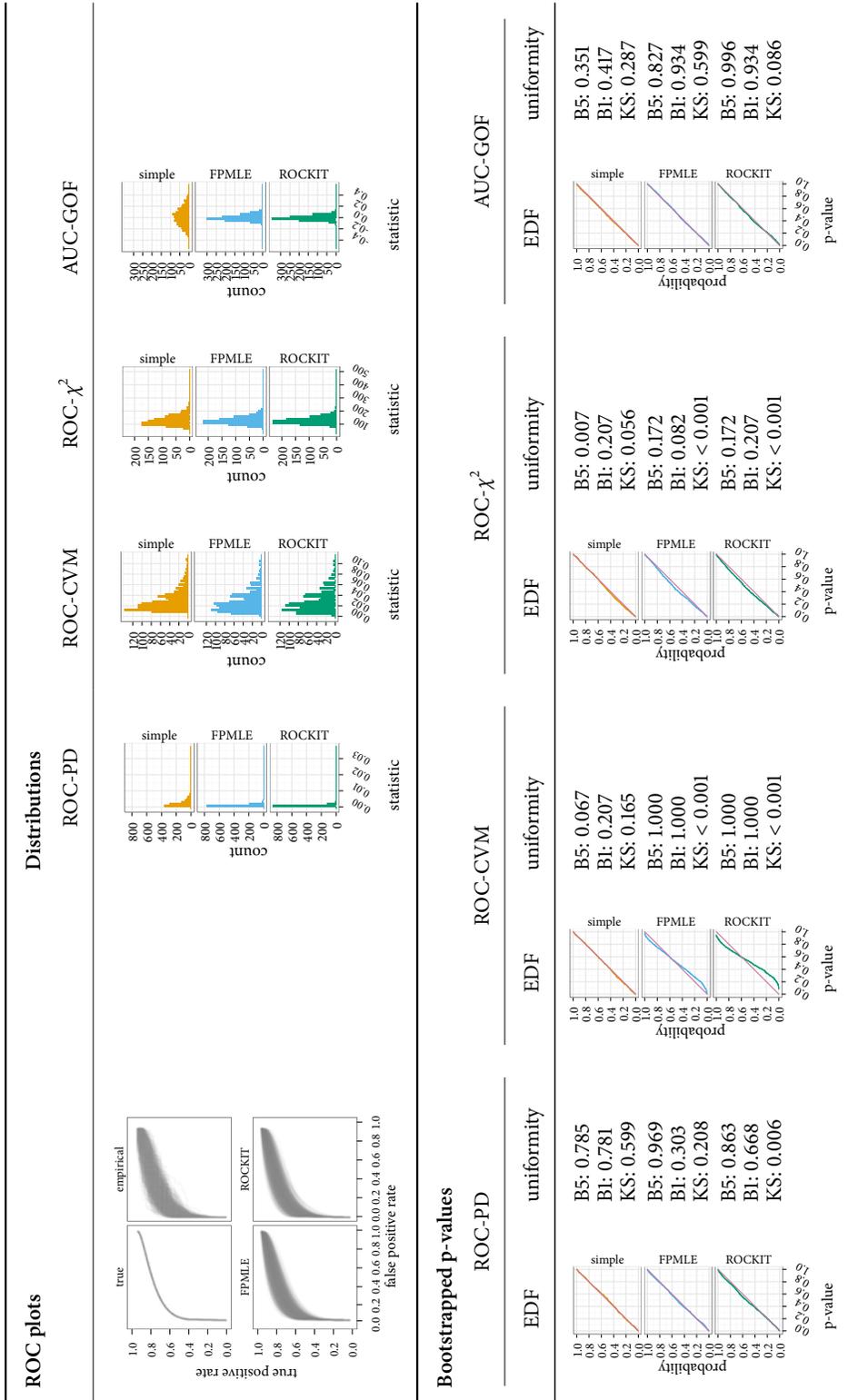


Figure A.42: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.5$, $\delta = 1$, $n_1 = 100$, $n_2 = 50$.

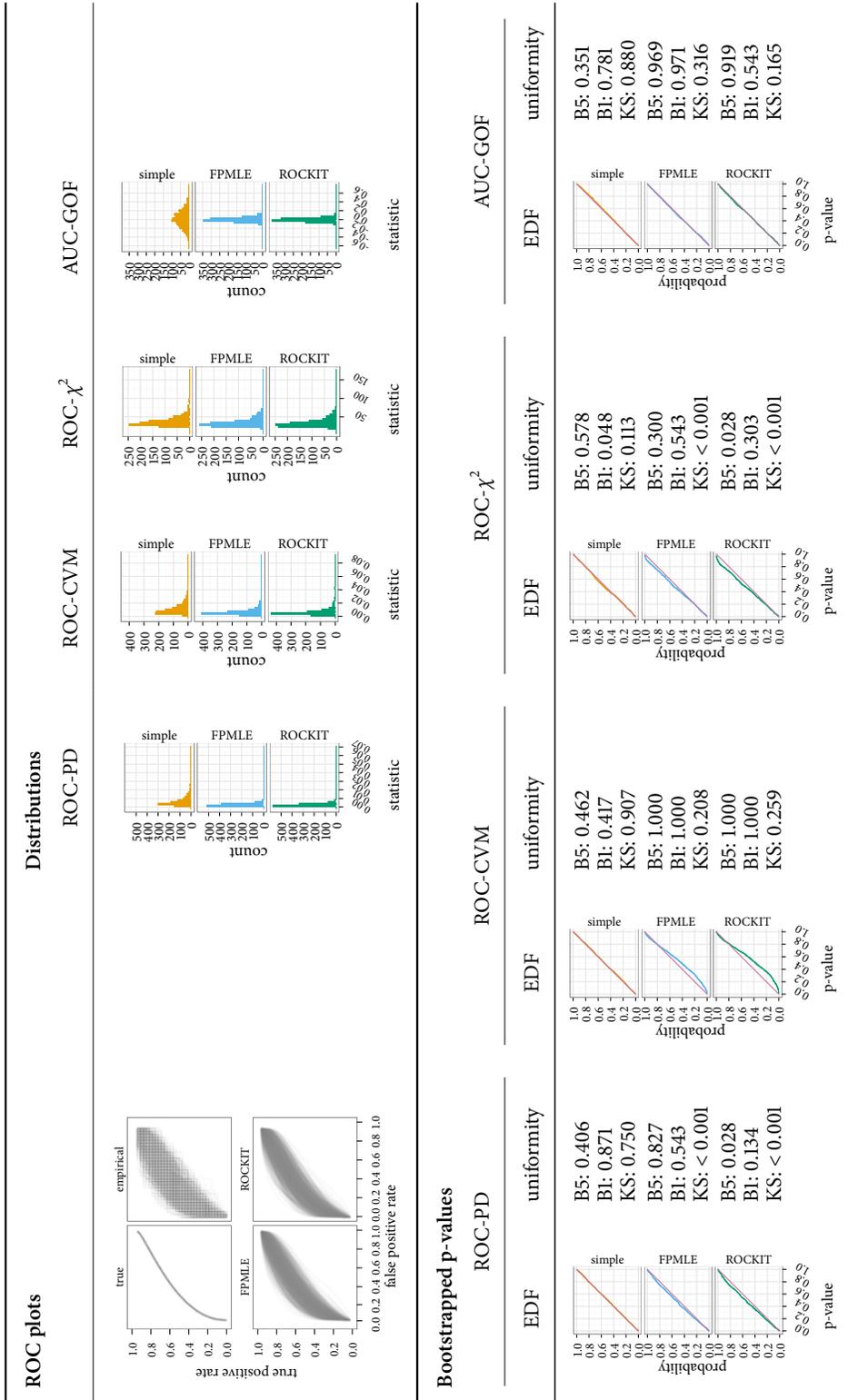


Figure A.43: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.75$, $\delta = 0.5$, $n_1 = 30$, $n_2 = 30$.

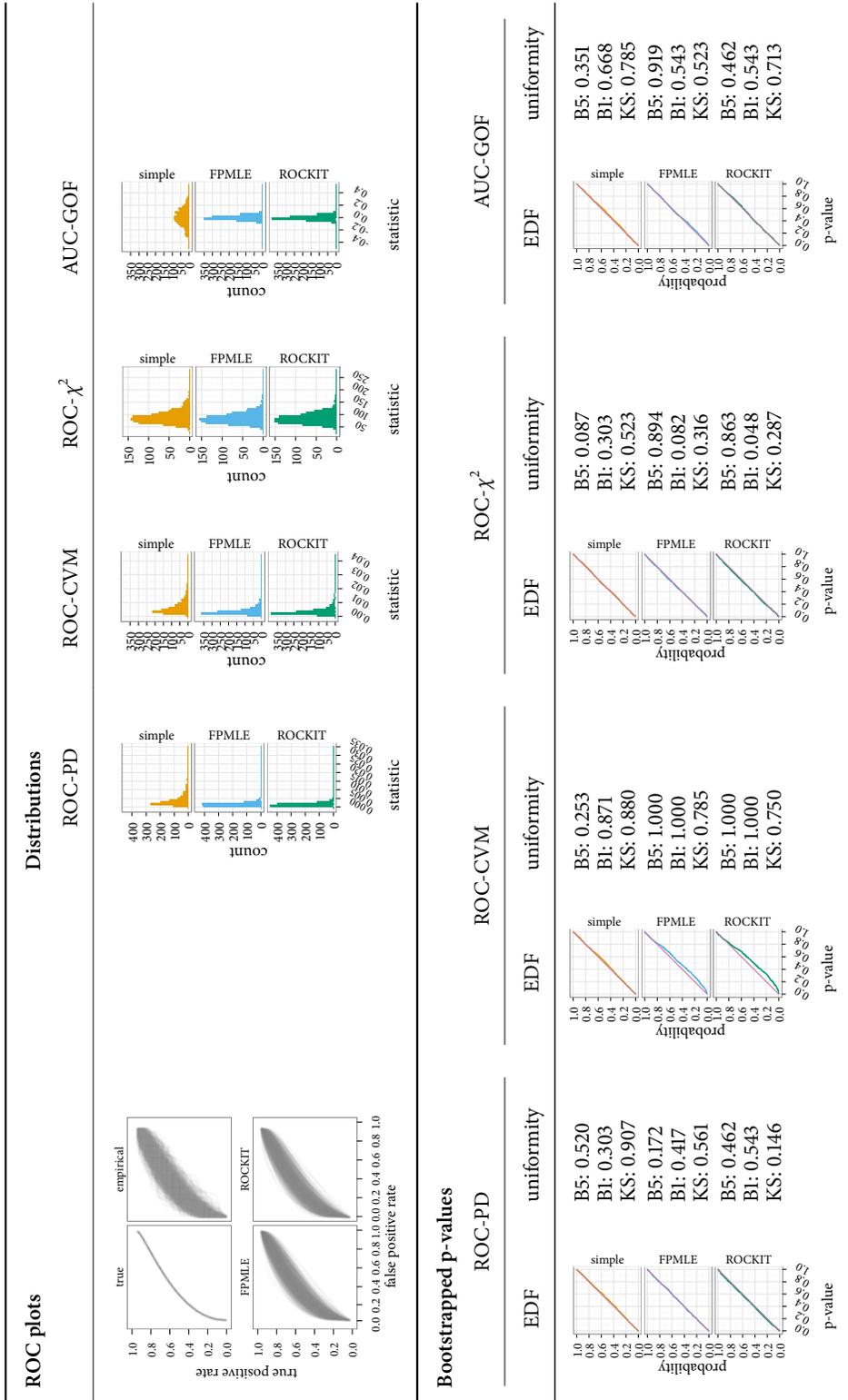


Figure A.44: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.75$, $\delta = 0.5$, $n_1 = 50$, $n_2 = 50$.

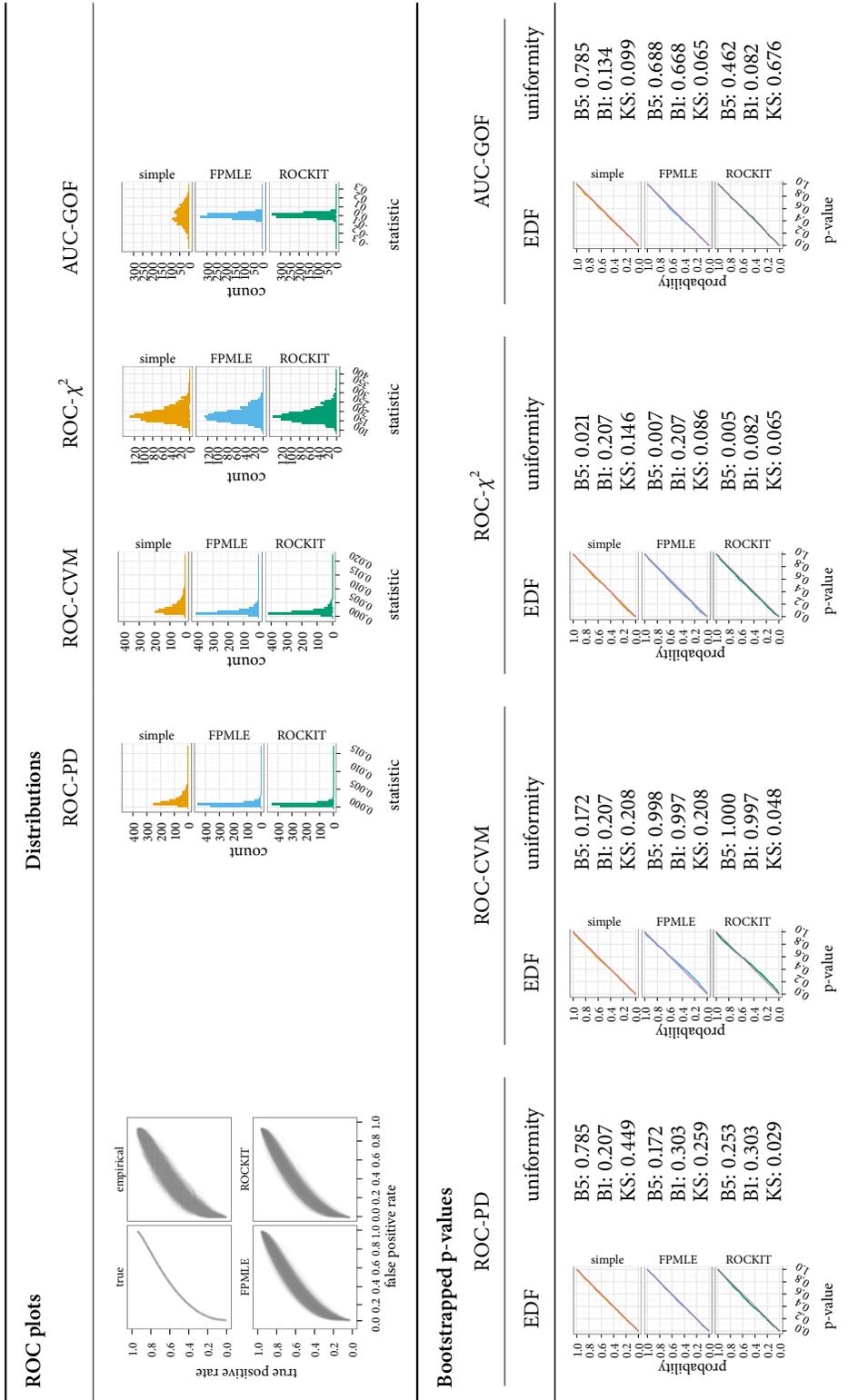


Figure A.45: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.75$, $\delta = 0.5$, $n_1 = 100$, $n_2 = 100$.

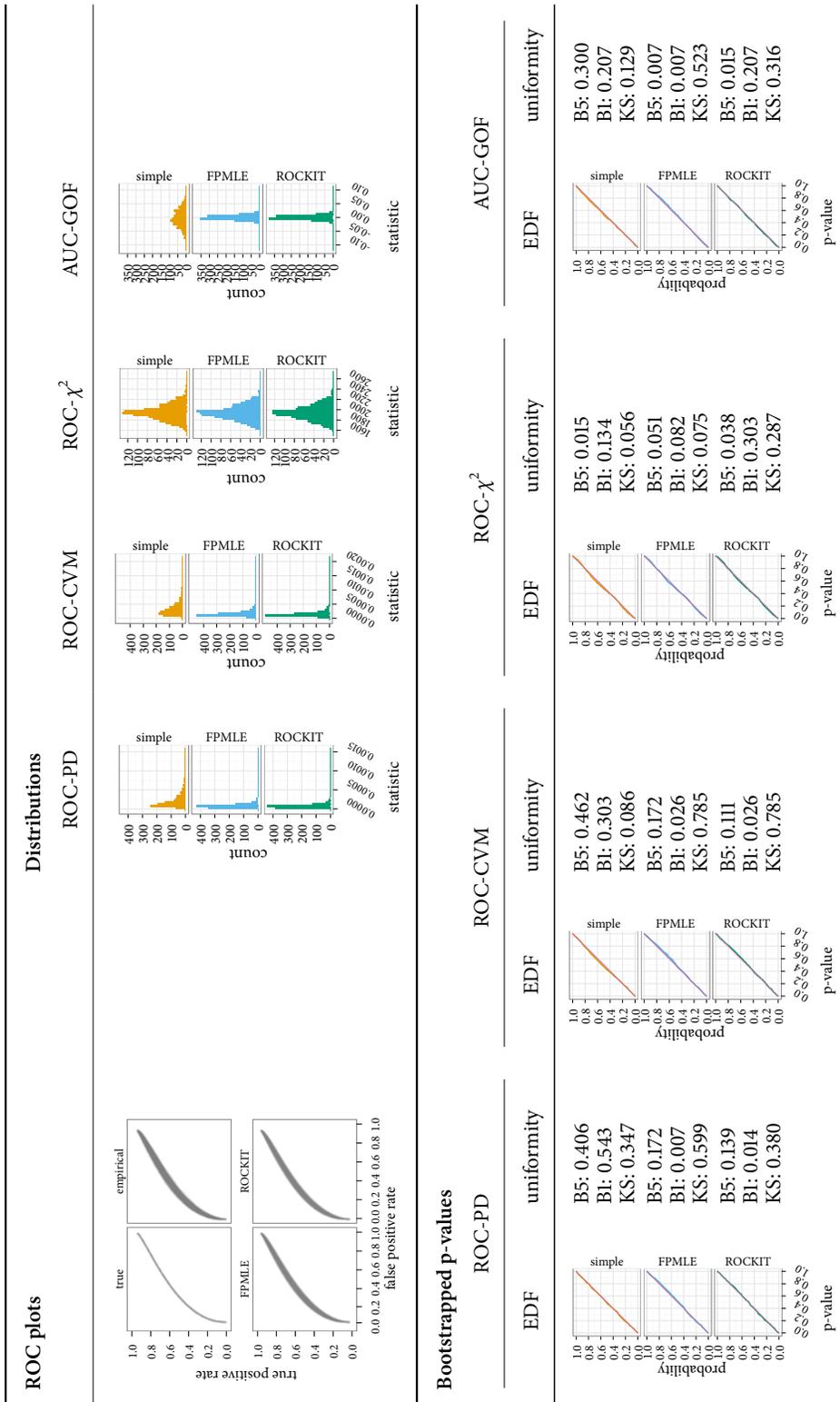


Figure A.46: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.75$, $\delta = 0.5$, $n_1 = 1000$, $n_2 = 1000$.

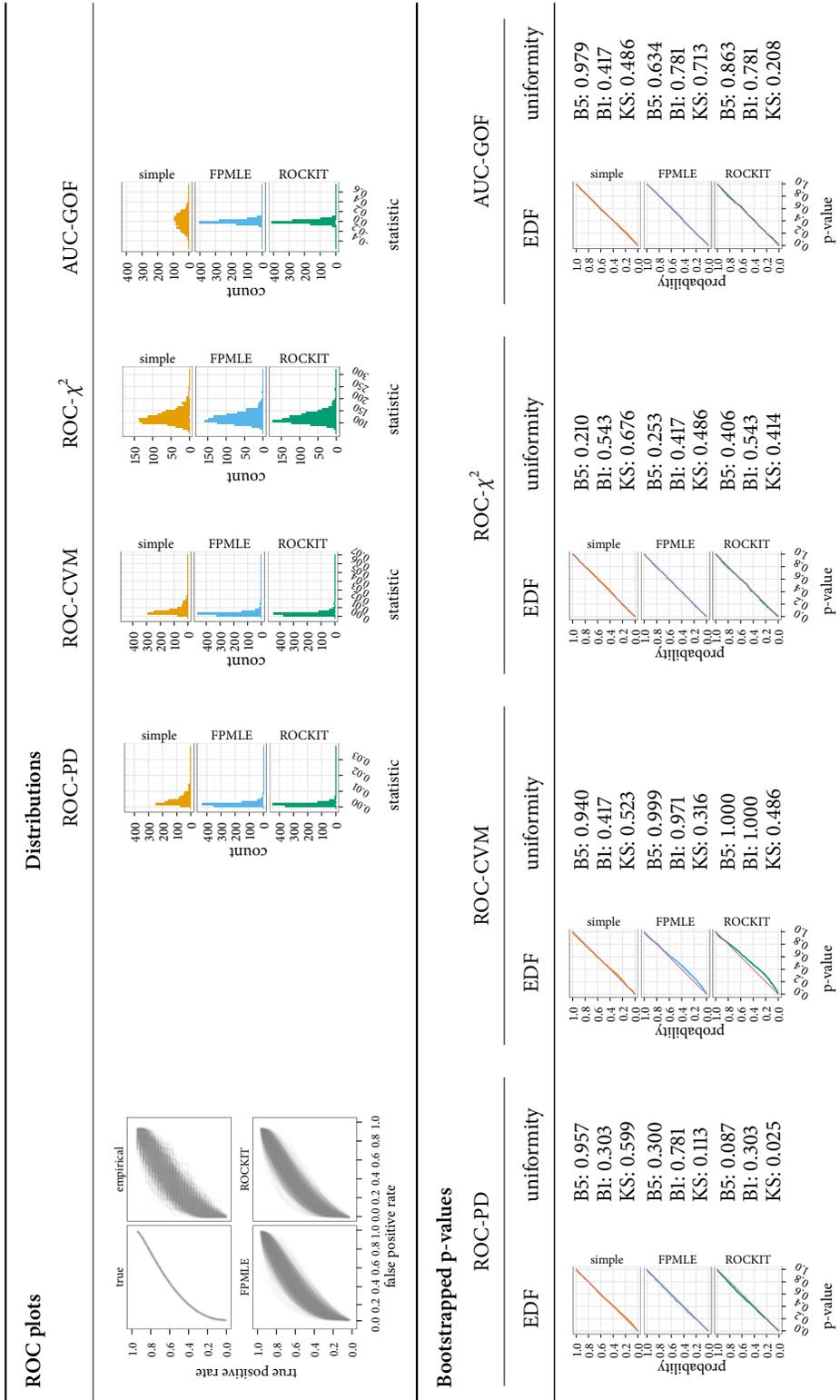


Figure A.47: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.75$, $\delta = 0.5$, $n_1 = 100$, $n_2 = 30$.

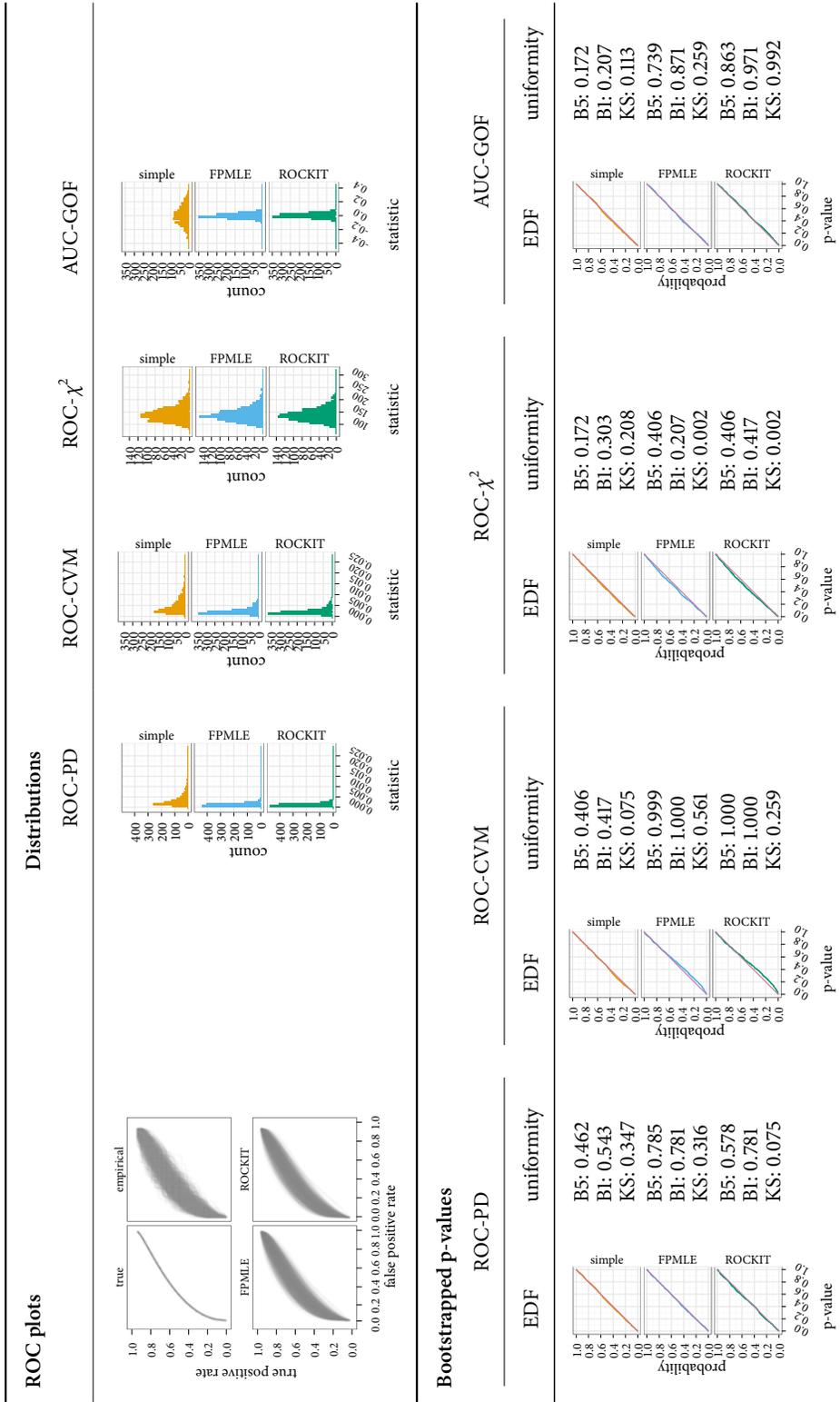


Figure A.48: Results for goodness-of-fit simulations for binormal ROCs. $\rho = 0.75$, $\delta = 0.5$, $n_1 = 100$, $n_2 = 50$.

A.2 Biexponential

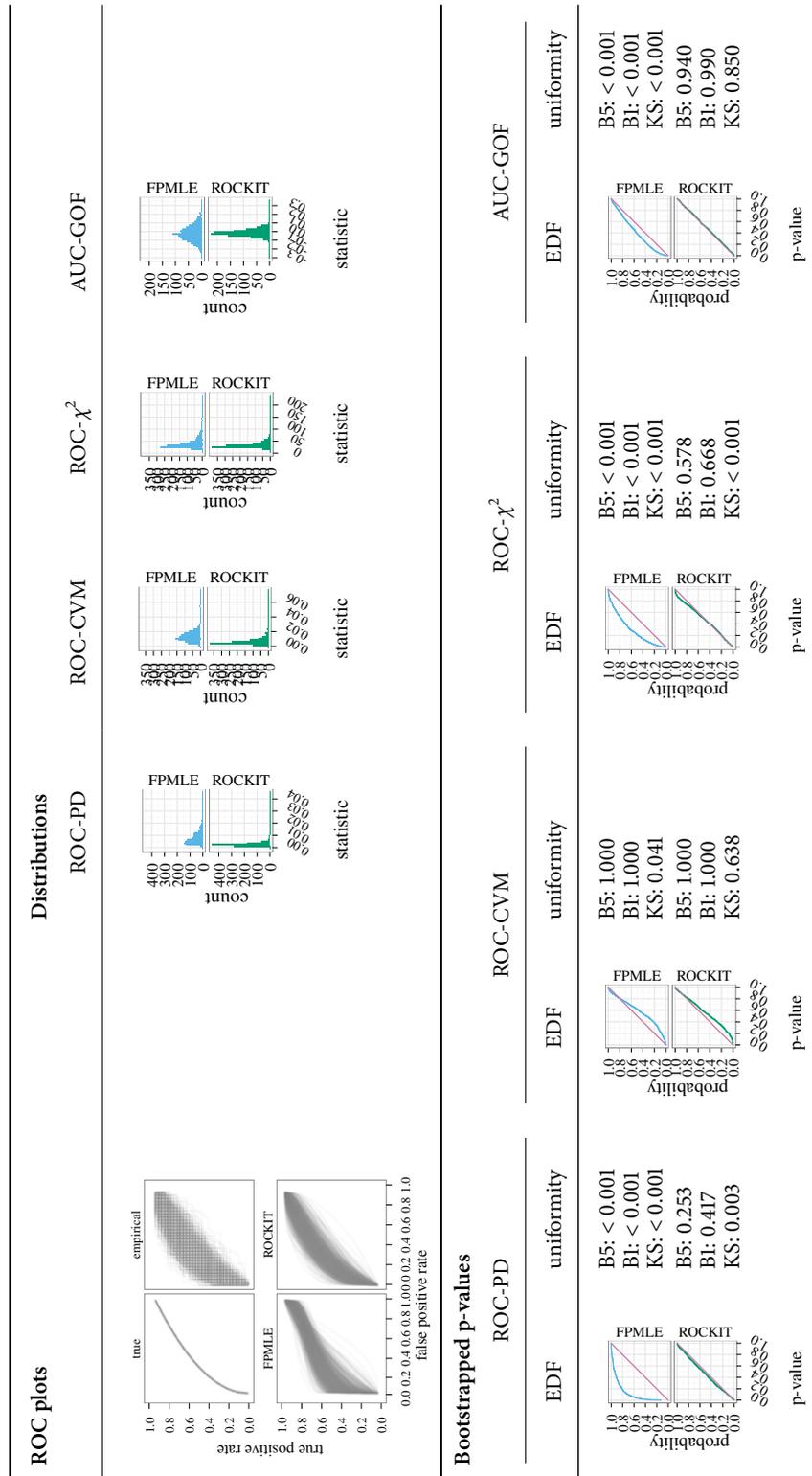


Figure A.49: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 30$, and $n_2 = 30$.

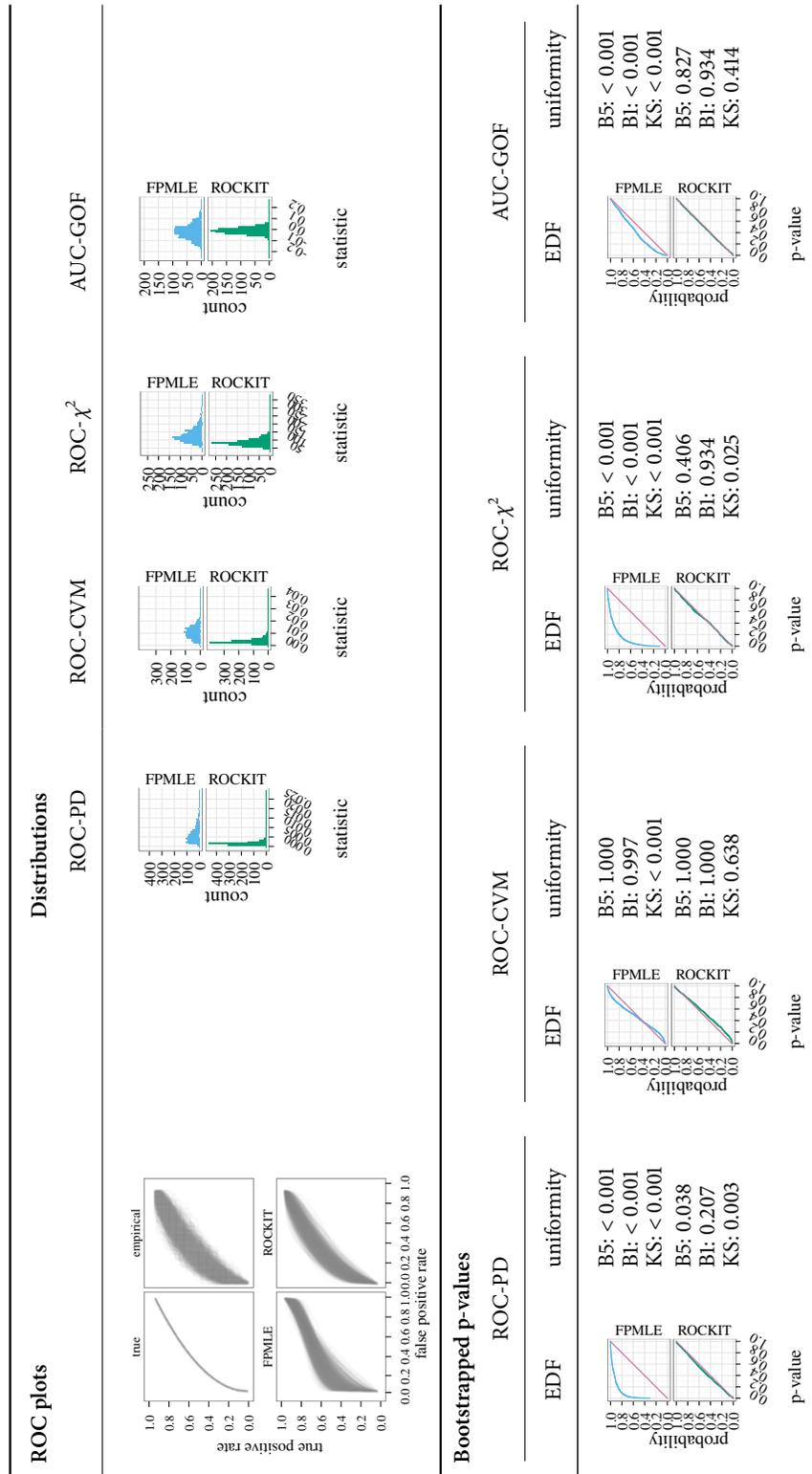


Figure A.50: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 50$, and $n_2 = 50$.

