Appendix A

For the purpose of completeness, in this section, we present the Kendall correlation values for all experimental results. Overall, we observe a similar pattern between the Spearman and Kendall correlation analyses, with both showing either strong or weak correlations in the same instances.

Table 8 reports the Kendall correlations for Independence, Separation, and Sufficiency across the three datasets for Probabilistic Classification-based fairness estimation methods with various probabilistic classifier cores. Overall, the findings are similar to those suggested by Spearman; i.e., some methods such as Logistic Regression, Ridge, and Lasso, show consistently high correlations, whereas kernel-based approaches often display low consistencies. For example, the correlations between Ridge and KLR-Polynomial for the Independence metric are (0.25, -0.08, -0.01) across the three datasets.

The Kendall results for studying the sensitivity of fairness measurement methods when utilizing density ratio estimation approaches beside the proposed Probabilistic Classification-based ones are presented in Table 9. Overall, we observe a high degree of inconsistency across different measurement approaches. For instance, the correlation between Logistic Regression and LSIF for measuring the Sufficiency metric are (-0.32*, -0.38*, -0.04) across the three datasets. Moreover, the level of inconsistency between the same pair of measurement methods varies across datasets. For example, the correlation between uLSIF α = 0.25 and LSIF when measuring the Independence metric is (0.65*, 0.24, 0.06).

We also conduct a Kendall correlation analysis on the generated synthetic datasets, aggregating results for every 10 datasets within specific mean intervals, resulting in four distinct intervals. Table 10 shows the Kendall correlation between different measurement methods approximating the Independence fairness metric at different mean intervals. The Kendall analysis shows a similar pattern to the Spearman analysis, where an increase in the mean value corresponded to a decrease in correlation. For example, the correlation between Lasso and KLR_Polynomial drops from 0.82 to 0.02 between the first and last mean intervals. A similar trend is observed when analyzing the Separation fairness metric. Table 11 presents the Kendall correlation values for Separation metric. It shows that correlations tend to decrease as the mean value increases. For instance, the correlation between KLR_Polynomial and KLR_Gaussian drops from 0.73 to -0.24 between the first and last mean intervals.

Moreover, we computed the Kendall correlation analysis between the predicted fairness metric and the actual fairness values of the synthetic datasets. Table 12 presents the Kendall correlations between the predicted Independence values produced by Probabilistic Classification-based approaches and the actual Independence values across different mean intervals. Overall, inconsistencies appear to increase with higher mean values. For example, the average correlation decreases from 0.87 to 0.57 between the first and last mean intervals.

Furthermore, we study the impact of applying a threshold on the consistency of fairness measurement methods. Tables 13 and 14 show the Kendall correlation values for the final mean interval before and after thresholding, for the Independence and Separation fairness metrics, respectively. The overall trend indicates that thresholding helps improve consistency across different fairness measurement methods on the synthetic datasets. For instance, the correlation between Ridge and KLR_Gaussian increased from 0.33 to 0.78 for the Independence metric, and from 0.29 to 0.73 for the Separation metric.

In conclusion, the Kendall correlation analysis highlights the sensitivity of density ratio estimation approaches both in terms of the underlying classifiers used in Probabilistic Classification–based methods and the specific density ratio techniques employed. These findings align with and reinforce the conclusions drawn from the Spearman correlation analysis.

