Towards Causal Discovery with Statistical Guarantees

Shreya Prakash
in collaboration with Fan Xia and Elena Erosheva

Talk Outline

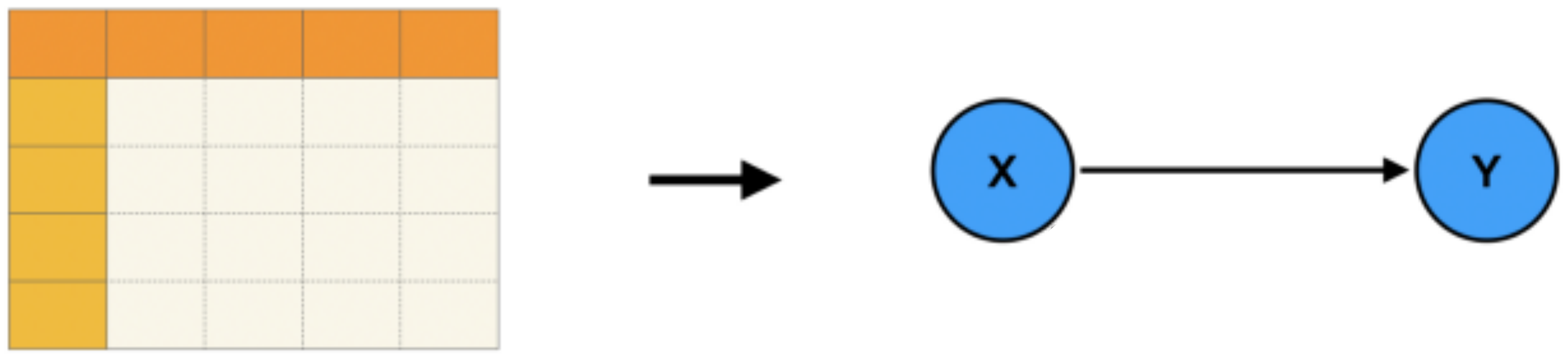
- Motivation
- Methods
- Simulations
- Real Data Results
- Conclusions

Talk Outline

- Motivation
- Methods
- Simulations
- Real Data Results
- Conclusions

Background

- Causal discovery methods aim to infer a causal directionality structure from the data
- In bivariate case, for example, can use observational data to understand whether sleep problems cause depression or vise versa (Rosenstrom et al. (2012))



Bivariate LiNGAM

- Shimizu et al. (2006) proposed a regression based causal discovery algorithm: Linear, Non-Gaussian, Acyclic causal Models (LiNGAM)
- Assumptions:
    1. Linearity
    2. Non-gaussian error terms
    3. Acyclicity
    4. No unobserved confounders

Bivariate LiNGAM

- In the bivariate case the goal is to decide between 2 possible linear causal models:

  1. $X \rightarrow Y$
  2. $Y \rightarrow X$

Bivariate LiNGAM

- In the bivariate case the goal is to decide between 2 possible linear causal models:

  1. $X \rightarrow Y$ $\quad (Y = \beta X + \eta_Y, X \perp \eta_Y)$

  2. $Y \rightarrow X$ $\quad (X = \rho Y + \eta_X, Y \perp \eta_X)$

Finite Sample Performance

- Shimizu et. al proved *identifiability* for LiNGAM but about LiNGAM’s *finite sample performance*
    - How does LiNGAM perform under **assumption violations**?
    - How does the **sample size** affect the discovery results?
- Currently this is not explored for LiNGAM and many other existing causal discovery algorithms

Talk Outline

- Motivation
- Methods
- Simulations
- Real Data Results
- Conclusions

Introducing the Test-based Method

- Introduce a framework to evaluate the finite sample performance of LiNGAM using hypothesis tests and a set of metrics related to statistical power

Introducing the Test-based Method

- Introduce a framework to evaluate the finite sample performance for LiNGAM using hypothesis tests and a set of metrics related to statistical power

- We re-purpose the goodness-of-fit and independence test introduced by Sen and Sen (2014) into a causal discovery algorithm using the same **modeling and distributional assumptions** as LiNGAM

Introducing the Test-based Method

- Introduce a framework to evaluate the finite sample performance for LiNGAM using hypothesis tests and a set of metrics related to statistical power

- We re-purpose the goodness-of-fit and independence test introduced by Sen and Sen (2014) into a causal discovery algorithm using the same **modeling and distributional assumptions** as LiNGAM

- The goodness-of-fit and independence test tests the following null hypothesis:

$$H_0 : X \perp \eta, \text{ relationship between } X \text{ and } Y \text{ is linear}$$