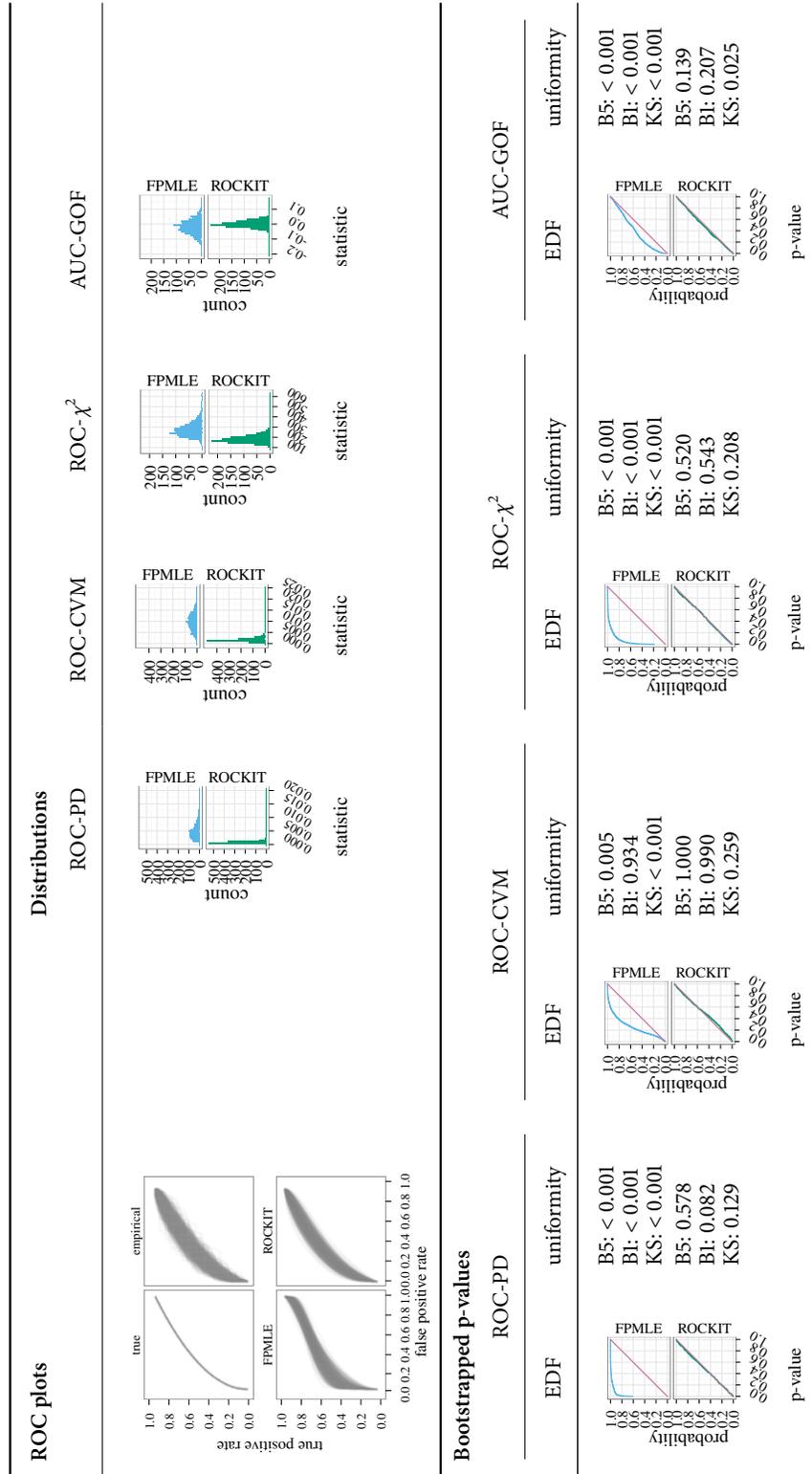


Figure A.51: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 100$, and $n_2 = 100$.

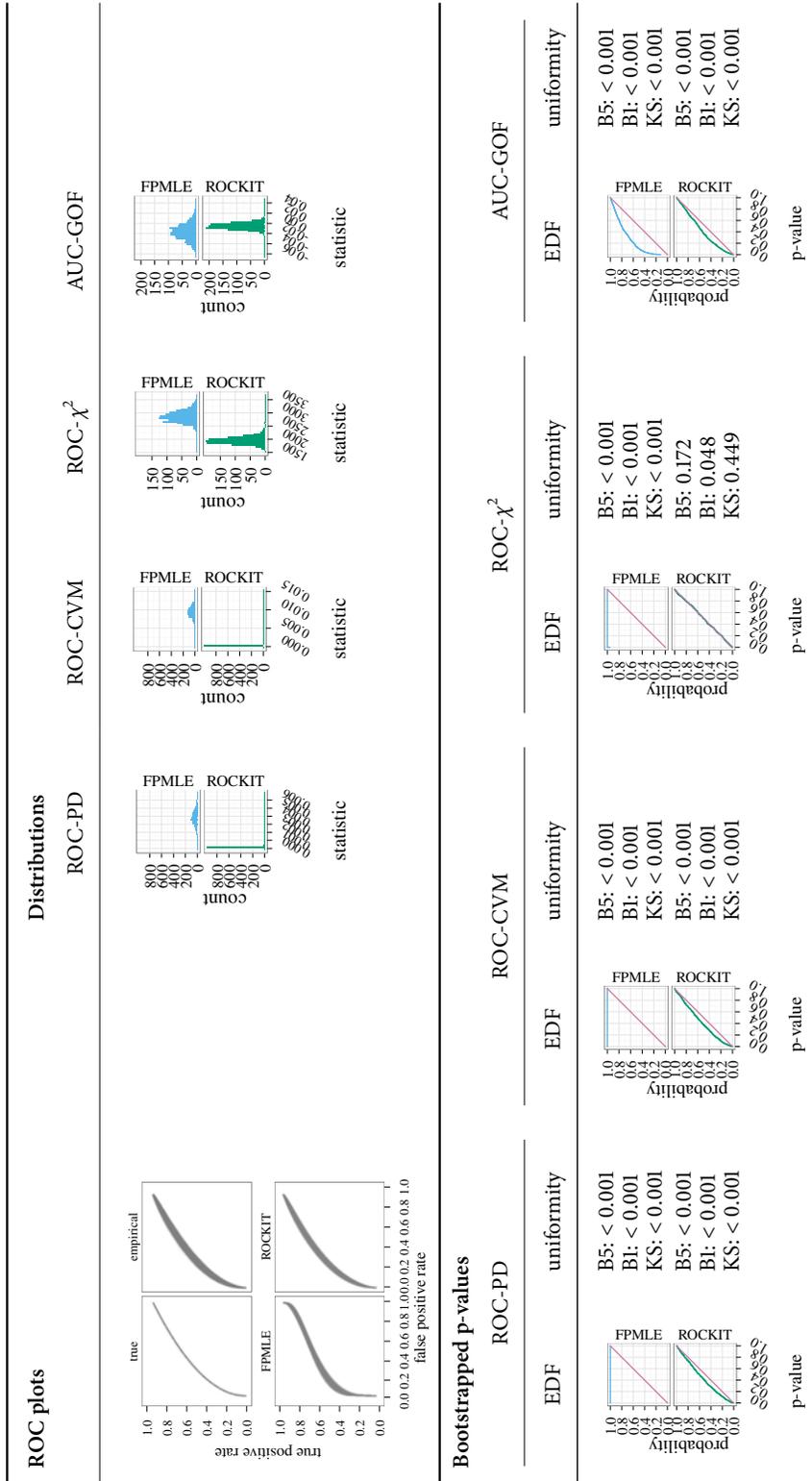


Figure A.52: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 1000$, and $n_2 = 1000$.

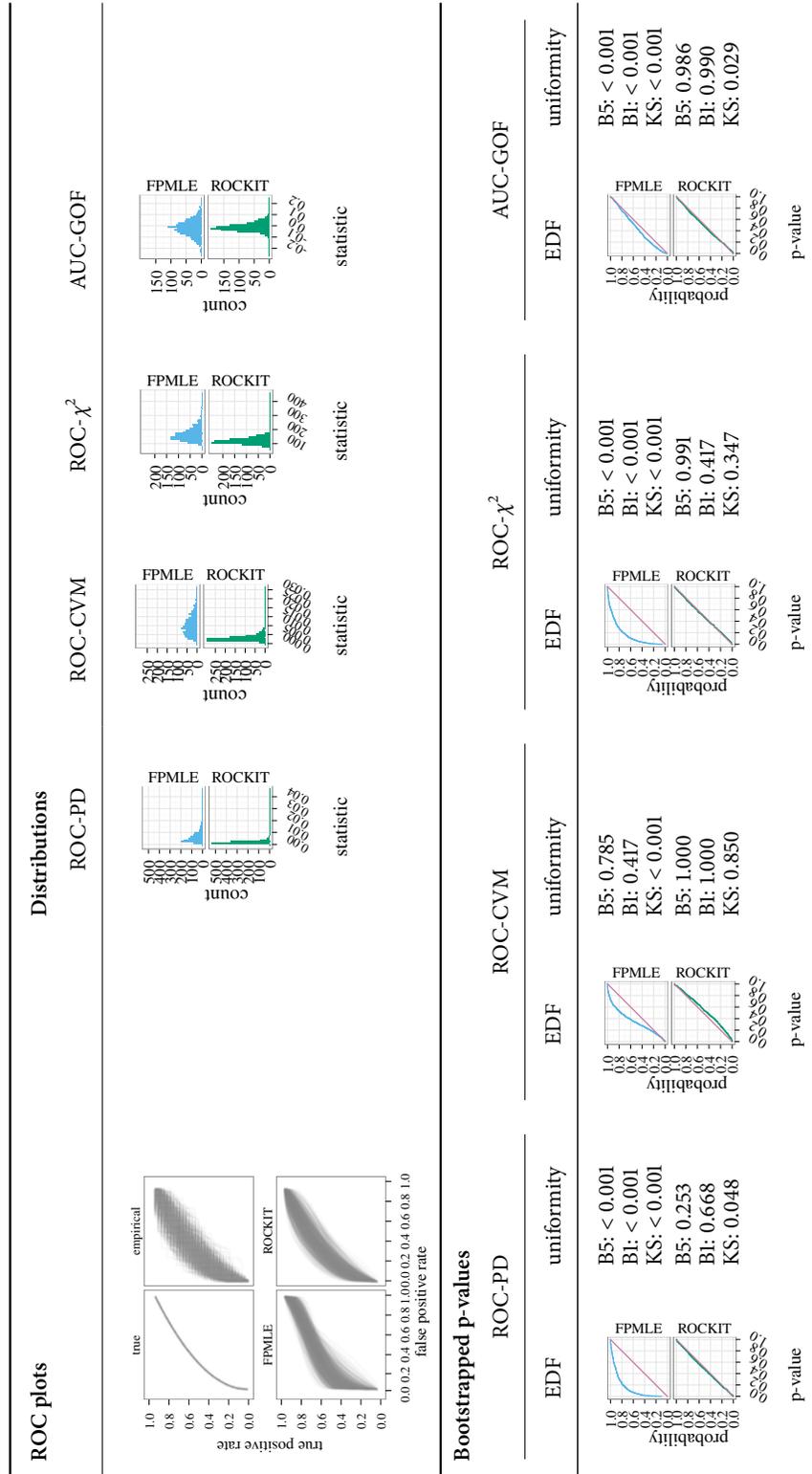


Figure A.53: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 100$, and $n_2 = 30$.

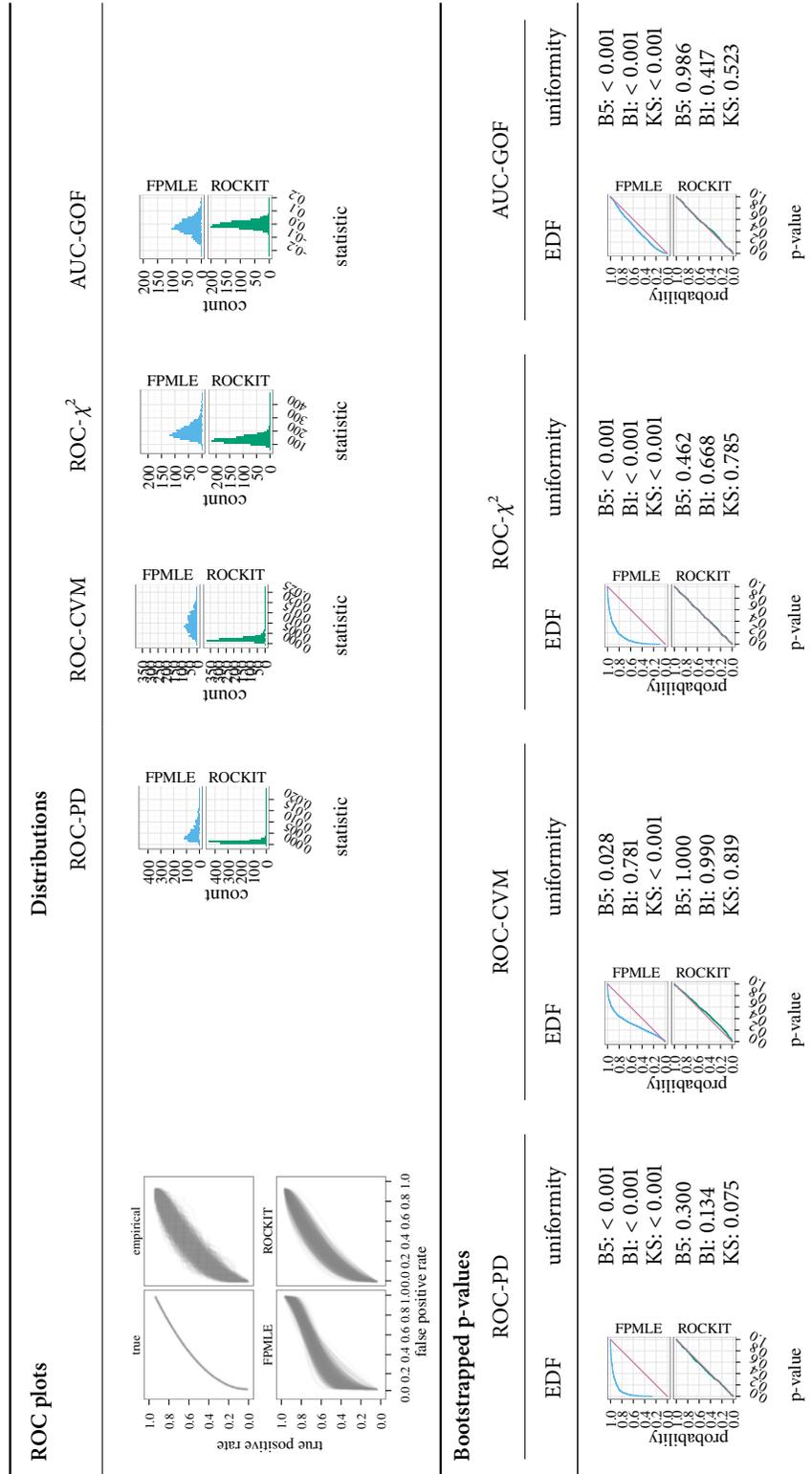


Figure A.5.4: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 100$, and $n_2 = 50$.

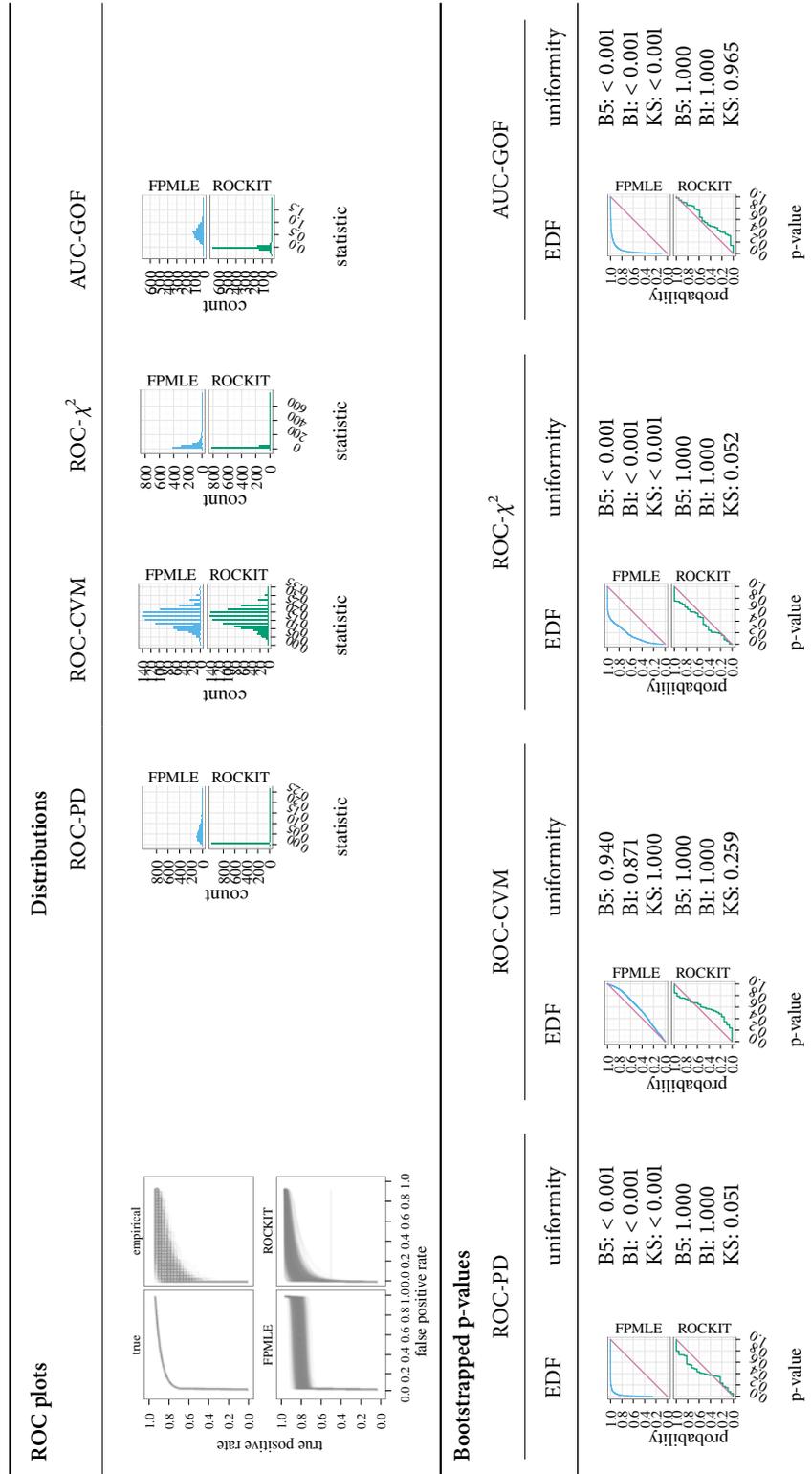


Figure A.55: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 15$, $n_1 = 30$, and $n_2 = 30$.

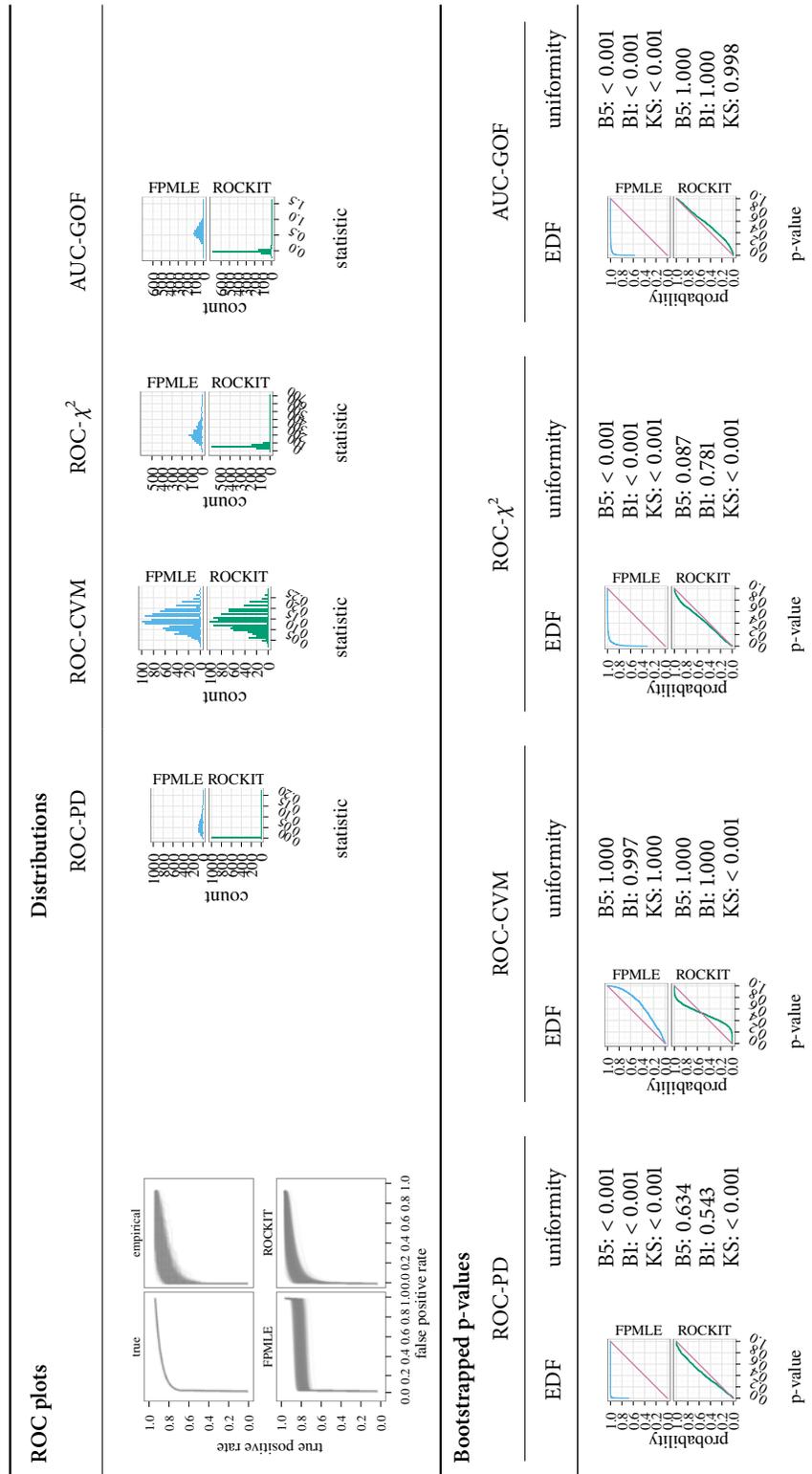


Figure A.56: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 15$, $n_1 = 50$, and $n_2 = 50$.

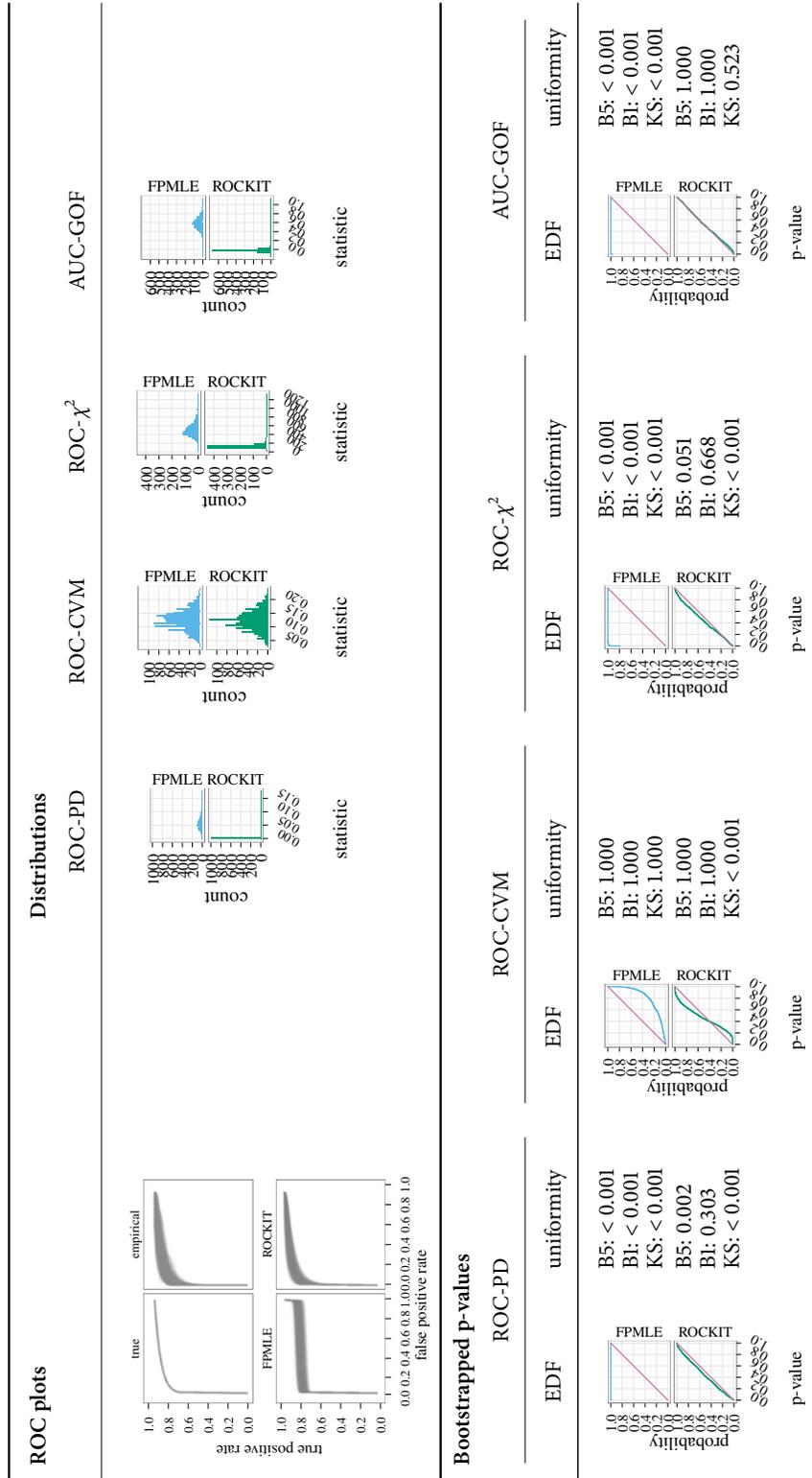


Figure A.57: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 15$, $n_1 = 100$, and $n_2 = 100$.

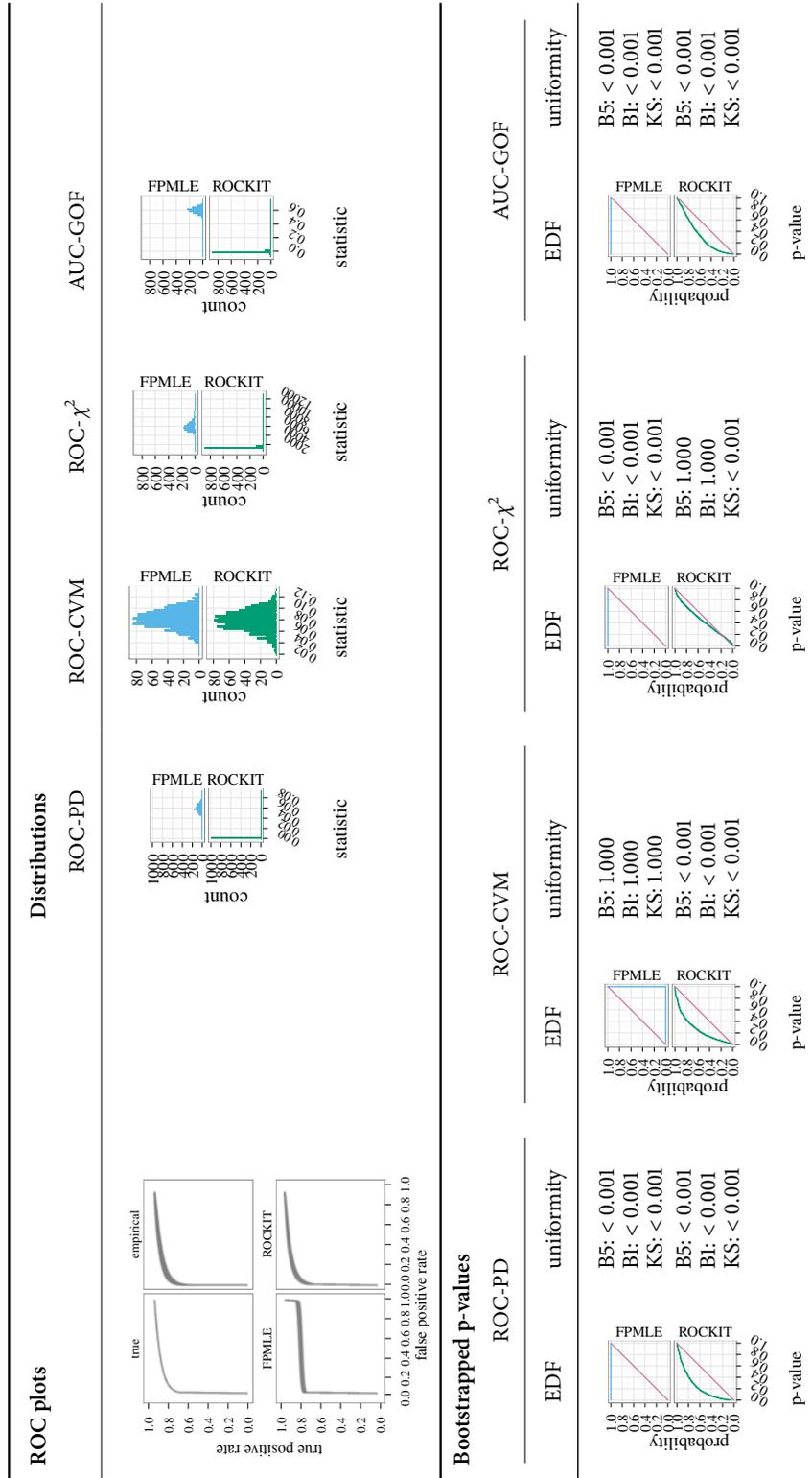


Figure A.58: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 15$, $n_1 = 1000$, and $n_2 = 1000$.

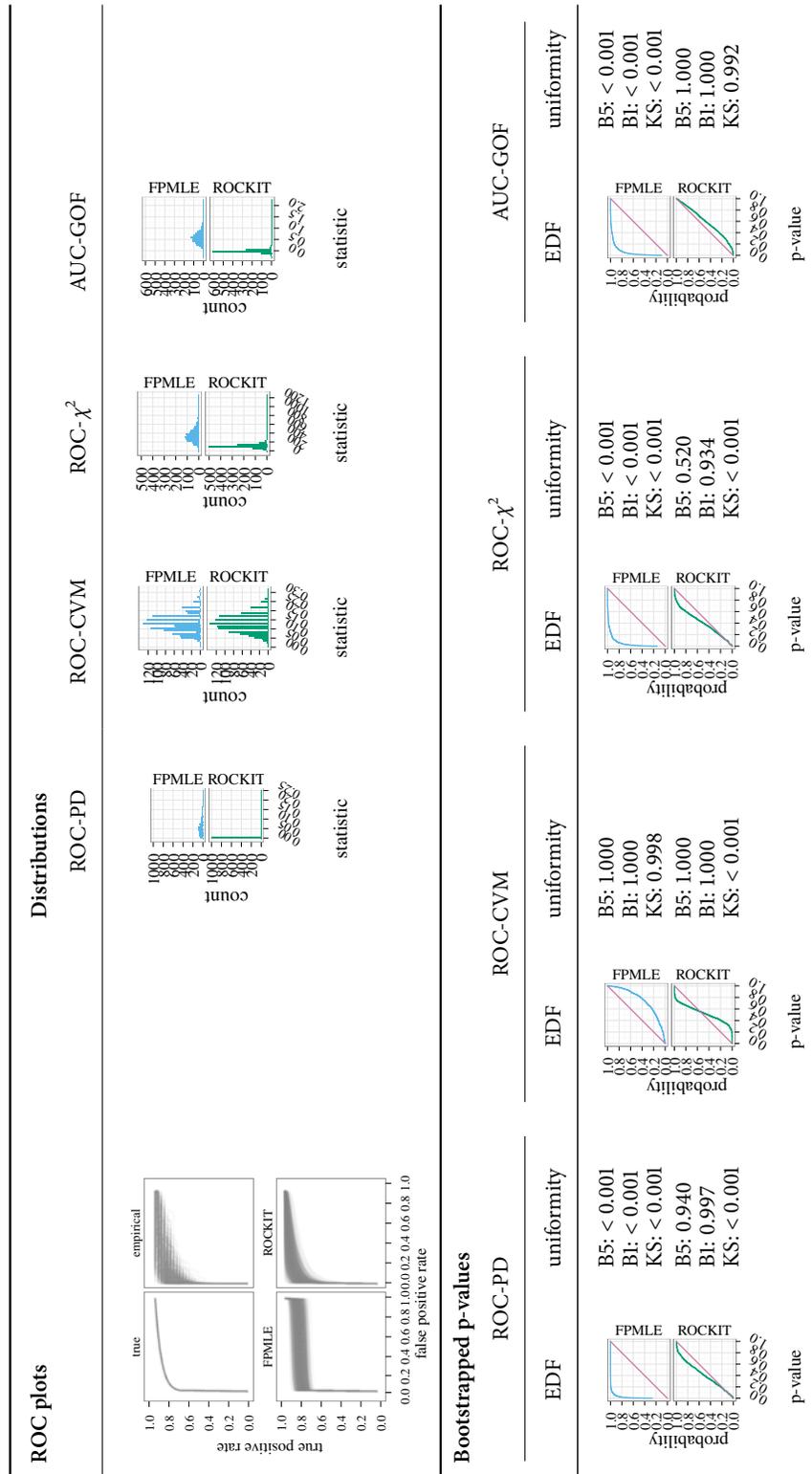


Figure A.59: Results for goodness-of-fit simulations for biexponential ROCs. $\lambda_1 = 1$, $\lambda_2 = 15$, $n_1 = 100$, and $n_2 = 30$.