Corr. Metric		Logistic Regression	Ridge	Lasso	KLR_Gaussian	KLR_Polynomial
Independence	Logistic Regression	-	(1.00*, 1.00*, 1.00*)	(1.00*, 0.90*, 0.91*)	(0.73*, 0.06, -0.39*)	(0.11, -0.29, 0.07)
	Ridge	(1.00*, 1.00*, 1.00*)	-	(1.00*, 0.93*, 0.89*)	(0.76*, 0.07, -0.31)	(0.25, -0.08, -0.01)
	Lasso	(1.00*, 0.90*, 0.91*)	(1.00*, 0.93*, 0.89*)	-	(0.77*, 0.02, -0.29)	(0.26, -0.1, -0.01)
	KLR_Gaussian	(0.73*, 0.06, -0.39*)	(0.76*, 0.07, -0.31)	(0.77*, 0.02, -0.29)	-	(0.21, 0.45*, -0.47*)
	KLR_Polynomial	(0.11, -0.29, 0.07)	(0.25, -0.08, -0.01)	(0.26, -0.1, -0.01)	(0.21, 0.45*, -0.47*	-
Separation	Logistic Regression	-	(1.00*, 1.00*, 1.00*)	(1.00*, 0.93*, 0.89*)	(0.76*, 0.07, -0.31)	(0.25, -0.08, -0.01)
	Ridge	(1.00*, 1.00*, 1.00*)	-	(1.00*, 0.93*, 0.89*)	(0.76*, 0.07, -0.31)	(0.25, -0.08, -0.01)
	Lasso	(1.00*, 0.93*, 0.89*)	(1.00*, 0.93*, 0.89*)	-	(0.77*, 0.02, -0.29)	(0.26, -0.10, -0.01)
	KLR_Gaussian	(0.76*, 0.07, -0.31)	(0.76*, 0.07, -0.31)	(0.77*, 0.02, -0.29)	-	(0.21, 0.45*, -0.47*)
	KLR_Polynomial	(0.25, -0.08, -0.01)	(0.25, -0.08, -0.01)	(0.26, -0.10, -0.01)	(0.21, 0.45*, -0.47*)	-
Sufficiency	Logistic Regression	-	(1.00*, 1.00*, 1.00*)	(0.98*, 1.00*, 0.87*)	(0.48*, 0.49*, 0.08)	(-0.28, -0.14, -0.32)
	Ridge	(1.00*, 1.00*, 1.00*)	-	(0.98*, 1.00*, 0.87*)	(0.48*, 0.49*, 0.08)	(-0.28, -0.14, -0.32)
	Lasso	(0.98*, 1.00*, 0.87*)	(0.98*, 1.00*, 0.87*)	-	(0.47*, 0.49*, 0.02)	(-0.28, -0.14, -0.28)
	KLR_Gaussian	(0.48*, 0.49*, 0.08)	(0.48*, 0.49*, 0.08)	(0.47*, 0.49*, 0.02)	-	(-0.59*, -0.32*, -0.64*)
	KLR_Polynomial	(-0.28, -0.14, -0.32)	(-0.28, -0.14, -0.32)	(-0.28, -0.14, -0.28)	(-0.59*, -0.32*, -0.64*)	-

Table 8: Kendall correlations between the output of pairs of fairness measurement methods with various underlying probabilistic classifiers when measuring Independence, Separation and Sufficiency (shown in each column/row). Results are presented across Law School, Communities and Crime, and Insurance datasets, respectively.