Introducing the Test-based Method

- Introduce a framework to evaluate the finite sample performance for LiNGAM using hypothesis tests and a set of metrics related to statistical power

- We re-purpose the goodness-of-fit and independence test introduced by Sen and Sen (2014) into a causal discovery algorithm using the same **modeling and distributional assumptions** as LiNGAM

- The goodness-of-fit and independence test tests the following null hypothesis:

$$H_0 : X \perp \eta, \text{ relationship between } X \text{ and } Y \text{ is linear}$$

- Which we re-purpose into a bivariate causal discovery algorithm:

$$H_1 = \begin{cases} H_Y^0 : X \to Y, H_Y^1 : Y \to X \\ H_X^0 : Y \to X, H_X^1 : X \to Y \end{cases} \Rightarrow \quad H_1 * = \begin{cases} H_Y^0 : X \perp \epsilon, H_Y^1 : X, \epsilon \text{ dependent} \\ H_X^0 : Y \perp \delta, H_X^1 : Y, \delta \text{ dependent} \end{cases}$$

Comparing Test-based method with LiNGAM

Test-based Method

- Tests independence as well as goodness of fit
- Sen & Sen test outputs a ***p-value***
- Compares p-values to a significance level
- More sensitivity to ***assumption violations***
- Impact of assumption violations are ***well-understood***

Both

- Same assumptions
- Estimate causal direction by running two regression models

LiNGAM

- Tests independence
- Only outputs "test statistic" (e.g mutual information)
- Compares test statistics of the two directions
- Less sensitive to assumption violations

Statistical Guarantees

P-values

- We estimate the **p-values** corresponding to the set of hypothesis tests

$$H_1* = \begin{cases} H_Y^0 : X \perp \epsilon, H_Y^1 : X, \epsilon \text{ dependent} \\ H_X^0 : Y \perp \delta, H_X^1 : Y, \delta \text{ dependent} \end{cases}$$

Statistical Guarantees

Power-related Metrics

- We introduce a set of metrics that are related to **power**, a relationship between sample size and our chances of determining the true causal direction
  - Allow us to assess how *sample size* and *assumption violations* affect the causal direction inferred in addition to added statistical guarantees that p-values give

Talk Outline

- Motivation
- Methods
- Simulations
- Real Data Results
- Conclusions

Simulation Setup

- Explore 3 levels of increasing linearity and Gaussianity
    - Evaluate how assumption violations affect the true direction detection rate
- Compare LiNGAM with the Hilbert Schmidt Independence Criteria as the independence measure with the Test-based Approach