A.3 Normal-gamma

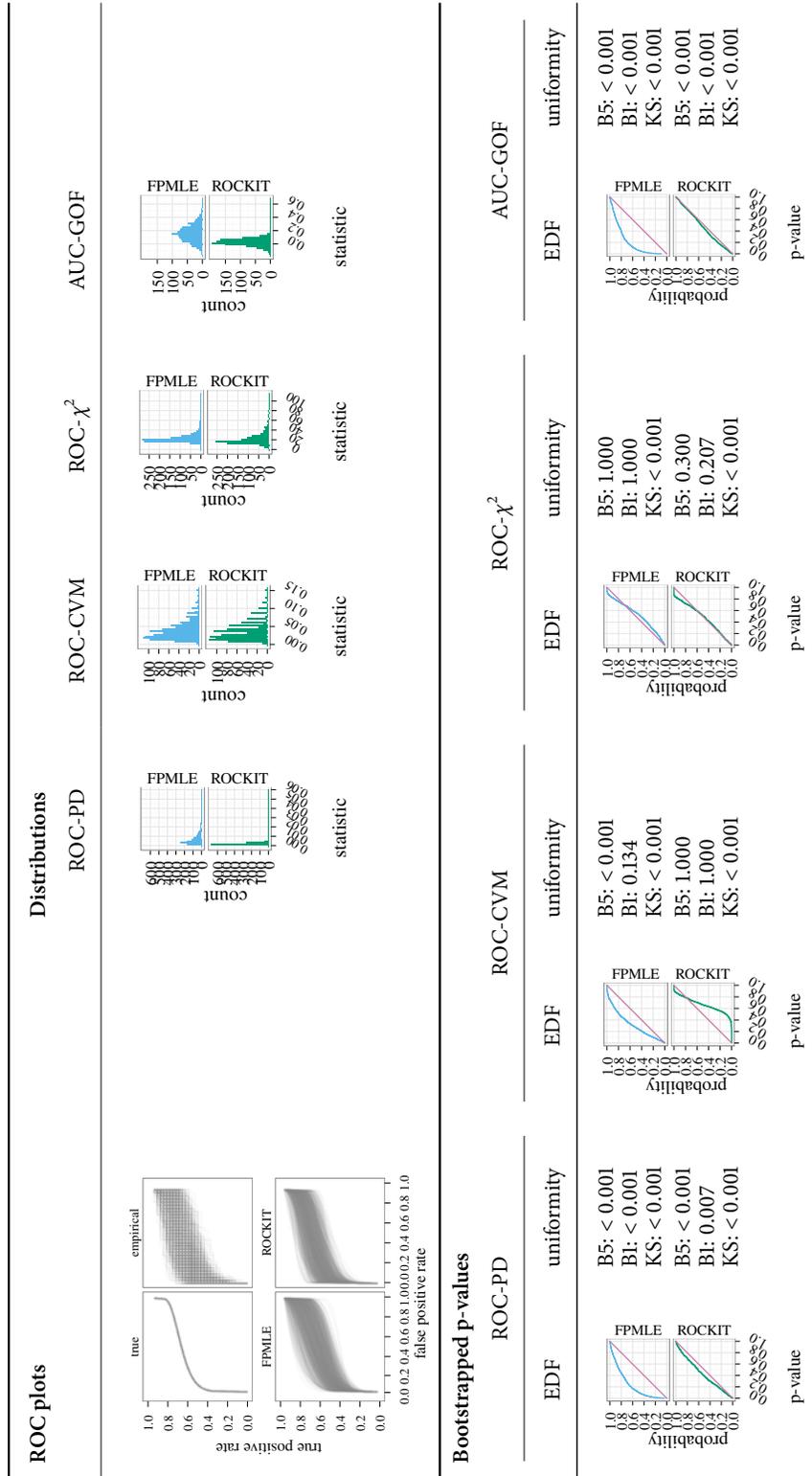


Figure A.61: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0, \sigma_1 = 1, \alpha_2 = 2, \beta_2 = 0.5, \theta_2 = 5, n_1 = 30, \text{ and } n_2 = 30.$

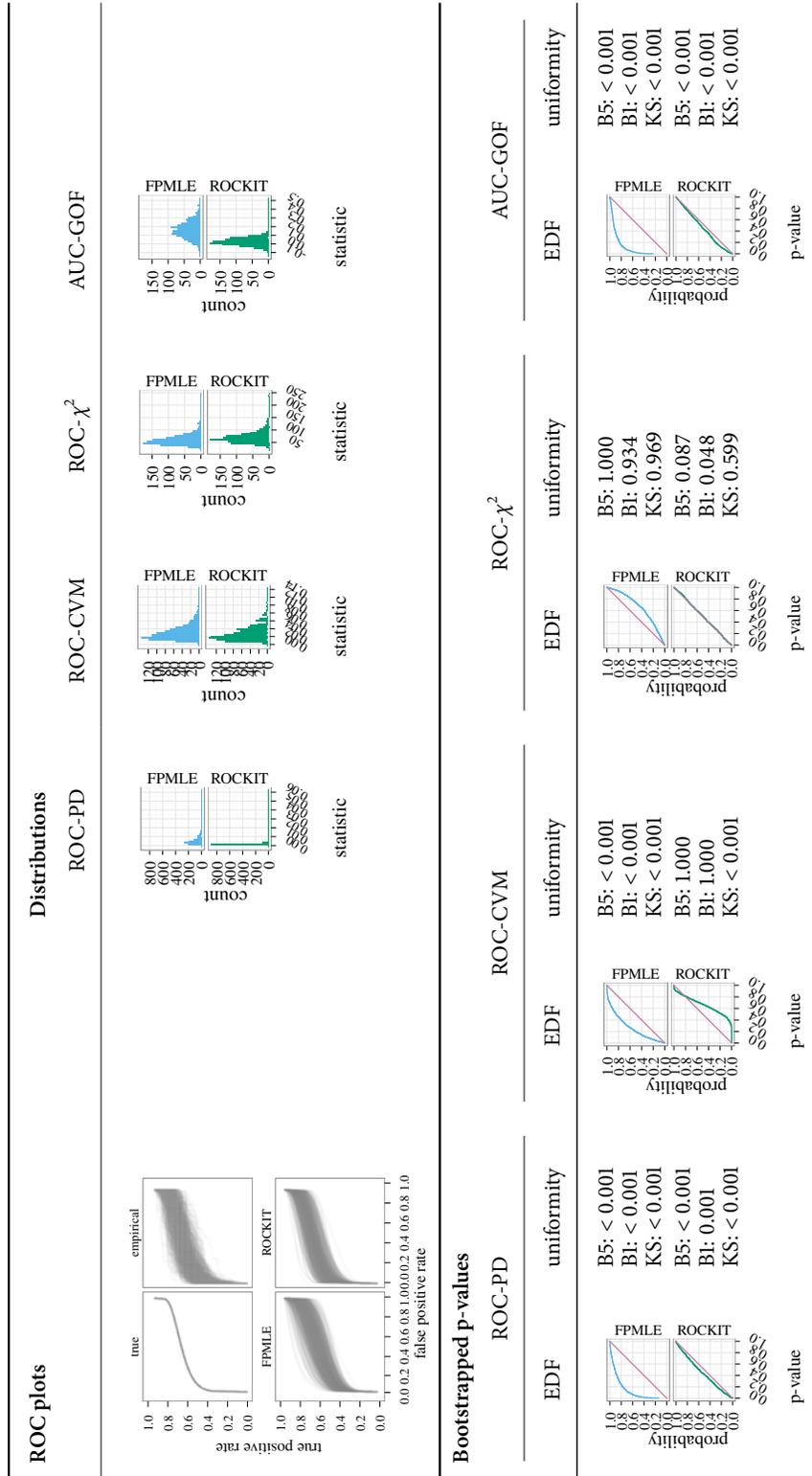


Figure A.62: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 2$, $\beta_2 = 0.5$, $\theta_2 = 5$, $n_1 = 50$, and $n_2 = 50$.

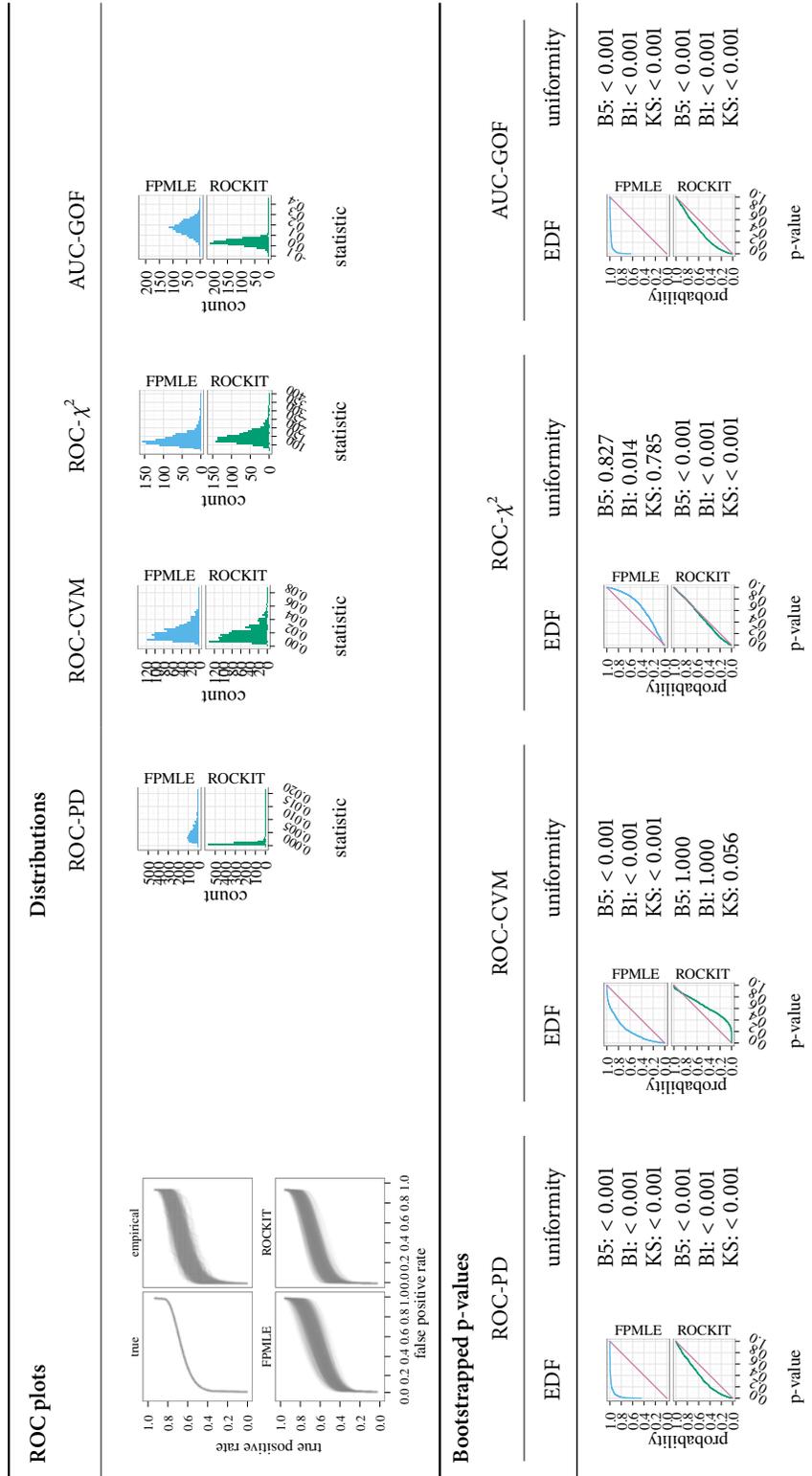


Figure A.63: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 2$, $\beta_2 = 0.5$, $\theta_2 = 5$, $n_1 = 100$, and $n_2 = 100$.

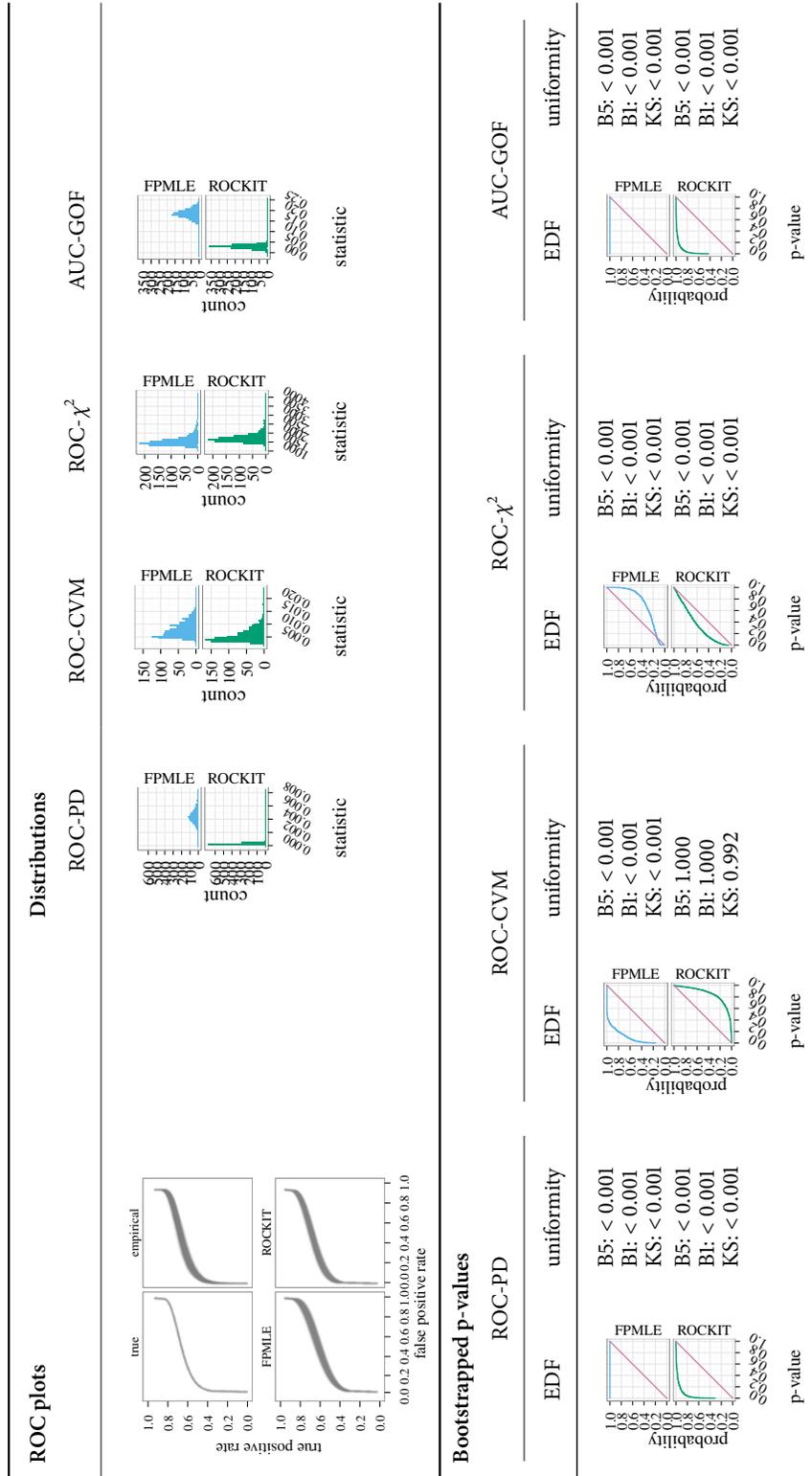


Figure A.64: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 2$, $\beta_2 = 0.5$, $\theta_2 = 5$, $n_1 = 1000$, and $n_2 = 1000$.

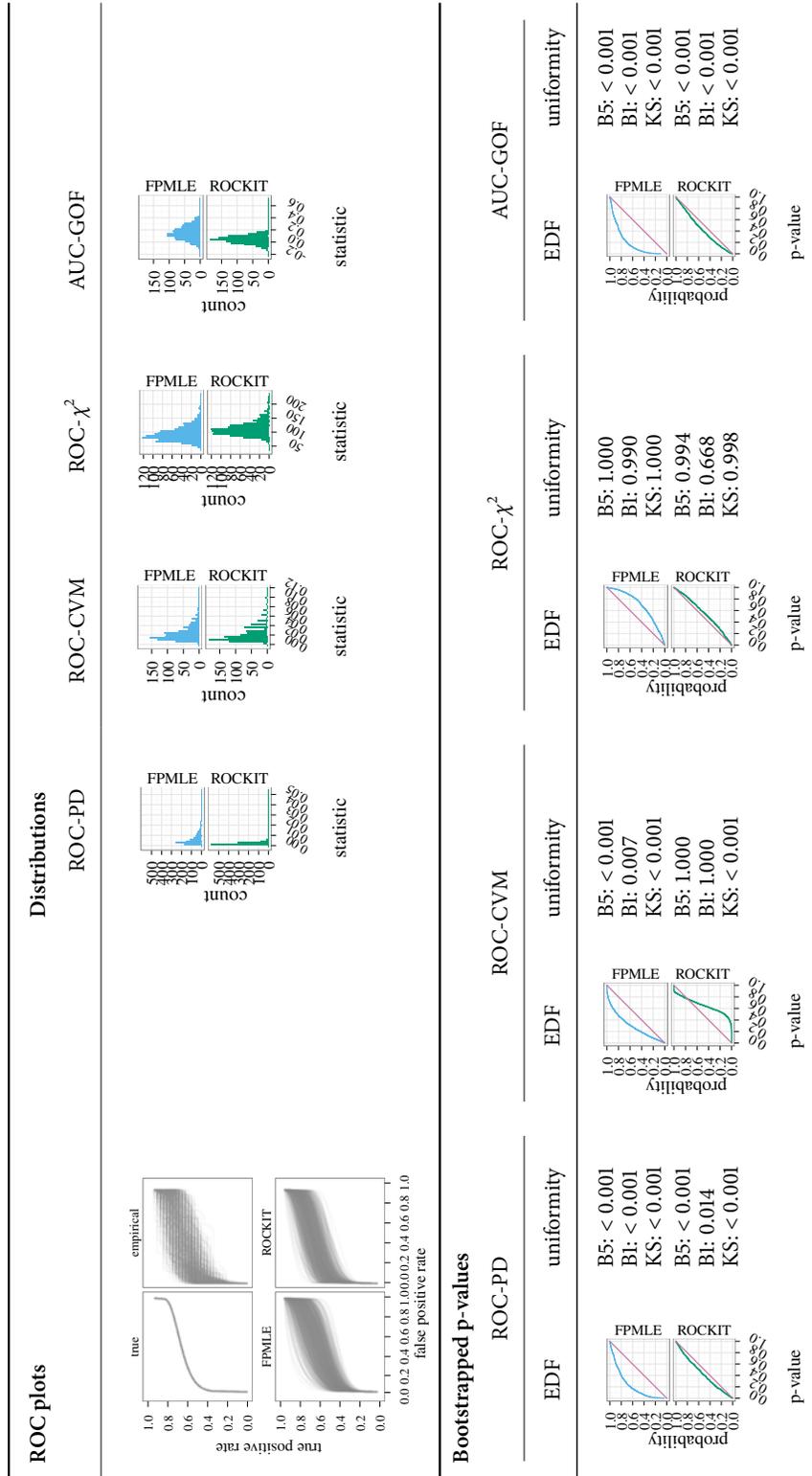


Figure A.65: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 2$, $\beta_2 = 0.5$, $\theta_2 = 5$, $n_1 = 100$, and $n_2 = 30$.

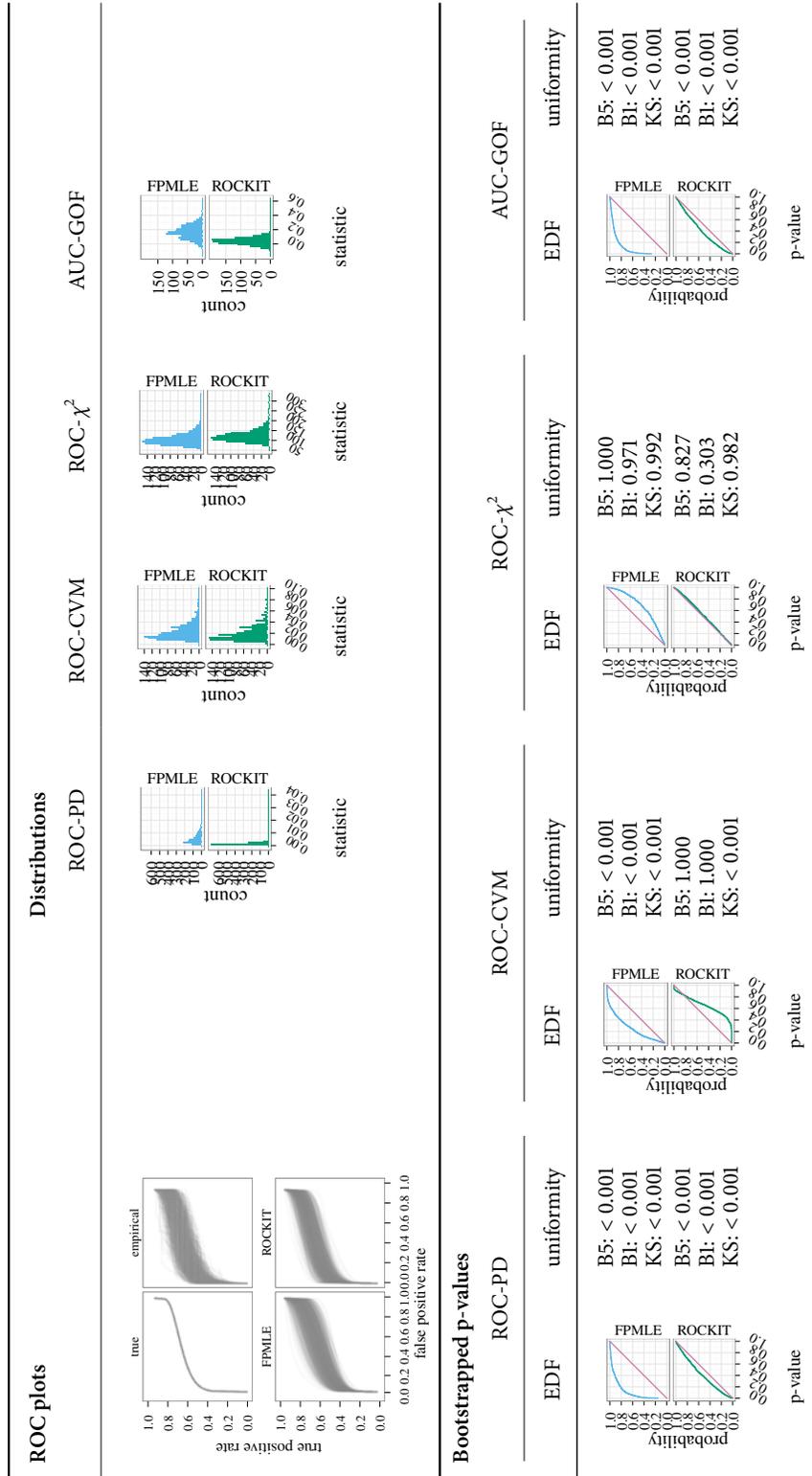


Figure A.66: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 2$, $\beta_2 = 0.5$, $\theta_2 = 5$, $n_1 = 100$, and $n_2 = 50$.

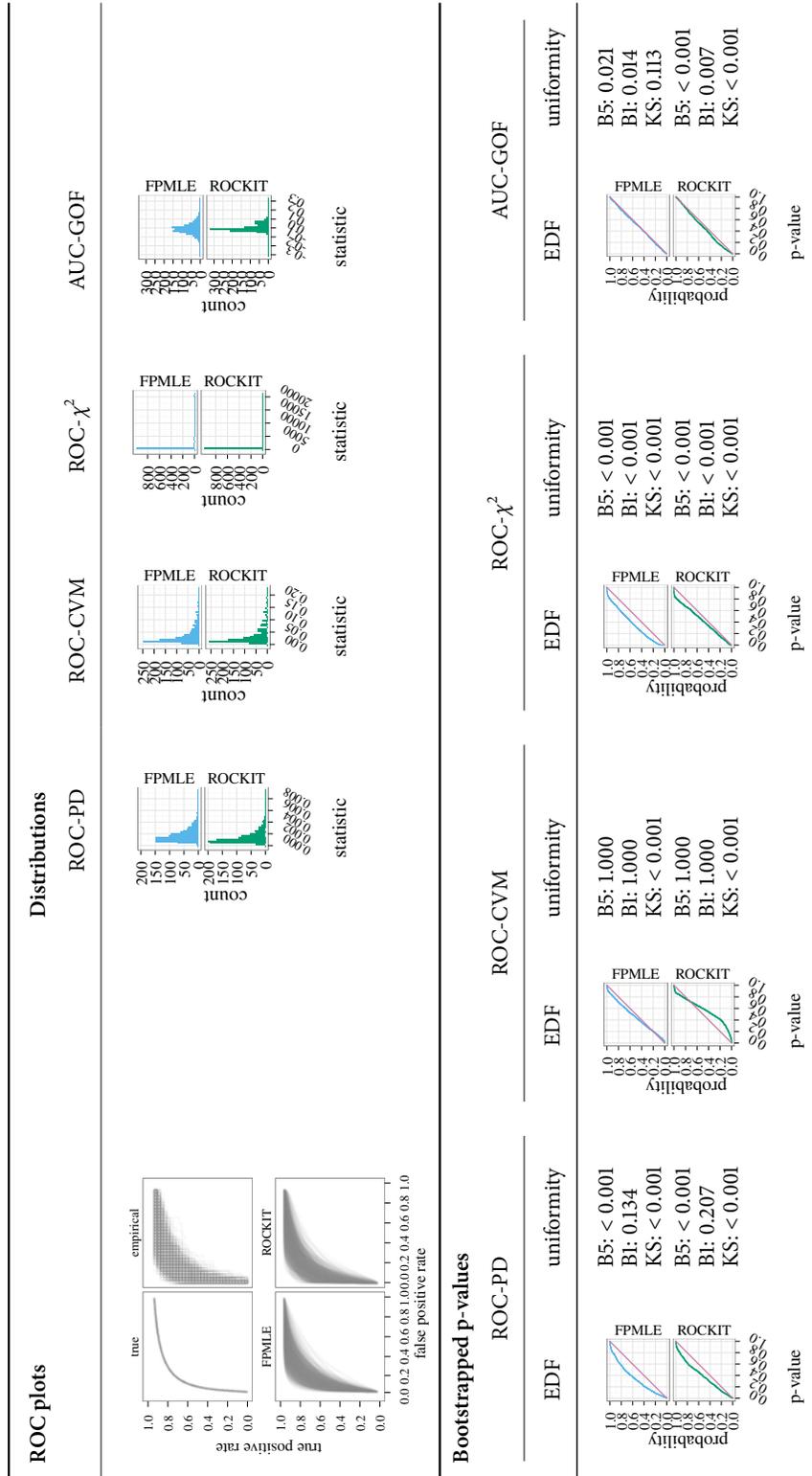


Figure A.67: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 3$, $\beta_2 = 2$, $\theta_2 = 3$, $n_1 = 30$, and $n_2 = 30$.

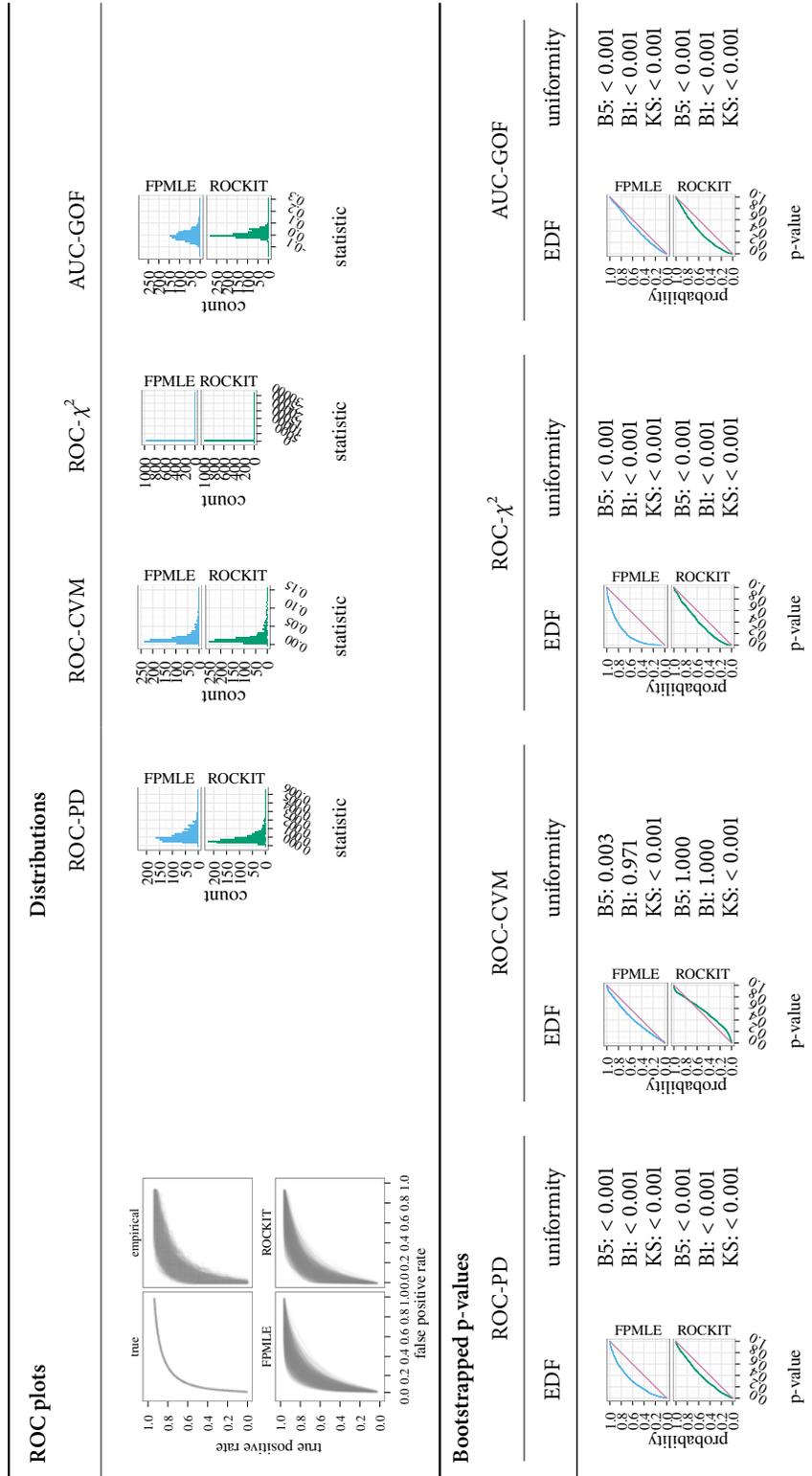


Figure A.68: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 3$, $\beta_2 = 2$, $\theta_2 = 3$, $n_1 = 50$, and $n_2 = 50$.