Corr. Metric	Fairness Measurement	Logistic Regression	KLR_Polynomial	LSIF	uLSIF $\alpha = 0.25$	uLSIF $\alpha = 0.5$	uLSIF $\alpha = 0.75$	uLSIF $\alpha = 1$
Independence	Logistic Regression	-	(0.03, -0.46*, 0.12)	(0.47*, 0.22, 0.04)	(0.60*, 0.13, -0.12)	(0.38*, 0.03, -0.32)	(0.21, 0.27, 0.27)	(-0.34*, 0.27, -0.01)
	KLR_Polynomial	(0.03, -0.46*, 0.12)	-	(0.04, -0.22, 0.35*)	(0.07, -0.3, 0.2)	(0.24, -0.04, 0.22)	(0.24, -0.27, 0.3)	(0.05, -0.27, 0.42*)
	LSIF	(0.47*, 0.22, 0.04)	(0.04, -0.22, 0.35*)	-	(0.65*, 0.24, 0.06)	(0.33*, -0.01, -0.03)	(0.17, 0.35, 0.04)	(-0.24, 0.35, 0.06)
	uLSIF $\alpha = 0.25$	(0.60*, 0.13, -0.12)	(0.07, -0.3, 0.2)	(0.65*, 0.24, 0.06)	-	(0.55*, 0.27, 0.4*)	(0.19, 0.72*, 0.3)	(-0.15, 0.72*, 0.24)
	uLSIF $\alpha = 0.5$	(0.38*, 0.03, -0.32)	(0.24, -0.04, 0.22)	(0.33*, -0.01, -0.03)	(0.55*, 0.27, 0.4*)	-	(0.32*, 0.47*, 0.36)	(-0.0, 0.47*, 0.39)
	uLSIF $\alpha = 0.75$	(0.21, 0.27, 0.27)	(0.24, -0.27, 0.3)	(0.17, 0.35, 0.04)	(0.19, 0.72*, 0.3)	(0.32*, 0.47*, 0.36)	-	(-0.11, 1.0*, 0.72*)
	uLSIF $\alpha = 1$	(-0.34*, 0.27, -0.01)	(0.05, -0.27, 0.42*)	(-0.24, 0.35, 0.06)	(-0.15, 0.72*, 0.24)	(-0.0, 0.47*, 0.39)	(-0.11, 1.0*, 0.72*)	-
Separation	Logistic Regression	-	(0.19, 0.54*, -0.01)	(0.67*, nan, 0.26)	(0.49*, nan, -0.05)	(0.28, nan, -0.26)	(0.26, nan, 0.27)	(0.03, nan, nan)
	KLR_Polynomial	(0.19, 0.54*, -0.01)	-	(0.04, nan, 0.17)	(0.4*, nan, 0.36*)	(0.35*, nan, 0.02)	(0.34*, nan, -0.27)	(0.07, nan, nan)
	LSIF	(0.67*, nan, 0.26)	(0.04, nan, 0.17)	-	(0.28, nan, 0.13)	(0.22, nan, 0.05)	(0.11, nan, 0.28)	(-0.17, nan, nan)
	uLSIF $\alpha = 0.25$	(0.49*, nan, -0.05)	(0.4*, nan, 0.36*)	(0.28, nan, 0.13)	-	(0.52*, nan, 0.22)	(0.34*, nan, 0.1)	(0.04, nan, nan)
	uLSIF $\alpha = 0.5$	(0.28, nan, -0.26)	(0.35*, nan, 0.02)	(0.22, nan, 0.05)	(0.52*, nan, 0.22)	-	(0.19, nan, -0.18)	(-0.04, nan, nan)
	uLSIF $\alpha = 0.75$	(0.26, nan, 0.27)	(0.34*, nan, -0.27)	(0.11, nan, 0.28)	(0.34*, nan, 0.1)	(0.19, nan, -0.18)	-	(0.12, nan, nan)
	uLSIF $\alpha = 1$	(0.03, nan, nan)	(0.07, nan, nan)	(-0.17, nan, nan)	(0.04, nan, nan)	(-0.04, nan, nan)	(0.12, nan, nan)	-
Sufficiency	Logistic Regression	-	(-0.26, -0.42*, -0.19)	(-0.32*, -0.38*, -0.04)	(-0.53*, -0.09, -0.43*)	(-0.43*, -0.19, -0.44*)	(-0.34*, -0.27, -0.17)	(0.01, -0.27, -0.22)
	KLR_Polynomial	(-0.26, -0.42*, -0.19)	-	(0.45*, 0.12, -0.13)	(0.32*, 0.15, 0.16)	(0.32*, 0.23, -0.02)	(0.1, 0.22, 0.05)	(0.31*, 0.22, -0.22)
	LSIF	(-0.32*, -0.38*, -0.04)	(0.45*, 0.12, -0.13)	-	(0.51*, 0.22, 0.05)	(0.38*, 0.15, -0.05)	(0.06, 0.35, -0.16)	(0.11, 0.35, -0.04)
	uLSIF $\alpha = 0.25$	(-0.53*, -0.09, -0.43*)	(0.32*, 0.15, 0.16)	(0.51*, 0.22, 0.05)	-	(0.54*, 0.26, 0.41*)	(0.36*, 0.72*, -0.24)	(-0.06, 0.72*, -0.29)
	uLSIF $\alpha = 0.5$	(-0.43*, -0.19, -0.44*)	(0.32*, 0.23, -0.02)	(0.38*, 0.15, -0.05)	(0.54*, 0.26, 0.41*)	-	(0.34*, 0.42*, 0.13)	(0.04, 0.42*, 0.2)
	uLSIF $\alpha = 0.75$	(-0.34*, -0.27, -0.17)	(0.1, 0.22, 0.05)	(0.06, 0.35, -0.16)	(0.36*, 0.72*, -0.24)	(0.34*, 0.42*, 0.13)	-	(0.02, 1.0*, 0.72*)
	uLSIF $\alpha = 1$	(0.01, -0.27, -0.22)	(0.31*, 0.22, -0.22)	(0.11, 0.35, -0.04)	(-0.06, 0.72*, -0.29)	(0.04, 0.42*, 0.2)	(0.02, 1.0*, 0.72*)	-

Table 9: Kendall correlations between the output of pairs of fairness measurement methods with various density ratio estimation approaches when measuring Independence, Separation and Sufficiency (shown in each column/row). Results are presented across Law School, Communities and Crime, and Insurance datasets, respectively.

Mean Interval	Fairness Measurement	Logistic Regression	Ridge	Lasso	KLR_Gaussian	KLR_Polynomial
0 - 0.9	Logistic Regression	-	1.00	1.00	0.73	0.82
	Ridge	1.00	-	1.00	0.73	0.82
	Lasso	1.00	1.00	-	0.73	0.82
	KLR_Gaussian	0.73	0.73	0.73	-	0.55
	KLR_Polynomial	0.82	0.82	0.82	0.55	-
1 - 1.9	Logistic Regression	-	1.00	1.00	0.91	0.73
	Ridge	1.00	-	1.00	0.91	0.73
	Lasso	1.00	1.00	-	0.91	0.73
	KLR_Gaussian	0.91	0.91	0.91	-	0.82
	KLR_Polynomial	0.73	0.73	0.73	0.82	-
2 - 2.9	Logistic Regression	-	0.96	1.00	0.78	0.38
	Ridge	0.96	-	0.96	0.73	0.33
	Lasso	1.00	0.96	-	0.78	0.38
	KLR_Gaussian	0.78	0.73	0.78	-	0.42
	KLR_Polynomial	0.38	0.33	0.38	0.42	-
3 - 3.9	Logistic Regression	-	0.78	0.96	0.56	-0.02
	Ridge	0.78	-	0.82	0.33	0.20
	Lasso	0.96	0.82	-	0.51	0.02
	KLR_Gaussian	0.56	0.33	0.51	-	-0.29
	KLR_Polynomial	-0.02	0.20	0.02	-0.29	-

Table 10: Kendall correlations between pairs of methods approximating the Independence fairness metric at different mean intervals.