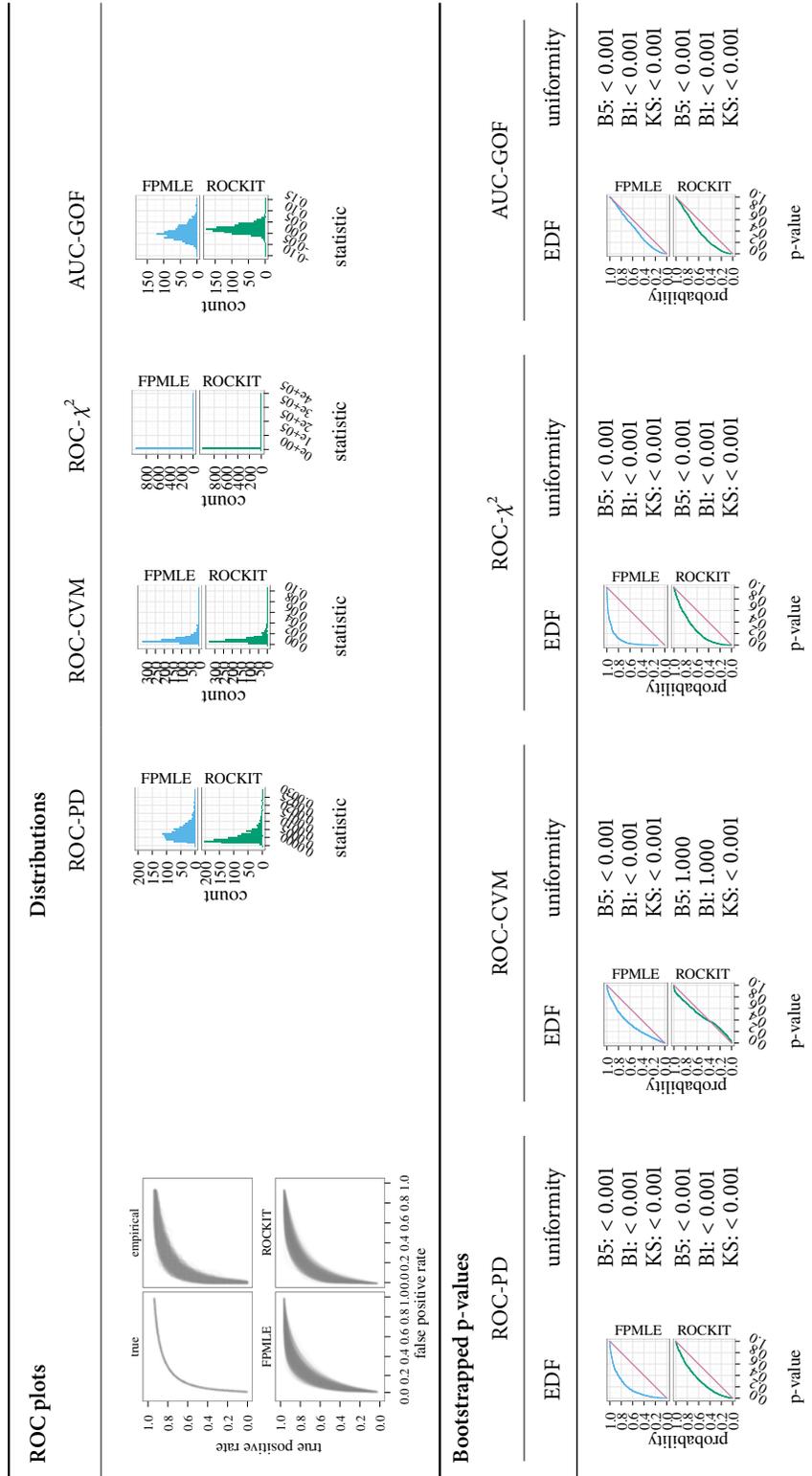


Figure A.69: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 3$, $\beta_2 = 2$, $\theta_2 = 3$, $n_1 = 100$, and $n_2 = 100$.

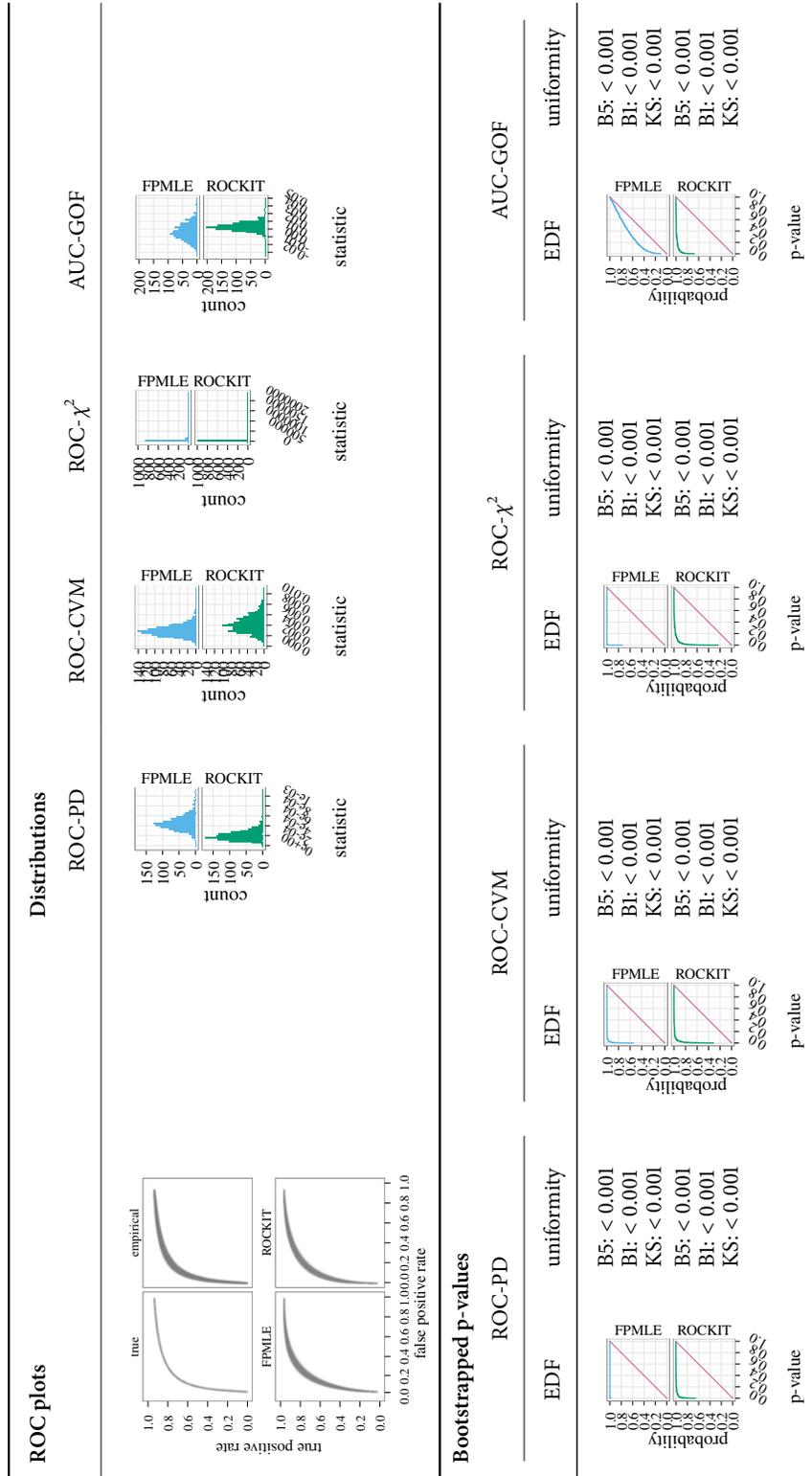


Figure A.70: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 3$, $\beta_2 = 2$, $\theta_2 = 3$, $n_1 = 1000$, and $n_2 = 1000$.

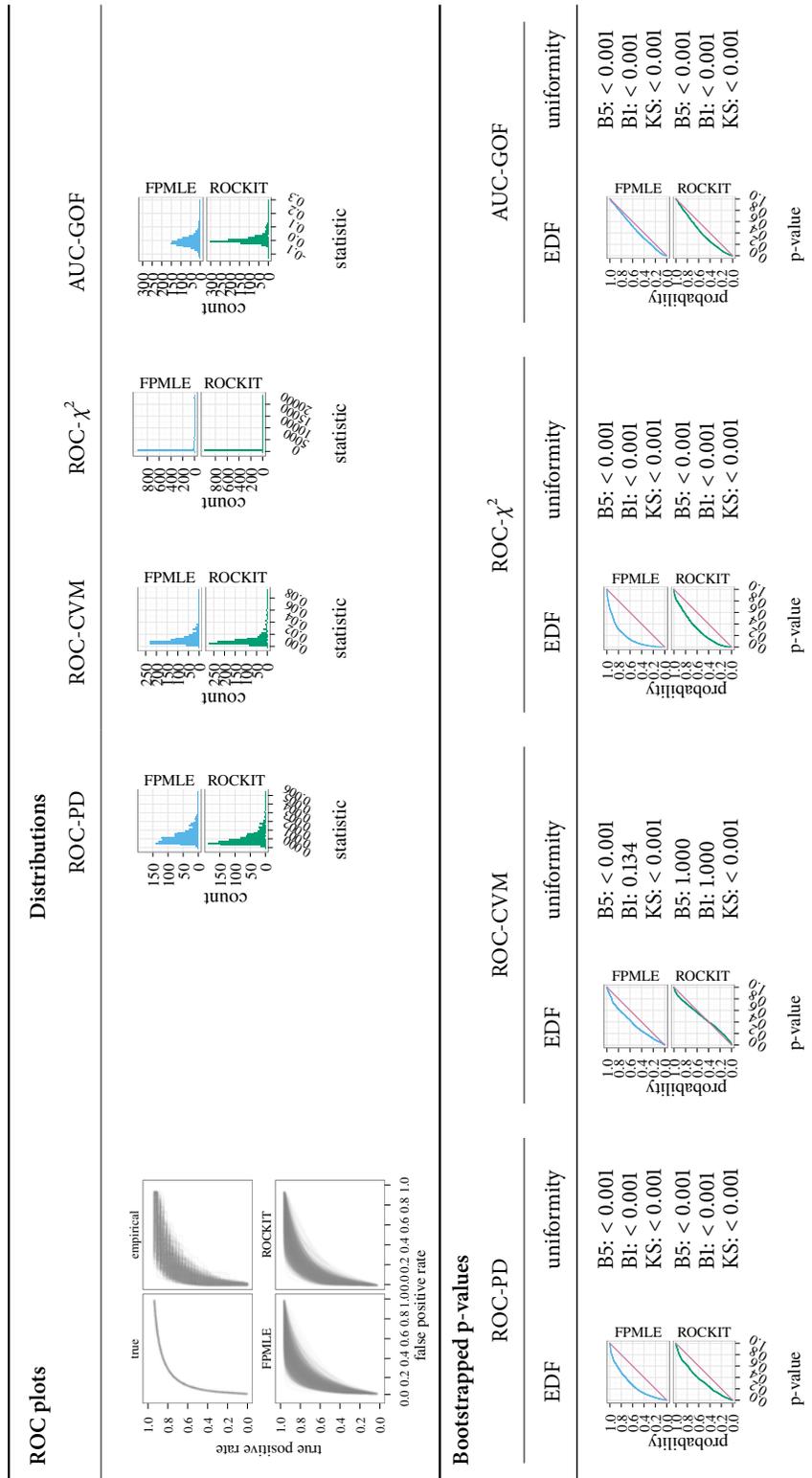


Figure A.71: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 3$, $\beta_2 = 2$, $\theta_2 = 3$, $n_1 = 100$, and $n_2 = 30$.

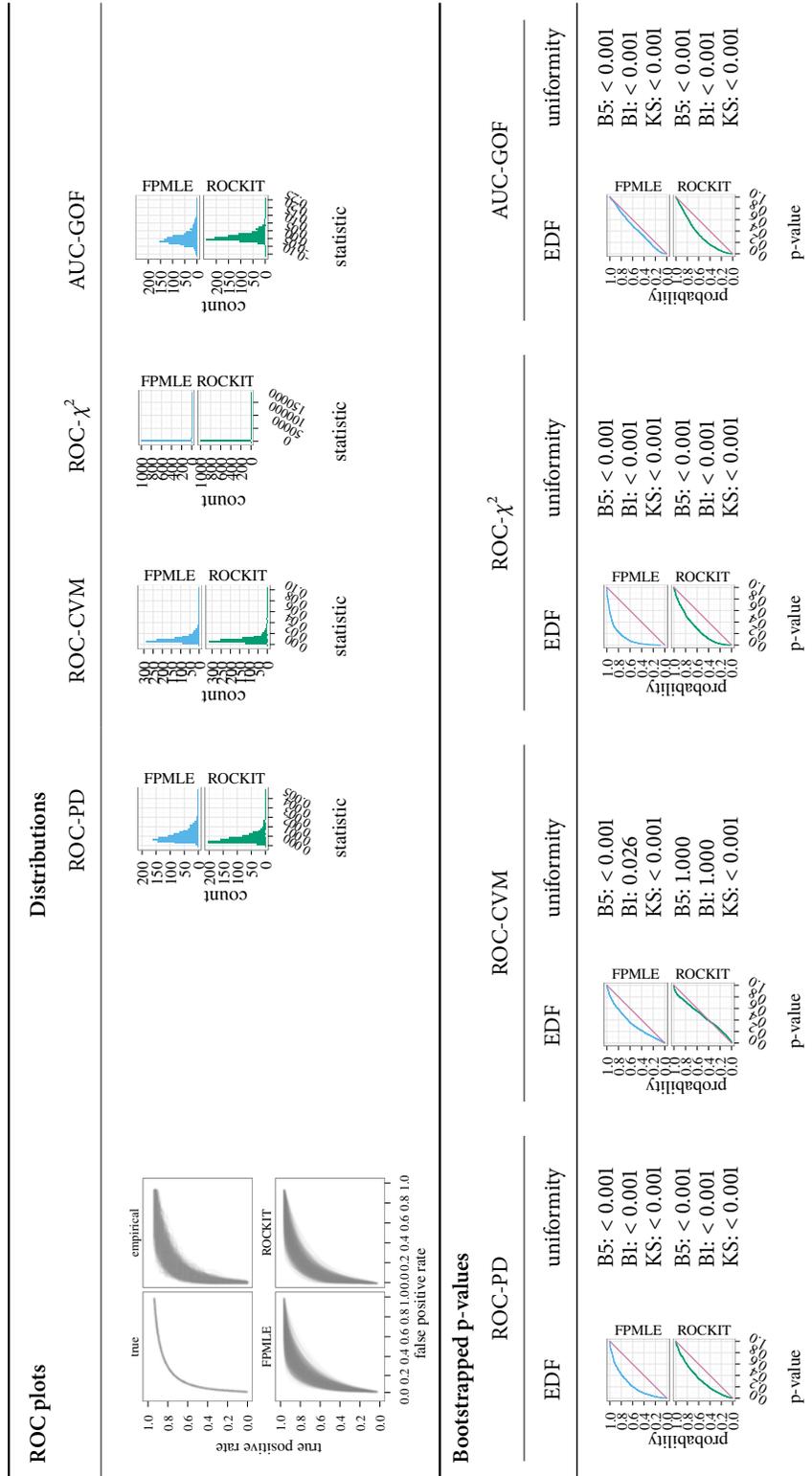


Figure A.72: Results for goodness-of-fit simulations for normal-gamma ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\alpha_2 = 3$, $\beta_2 = 2$, $\theta_2 = 3$, $n_1 = 100$, and $n_2 = 50$.

A.4 Normal-mixture

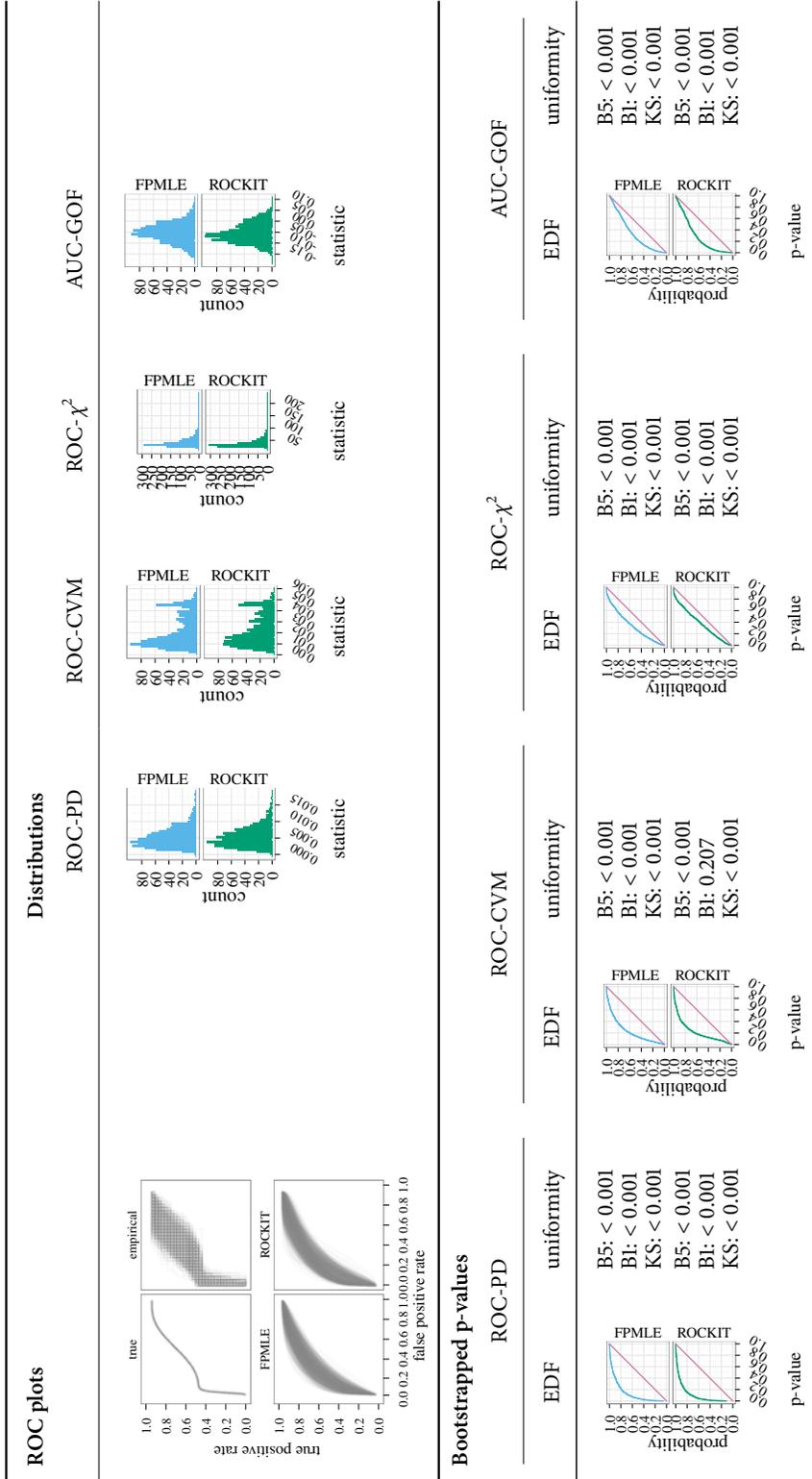


Figure A.73: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 0$, $\sigma_2 = 0.5$, $\mu_3 = 2$, $\sigma_3 = 0.25$, $\alpha = 0.5$, $n_1 = 30$, $n_2 = 30$, and $n_3 = 30$.

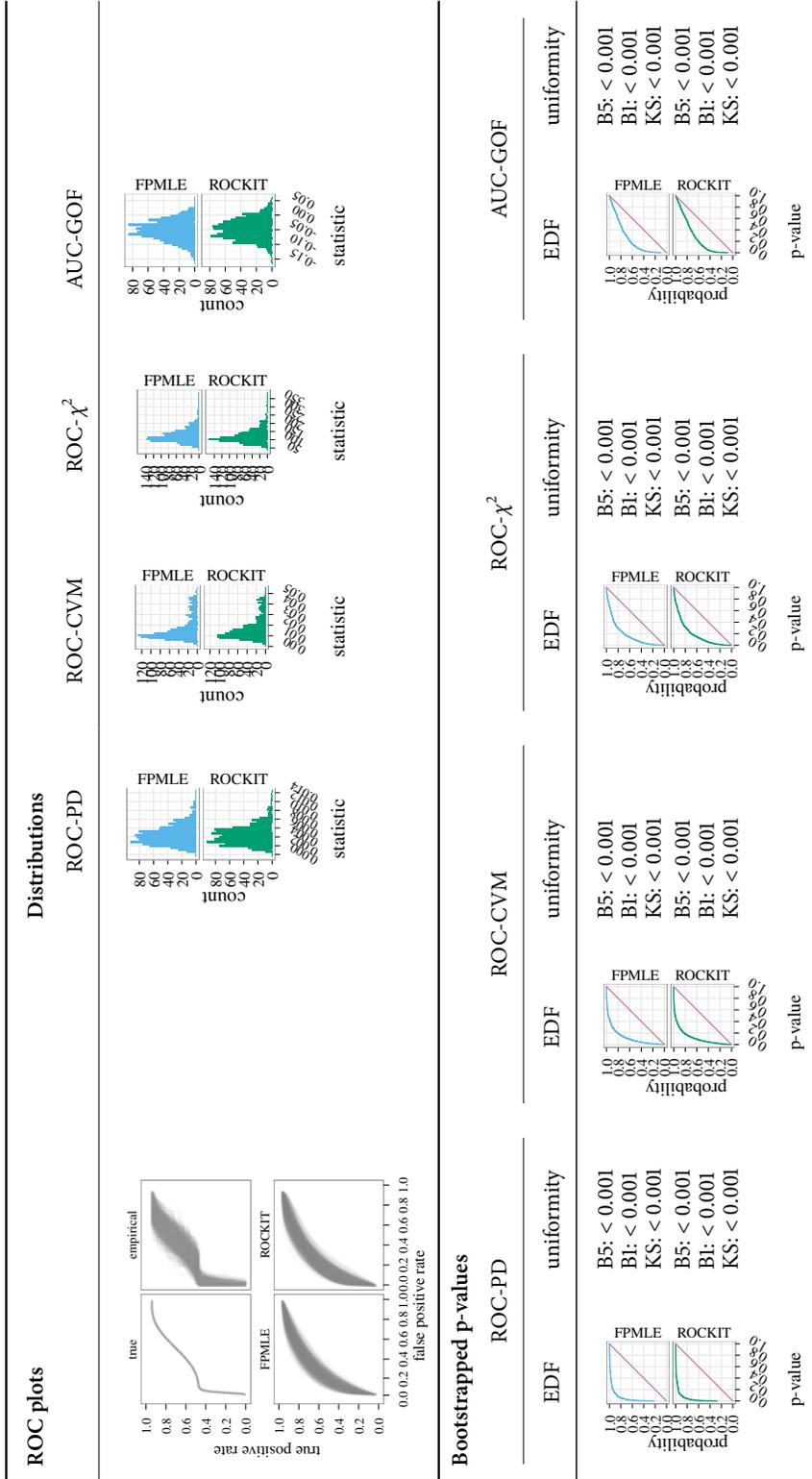


Figure A.74: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 0$, $\sigma_2 = 0.5$, $\mu_3 = 2$, $\sigma_3 = 0.25$, $\alpha = 0.5$, $n_1 = 50$, $n_2 = 50$, and $n_3 = 50$.

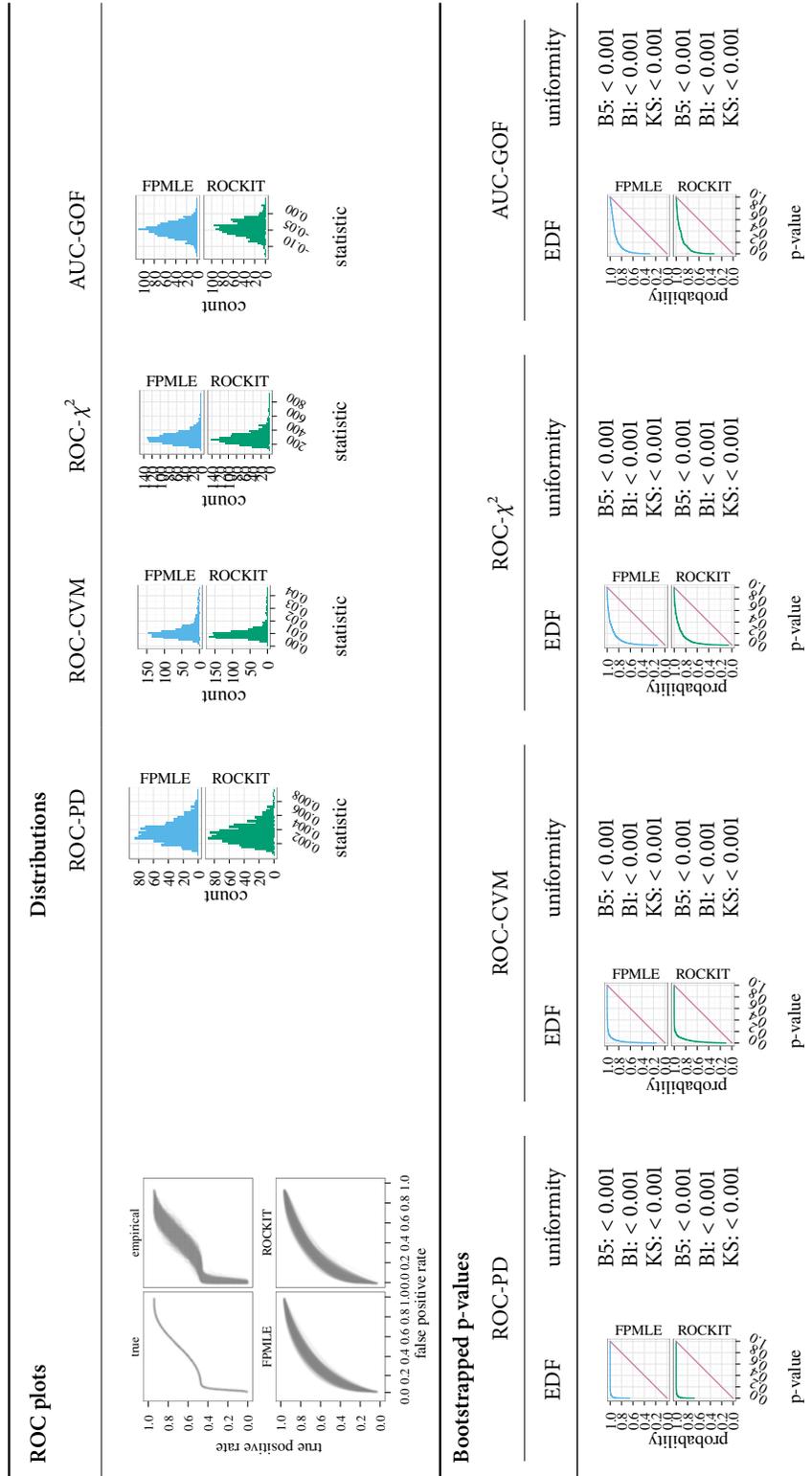


Figure A.75: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 0$, $\sigma_2 = 0.5$, $\mu_3 = 2$, $\sigma_3 = 0.25$, $\alpha = 0.5$, $n_1 = 100$, and $n_2 = 100$.

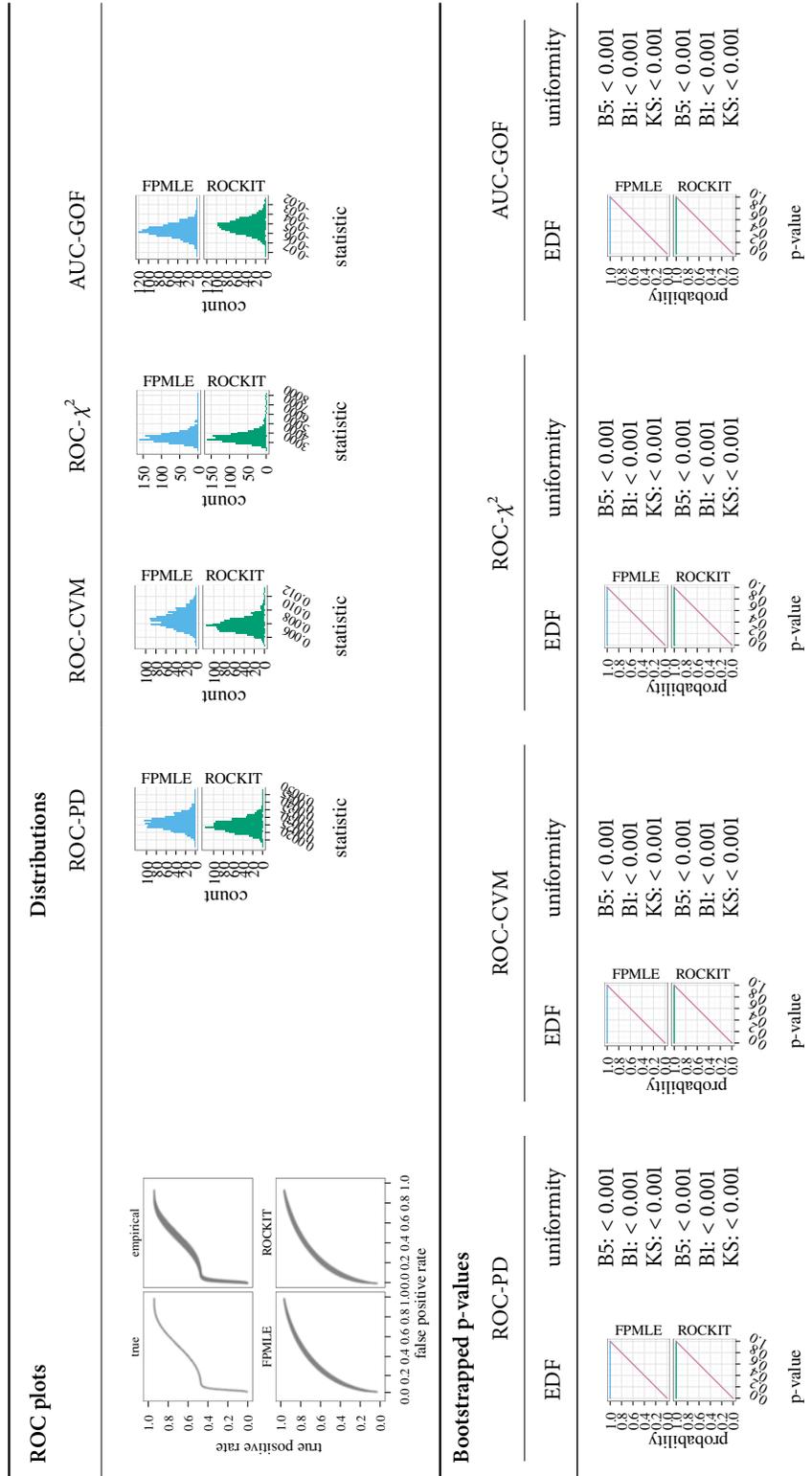


Figure A.76: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 0$, $\sigma_2 = 0.5$, $\mu_3 = 2$, $\sigma_3 = 0.25$, $\alpha = 0.5$, $n_1 = 1000$, $n_2 = 1000$, and $n_3 = 1000$.

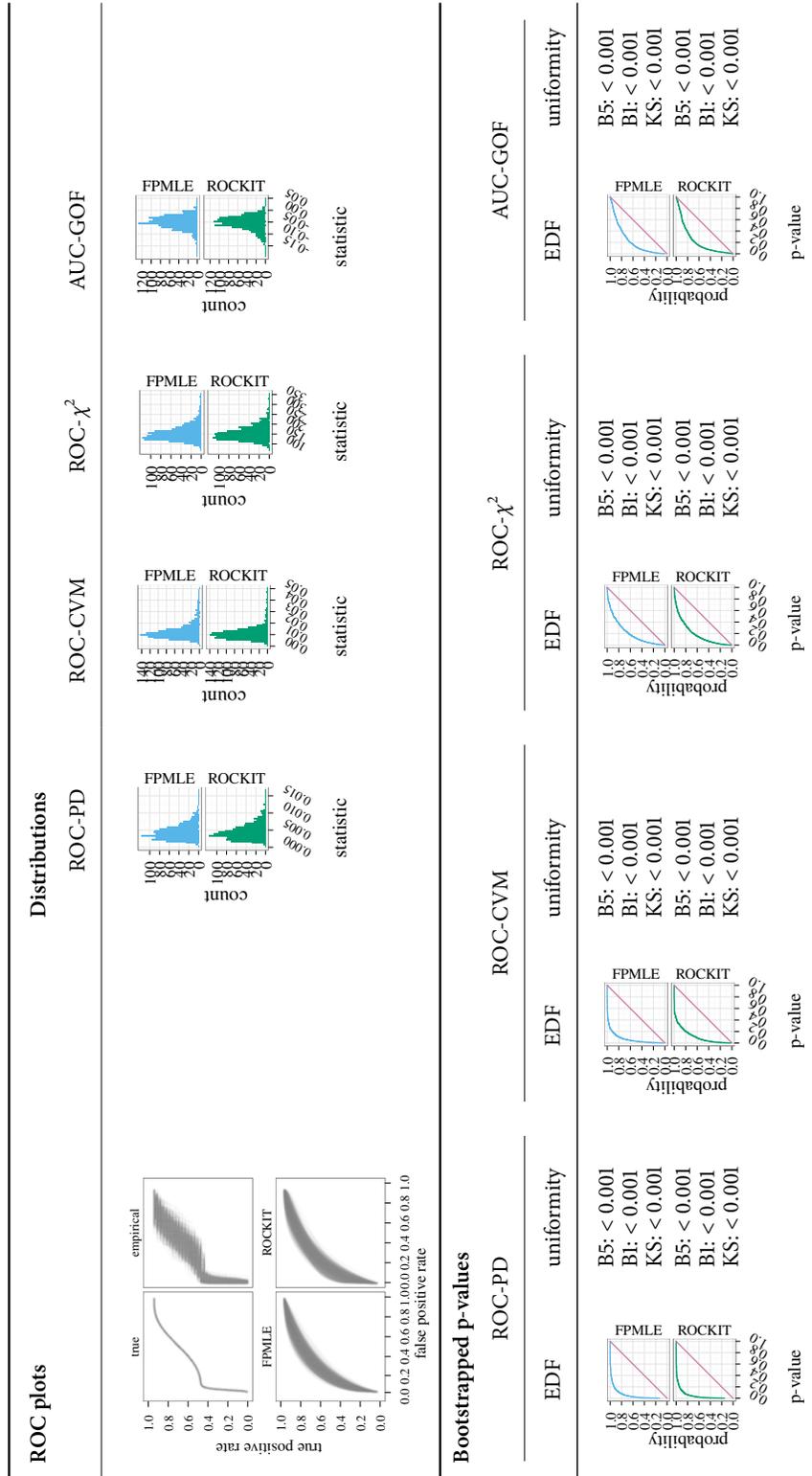


Figure A.77: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 0$, $\sigma_2 = 0.5$, $\mu_3 = 2$, $\sigma_3 = 0.25$, $\alpha = 0.5$, $n_1 = 100$, and $n_2 = 30$.

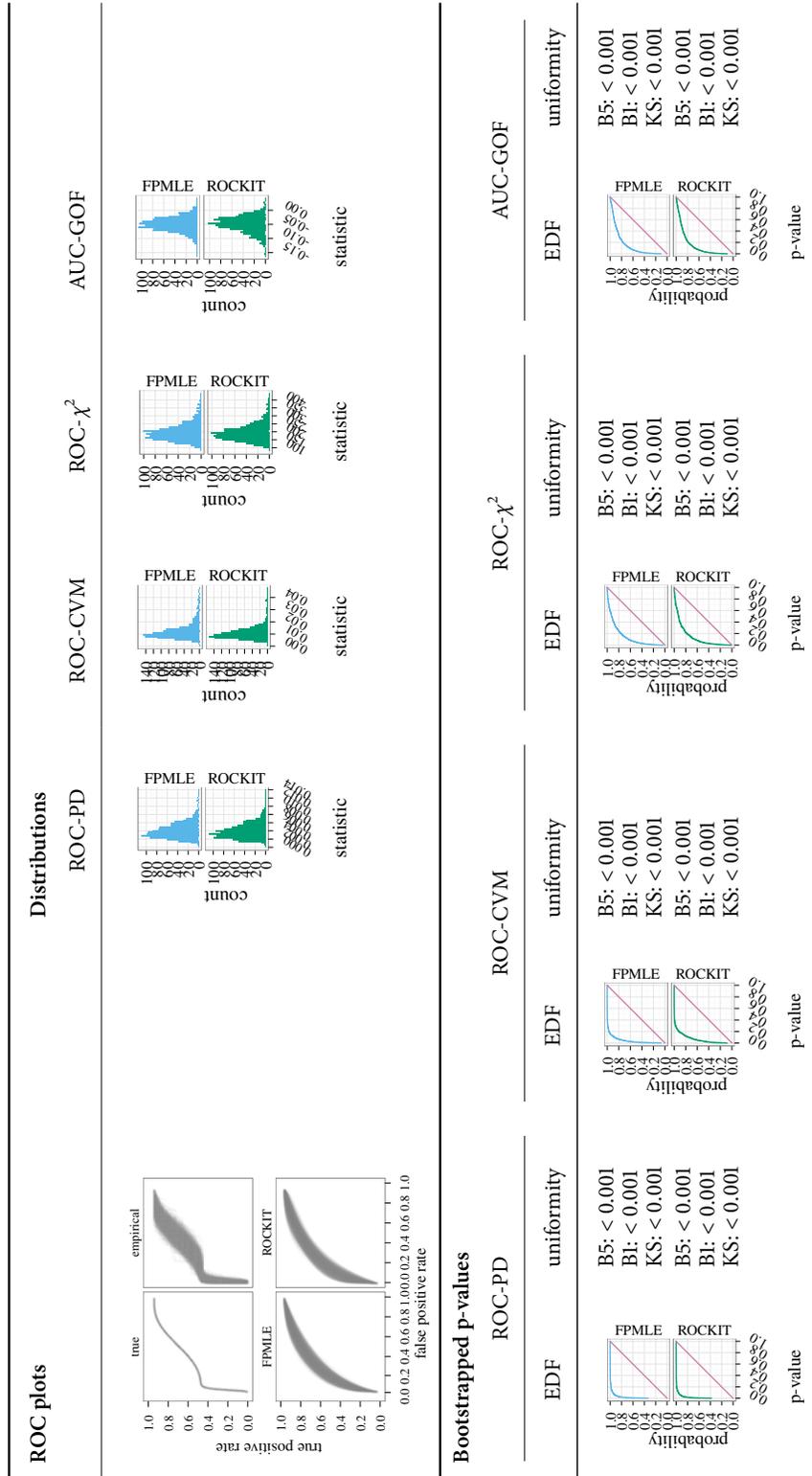


Figure A.78: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 0$, $\sigma_2 = 0.5$, $\mu_3 = 2$, $\sigma_3 = 0.25$, $\alpha = 0.5$, $n_1 = 100$, and $n_2 = 50$.

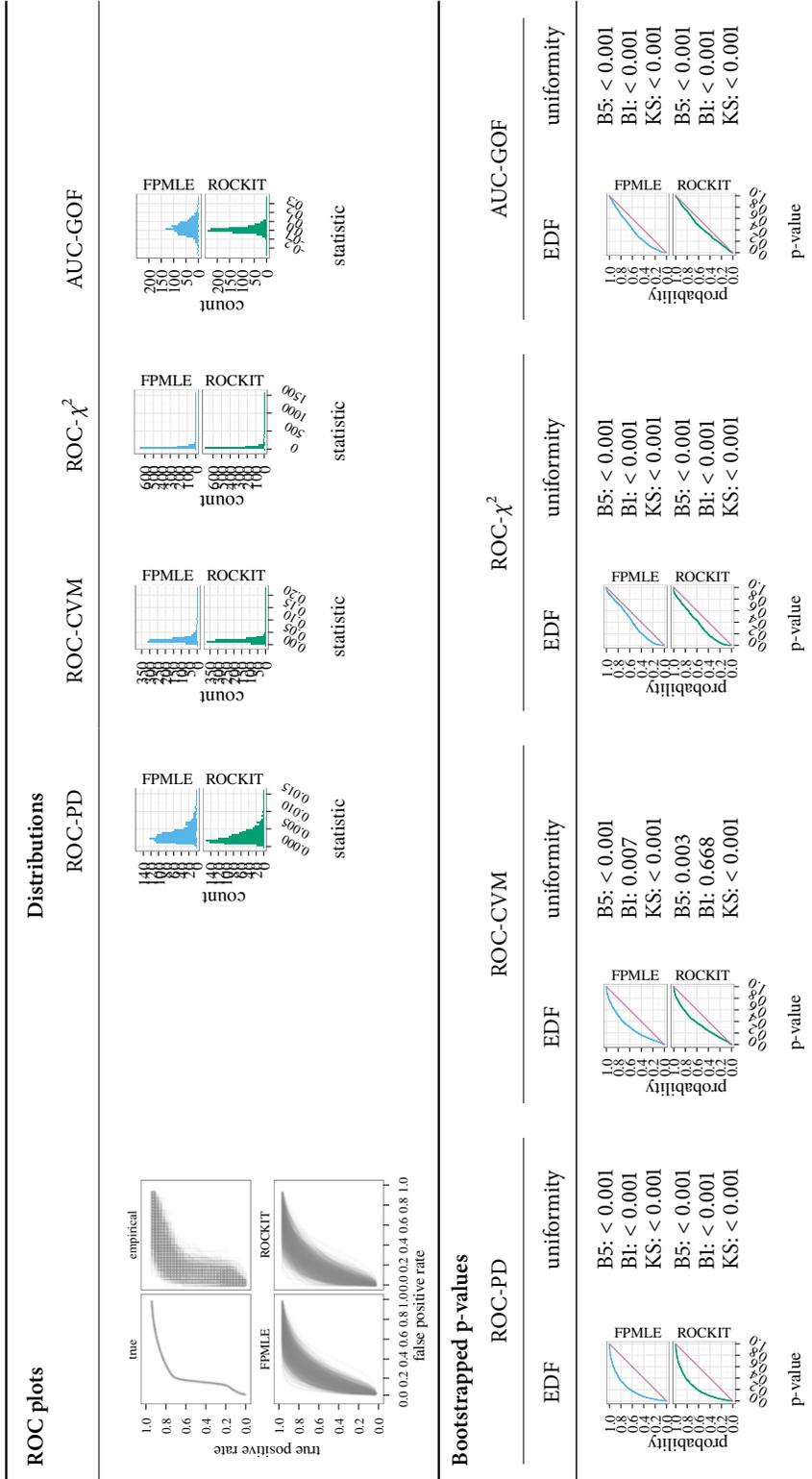


Figure A.79: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0, \sigma_1 = 1, \mu_2 = 1, \sigma_2 = 1, \mu_3 = 1.1, \sigma_3 = 0.1, \alpha = 0.5, n_1 = 30, n_2 = 30, n_3 = 30$.

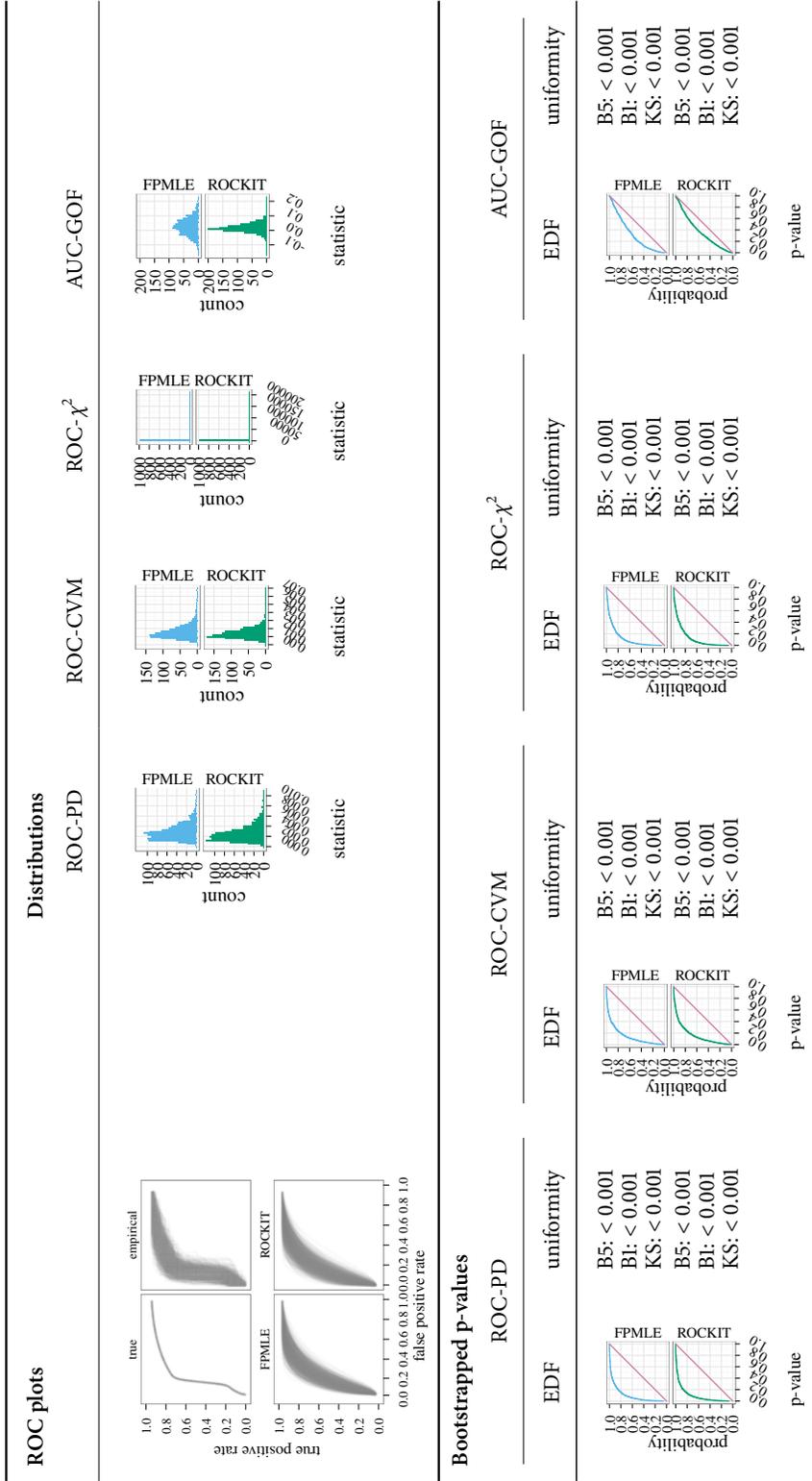


Figure A.80: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0, \sigma_1 = 1, \mu_2 = 1, \sigma_2 = 1, \mu_3 = 1.1, \sigma_3 = 0.1, \alpha = 0.5, n_1 = 50, n_2 = 50,$ and $n_3 = 50$.

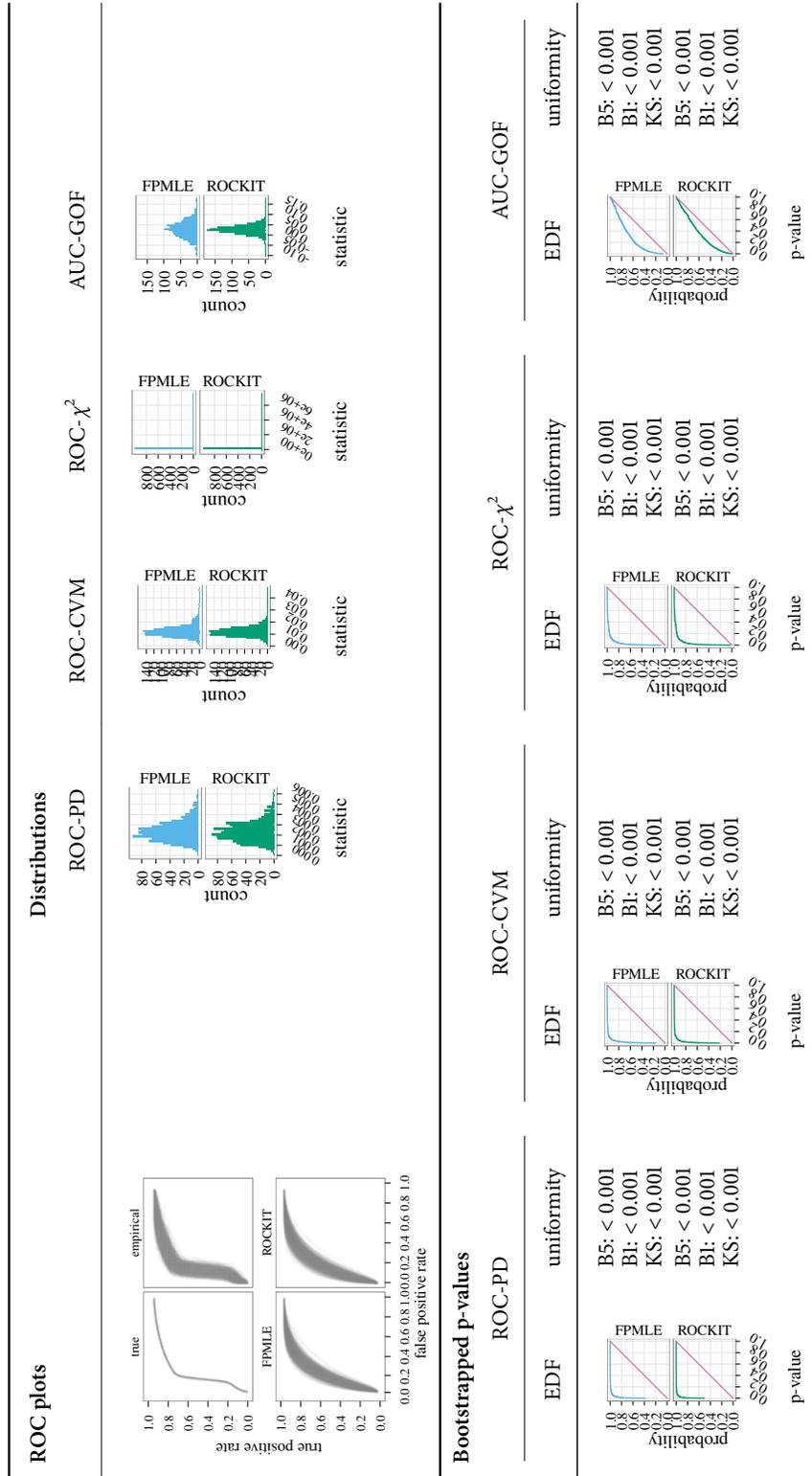


Figure A.81: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 1$, $\sigma_2 = 1$, $\mu_3 = 1.1$, $\sigma_3 = 0.1$, $\alpha = 0.5$, $n_1 = 100$, and $n_2 = 100$.

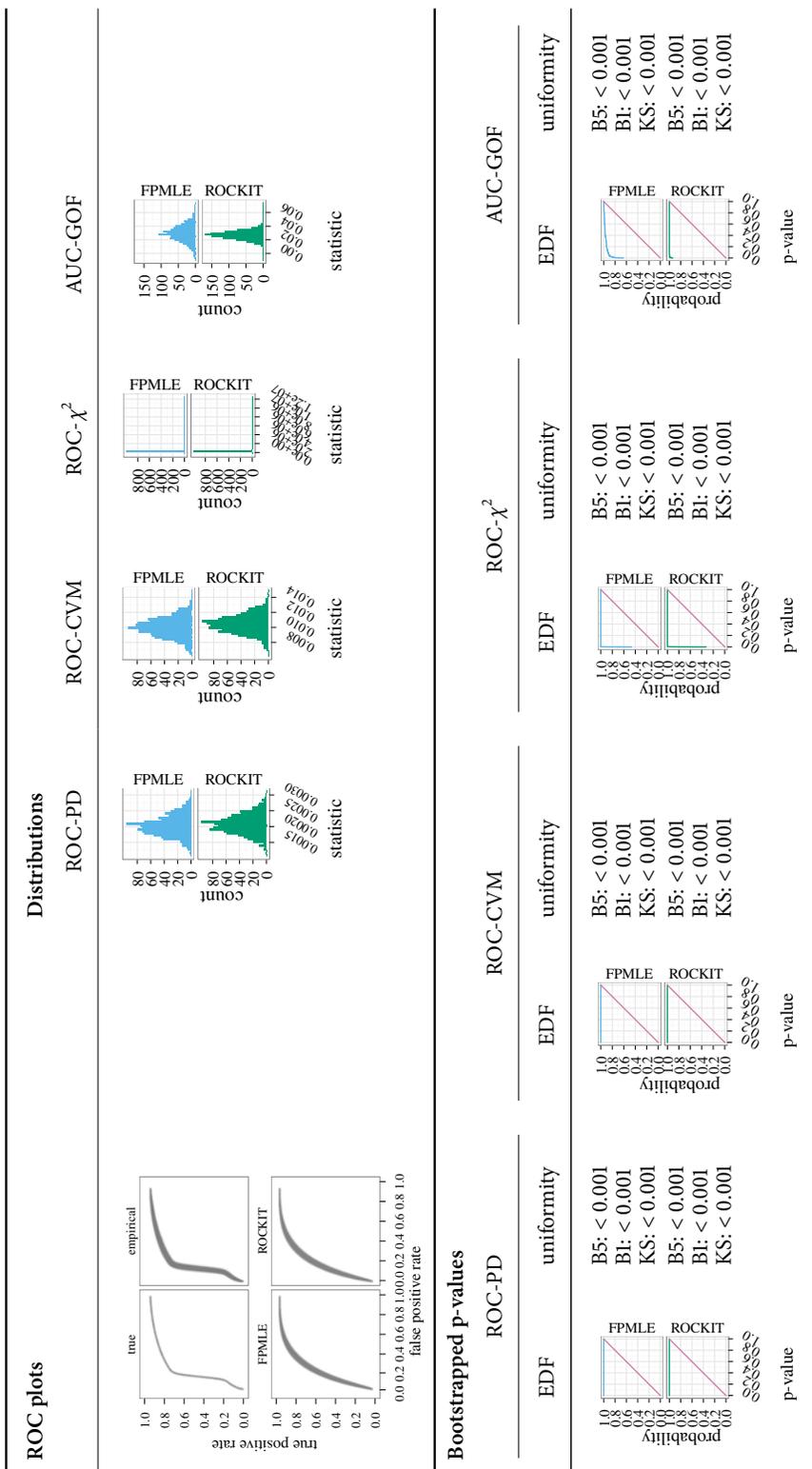


Figure A.82: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0, \sigma_1 = 1, \mu_2 = 1, \sigma_2 = 1, \mu_3 = 1.1, \sigma_3 = 0.1, \alpha = 0.5, n_1 = 1000, n_2 = 1000, n_3 = 1000$.

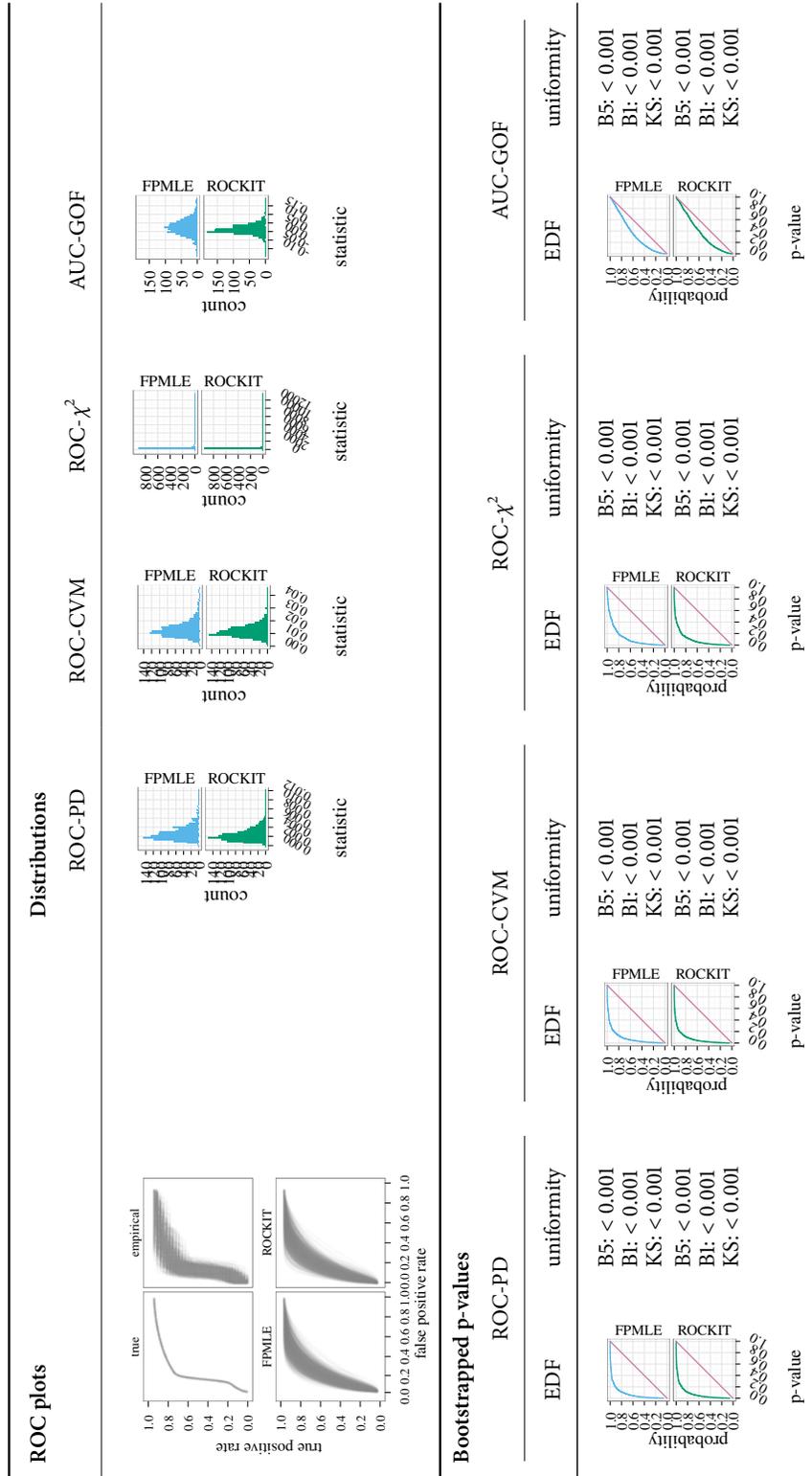


Figure A.83: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 1$, $\sigma_2 = 1$, $\mu_3 = 1.1$, $\sigma_3 = 0.1$, $\alpha = 0.5$, $n_1 = 100$, and $n_2 = 30$.

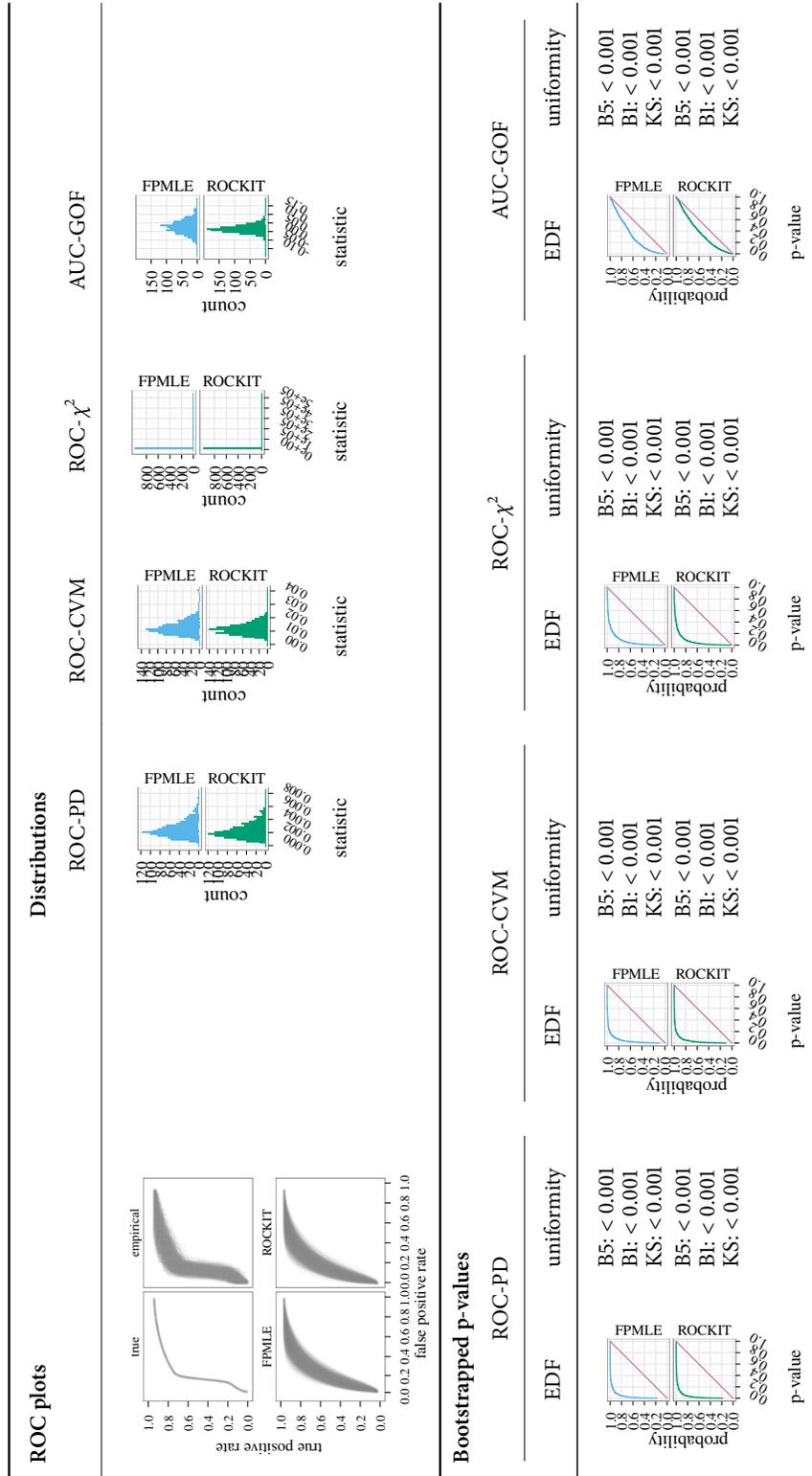


Figure A.84: Results for goodness-of-fit simulations for normal-mixture ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 1$, $\sigma_2 = 1$, $\mu_3 = 1.1$, $\sigma_3 = 0.1$, $\alpha = 0.5$, $n_1 = 100$, and $n_2 = 50$.

A.5 Normal-Cauchy

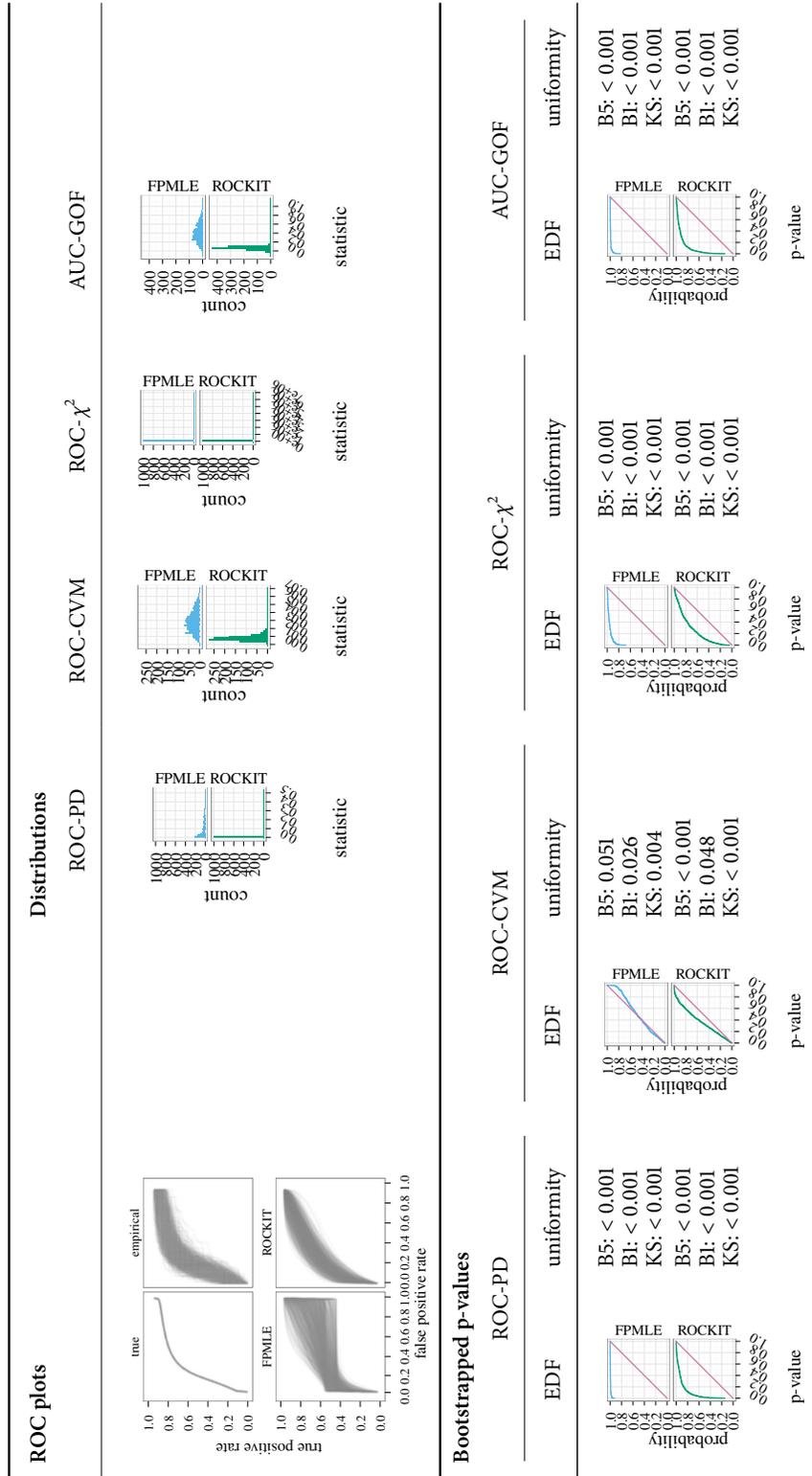
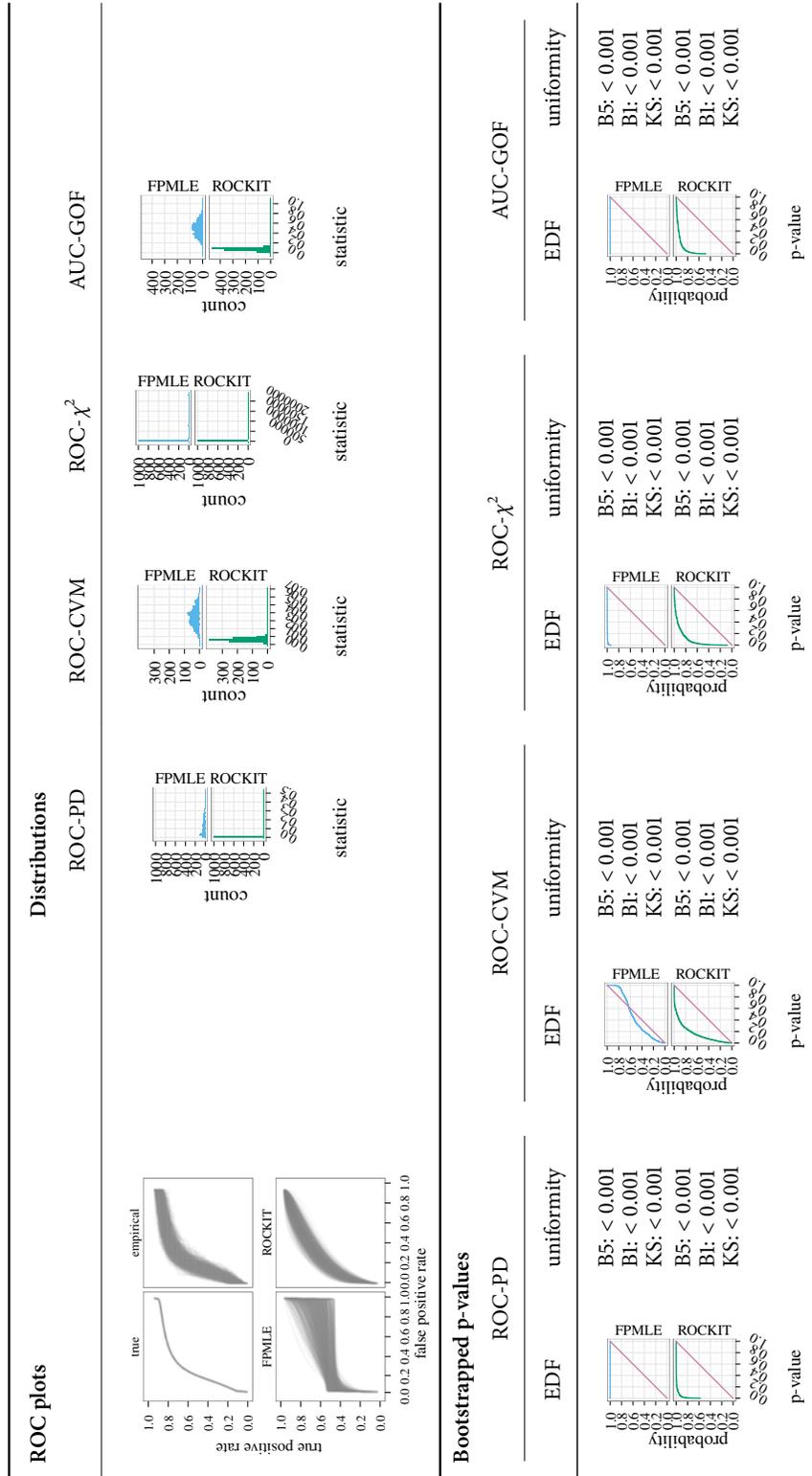


Figure A.85: Results for goodness-of-fit simulations for normal-Cauchy ROC curves. $\mu_1 = 0, \sigma_1 = 1, \theta_2 = 1, \sigma_2 = 0.5, n_1 = 50, \text{ and } n_2 = 50.$



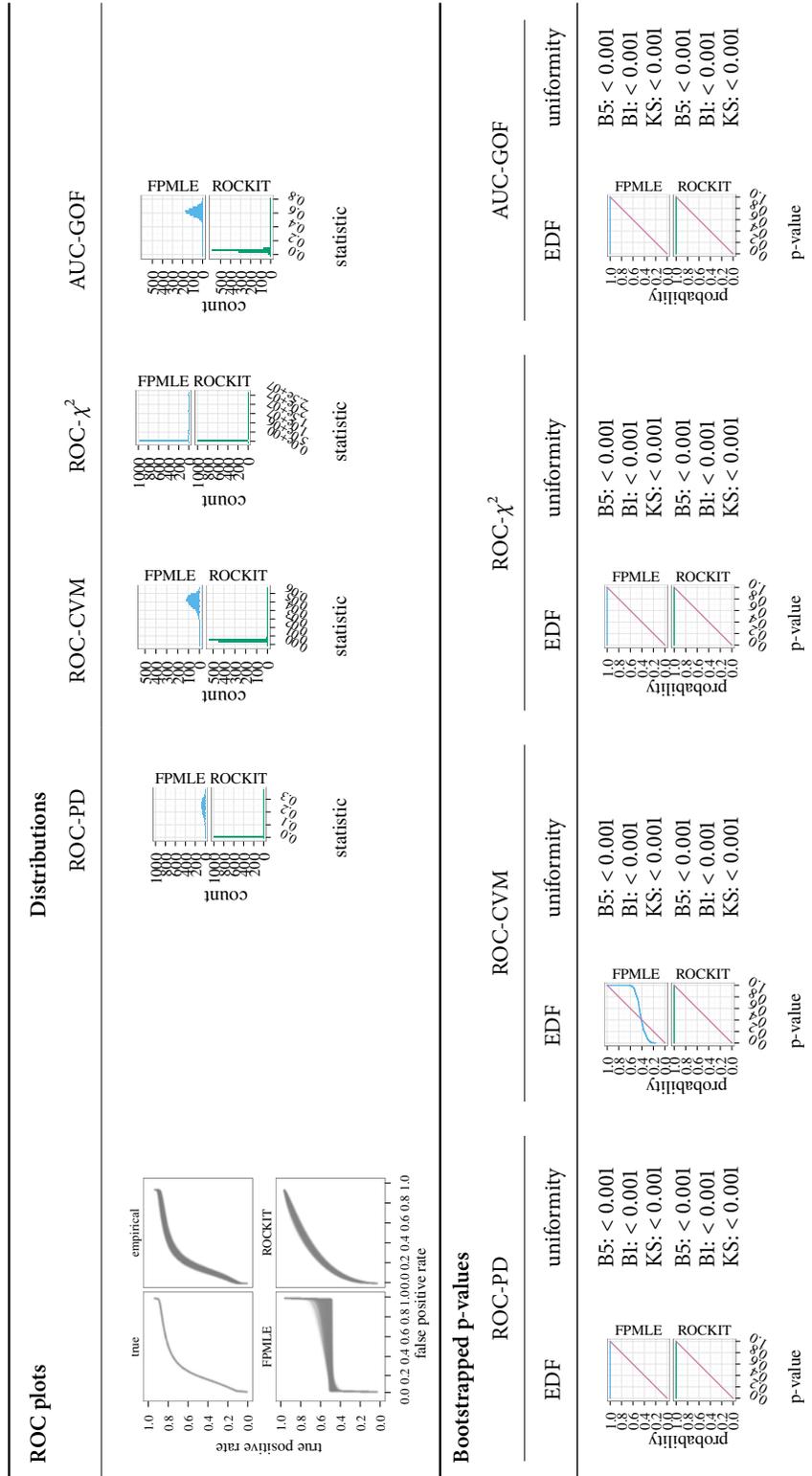


Figure A.87: Results for goodness-of-fit simulations for normal-Cauchy ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\theta_2 = 1$, $\sigma_2 = 0.5$, $n_1 = 1000$, and $n_2 = 1000$.