Mean Interval	Fairness Measurement	Logistic Regression	Ridge	Lasso	KLR_Gaussian	KLR_Polynomial
0 - 0.9	Logistic Regression	-	0.99	0.9	0.72	0.81
	Ridge	0.99	-	1.00	0.73	0.82
	Lasso	0.99	1.00	-	0.73	0.82
	KLR_Gaussian	0.72	0.73	0.73	-	0.73
	KLR_Polynomial	0.81	0.82	0.82	0.73	-
1 - 1.9	Logistic Regression	-	1.00	1.00	0.96	0.78
	Ridge	1.00	-	1.00	0.96	0.78
	Lasso	1.00	1.00	-	0.96	0.78
	KLR_Gaussian	0.96	0.96	0.96	-	0.73
	KLR_Polynomial	0.78	0.78	0.78	0.73	-
2 - 2.9	Logistic Regression	-	0.96	1.00	0.87	0.33
	Ridge	0.96	-	0.96	0.82	0.38
	Lasso	1.00	0.96	-	0.87	0.33
	KLR_Gaussian	0.87	0.82	0.87	-	0.38
	KLR_Polynomial	0.33	0.38	0.33	0.38	-
3 - 3.9	Logistic Regression	-	0.78	0.96	0.51	0.07
	Ridge	0.78	-	0.82	0.29	0.29
	Lasso	0.96	0.82	-	0.47	0.11
	KLR_Gaussian	0.51	0.29	0.47	-	-0.24
	KLR_Polynomial	0.07	0.29	0.11	-0.24	-

Table 11: Kendall correlations between pairs of methods approximating the Separation fairness metric at different mean intervals.

Mean Interval	Logistic Regression	Ridge	Lasso	KLR_Gaussian	KLR_Polynomial
0-0.9	0.94	0.94	0.94	0.67	0.85
1-1.9	1.00	1.00	1.00	0.91	0.73
2-2.9	0.91	0.96	0.91	0.78	0.38
3-3.9	0.87	0.64	0.82	0.69	-0.16

Table 12: Kendall correlations between actual and estimated Independence metric (with various core probabilistic classifiers) across various mean intervals.

Corr. Metric		Logistic Regression	Ridge	Lasso	KLR_Gaussian	KLR_Polynomial
Kendall	Logistic Regression	-	0.78	0.96	0.56	-0.02
(before thresholding)	Ridge	0.78	-	0.82	0.33	0.20
	Lasso	0.96	0.82	-	0.51	0.02
	KLR_Gaussian	0.56	0.33	0.51	-	-0.29
	KLR_Polynomial	-0.02	0.20	0.02	-0.29	-
Kendall	Logistic Regression	-	1.00	1.00	0.78	0.91
(after thresholding)	Ridge	1.00	-	1.00	0.78	0.91
	Lasso	1.00	1.00	-	0.78	0.91
	KLR_Gaussian	0.78	0.78	0.78	-	0.87
	KLR_Polynomial	0.91	0.91	0.91	0.87	-

Table 13: Kendall correlations for the Independence fairness metric, computed using various measurement methods (relying on different classifiers), before (top) and after (bottom) thresholding their predicted probabilities.

Corr. Metric		Logistic Regression	Ridge	Lasso	KLR_Gaussian	KLR_Polynomial
Kendall	Logistic Regression	-	0.78	0.96	0.51	0.07
(before thresholding)	Ridge	0.78	-	0.82	0.29	0.29
	Lasso	0.96	0.82	-	0.47	0.11
	KLR_Gaussian	0.51	0.29	0.47	-	-0.24
	KLR_Polynomial	0.07	0.29	0.11	-0.24	-
Kendall	Logistic Regression	-	1.00	1.00	0.73	0.56
(after thresholding)	Ridge	1.00	-	1.00	0.73	0.56
	Lasso	1.00	1.00	-	0.73	0.56
	KLR_Gaussian	0.73	0.73	0.73	-	0.29
	KLR_Polynomial	0.56	0.56	0.56	0.29	-

Table 14: Kendall correlations for the Separation fairness metric, computed using various measurement methods (relying on different classifiers), before (top) and after (bottom) thresholding their predicted probabilities.