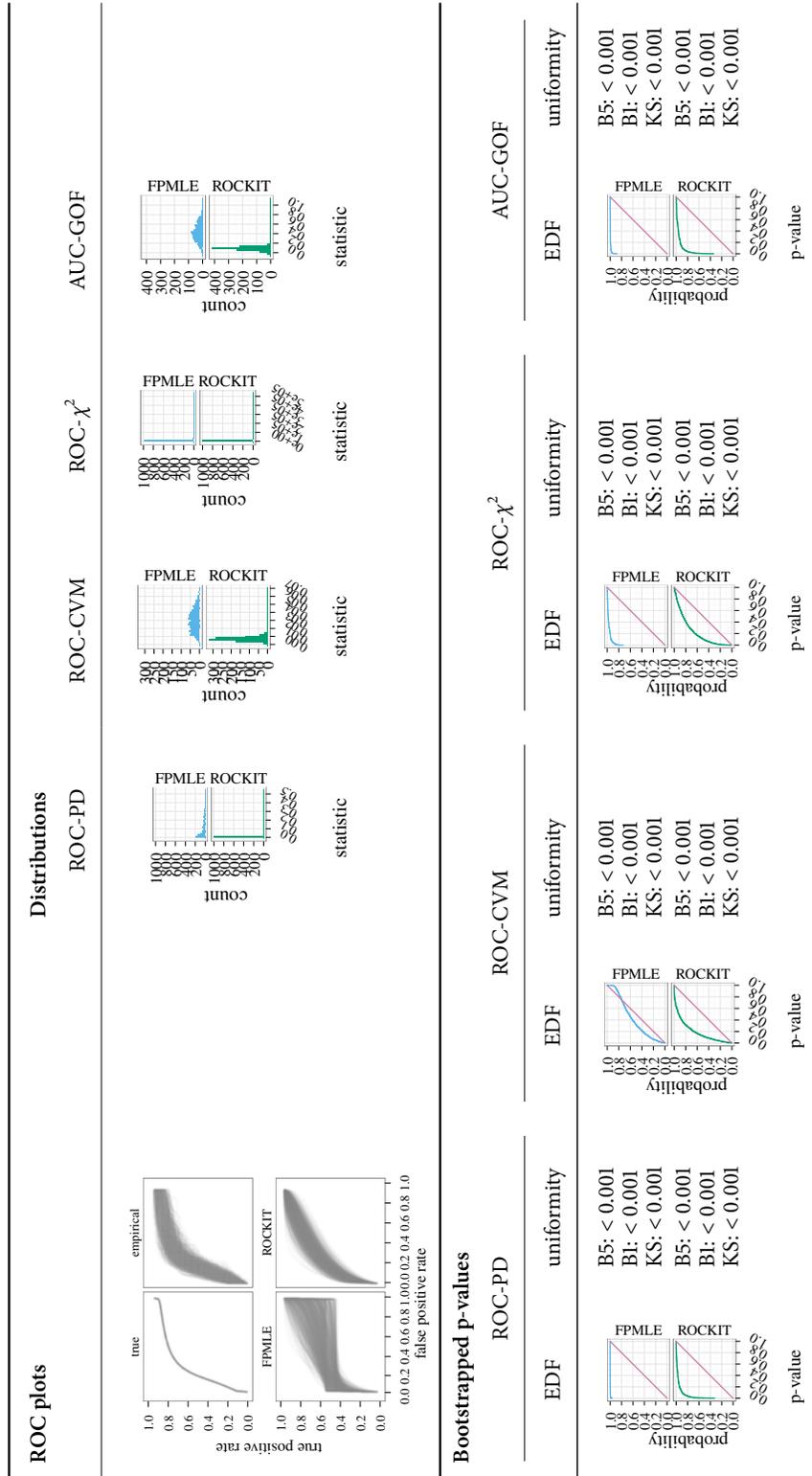


Figure A.88: Results for goodness-of-fit simulations for normal-Cauchy ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\theta_2 = 1$, $\sigma_2 = 0.5$, $n_1 = 100$, and $n_2 = 50$.

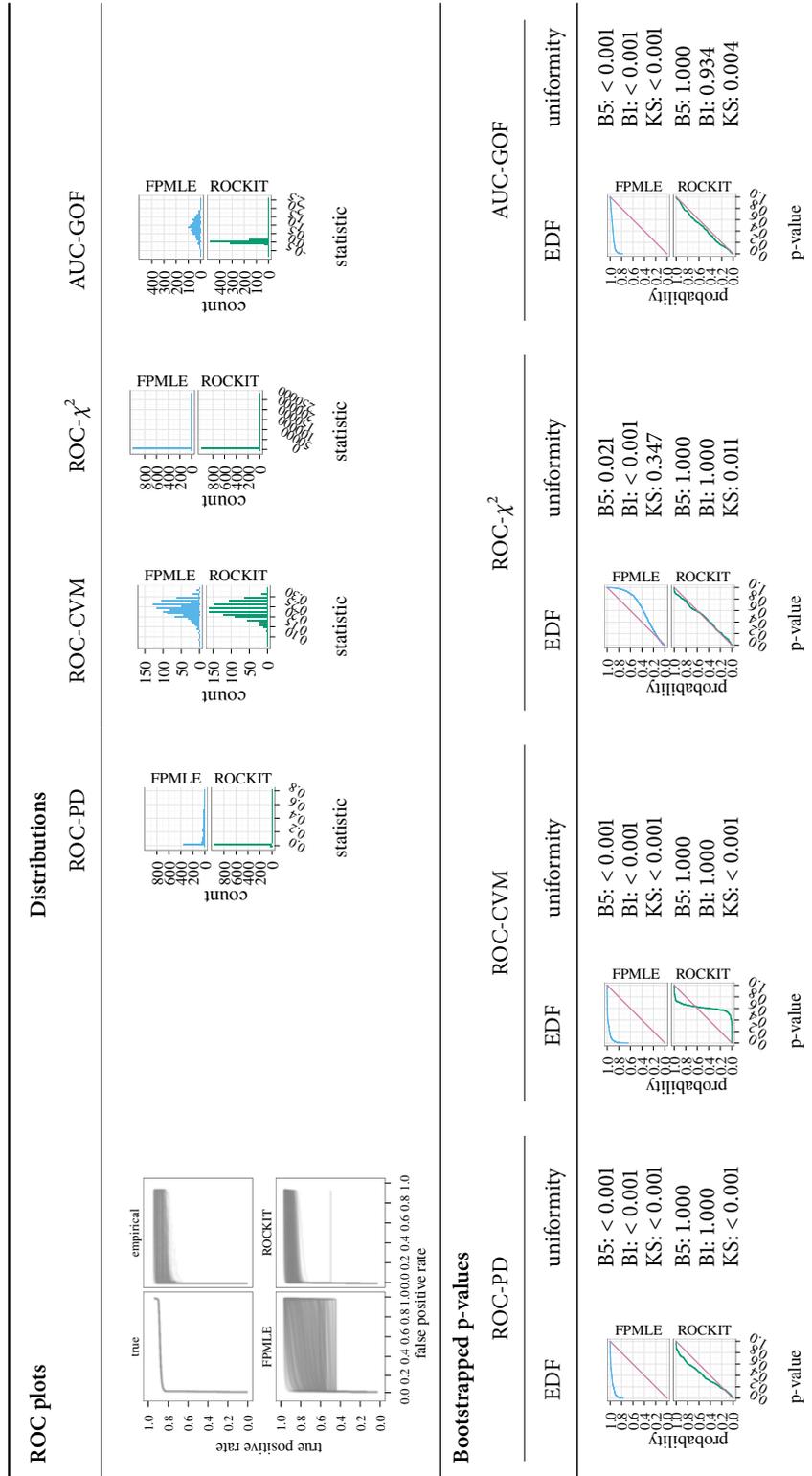


Figure A.89: Results for goodness-of-fit simulations for normal-Cauchy ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\theta_2 = 5$, $\sigma_2 = 1$, $n_1 = 50$, and $n_2 = 50$.

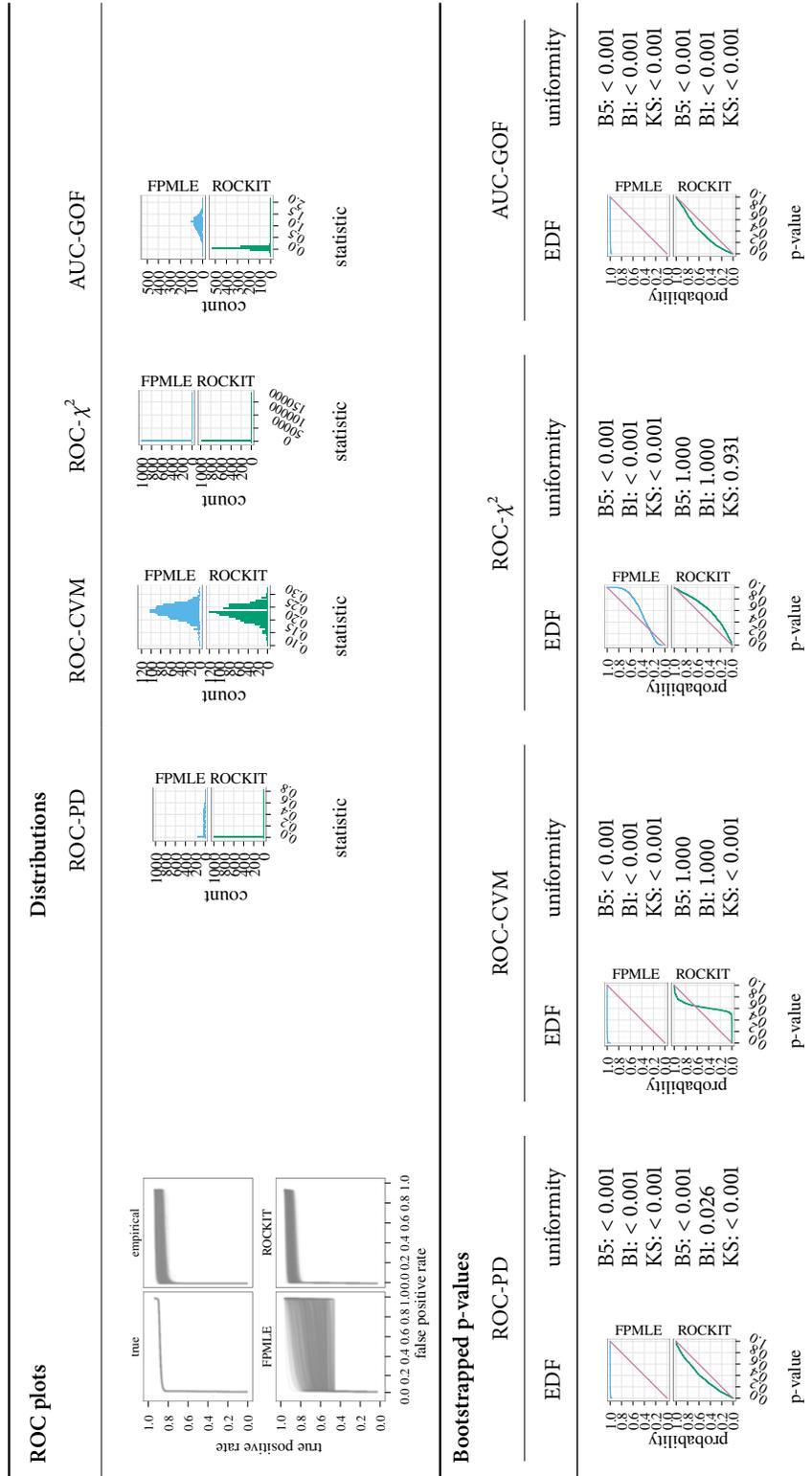


Figure A.90: Results for goodness-of-fit simulations for normal-Cauchy ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\theta_2 = 5$, $\sigma_2 = 1$, $n_1 = 100$, and $n_2 = 100$.

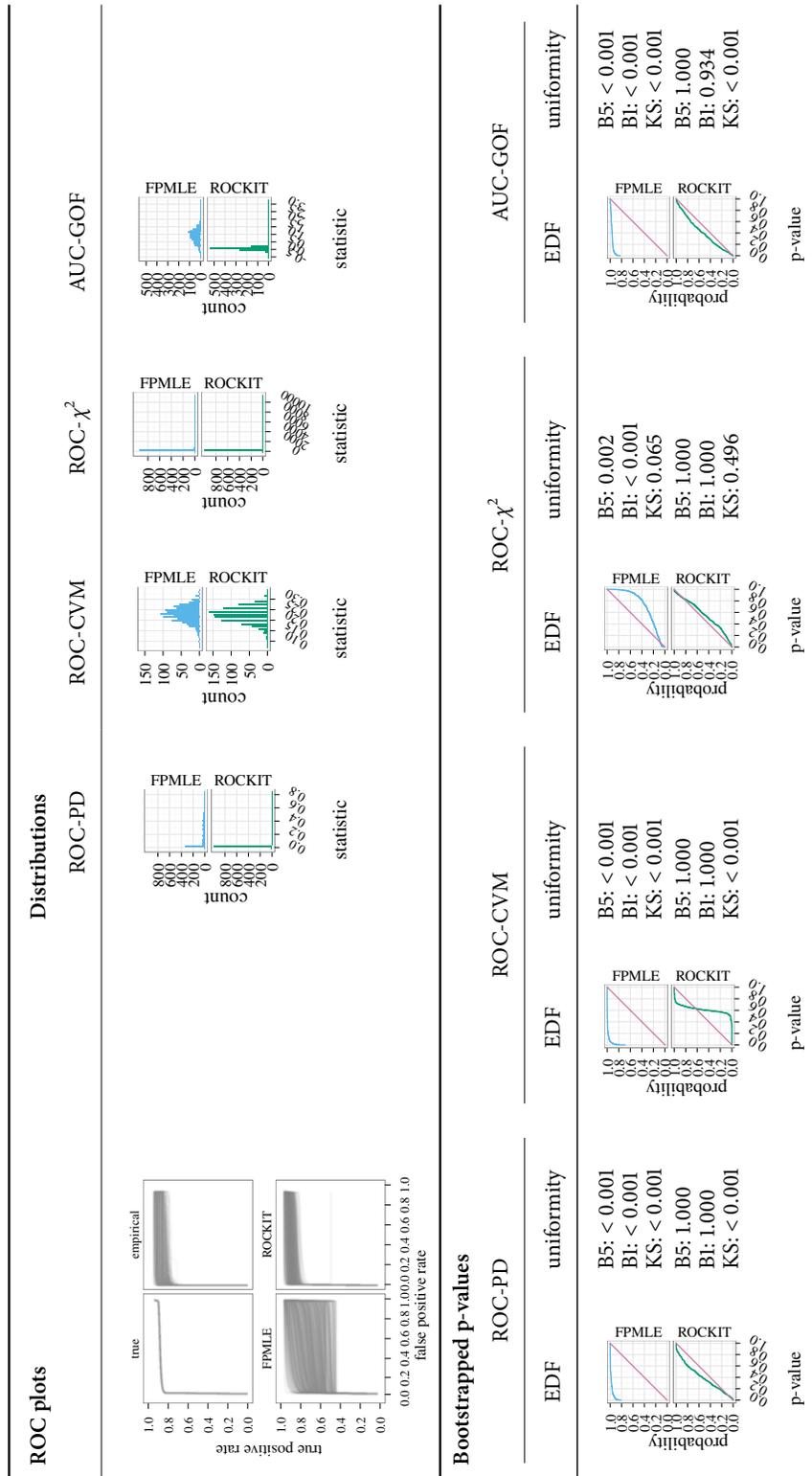


Figure A.91: Results for goodness-of-fit simulations for normal-Cauchy ROC curves. $\mu_1 = 0$, $\sigma_1 = 1$, $\theta_2 = 5$, $\sigma_2 = 1$, $n_1 = 100$, and $n_2 = 50$.

Appendix B: Results of simulations for estimation methods

This appendix contains all of the results of the simulations to study performance of binormal ROC function estimators. To save space, we abbreviate the labels of measures of global squared error in the perpendicular, vertical, and horizontal directions with the letters p , v , and h .

We took the square root of the mean squared error and variance to put the values on the same scale as the original estimands. We also multiplied all results by 100 to make them percentages.

B.1 Binormal

Table B.1: Global errors for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 0$.

method	$n_1 = 30, n_2 = 30$			$n_1 = 50, n_2 = 50$			$n_1 = 100, n_2 = 100$			$n_1 = 100, n_2 = 30$			$n_1 = 100, n_2 = 50$		
	p	v	h	p	v	h	p	v	h	p	v	h	p	v	h
ROC-CVM	8.10 ^g	9.38 ^g	9.38 ^g	6.43 ^g	7.50 ^g	7.50 ^g	4.26 ^g	5.04 ^g	5.04 ^g	6.37 ^g	7.45 ^g	7.45 ^g	5.38 ^g	6.32 ^g	6.32 ^g
ROC-MPD	7.92 ^e	9.20 ^e	9.20 ^e	6.33 ^f	7.40 ^f	7.40 ^f	4.24 ^f	5.01 ^f	5.01 ^f	6.26 ^f	7.33 ^f	7.33 ^f	5.31 ^f	6.25 ^f	6.25 ^f
ROC-TLS	7.33 ^a	8.55 ^a	8.55 ^a	5.90 ^a	6.92 ^a	6.92 ^a	3.97 ^b	4.69 ^b	4.69 ^b	5.80 ^a	6.81 ^a	6.81 ^a	4.91 ^a	5.79 ^a	5.79 ^a
FPML	7.40 ^b	8.64 ^b	8.64 ^b	5.96 ^b	6.99 ^b	6.99 ^b	3.96 ^a	4.69 ^a	4.69 ^a	5.85 ^b	6.87 ^b	6.87 ^b	4.95 ^b	5.84 ^b	5.84 ^b
ROC-GLM	7.99 ^f	9.27 ^f	9.27 ^f	6.26 ^e	7.31 ^e	7.31 ^e	4.14 ^e	4.89 ^e	4.89 ^e	6.08 ^e	7.13 ^e	7.13 ^e	5.15 ^e	6.07 ^e	6.07 ^e
ROC-SLS	7.72 ^c	8.99 ^c	8.99 ^c	6.08 ^c	7.12 ^c	7.12 ^c	4.03 ^c	4.77 ^c	4.77 ^c	5.89 ^c	6.92 ^c	6.92 ^c	4.98 ^c	5.87 ^c	5.87 ^c
ROCKIT	7.76 ^d	9.03 ^d	9.03 ^d	6.16 ^d	7.20 ^d	7.20 ^d	4.11 ^d	4.85 ^d	4.85 ^d	6.01 ^d	7.04 ^d	7.04 ^d	5.06 ^d	5.96 ^d	5.96 ^d

Table B.2: Global errors for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$			$n_1 = 50, n_2 = 50$			$n_1 = 100, n_2 = 100$			$n_1 = 100, n_2 = 30$			$n_1 = 100, n_2 = 50$		
	p	v	h	p	v	h	p	v	h	p	v	h	p	v	h
ROC-CVM	7.77 ^s	9.12 ^s	9.20 ^s	5.96 ^s	7.10 ^s	7.16 ^s	4.10 ^s	4.92 ^s	4.97 ^s	6.15 ^s	7.29 ^s	7.39 ^s	5.09 ^s	6.07 ^s	6.15 ^s
ROC-MPD	7.56 ^e	8.95 ^e	8.96 ^e	5.84 ^f	6.99 ^f	6.99 ^f	4.03 ^f	4.86 ^f	4.86 ^f	6.01 ^f	7.18 ^f	7.21 ^f	4.99 ^f	5.99 ^f	6.02 ^f
ROC-TLS	7.03 ^b	8.35 ^a	8.37 ^b	5.45 ^b	6.53 ^b	6.53 ^b	3.75 ^b	4.53 ^b	4.53 ^b	5.60 ^a	6.69 ^a	6.73 ^a	4.65 ^a	5.57 ^a	5.60 ^a
FPML	7.02 ^a	8.35 ^b	8.36 ^a	5.43 ^a	6.52 ^a	6.51 ^a	3.75 ^a	4.52 ^a	4.52 ^a	5.61 ^b	6.70 ^b	6.75 ^b	4.65 ^b	5.58 ^b	5.61 ^b
ROC-GLM	7.58 ^f	8.97 ^f	8.99 ^f	5.76 ^e	6.90 ^e	6.89 ^e	3.92 ^e	4.73 ^e	4.72 ^e	5.85 ^e	6.98 ^e	7.02 ^e	4.85 ^e	5.82 ^e	5.85 ^e
ROC-SLS	7.37 ^c	8.72 ^c	8.77 ^c	5.63 ^d	6.74 ^c	6.74 ^d	3.82 ^c	4.60 ^c	4.60 ^c	5.69 ^c	6.80 ^c	6.84 ^c	4.72 ^c	5.67 ^c	5.70 ^c
ROCKIT	7.41 ^d	8.78 ^d	8.80 ^d	5.63 ^c	6.74 ^d	6.74 ^c	3.82 ^d	4.62 ^d	4.61 ^d	5.77 ^d	6.89 ^d	6.94 ^d	4.75 ^d	5.69 ^d	5.73 ^d

Table B.3: Global errors for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 1$.

method	$n_1 = 30, n_2 = 30$			$n_1 = 50, n_2 = 50$			$n_1 = 100, n_2 = 100$			$n_1 = 100, n_2 = 30$			$n_1 = 100, n_2 = 50$		
	p	v	h	p	v	h	p	v	h	p	v	h	p	v	h
ROC-CVM	6.90 ^g	8.49 ^g	8.75 ^g	5.39 ^g	6.67 ^g	6.93 ^g	3.71 ^g	4.63 ^g	4.83 ^g	5.45 ^g	6.69 ^g	7.06 ^g	4.56 ^g	5.64 ^g	5.95 ^g
ROC-MPD	6.61 ^d	8.29 ^f	8.31 ^d	5.12 ^f	6.49 ^f	6.49 ^f	3.54 ^f	4.52 ^f	4.52 ^f	5.27 ^f	6.60 ^f	6.76 ^f	4.39 ^f	5.54 ^f	5.63 ^f
ROC-TLS	6.31 ^b	7.90 ^b	7.93 ^b	4.84 ^b	6.12 ^b	6.12 ^b	3.33 ^b	4.23 ^b	4.23 ^b	4.95 ^b	6.17 ^b	6.36 ^b	4.12 ^b	5.17 ^b	5.28 ^b
FPML	6.10 ^a	7.63 ^a	7.64 ^a	4.74 ^a	5.98 ^a	5.96 ^a	3.28 ^a	4.16 ^a	4.15 ^a	4.88 ^a	6.06 ^a	6.24 ^a	4.06 ^a	5.08 ^a	5.19 ^a
ROC-GLM	6.63 ^f	8.26 ^d	8.33 ^e	5.05 ^d	6.38 ^d	6.39 ^e	3.45 ^e	4.40 ^e	4.39 ^e	5.15 ^e	6.41 ^e	6.61 ^e	4.28 ^e	5.38 ^e	5.48 ^e
ROC-SLS	6.62 ^e	8.29 ^e	8.33 ^f	5.06 ^e	6.40 ^e	6.39 ^d	3.43 ^d	4.37 ^d	4.35 ^d	5.08 ^d	6.36 ^d	6.52 ^d	4.24 ^c	5.35 ^d	5.43 ^c
ROCKIT	6.47 ^c	8.09 ^c	8.11 ^c	4.95 ^c	6.26 ^c	6.25 ^c	3.34 ^c	4.25 ^c	4.25 ^c	5.06 ^c	6.31 ^c	6.47 ^c	4.25 ^d	5.33 ^c	5.46 ^d

Table B.4: Global errors for fits to observations from a binormal ROC curve. $\rho = 0.5$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$			$n_1 = 50, n_2 = 50$			$n_1 = 100, n_2 = 100$			$n_1 = 100, n_2 = 30$			$n_1 = 100, n_2 = 50$		
	p	v	h	p	v	h	p	v	h	p	v	h	p	v	h
ROC-CVM	7.70 ^e	8.37 ^e	10.70 ^e	5.85 ^e	6.41 ^d	8.23 ^e	4.02 ^e	4.40 ^e	5.69 ^e	6.96 ^f	7.55 ^f	9.80 ^f	5.48 ^e	5.98 ^e	7.77 ^f
ROC-MPD	7.74 ^f	8.40 ^g	10.73 ^f	5.93 ^g	6.51 ^g	8.30 ^g	4.11 ^g	4.51 ^g	5.77 ^g	7.06 ^g	7.70 ^g	9.87 ^g	5.59 ^g	6.13 ^g	7.87 ^g
ROC-TLS	7.41 ^b	8.11 ^b	10.27 ^b	5.64 ^b	6.22 ^b	7.92 ^b	3.90 ^b	4.27 ^b	5.49 ^b	6.68 ^b	7.30 ^c	9.38 ^b	5.28 ^b	5.79 ^b	7.46 ^b
FPML	6.93 ^a	7.51 ^a	9.76 ^a	5.36 ^a	5.85 ^a	7.58 ^a	3.73 ^a	4.05 ^a	5.28 ^a	6.33 ^a	6.87 ^a	8.98 ^a	5.03 ^a	5.46 ^a	7.15 ^a
ROC-GLM	7.82 ^g	8.39 ^f	10.82 ^g	5.91 ^f	6.44 ^f	8.28 ^f	4.04 ^f	4.42 ^f	5.69 ^f	6.92 ^e	7.51 ^e	9.71 ^e	5.49 ^f	5.99 ^f	7.74 ^e
ROC-SLS	7.68 ^d	8.37 ^d	10.63 ^d	5.84 ^d	6.43 ^e	8.18 ^d	3.99 ^d	4.38 ^d	5.61 ^d	6.83 ^d	7.46 ^d	9.57 ^d	5.42 ^d	5.94 ^d	7.63 ^d
ROCKIT	7.56 ^c	8.12 ^c	10.51 ^c	5.69 ^c	6.22 ^c	8.01 ^c	3.90 ^c	4.28 ^c	5.49 ^c	6.69 ^c	7.25 ^b	9.43 ^c	5.33 ^c	5.82 ^c	7.53 ^c

Table B.5: Global errors for fits to observations from a binormal ROC curve. $\rho = 0.75$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$			$n_1 = 50, n_2 = 50$			$n_1 = 100, n_2 = 100$			$n_1 = 100, n_2 = 30$			$n_1 = 100, n_2 = 50$		
	p	v	h	p	v	h	p	v	h	p	v	h	p	v	h
ROC-CVM	7.59 ^g	8.63 ^g	9.56 ^g	5.82 ^g	6.69 ^f	7.42 ^g	3.99 ^f	4.61 ^f	5.13 ^g	6.43 ^g	7.34 ^f	8.20 ^g	5.19 ^f	5.96 ^f	6.66 ^g
ROC-MPD	7.53 ^e	8.61 ^e	9.45 ^e	5.80 ^f	6.71 ^g	7.35 ^f	3.99 ^g	4.65 ^g	5.10 ^f	6.42 ^f	7.38 ^g	8.14 ^f	5.20 ^g	6.01 ^g	6.63 ^f
ROC-TLS	7.07 ^b	8.12 ^b	8.89 ^b	5.43 ^b	6.30 ^b	6.89 ^b	3.74 ^b	4.35 ^b	4.79 ^b	5.98 ^b	6.88 ^b	7.61 ^b	4.84 ^b	5.60 ^b	6.19 ^b
FPML	6.94 ^a	7.94 ^a	8.78 ^a	5.37 ^a	6.19 ^a	6.83 ^a	3.70 ^a	4.29 ^a	4.75 ^a	5.92 ^a	6.78 ^a	7.56 ^a	4.79 ^a	5.52 ^a	6.15 ^a
ROC-GLM	7.57 ^f	8.61 ^f	9.50 ^f	5.74 ^e	6.62 ^e	7.27 ^e	3.90 ^e	4.54 ^e	4.98 ^e	6.25 ^e	7.17 ^e	7.95 ^e	5.06 ^e	5.84 ^e	6.46 ^e
ROC-SLS	7.40 ^d	8.48 ^d	9.29 ^c	5.64 ^d	6.54 ^d	7.14 ^d	3.82 ^d	4.45 ^d	4.88 ^d	6.10 ^c	7.02 ^d	7.74 ^c	4.95 ^d	5.73 ^d	6.31 ^c
ROCKIT	7.38 ^c	8.41 ^c	9.29 ^d	5.59 ^c	6.45 ^c	7.10 ^c	3.79 ^c	4.40 ^c	4.87 ^c	6.12 ^d	7.02 ^c	7.80 ^d	4.93 ^c	5.68 ^c	6.31 ^d

Table B.6: Errors for ρ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 0$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	31.80 ^g	5.08 ^g	20.80 ^g	2.74 ^g	13.40 ^g	1.33 ^g	22.69 ^g	4.91 ^g	17.59 ^g	2.84 ^g
ROC-MPD	25.66 ^f	2.30 ^e	18.56 ^f	1.43 ^e	12.74 ^f	0.82 ^e	20.10 ^f	3.48 ^f	16.07 ^f	1.92 ^f
ROC-TLS	20.59 ^c	1.60 ^b	14.94 ^c	1.04 ^c	10.31 ^c	0.54 ^c	16.08 ^b	1.46 ^c	12.90 ^b	1.03 ^b
FPML	19.67 ^a	1.43 ^a	14.78 ^b	0.92 ^a	10.21 ^b	0.47 ^b	16.55 ^c	3.42 ^e	13.04 ^c	1.68 ^e
ROC-GLM	22.25 ^d	-5.05 ^f	16.28 ^d	-3.54 ^f	11.20 ^d	-1.99 ^f	17.41 ^d	0.27 ^a	13.91 ^d	-1.08 ^c
ROC-SLS	19.74 ^b	-2.19 ^d	14.68 ^a	-1.01 ^b	10.17 ^a	-0.16 ^a	15.57 ^a	-0.65 ^b	12.64 ^a	-0.32 ^a
ROCKIT	23.68 ^e	2.11 ^c	16.78 ^e	1.23 ^d	11.95 ^e	0.63 ^d	17.98 ^e	3.05 ^d	14.76 ^e	1.31 ^d

Table B.7: Errors for ρ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	31.83 ^g	5.37 ^g	21.29 ^g	2.80 ^g	14.05 ^g	1.22 ^f	23.46 ^g	4.74 ^g	17.98 ^g	2.62 ^g
ROC-MPD	26.64 ^f	2.63 ^e	19.05 ^f	1.53 ^e	13.20 ^f	0.81 ^e	21.08 ^f	3.57 ^f	16.77 ^f	1.88 ^f
ROC-TLS	22.08 ^c	1.90 ^c	15.79 ^c	1.27 ^c	10.71 ^b	0.61 ^c	16.85 ^c	1.45 ^c	13.52 ^b	1.01 ^b
FPML	19.67 ^a	1.43 ^b	14.78 ^a	0.92 ^b	10.21 ^a	0.47 ^b	16.55 ^a	3.42 ^e	13.04 ^a	1.68 ^e
ROC-GLM	23.09 ^d	-4.53 ^f	16.75 ^d	-3.34 ^f	11.54 ^d	-1.99 ^g	18.19 ^d	0.45 ^b	14.46 ^d	-1.05 ^c
ROC-SLS	21.23 ^b	-1.11 ^a	15.67 ^b	-0.33 ^a	10.81 ^c	0.18 ^a	16.67 ^b	-0.41 ^a	13.52 ^c	-0.09 ^a
ROCKIT	24.83 ^e	2.42 ^d	17.12 ^e	1.39 ^d	11.95 ^e	0.79 ^d	18.51 ^e	3.17 ^d	14.69 ^e	1.11 ^d

Table B.8: Errors for ρ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 1$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	36.60 ^g	7.08 ^g	24.85 ^g	3.65 ^g	15.89 ^g	1.29 ^f	25.22 ^g	4.93 ^g	20.02 ^g	2.79 ^g
ROC-MPD	29.57 ^f	3.35 ^f	20.88 ^f	1.72 ^e	14.41 ^f	0.96 ^e	23.25 ^f	4.09 ^f	18.39 ^f	2.14 ^f
ROC-TLS	28.59 ^e	3.28 ^e	17.90 ^b	1.58 ^c	11.97 ^b	0.74 ^d	18.94 ^b	1.45 ^c	15.22 ^b	1.01 ^c
FPML	19.67 ^a	1.43 ^b	14.78 ^a	0.92 ^b	10.21 ^a	0.47 ^b	16.55 ^a	3.42 ^e	13.04 ^a	1.68 ^d
ROC-GLM	25.62 ^c	-3.12 ^d	18.25 ^c	-2.89 ^f	12.54 ^d	-1.93 ^g	20.33 ^e	1.27 ^b	16.00 ^d	-0.72 ^b
ROC-SLS	25.39 ^b	-0.41 ^a	18.30 ^d	-0.11 ^a	12.64 ^e	0.15 ^a	19.34 ^c	-0.57 ^a	15.87 ^c	-0.24 ^a
ROCKIT	27.65 ^d	2.91 ^c	18.75 ^e	1.65 ^d	12.28 ^c	0.71 ^c	20.06 ^d	2.94 ^d	16.14 ^e	1.91 ^e

Table B.9: Errors for ρ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 1.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	45.49 ^f	11.08 ^g	32.66 ^g	6.14 ^g	21.11 ^g	2.74 ^g	32.51 ^g	7.31 ^g	25.86 ^g	4.51 ^g
ROC-MPD	33.98 ^e	4.65 ^e	24.05 ^f	2.57 ^f	16.40 ^f	1.13 ^d	26.87 ^f	5.11 ^f	21.06 ^f	2.65 ^f
ROC-TLS	57.64 ^g	9.88 ^f	21.94 ^c	2.34 ^e	14.32 ^b	1.05 ^c	22.82 ^b	1.54 ^b	18.29 ^b	1.14 ^c
FPML	19.67 ^a	1.43 ^c	14.78 ^a	0.92 ^b	10.21 ^a	0.47 ^b	16.55 ^a	3.42 ^e	13.04 ^a	1.68 ^d
ROC-GLM	30.20 ^b	-0.07 ^a	21.18 ^b	-1.47 ^c	14.43 ^c	-1.69 ^f	24.27 ^e	3.19 ^d	18.85 ^d	0.19 ^a
ROC-SLS	31.72 ^c	0.10 ^b	22.59 ^e	-0.02 ^a	15.55 ^e	-0.15 ^a	23.79 ^d	-1.18 ^a	19.65 ^e	-0.68 ^b
ROCKIT	32.61 ^d	4.05 ^d	22.06 ^d	2.27 ^d	15.10 ^d	1.37 ^e	23.02 ^c	2.95 ^c	18.60 ^c	1.93 ^e

Table B.10: Errors for ρ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 2$.

method	$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	43.76 ^g	10.28 ^g	31.61 ^g	6.53 ^g	44.45 ^g	12.48 ^g	37.21 ^g	8.77 ^g
ROC-MPD	29.37 ^d	4.13 ^e	19.51 ^e	1.63 ^d	32.55 ^e	7.37 ^e	25.33 ^f	3.95 ^f
ROC-TLS	30.02 ^f	4.29 ^f	18.33 ^c	1.70 ^e	39.92 ^f	5.05 ^d	23.65 ^d	1.88 ^c
FPML	14.78 ^a	0.92 ^b	10.21 ^a	0.47 ^a	16.55 ^e	3.42 ^b	13.04 ^a	1.68 ^b
ROC-GLM	26.73 ^b	1.83 ^c	17.56 ^b	-0.73 ^c	30.32 ^c	-1.36 ^a	23.54 ^c	2.65 ^e
ROC-SLS	29.52 ^e	0.56 ^a	19.86 ^f	-0.53 ^b	30.54 ^d	7.44 ^f	25.03 ^e	-1.06 ^a
ROCKIT	28.19 ^c	3.70 ^d	18.52 ^d	2.01 ^f	28.34 ^b	3.91 ^c	22.85 ^b	2.44 ^d

Table B.11: Errors for ρ for fits to observations from a binormal ROC curve. $\rho = 0.5$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	16.36 ^g	3.88 ^g	10.77 ^g	2.13 ^f	7.09 ^f	1.14 ^f	12.88 ^f	3.29 ^g	9.56 ^f	1.96 ^g
ROC-MPD	14.84 ^f	1.32 ^b	10.59 ^f	0.64 ^b	7.37 ^g	0.43 ^c	13.12 ^g	2.06 ^e	10.00 ^g	0.99 ^d
ROC-TLS	14.56 ^e	3.50 ^f	9.76 ^c	2.17 ^g	6.44 ^b	1.27 ^g	11.23 ^c	2.64 ^f	8.67 ^c	1.90 ^f
FPML	9.83 ^a	0.72 ^a	7.39 ^a	0.46 ^a	5.11 ^a	0.23 ^b	8.28 ^a	1.71 ^d	6.52 ^a	0.84 ^c
ROC-GLM	13.71 ^d	-2.30 ^e	9.80 ^d	-1.87 ^e	6.67 ^e	-1.11 ^e	11.51 ^e	0.09 ^a	8.90 ^e	-0.70 ^b
ROC-SLS	13.61 ^c	1.44 ^c	9.84 ^e	0.91 ^c	6.62 ^d	0.56 ^d	11.24 ^d	1.56 ^c	8.89 ^d	1.05 ^e
ROCKIT	13.14 ^b	-1.51 ^d	9.35 ^b	-1.12 ^d	6.49 ^c	0.17 ^a	10.43 ^b	-0.34 ^b	8.47 ^b	-0.17 ^a

Table B.12: Errors for ρ for fits to observations from a binormal ROC curve. $\rho = 0.75$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	22.36 ^g	4.23 ^g	15.11 ^g	2.27 ^f	10.04 ^f	0.95 ^d	17.17 ^g	3.62 ^g	13.16 ^f	2.02 ^f
ROC-MPD	20.22 ^f	1.96 ^d	14.44 ^f	0.98 ^d	10.07 ^g	0.59 ^g	16.89 ^f	2.82 ^f	13.17 ^g	1.41 ^g
ROC-TLS	17.88 ^c	3.24 ^e	12.50 ^b	2.16 ^e	8.40 ^b	1.22 ^e	13.67 ^c	2.33 ^d	10.87 ^b	1.76 ^e
FPML	14.75 ^a	1.07 ^c	11.08 ^a	0.69 ^b	7.66 ^a	0.35 ^a	12.41 ^a	2.56 ^e	9.78 ^a	1.26 ^d
ROC-GLM	18.09 ^d	-3.41 ^f	13.03 ^e	-2.70 ^g	8.92 ^d	-1.63 ^f	14.57 ^e	0.12 ^a	11.48 ^e	-0.98 ^c
ROC-SLS	17.19 ^b	1.07 ^b	12.53 ^c	0.91 ^c	8.64 ^c	0.78 ^c	13.58 ^b	1.01 ^b	10.98 ^c	0.93 ^b
ROCKIT	18.35 ^e	-0.13 ^a	12.80 ^d	-0.13 ^a	8.94 ^e	0.48 ^b	14.07 ^d	1.08 ^c	11.29 ^d	0.79 ^a

Table B.13: Errors for δ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 0$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	29.94 ^g	-0.28 ^g	22.62 ^g	-0.25 ^g	15.15 ^g	0.53 ^g	23.64 ^g	0.66 ^g	19.13 ^g	0.60 ^g
ROC-MPD	28.29 ^f	-0.28 ^e	21.93 ^f	-0.23 ^b	14.95 ^f	0.51 ^d	22.74 ^f	0.63 ^e	18.68 ^f	0.56 ^e
ROC-TLS	26.67 ^a	-0.17 ^b	20.73 ^a	-0.32 ^e	14.23 ^a	0.47 ^a	21.41 ^a	0.56 ^c	17.67 ^a	0.45 ^b
FPML	27.08 ^b	-0.24 ^d	20.88 ^b	-0.25 ^c	14.29 ^b	0.52 ^e	21.84 ^c	0.55 ^a	17.86 ^d	0.47 ^c
ROC-GLM	27.40 ^d	-0.22 ^c	21.18 ^d	-0.26 ^d	14.48 ^e	0.51 ^c	22.02 ^d	0.60 ^d	18.06 ^e	0.53 ^d
ROC-SLS	27.22 ^c	-6.03 ^f	21.17 ^c	-4.26 ^f	14.35 ^c	-1.78 ^f	21.42 ^b	-1.45 ^f	17.76 ^b	-1.67 ^f
ROCKIT	28.11 ^e	-0.12 ^a	21.30 ^e	-0.22 ^a	14.43 ^d	0.50 ^b	22.13 ^e	0.56 ^b	17.85 ^c	0.43 ^a

Table B.14: Errors for δ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	33.66 ^g	3.05 ^e	24.35 ^g	1.56 ^g	16.32 ^g	1.35 ^f	26.14 ^g	3.20 ^g	20.87 ^g	2.10 ^g
ROC-MPD	30.25 ^f	1.63 ^c	22.79 ^f	0.75 ^c	15.55 ^f	1.01 ^d	24.15 ^f	2.29 ^d	19.71 ^f	1.51 ^e
ROC-TLS	28.33 ^b	0.38 ^a	21.53 ^a	-0.03 ^a	14.74 ^a	0.73 ^a	22.47 ^b	1.00 ^a	18.48 ^b	0.78 ^a
FPML	28.15 ^a	1.92 ^d	21.56 ^b	1.01 ^d	14.81 ^d	1.12 ^e	23.11 ^c	2.62 ^e	18.80 ^c	1.67 ^f
ROC-GLM	29.11 ^d	1.38 ^b	21.99 ^d	0.51 ^b	14.97 ^e	0.88 ^c	23.28 ^d	2.18 ^c	18.96 ^e	1.36 ^c
ROC-SLS	28.50 ^c	-5.78 ^g	21.79 ^c	-4.01 ^f	14.81 ^b	-1.51 ^g	22.34 ^a	-1.15 ^b	18.48 ^a	-1.37 ^d
ROCKIT	30.20 ^e	3.24 ^f	22.25 ^e	1.53 ^e	14.81 ^c	0.87 ^b	23.50 ^e	3.01 ^f	18.85 ^d	1.20 ^b

Table B.15: Errors for δ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 1$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	44.30 ^g	7.44 ^g	31.77 ^g	4.06 ^g	20.06 ^g	2.22 ^g	33.43 ^g	5.95 ^g	26.15 ^g	3.74 ^g
ROC-MPD	35.68 ^e	3.93 ^c	25.84 ^f	1.87 ^c	17.51 ^f	1.61 ^e	28.76 ^f	4.29 ^c	22.84 ^f	2.59 ^d
ROC-TLS	34.56 ^d	1.71 ^a	24.23 ^c	0.41 ^a	16.36 ^c	0.94 ^b	26.10 ^b	1.38 ^b	21.11 ^b	0.99 ^a
FPML	31.05 ^a	4.08 ^d	23.53 ^a	2.27 ^d	16.17 ^a	1.71 ^f	26.48 ^c	4.69 ^e	21.15 ^c	2.87 ^e
ROC-GLM	34.29 ^c	3.91 ^b	24.78 ^d	1.70 ^b	16.62 ^e	1.35 ^d	27.77 ^e	4.42 ^d	21.88 ^d	2.48 ^c
ROC-SLS	32.17 ^b	-5.46 ^e	24.20 ^b	-3.69 ^f	16.41 ^d	-1.20 ^c	25.68 ^a	-1.03 ^a	20.98 ^a	-1.17 ^b
ROCKIT	36.05 ^f	6.72 ^f	25.31 ^e	3.54 ^e	16.36 ^b	0.61 ^a	27.55 ^d	5.20 ^f	22.37 ^e	2.88 ^f

Table B.16: Errors for δ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 1.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	62.62 ^f	14.63 ^g	47.33 ^g	8.82 ^g	30.26 ^g	4.97 ^g	51.88 ^g	11.83 ^g	39.85 ^g	7.67 ^g
ROC-MPD	45.03 ^d	7.30 ^c	31.77 ^f	3.77 ^d	21.36 ^f	2.40 ^f	37.53 ^f	7.28 ^d	28.73 ^f	4.22 ^d
ROC-TLS	68.64 ^g	8.49 ^d	29.98 ^c	1.35 ^a	19.79 ^c	1.15 ^b	33.50 ^c	1.88 ^b	26.53 ^c	1.26 ^b
FPML	35.35 ^a	6.23 ^b	26.50 ^a	3.52 ^c	18.17 ^a	2.31 ^e	31.29 ^a	6.76 ^c	24.49 ^a	4.08 ^c
ROC-GLM	45.00 ^c	9.08 ^e	30.87 ^d	4.26 ^e	20.13 ^d	2.23 ^d	37.14 ^e	8.58 ^f	28.00 ^e	4.65 ^e
ROC-SLS	39.28 ^b	-4.97 ^a	29.23 ^b	-3.29 ^b	19.75 ^b	-1.05 ^a	32.50 ^b	-1.26 ^a	26.19 ^b	-1.18 ^a
ROCKIT	46.62 ^e	11.20 ^f	31.39 ^e	6.06 ^f	20.88 ^e	1.62 ^c	35.35 ^d	7.70 ^e	27.66 ^d	5.16 ^f

Table B.17: Errors for δ for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 2$.

method	$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	69.60 ^g	16.42 ^g	52.80 ^g	12.13 ^g	81.01 ^g	22.84 ^g	65.30 ^g	16.13 ^g
ROC-MPD	43.42 ^c	7.43 ^d	27.98 ^f	3.78 ^d	51.94 ^f	12.49 ^e	39.03 ^e	7.19 ^d
ROC-TLS	43.63 ^d	4.30 ^b	26.45 ^c	1.75 ^b	60.96 ^e	6.59 ^b	36.84 ^c	2.22 ^b
FPML	30.18 ^a	4.78 ^c	20.63 ^a	2.91 ^c	36.98 ^a	8.82 ^c	28.49 ^a	5.28 ^c
ROC-GLM	44.39 ^f	10.48 ^e	26.84 ^d	4.37 ^e	44.78 ^b	-1.58 ^a	39.59 ^f	9.53 ^f
ROC-SLS	39.81 ^b	-2.38 ^a	26.08 ^b	-1.10 ^a	53.85 ^d	17.17 ^f	35.63 ^b	-1.44 ^a
ROCKIT	44.11 ^e	10.50 ^f	27.53 ^e	5.24 ^f	48.92 ^c	11.75 ^d	37.54 ^d	7.85 ^e

Table B.18: Errors for δ for fits to observations from a binormal ROC curve. $\rho = 0.5$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	26.37 ^g	2.33 ^g	19.16 ^g	1.14 ^f	12.79 ^g	1.11 ^g	24.12 ^g	3.07 ^g	18.19 ^g	1.89 ^g
ROC-MPD	24.57 ^f	1.16 ^b	18.48 ^f	0.45 ^c	12.67 ^f	0.77 ^d	22.73 ^f	2.01 ^e	17.66 ^f	1.21 ^d
ROC-TLS	23.68 ^b	1.19 ^c	17.89 ^c	0.51 ^d	12.26 ^c	0.80 ^e	21.75 ^c	1.91 ^d	17.00 ^c	1.29 ^e
FPML	22.81 ^a	1.86 ^e	17.36 ^a	0.88 ^e	11.94 ^a	0.94 ^f	21.29 ^a	2.45 ^f	16.55 ^a	1.49 ^f
ROC-GLM	24.18 ^e	1.02 ^a	18.15 ^e	0.35 ^a	12.42 ^e	0.70 ^c	22.17 ^e	1.87 ^b	17.26 ^e	1.12 ^c
ROC-SLS	23.84 ^c	-2.25 ^f	18.11 ^d	-1.70 ^g	12.34 ^d	-0.42 ^b	21.82 ^d	0.76 ^a	17.10 ^d	0.09 ^a
ROCKIT	23.87 ^d	1.31 ^d	17.79 ^b	0.42 ^b	12.10 ^b	-0.23 ^a	21.67 ^b	1.89 ^c	16.91 ^b	0.75 ^b

Table B.19: Errors for δ for fits to observations from a binormal ROC curve. $\rho = 0.75$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	29.09 ^g	2.62 ^f	21.40 ^g	1.38 ^f	14.26 ^g	1.21 ^g	24.58 ^g	3.05 ^g	19.15 ^g	1.95 ^g
ROC-MPD	26.77 ^f	1.37 ^c	20.20 ^f	0.56 ^c	13.81 ^f	0.88 ^e	23.13 ^f	2.14 ^d	18.37 ^f	1.35 ^e
ROC-TLS	25.37 ^b	0.78 ^a	19.25 ^b	0.26 ^a	13.22 ^b	0.85 ^c	21.65 ^b	1.38 ^b	17.37 ^a	1.05 ^c
FPML	25.17 ^a	1.89 ^d	19.23 ^a	0.95 ^d	13.21 ^a	1.03 ^f	22.06 ^c	2.53 ^f	17.53 ^c	1.58 ^f
ROC-GLM	26.08 ^d	1.21 ^b	19.64 ^d	0.41 ^b	13.40 ^e	0.78 ^b	22.35 ^e	1.98 ^c	17.77 ^e	1.22 ^d
ROC-SLS	25.64 ^c	-4.08 ^g	19.53 ^c	-2.87 ^g	13.27 ^c	-0.91 ^d	21.61 ^a	-0.26 ^a	17.41 ^b	-0.62 ^a
ROCKIT	26.40 ^e	2.32 ^e	19.68 ^e	0.98 ^e	13.34 ^d	0.38 ^a	22.23 ^d	2.51 ^e	17.63 ^d	0.93 ^b

Table B.20: Errors for AUC for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 0$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	7.62 ^g	-0.57 ^g	6.06 ^g	0.31 ^e	3.91 ^a	0.09 ^e	5.83 ^a	0.17 ^e	4.81 ^a	0.17 ^g
ROC-MPD	7.84 ^h	-0.84 ^h	5.67 ^a	-0.39 ^g	4.19 ^g	0.04 ^b	6.34 ^h	0.19 ^f	5.25 ^g	0.11 ^c
ROC-TLS	7.47 ^c	-0.29 ^d	5.68 ^b	-0.15 ^b	4.00 ^c	0.12 ^f	5.87 ^b	0.45 ^h	4.88 ^c	0.21 ^h
FPML	7.58 ^f	-0.46 ^f	6.11 ^h	-0.46 ^h	4.21 ^h	-0.06 ^c	6.11 ^f	-0.22 ^g	5.31 ^h	-0.16 ^f
ROC-GLM	7.34 ^a	-0.30 ^e	6.05 ^f	-0.25 ^c	4.04 ^d	0.19 ^h	5.92 ^c	0.11 ^b	4.92 ^d	0.13 ^d
ROC-SLS	7.38 ^b	0.04 ^a	5.90 ^e	-0.35 ^f	4.17 ^f	0.03 ^a	6.12 ^g	-0.11 ^c	5.07 ^f	-0.08 ^b
ROCKIT	7.55 ^e	0.24 ^c	5.71 ^c	-0.29 ^d	3.94 ^b	-0.08 ^d	6.02 ^d	0.00 ^a	4.87 ^b	-0.05 ^a
empirical	7.49 ^d	-0.08 ^b	5.88 ^d	-0.06 ^a	4.08 ^e	0.14 ^g	6.03 ^e	0.17 ^d	5.03 ^e	0.14 ^e

Table B.21: Errors for AUC for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	7.22 ^f	-0.40 ^g	5.77 ^g	0.36 ^f	3.74 ^a	0.10 ^f	5.59 ^b	0.21 ^g	4.61 ^a	0.18 ^h
ROC-MPD	7.43 ^h	-0.91 ^h	5.42 ^a	-0.42 ^g	4.00 ^g	0.03 ^a	6.02 ^h	0.10 ^c	5.01 ^h	0.04 ^a
ROC-TLS	7.05 ^b	-0.20 ^c	5.44 ^b	-0.13 ^b	3.79 ^c	0.10 ^e	5.61 ^c	0.38 ^h	4.64 ^b	0.16 ^g
FPML	7.18 ^e	-0.40 ^f	5.83 ^h	-0.42 ^h	3.97 ^f	-0.04 ^b	5.76 ^f	-0.17 ^f	4.99 ^g	-0.16 ^f
ROC-GLM	7.07 ^c	-0.28 ^d	5.75 ^f	-0.22 ^c	3.82 ^d	0.17 ^h	5.56 ^a	0.08 ^b	4.65 ^c	0.11 ^d
ROC-SLS	7.04 ^a	0.07 ^b	5.61 ^e	-0.32 ^e	4.01 ^h	0.05 ^c	5.81 ^g	-0.11 ^d	4.89 ^f	-0.06 ^c
ROCKIT	7.27 ^g	0.36 ^e	5.49 ^c	-0.23 ^d	3.77 ^b	-0.07 ^d	5.71 ^d	0.03 ^a	4.68 ^d	-0.04 ^b
empirical	7.11 ^d	-0.06 ^a	5.59 ^d	-0.07 ^a	3.87 ^e	0.13 ^g	5.71 ^e	0.14 ^e	4.78 ^e	0.12 ^e

Table B.22: Errors for AUC for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 1$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	6.19 ^e	-0.28 ^f	4.94 ^f	0.31 ^f	3.25 ^a	0.06 ^e	4.84 ^b	0.22 ^g	6.19 ^e	-0.28 ^f
ROC-MPD	6.37 ^g	-0.80 ^h	6.37 ^g	-0.80 ^h	3.47 ^g	-0.03 ^c	5.18 ^h	-0.03 ^a	6.37 ^g	-0.80 ^h
ROC-TLS	6.08 ^a	-0.23 ^d	6.08 ^a	-0.23 ^d	3.26 ^b	0.02 ^a	4.86 ^c	0.22 ^h	6.08 ^a	-0.23 ^d
FPML	6.19 ^f	-0.34 ^g	6.19 ^f	-0.34 ^g	3.41 ^f	-0.05 ^d	4.91 ^e	-0.19 ^f	6.19 ^f	-0.34 ^g
ROC-GLM	6.11 ^b	-0.24 ^e	6.11 ^b	-0.24 ^e	3.28 ^d	0.10 ^g	4.78 ^a	0.05 ^c	6.11 ^b	-0.24 ^e
ROC-SLS	6.12 ^c	-0.04 ^a	6.12 ^c	-0.04 ^a	3.51 ^h	0.03 ^b	5.03 ^g	-0.11 ^e	6.12 ^c	-0.04 ^a
ROCKIT	6.40 ^h	0.22 ^c	6.40 ^h	0.22 ^c	3.28 ^c	-0.09 ^f	4.96 ^f	-0.05 ^b	6.40 ^h	0.22 ^c
empirical	6.13 ^d	-0.04 ^b	6.13 ^d	-0.04 ^b	3.33 ^e	0.11 ^h	4.90 ^d	0.09 ^d	6.13 ^d	-0.04 ^b

Table B.23: Errors for AUC for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 1.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 30, n_2 = 50$		$n_1 = 50, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	4.86 ^e	-0.24 ^d	3.83 ^e	0.18 ^d	2.60 ^d	0.01 ^a	3.81 ^b	0.10 ^d	3.19 ^c	0.09 ^f	3.19 ^c	0.09 ^f
ROC-MPD	4.99 ^h	-0.76 ^h	3.70 ^b	-0.38 ^h	2.75 ^g	-0.06 ^f	4.09 ^h	-0.11 ^f	3.44 ^h	-0.05 ^d	3.44 ^h	-0.05 ^d
ROC-TLS	4.77 ^a	-0.32 ^f	3.70 ^a	-0.12 ^b	2.59 ^b	-0.02 ^d	3.83 ^c	0.07 ^c	3.15 ^b	0.02 ^b	3.15 ^b	0.02 ^b
FPML	4.84 ^d	-0.37 ^g	4.00 ^h	-0.35 ^g	2.68 ^f	-0.04 ^e	3.85 ^e	-0.21 ^h	3.31 ^f	-0.13 ^h	3.31 ^f	-0.13 ^h
ROC-GLM	4.86 ^f	-0.31 ^e	3.85 ^g	-0.24 ^e	2.57 ^a	0.02 ^c	3.73 ^a	-0.04 ^a	3.10 ^a	0.01 ^a	3.10 ^a	0.01 ^a
ROC-SLS	4.80 ^c	-0.13 ^c	3.84 ^f	-0.24 ^f	2.80 ^h	-0.01 ^b	3.95 ^g	-0.15 ^g	3.39 ^g	-0.07 ^e	3.39 ^g	-0.07 ^e
ROCKIT	4.98 ^g	0.06 ^b	3.71 ^c	-0.18 ^c	2.63 ^e	-0.08 ^h	3.91 ^f	-0.10 ^e	3.28 ^e	-0.10 ^g	3.28 ^e	-0.10 ^g
empirical	4.80 ^b	-0.02 ^a	3.74 ^d	-0.05 ^a	2.60 ^c	0.08 ^g	3.83 ^d	0.05 ^b	3.21 ^d	0.05 ^c	3.21 ^d	0.05 ^c

Table B.24: Errors for AUC for fits to observations from a binormal ROC curve. $\rho = 1$, $\delta = 2$.

method	$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	2.70 ^d	0.02 ^a	1.86 ^c	-0.02 ^c	2.70 ^b	-0.05 ^b	2.27 ^c	-0.02 ^b
ROC-MPD	2.71 ^e	-0.37 ^h	1.98 ^g	-0.06 ^h	2.94 ^h	-0.21 ^g	2.46 ^h	-0.13 ^f
ROC-TLS	2.65 ^b	-0.16 ^c	1.86 ^d	-0.04 ^e	2.77 ^e	-0.08 ^c	2.25 ^b	-0.07 ^c
FPML	2.83 ^h	-0.34 ^g	1.91 ^f	-0.03 ^d	2.75 ^d	-0.24 ^h	2.33 ^e	-0.14 ^h
ROC-GLM	2.72 ^f	-0.25 ^f	1.82 ^a	-0.01 ^b	2.81 ^g	-0.21 ^f	2.19 ^a	-0.08 ^d
ROC-SLS	2.73 ^g	-0.24 ^e	2.00 ^h	-0.01 ^a	2.65 ^a	-0.14 ^d	2.40 ^g	-0.10 ^e
ROCKIT	2.62 ^a	-0.18 ^d	1.91 ^e	-0.06 ^g	2.80 ^f	-0.18 ^e	2.37 ^f	-0.13 ^g
empirical	2.66 ^c	-0.03 ^b	1.84 ^b	0.05 ^f	2.72 ^c	0.01 ^a	2.28 ^d	0.01 ^a

Table B.25: Errors for AUC for fits to observations from a binormal ROC curve. $\rho = 0.5$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	7.23 ^d	-0.52 ^g	5.77 ^g	0.30 ^f	3.77 ^b	0.06 ^d	6.59 ^c	0.26 ^f	5.19 ^d	0.22 ^g
ROC-MPD	7.46 ^h	-0.66 ^h	5.41 ^a	-0.34 ^g	3.99 ^g	0.13 ^g	6.89 ^h	0.33 ^g	5.56 ^h	0.24 ^h
ROC-TLS	7.28 ^f	-0.33 ^e	5.49 ^b	-0.21 ^b	3.83 ^c	0.00 ^a	6.55 ^b	0.39 ^h	5.14 ^a	0.10 ^d
FPML	7.09 ^a	-0.26 ^c	5.97 ^h	-0.24 ^c	3.99 ^h	0.02 ^b	6.75 ^g	-0.06 ^b	5.55 ^g	-0.08 ^b
ROC-GLM	7.09 ^b	-0.45 ^f	5.54 ^d	-0.37 ^h	3.87 ^d	0.09 ^e	6.50 ^a	0.08 ^c	5.17 ^b	0.11 ^e
ROC-SLS	7.11 ^c	0.15 ^a	5.63 ^e	-0.28 ^d	3.92 ^f	0.03 ^c	6.63 ^d	-0.09 ^d	5.30 ^e	-0.06 ^a
ROCKIT	7.45 ^g	0.31 ^d	5.49 ^c	-0.28 ^e	3.73 ^a	-0.21 ^h	6.71 ^f	-0.03 ^a	5.19 ^c	-0.12 ^f
empirical	7.25 ^e	-0.15 ^b	5.64 ^f	-0.14 ^a	3.90 ^e	0.10 ^f	6.69 ^e	0.11 ^e	5.34 ^f	0.09 ^c

Table B.26: Errors for AUC for fits to observations from a binormal ROC curve. $\rho = 0.75$, $\delta = 0.5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	7.17 ^f	-0.46 ^g	5.73 ^g	0.34 ^f	3.71 ^b	0.08 ^e	6.00 ^b	0.23 ^g	4.83 ^a	0.19 ^h
ROC-MPD	7.36 ^h	-0.86 ^h	5.37 ^a	-0.38 ^h	3.96 ^h	0.08 ^d	6.41 ^h	0.18 ^f	5.23 ^h	0.10 ^d
ROC-TLS	7.08 ^c	-0.26 ^c	5.40 ^b	-0.17 ^b	3.77 ^c	0.05 ^b	6.01 ^c	0.37 ^h	4.84 ^b	0.12 ^f
FPML	7.11 ^e	-0.34 ^e	5.86 ^h	-0.34 ^g	3.94 ^g	-0.01 ^a	6.16 ^g	-0.13 ^e	5.21 ^g	-0.12 ^g
ROC-GLM	6.98 ^b	-0.35 ^f	5.61 ^f	-0.31 ^d	3.82 ^d	0.13 ^g	5.97 ^a	0.07 ^b	4.85 ^c	0.09 ^b
ROC-SLS	6.97 ^a	0.10 ^a	5.57 ^e	-0.32 ^e	3.94 ^f	0.05 ^c	6.15 ^f	-0.08 ^c	5.03 ^f	-0.06 ^a
ROCKIT	7.31 ^g	0.32 ^d	5.45 ^c	-0.28 ^c	3.70 ^a	-0.14 ^h	6.11 ^d	-0.00 ^a	4.86 ^d	-0.10 ^c
empirical	7.09 ^d	-0.10 ^b	5.55 ^d	-0.11 ^a	3.84 ^e	0.12 ^f	6.12 ^e	0.12 ^d	4.99 ^e	0.11 ^e

B.2 Biexponential

Table B.27: Global errors for fits to observations from a biexponential ROC curve. $\lambda_1 = 1, \lambda_2 = 2$.

method	$n_1 = 30, n_2 = 30$			$n_1 = 50, n_2 = 50$			$n_1 = 100, n_2 = 100$			$n_1 = 100, n_2 = 30$			$n_1 = 100, n_2 = 50$		
	p	v	h	p	v	h	p	v	h	p	v	h	p	v	h
ROC-CVM	7.55 ^f	8.68 ^f	9.36 ^f	5.65 ^f	6.58 ^f	7.08 ^f	3.94 ^e	4.61 ^e	4.98 ^f	6.22 ^f	7.20 ^e	7.80 ^f	5.04 ^e	5.88 ^e	6.37 ^e
ROC-MPD	7.43 ^d	8.61 ^e	9.20 ^d	5.60 ^e	6.57 ^e	7.00 ^e	3.94 ^f	4.65 ^f	4.97 ^e	6.21 ^e	7.23 ^f	7.76 ^e	5.07 ^f	5.94 ^f	6.37 ^f
ROC-TLS	6.95 ^a	8.08 ^a	8.62 ^a	5.26 ^a	6.17 ^a	6.58 ^a	3.70 ^a	4.36 ^a	4.67 ^a	5.78 ^a	6.73 ^a	7.24 ^a	4.73 ^a	5.54 ^a	5.96 ^a
FPML	9.29 ^g	10.67 ^g	10.73 ^g	8.13 ^g	9.50 ^g	9.55 ^g	7.21 ^g	8.55 ^g	8.62 ^g	7.95 ^g	9.29 ^g	9.39 ^g	7.50 ^g	8.83 ^g	8.92 ^g
ROC-GLM	7.46 ^e	8.60 ^d	9.25 ^e	5.55 ^d	6.48 ^d	6.95 ^d	3.84 ^d	4.52 ^d	4.85 ^d	6.03 ^d	7.00 ^d	7.56 ^d	4.93 ^d	5.76 ^d	6.21 ^d
ROC-SLS	7.22 ^c	8.39 ^c	8.95 ^b	5.40 ^b	6.34 ^c	6.75 ^b	3.76 ^c	4.44 ^c	4.73 ^c	5.88 ^c	6.86 ^c	7.37 ^c	4.82 ^c	5.66 ^c	6.06 ^c
ROCKIT	7.22 ^b	8.33 ^b	8.97 ^c	5.42 ^c	6.34 ^b	6.79 ^c	3.74 ^b	4.39 ^b	4.72 ^b	5.87 ^b	6.83 ^b	7.36 ^b	4.79 ^b	5.60 ^b	6.04 ^b

Table B.28: Global errors for fits to observations from a biexponential ROC curve. $\lambda_1 = 1, \lambda_2 = 15$.

method	$n_1 = 30, n_2 = 30$			$n_1 = 50, n_2 = 50$			$n_1 = 100, n_2 = 100$			$n_1 = 100, n_2 = 30$			$n_1 = 100, n_2 = 50$		
	p	v	h	p	v	h	p	v	h	p	v	h	p	v	h
ROC-CVM	4.57 ^e	5.82 ^e	8.45 ^c	3.05 ^f	3.64 ^f	6.63 ^f	2.20 ^f	2.49 ^c	4.97 ^f	3.78 ^e	4.38 ^e	7.98 ^c	2.95 ^f	3.31 ^c	6.50 ^f
ROC-MPD	3.67 ^a	4.40 ^b	7.60 ^a	2.93 ^c	3.44 ^c	6.20 ^d	2.10 ^d	2.49 ^e	4.54 ^e	3.57 ^b	4.19 ^c	7.44 ^b	2.88 ^c	3.38 ^d	6.12 ^d
ROC-TLS	7.14 ^f	11.04 ^f	9.74 ^d	2.89 ^b	3.43 ^b	5.94 ^a	2.02 ^b	2.46 ^b	4.27 ^a	4.66 ^f	7.36 ^f	8.32 ^e	2.79 ^b	3.25 ^b	5.80 ^b
FPML	23.16 ^g	11.53 ^g	26.40 ^g	23.57 ^g	11.71 ^g	27.00 ^g	24.07 ^g	11.92 ^g	27.59 ^g	22.87 ^g	11.48 ^g	26.32 ^g	23.50 ^g	11.72 ^g	27.00 ^g
ROC-GLM	4.18 ^c	4.83 ^d	11.65 ^f	3.01 ^d	3.52 ^d	6.39 ^e	2.10 ^c	2.49 ^d	4.54 ^d	3.71 ^d	4.28 ^d	8.36 ^f	2.90 ^d	3.41 ^f	6.25 ^e
ROC-SLS	4.22 ^d	4.70 ^c	10.89 ^e	3.01 ^e	3.56 ^e	6.12 ^c	2.14 ^e	2.64 ^f	4.40 ^c	3.69 ^c	4.08 ^b	8.09 ^d	2.91 ^e	3.39 ^e	6.09 ^c
ROCKIT	3.95 ^b	4.21 ^a	7.92 ^b	2.82 ^a	3.16 ^a	5.98 ^b	2.00 ^a	2.28 ^a	4.30 ^b	3.42 ^a	3.64 ^a	7.08 ^a	2.70 ^a	2.99 ^a	5.76 ^a

Table B.29: Errors for AUC for fits to observations from a biexponential ROC curve. $\lambda_1 = 1, \lambda_2 = 2$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	7.03 ^g	-0.26 ^h	5.26 ^d	0.07 ^f	3.84 ^g	-0.11 ^e	6.07 ^h	0.03 ^a	4.99 ^h	-0.10 ^b
ROC-MPD	6.84 ^c	0.23 ^g	5.11 ^a	0.05 ^d	3.61 ^b	0.00 ^a	5.84 ^e	-0.15 ^c	4.73 ^c	-0.14 ^c
ROC-TLS	7.20 ^h	-0.22 ^f	5.28 ^e	-0.10 ^g	3.84 ^h	0.03 ^c	5.67 ^b	-0.17 ^e	4.82 ^f	-0.20 ^e
FPML	6.63 ^a	0.01 ^b	5.23 ^c	0.14 ^h	3.52 ^a	0.24 ^h	5.56 ^a	0.37 ^f	4.57 ^a	0.31 ^g
ROC-GLM	6.81 ^b	0.01 ^a	5.31 ^f	0.04 ^c	3.70 ^e	0.12 ^f	5.72 ^d	0.37 ^g	4.77 ^e	0.23 ^f
ROC-SLS	6.89 ^d	-0.09 ^e	5.37 ^h	-0.04 ^b	3.68 ^d	0.18 ^g	5.71 ^c	0.15 ^d	4.73 ^d	0.19 ^d
ROCKIT	7.01 ^e	0.06 ^d	5.21 ^b	-0.06 ^e	3.65 ^c	-0.11 ^d	5.93 ^g	-0.44 ^h	4.65 ^b	-0.35 ^h
empirical	7.01 ^f	0.02 ^c	5.34 ^g	-0.00 ^a	3.77 ^f	-0.01 ^b	5.92 ^f	-0.07 ^b	4.87 ^g	-0.08 ^a

Table B.30: Errors for AUC for fits to observations from a biexponential ROC curve. $\lambda_1 = 1, \lambda_2 = 15$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	5.53 ^h	-2.21 ^h	4.76 ^g	-1.76 ^h	4.58 ^h	-1.71 ^h	5.04 ^g	-1.69 ^d	4.82 ^h	-1.61 ^e
ROC-MPD	5.21 ^b	-1.85 ^b	4.66 ^b	-1.58 ^b	4.52 ^d	-1.64 ^e	5.04 ^h	-1.86 ^h	4.71 ^e	-1.68 ^f
ROC-TLS	5.43 ^f	-2.09 ^f	4.72 ^e	-1.69 ^g	4.54 ^f	-1.65 ^g	5.00 ^e	-1.85 ^g	4.79 ^g	-1.73 ^h
FPML	5.51 ^g	-2.12 ^g	4.76 ^f	-1.65 ^d	4.49 ^c	-1.50 ^b	4.88 ^c	-1.49 ^b	4.65 ^c	-1.40 ^b
ROC-GLM	5.36 ^d	-2.06 ^e	4.70 ^c	-1.60 ^c	4.52 ^e	-1.59 ^d	4.82 ^b	-1.51 ^c	4.61 ^b	-1.48 ^c
ROC-SLS	5.39 ^e	-2.02 ^d	4.79 ^h	-1.68 ^e	4.47 ^b	-1.53 ^c	4.96 ^d	-1.81 ^e	4.66 ^d	-1.57 ^d
ROCKIT	5.31 ^c	-1.98 ^c	4.70 ^d	-1.69 ^f	4.55 ^g	-1.64 ^f	5.02 ^f	-1.82 ^f	4.73 ^f	-1.71 ^g
empirical	3.34 ^a	-0.02 ^a	2.50 ^a	-0.04 ^a	1.79 ^a	0.01 ^a	3.11 ^a	-0.01 ^a	2.48 ^a	0.01 ^a

B.3 Normal-gamma

Table B.31: Global errors for fits to observations from a normal-gamma ROC curve. $\mu_1 = 0, \sigma_1 = 1, \alpha_2 = 2, \beta_2 = 0.5, \theta_2 = 5$.

method	$n_1 = 30, n_2 = 30$			$n_1 = 50, n_2 = 50$			$n_1 = 100, n_2 = 100$			$n_1 = 100, n_2 = 30$			$n_1 = 100, n_2 = 50$		
	p	v	h	p	v	h	p	v	h	p	v	h	p	v	h
ROC-CVM	8.29 ^e	8.54 ^f	13.20 ^f	6.43 ^f	6.67 ^f	10.44 ^f	4.75 ^f	4.88 ^f	7.66 ^f	8.04 ^f	8.28 ^f	13.00 ^f	6.31 ^f	6.53 ^f	10.30 ^f
ROC-MPD	8.11 ^b	8.24 ^c	12.91 ^b	6.25 ^d	6.45 ^d	10.18 ^c	4.54 ^d	4.66 ^d	7.35 ^d	7.84 ^e	8.01 ^e	12.62 ^e	6.12 ^e	6.30 ^e	9.98 ^e
ROC-TLS	7.95 ^a	8.16 ^a	12.64 ^a	6.10 ^d	6.34 ^a	9.93 ^a	4.46 ^a	4.57 ^a	7.19 ^a	7.57 ^a	7.79 ^a	12.27 ^a	5.94 ^a	6.13 ^a	9.70 ^d
FPML	10.55 ^g	10.81 ^g	16.60 ^g	9.17 ^g	9.55 ^g	14.72 ^g	8.04 ^g	8.45 ^g	13.06 ^g	10.30 ^g	10.60 ^g	16.30 ^g	9.03 ^g	9.40 ^g	14.51 ^g
ROC-GLM	8.31 ^f	8.27 ^d	13.12 ^e	6.27 ^e	6.43 ^c	10.23 ^e	4.50 ^c	4.61 ^c	7.30 ^c	7.77 ^d	7.89 ^c	12.54 ^d	6.06 ^c	6.22 ^b	9.92 ^c
ROC-SLS	8.17 ^c	8.31 ^e	12.95 ^c	6.24 ^c	6.46 ^e	10.17 ^b	4.50 ^b	4.61 ^b	7.28 ^b	7.73 ^c	7.91 ^d	12.48 ^b	6.04 ^b	6.22 ^c	9.87 ^b
ROCKIT	8.18 ^d	8.19 ^b	12.98 ^d	6.22 ^b	6.40 ^b	10.19 ^d	4.60 ^e	4.72 ^e	7.45 ^e	7.70 ^b	7.86 ^b	12.52 ^c	6.07 ^d	6.24 ^d	9.96 ^d

Table B.32: Global errors for fits to observations from a normal-gamma ROC curve. $\mu_1 = 0, \sigma_1 = 1, \alpha_2 = 3, \beta_2 = 2, \theta_2 = 3$.

method	$n_1 = 30, n_2 = 30$			$n_1 = 50, n_2 = 50$			$n_1 = 100, n_2 = 100$			$n_1 = 100, n_2 = 30$			$n_1 = 100, n_2 = 50$		
	p	v	h	p	v	h	p	v	h	p	v	h	p	v	h
ROC-CVM	5.77 ^e	7.98 ^d	9.07 ^g	4.95 ^g	6.64 ^f	8.22 ^g	4.09 ^g	5.24 ^g	7.38 ^g	5.15 ^g	6.42 ^f	8.69 ^g	4.58 ^g	5.78 ^f	7.99 ^g
ROC-MPD	5.34 ^a	7.77 ^b	8.03 ^a	4.13 ^a	6.19 ^a	6.39 ^b	2.98 ^a	4.68 ^a	4.72 ^b	4.33 ^a	6.11 ^a	7.03 ^b	3.61 ^a	5.33 ^a	5.83 ^b
ROC-TLS	5.85 ^g	8.37 ^f	8.50 ^e	4.53 ^f	6.58 ^e	6.65 ^e	3.61 ^f	5.22 ^e	5.19 ^e	4.55 ^e	6.17 ^c	7.14 ^d	4.01 ^e	5.60 ^e	6.09 ^e
FPML	5.59 ^c	7.60 ^a	8.54 ^f	4.53 ^e	6.22 ^b	7.16 ^f	3.52 ^e	4.88 ^d	5.80 ^f	4.76 ^f	6.22 ^d	7.74 ^f	4.10 ^f	5.50 ^c	6.74 ^f
ROC-GLM	5.45 ^b	7.92 ^c	8.27 ^b	4.17 ^b	6.37 ^c	6.32 ^a	3.01 ^b	4.85 ^b	4.57 ^a	4.35 ^b	6.13 ^b	7.02 ^a	3.62 ^b	5.42 ^b	5.74 ^a
ROC-SLS	5.81 ^f	8.38 ^g	8.48 ^d	4.48 ^d	6.68 ^g	6.63 ^d	3.31 ^d	5.23 ^f	4.91 ^c	4.53 ^c	6.52 ^g	7.10 ^c	3.87 ^d	5.83 ^g	5.95 ^c
ROCKIT	5.67 ^d	8.22 ^e	8.35 ^c	4.43 ^c	6.54 ^d	6.62 ^c	3.28 ^c	4.85 ^c	4.93 ^d	4.55 ^d	6.24 ^e	7.23 ^e	3.87 ^c	5.50 ^d	6.06 ^d

Table B.33: Errors for AUC for fits to observations from a normal-gamma ROC curve. $\mu_1 = 0, \sigma_1 = 1, \alpha_2 = 2, \beta_2 = 0.5, \theta_2 = 5$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	7.81 ^c	-0.99 ^b	6.42 ^f	-1.62 ^h	4.80 ^b	-1.18 ^e	7.94 ^f	-1.22 ^g	6.53 ^h	-1.29 ^f
ROC-MPD	7.91 ^d	-1.48 ^g	6.36 ^b	-1.16 ^d	4.92 ^e	-1.21 ^f	7.71 ^d	-0.91 ^d	6.26 ^e	-1.03 ^e
ROC-TLS	8.16 ^h	-1.36 ^f	6.37 ^c	-1.31 ^f	4.85 ^c	-1.05 ^c	7.70 ^c	-1.01 ^e	6.16 ^b	-0.97 ^d
FPML	8.04 ^f	-1.05 ^c	6.41 ^e	-1.16 ^e	4.89 ^d	-0.95 ^b	7.56 ^b	-0.70 ^b	6.26 ^d	-0.94 ^c
ROC-GLM	7.64 ^b	-1.21 ^e	6.43 ^g	-1.33 ^g	5.05 ^h	-1.49 ^g	8.03 ^g	-1.17 ^f	6.41 ^g	-1.31 ^g
ROC-SLS	8.09 ^g	-1.20 ^d	6.57 ^h	-1.04 ^c	5.02 ^g	-1.54 ^h	8.04 ^h	-1.78 ^h	6.39 ^f	-1.49 ^h
ROCKIT	8.04 ^e	-1.49 ^h	6.38 ^d	-0.98 ^b	4.94 ^f	-1.05 ^d	7.77 ^e	-0.82 ^c	6.20 ^c	-0.92 ^b
empirical	7.40 ^a	-0.28 ^a	5.79 ^a	-0.19 ^a	4.13 ^a	-0.07 ^a	7.27 ^a	-0.09 ^a	5.68 ^a	-0.06 ^a

Table B.34: Errors for AUC for fits to observations from a normal-gamma ROC curve. $\mu_1 = 0, \sigma_1 = 1, \alpha_2 = 3, \beta_2 = 2, \theta_2 = 3$.

method	$n_1 = 30, n_2 = 30$		$n_1 = 50, n_2 = 50$		$n_1 = 100, n_2 = 100$		$n_1 = 100, n_2 = 30$		$n_1 = 100, n_2 = 50$	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
ROC-CVM	4.72 ^a	-0.16 ^b	3.90 ^f	-0.03 ^d	2.83 ^c	0.10 ^e	3.91 ^c	0.01 ^a	3.33 ^c	0.10 ^e
ROC-MPD	4.86 ^e	-0.35 ^e	3.74 ^b	-0.03 ^c	2.86 ^d	0.14 ^f	3.87 ^b	0.14 ^g	3.32 ^b	0.22 ^h
ROC-TLS	4.81 ^d	-0.40 ^f	3.93 ^g	-0.13 ^g	2.92 ^f	0.00 ^a	4.05 ^f	-0.12 ^f	3.38 ^e	0.08 ^d
FPML	4.80 ^c	0.06 ^a	3.81 ^d	0.00 ^a	2.95 ^h	0.01 ^b	4.07 ^g	0.07 ^e	3.56 ^h	0.00 ^a
ROC-GLM	5.02 ^g	-0.48 ^g	4.06 ^h	-0.32 ^h	2.91 ^e	-0.14 ^h	3.92 ^d	-0.04 ^c	3.40 ^f	-0.01 ^b
ROC-SLS	5.16 ^h	-0.52 ^h	3.90 ^e	-0.03 ^b	2.93 ^g	-0.14 ^g	4.08 ^h	-0.07 ^d	3.41 ^g	-0.15 ^f
ROCKIT	4.90 ^f	-0.22 ^d	3.79 ^c	0.08 ^e	2.70 ^b	0.05 ^c	3.97 ^e	0.35 ^h	3.35 ^d	0.15 ^g
empirical	4.73 ^b	-0.19 ^c	3.66 ^a	-0.09 ^f	2.58 ^a	-0.06 ^d	3.77 ^a	-0.04 ^b	3.14 ^a	-0.03 ^c

Appendix C: A chi-squared goodness-of-fit test for continuous ROC curves

Our chi-squared goodness-of-fit test for continuous ROC curves depends on a difference between counts observed with the empirical ROC curve and counts expected with the parametric one. In particular, this chi-squared test compares the observed number of true positive scores between pairs of false positive scores to the expected number. On an empirical ROC curve, the observed number of true positive scores between a pair of false positive scores p_1 and p_2 with $p_1 < p_2$ is $\widehat{ROC}(p_2) - \widehat{ROC}(p_1)$, which is the length of the vertical jump at p_2 . The expected number of true positive scores at p_2 comes from evaluating the parametric ROC curve at the boundaries of a bin centered on p_2 and taking the difference.

To define the bins, we divide the horizontal axis according to the total number n_1 of observations from the negative class. The widths of the bins would all ideally equal $1/n_1$, and the bins' centers would be integer multiples of $1/n_1$ (which allow for all possible false positive rates). However, the domain of the ROC curve hits its lower bound at the false positive rate of zero and its upper bound at the false positive rate of one. So the first and last bins cannot be centered about the first and last false positive rates. To handle this matter, we set the widths of the first and last bins to $1/2n_1$.

The specific boundaries of the bins are then:

$$\begin{aligned}
 \text{bin}_1 &: \left(0, \frac{1}{2n_1}\right) \\
 \text{bin}_2 &: \left(\frac{1}{2n_1}, \frac{3}{2n_1}\right) \\
 \text{bin}_3 &: \left(\frac{3}{2n_1}, \frac{5}{2n_1}\right) \\
 &\vdots \\
 \text{bin}_{n_1} &: \left(\frac{n_1-3}{2n_1}, \frac{n_1-1}{2n_1}\right) \\
 \text{bin}_{n_1+1} &: \left(\frac{n_1-1}{2n_1}, 1\right).
 \end{aligned} \tag{C.1}$$

With the bins so defined, the observed counts in each bin are:

$$\begin{aligned}
 o_1 &= \widehat{ROC}(0) \\
 o_2 &= \widehat{ROC}(1/n_1) - \widehat{ROC}(0) \\
 o_3 &= \widehat{ROC}(2/n_1) - \widehat{ROC}(1/n_1) \\
 &\vdots \\
 o_{n_1} &= \widehat{ROC}(1) - \widehat{ROC}(1 - 1/n_1) \\
 o_{n_1+1} &= \widehat{ROC}(1)
 \end{aligned} \tag{C.2}$$

The expected counts follow from the difference between the values of the parametric ROC curve at

the upper and lower boundaries of each bin:

$$\begin{aligned}
 e_1 &= ROC_\theta\left(\frac{1}{2n_1}\right) - ROC_\theta(0) \\
 e_2 &= ROC_\theta\left(\frac{3}{2n_1}\right) - ROC_\theta\left(\frac{1}{2n_1}\right) \\
 e_3 &= ROC_\theta\left(\frac{5}{2n_1}\right) - ROC_\theta\left(\frac{3}{2n_1}\right) \\
 &\vdots \\
 e_{n_1} &= ROC_\theta\left(\frac{n_1-1}{2n_1}\right) - ROC_\theta\left(\frac{n_1-3}{2n_1}\right) \\
 e_{n_1+1} &= ROC_\theta(1) - ROC_\theta\left(\frac{n_1-1}{2n_1}\right).
 \end{aligned} \tag{C.3}$$

Now the value of our chi-squared goodness-of-fit test for a continuous ROC is:

$$\chi^2 = \sum_{i=1}^{n_1+1} \frac{(o_i - e_i)^2}{e_i}. \tag{C.4}$$

We must again caution that we do not recommend applying this test in general without further study.

Appendix D: Computing environment

D.1 Software

We conducted the majority of simulations and analyses with the statistical language and environment R (R Development Core Team, 2011). Much of our computer code is new. R also has numerous contributed packages for specific tasks. We used the `deSolve` package by Soetaert et al. (2010) for solving differential equations numerically and the `multicore` package by Urbanek (2011) for running multiple processor cores in parallel. In addition, we created figures with the `ggplot2` package by Wickham (2009) and investigated empirical ROCs with the `ROCR` package by Sing et al. (2009).

We wrote the core code for the studentized bootstrap confidence intervals and regions in the computer language C with additional functions from the GNU Scientific Library (Galassi et al., 2009).

All of the computer code ran on the GNU/Linux operating system.

D.2 Hardware

The two primary computers on which we ran our code had the following processors:

1. Intel®Core™i7-2600K, 3.40GHz, 4 cores with hyper-threading, and
2. Intel®Xeon®X5450, 3.00GHz, 8 cores.

D.3 Times

As a reference for further studies, we report the time to run a few simulations. Table D.1 lists the times to complete a simulation for studying the goodness-of-fit statistics. Each simulation consists of the steps outlined in Section 4.1.

Table D.1: Time to complete simulations for goodness-of-fit statistics applied to binormal ROC curves.

ρ	δ	n_1	n_2	machine	cores	time (hr)	time on single core (days)
1	1	50	50	Intel®Xeon®X5450	6	9.98	2.5
1	1	1000	1000	Intel®Xeon®X5450	6	27.74	6.9
0.5	1	50	50	Intel®Core™i7-2600K	4	9.16	1.5
0.5	1	1000	1000	Intel®Core™i7-2600K	2	53.24	4.4

Bibliography

Bibliography

- T. A. Alonzo and M. S. Pepe. Distribution-free ROC analysis using binary regression techniques. *Biostatistics*, 3(3):421–432, 2002.
- E. Anderson. The Irises of the Gaspé Peninsula. *Bulletin of the American Iris Society*, 59:2–5, 1935.
- T. W. Anderson and D. A. Darling. Asymptotic theory of certain ‘goodness of fit’ criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):193–212, June 1952.
- T. W. Anderson and D. A. Darling. A test of goodness of fit. *Journal of the American Statistical Association*, 49(268):pp. 765–769, 1954.
- D. Bamber. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12(4):387 – 415, 1975.
- T. Cai. Semi-parametric ROC regression analysis with placement values. *Biostatistics*, 5(1):45–60, Jan. 2004.
- T. Cai and C. S. Moskowitz. Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test. *Biostatistics*, 5(4):573 –586, Oct. 2004.
- T. Cai and M. S. Pepe. Semiparametric receiver operating characteristic analysis to evaluate biomarkers for disease. *Journal of the American Statistical Association*, 97(460):1099–1107, Dec. 2002.
- T. Cai and Y. Zheng. Model checking for ROC regression analysis. *Biometrics*, 63(1):152–163, Mar. 2007.
- G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, 2 edition, 2002.
- H. Cramér. On the composition of elementary errors. *Skandinavisk Aktuarietidskrift*, 11:pp. 141–180, 1928.
- M. Csörgö. *Quantile Processes with Statistical Applications*. SIAM, 1983.
- R. B. D’Agostino and M. A. Stephens. *Goodness-of-fit Techniques*, chapter 1. CRC Press, New York, 1986.
- A. DasGupta. *Asymptotic Theory of Statistics and Probability*. Springer, 2008.
- O. Davidov and Y. Nov. Minimum-norm estimation for binormal receiver operating characteristic (ROC) curves. *Biometrical Journal*, 51(6):1030–1046, 2009.

- A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, New York, 1997.
- D. D. Dorfman and E. Alf. Maximum likelihood estimation of parameters of signal detection theory—a direct solution. *Psychometrika*, 33(1):117–124, Mar. 1968.
- D. D. Dorfman and E. Alf. Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating-method data. *Journal of Mathematical Psychology*, 6(3):487 – 496, Oct. 1969.
- J. Durbin. *Distribution Theory for Tests Based on the Sample Distribution Function*. SIAM, 1973.
- T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7(2):179–188, 1936.
- M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Rossi. *GNU Scientific Library Reference Manual - Third Edition*. Network Theory Ltd., 3rd revised edition edition, Jan. 2009.
- M. J. Goddard and I. Hinberg. Receiver operator characteristic (ROC) curves and non-normal data: An empirical study. *Statistics in Medicine*, 9(3):325–337, 1990.
- D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics*. John Wiley and Sons, New York, 1966.
- P. E. Greenwood and M. S. Nikulin. *A Guide to Chi-squared Testing*. John Wiley & Sons, New York, 1996.
- U. Grenander. Stochastic processes and statistical inference. *Arkiv for Matematik*, 1:195–277, 1950.
- J. Gu and S. Ghosal. Bayesian ROC curve estimation under binormality using a rank likelihood. *Journal of Statistical Planning and Inference*, 139(6):2076–2083, June 2009.
- J. Gu, S. Ghosal, and A. Roy. Bayesian bootstrap estimation of ROC curve. *Statistics in Medicine*, 27(26):5407–5420, Nov. 2008.
- W. Gu and M. S. Pepe. Estimating the diagnostic likelihood ratio of a continuous marker. *Bio-statistics (Oxford, England)*, July 2010.
- J. A. Hanley. The robustness of the ‘binormal’ assumptions used in fitting ROC curves. *Medical Decision Making*, 8(3):197 –203, 1988.
- J. A. Hanley. The use of the ‘binormal’ model for parametric roc analysis of quantitative diagnostic tests. *Statistics in Medicine*, 15(14):1575–1585, 1996.
- J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, 1982.
- F. Hsieh and B. W. Turnbull. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The Annals of Statistics*, 24(1):25–40, 1996.

- W. J. Krzanowski and D. J. Hand. *ROC Curves for Continuous Data*. Chapman & Hall, 1 edition, 2009.
- T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado. The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, 38(5):404 – 415, 2005.
- P. A. W. Lewis. Distribution of the anderson-darling statistic. *The Annals of Mathematical Statistics*, 32(4):pp. 1118–1124, 1961.
- A. Mattuck. *Introduction to Analysis*. Prentice Hall, 1999.
- C. E. Metz, B. A. Herman, and J. H. Shen. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine*, 17(9): 1033–1053, May 1998.
- P. W. Millar. A general approach to the optimality of minimum distance estimators. *Transactions of the American Mathematical Society*, 286(1):377–418, Nov. 1984.
- D. S. Moore. *Goodness-of-fit Techniques*, chapter 3. CRC Press, New York, 1986.
- J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:pp. 289–337, 1933.
- K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 50:157–172, 1900.
- M. Pepe and T. Cai. The analysis of placement values for evaluating discriminatory measures. *Biometrics*, 60(2):528–535, 2004.
- M. Pepe, G. Longton, and H. Janes. Estimation and comparison of receiver operating characteristic curves. *The Stata journal*, 9(1):1–1, Mar. 2009.
- M. S. Pepe. A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika*, 84(3):595–608, 1997.
- M. S. Pepe. Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. *Biometrics*, 54:124–135, Mar. 1998.
- M. S. Pepe. An interpretation for the ROC curve and inference using GLM procedures. *Biometrics*, 56(2):352–359, 2000.
- M. S. Pepe. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, 1 edition, 2004.
- M. S. Pepe and G. Longton. Standardizing diagnostic markers to evaluate and compare their performance. *Epidemiology*, 16:598–603, Sep 2005.
- W. Peterson, T. Birdsall, and W. Fox. The theory of signal detectability. *Information Theory, IRE Professional Group on*, 4(4):171 –212, september 1954.

- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.
- J. Shao. *Mathematical Statistics*. Springer-Verlag, New York, 2 edition, 2003.
- T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer. *ROCR: Visualizing the performance of scoring classifiers.*, 2009. R package version 1.0-4.
- N. V. Smirnov. Sur la distribution de w^2 . *C. R. Acad. Sci. Paris*, 202:449–452, 1936.
- N. V. Smirnov. On the distribution of the w^2 criterion of von mises. *Rec. Math.*, 2:973–993, 1937.
- K. Soetaert, T. Petzoldt, and R. W. Setzer. Solving differential equations in R: Package desolve. *Journal of Statistical Software*, 33(9):1–25, 2010.
- M. A. Stephens. *Goodness-of-fit Techniques*, chapter 4. CRC Press, New York, 1986.
- L. Stover, M. P. Gorga, S. T. Neely, and D. Montoya. Toward optimizing the clinical utility of distortion product otoacoustic emission measurements. *J. Acoust. Soc. Am.*, 100:956–967, Aug. 1996.
- J. Swets and R. Pickett. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press series in cognition and perception. Academic Press, New York, 1982.
- J. A. Swets. The relative operating characteristic in psychology. *Science*, 182(4116):990–1000, 1973.
- J. A. Swets. Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychological Bulletin*, 99(1):100–117, Jan. 1986.
- J. A. Swets, R. M. Dawes, and J. Monahan. Better decisions through science. *Scientific American*, 283(4):82–87, 2000.
- L. Tang and X. Zhou. Semiparametric least squares analysis of clustered roc curve data. *Journal of Computational and Graphical Statistics*, 2011. Accepted.
- O. Thas. *Comparing Distributions*. Springer, Oct. 2009.
- L. L. Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273 – 286, 1927a.
- L. L. Thurstone. Psychophysical analysis. *The American Journal of Psychology*, 38(3):pp. 368–389, 1927b.
- S. Urbanek. *multicore: Parallel processing of R code on machines with multiple cores or CPUs*, 2011. R package version 0.1-7.
- D. van Meter and D. Middleton. Modern statistical approaches to reception in communication theory. *Information Theory, IRE Professional Group on*, 4(4):119 –145, september 1954.
- R. von Mises. *Wahrscheinlichkeitsrechnung*. Wein, Leipzig, 1931.
- A. Wald. *Sequential Analysis*. Wiley, New York, 1947.

- A. Wald. *Statistical Decision Functions*. Wiley, New York, 1950.
- L. Wasserman. *All of Nonparametric Statistics (Springer Texts in Statistics)*. Springer, New York, 2005.
- H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.
- X. Zhou, D. K. McClish, and N. A. Obuchowski. *Statistical Methods in Diagnostic Medicine*. Wiley-Interscience, New York, 1 edition, July 2002.
- K. H. Zou, J. L. Gastwirth, and B. J. McNeil. A goodness-of-fit test for a receiver operating characteristic curve from continuous diagnostic test data. In J. E. Kolassa and D. Oakes, editors, *Lecture Notes Monograph Series*, volume 43, page 59–68. Institute of Mathematical Statistics, 2003.
- K. H. Zou and W. J. Hall. Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics*, 27(5):621–631, 2000.
- K. H. Zou, A. J. O’Malley, and L. Mauri. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115(5):654–657, 2007.
- K. H. Zou, F. S. Resnic, I. Talos, D. Goldberg-Zimring, J. G. Bhagwat, S. J. Haker, R. Kikinis, F. A. Jolesz, and L. Ohno-Machado. A global goodness-of-fit test for receiver operating characteristic curve analysis via the bootstrap method. *Journal of Biomedical Informatics*, 38(5):395–403, 2005.

Curriculum Vitae

Brad earned his B.S. degree in physics from Stanford University in 2000. He received his Master's degree in physics from the University of California, Santa Barbara, in 2004. From 2005 to 2007, he worked as a development editor in the physics group at the publisher Pearson Addison-Wesley. He then joined Archimedes, Inc., a healthcare modeling company in 2008. His publications and presentations include the following:

1. B. Patterson, J. Miller, and C. Saunders (2011). "ROC Curves for Methods of Evaluating Evidence: Common Performance Measures Based on Similarity Scores." Presented at National Institute of Justice Trace Evidence Symposium, Kansas City, MO.
2. D. Eddy, J. Adler, B. Patterson, D. Lucas, and M. Morris (2011). "Individualized guidelines: The potential for increasing quality and reducing costs." *Annals of Internal Medicine* 154(9), 627-634.
3. A. Affolder et al. (2004). "Test of CMS tracker silicon detector modules with the ARC system." *Nuclear Instruments and Methods A* 535, 374-378.

He is the recipient of the following awards, fellowships, and scholarships:

- GMU Volgenau School of Engineering Outstanding Graduate Student, 2012.
- Distinguished Academic Achievement, Ph.D. Program in Statistical Science, GMU, 2012.
- National Institute of Justice Graduate Fellowship, 2011-2012.
- IMC, Inc. Suneeth Nayak Graduate Scholarship, 2011.
- Outstanding Graduate Student, Department of Statistics, GMU, 2010.