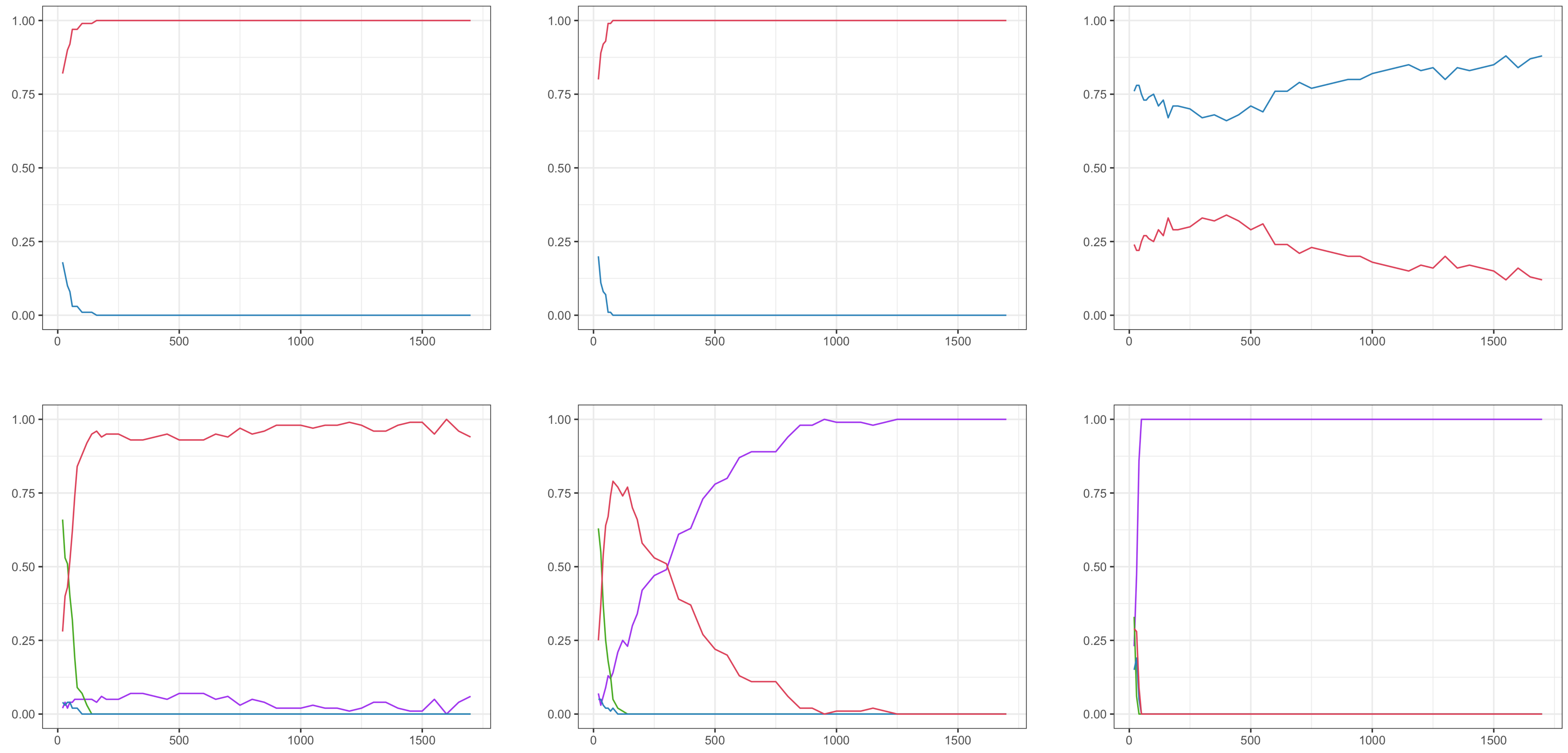
Simulation Setup

Key Takeaways

- Our LiNGAM simulations will show us the chance of choosing the (in)correct direction as a function of sample size
- Our Test-based approach simulations will show us the chance of choosing the (in)correct direction as a function of sample size as well as indicate if there are ***any assumption violations***

Linearity Simulation Results

Linearity Simulations



Polynomial = 1 | Polynomial = 1.5 | Polynomial = 5

Test Approach	Line Color	Description
Lingam with HSIC	red	% of choosing $X \to Y$
	blue	% of choosing $Y \to X$
Test-based approach	purple	% of rejecting in both directions
	green	% of failing to reject in both directions
	blue	% of reject only $X \to Y$
	red	% of reject only $Y \to X$

Gaussianity Simulation Results

Gaussianity Simulations

GMM with 3 mixtures | GMM with 2 mixtures | Gaussian

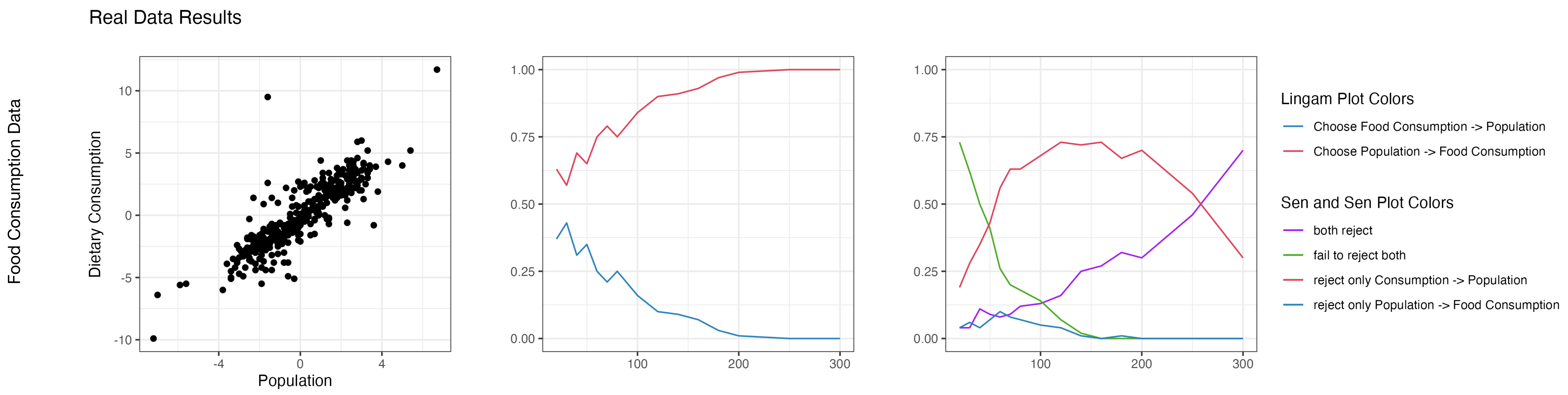
Test Approach	Line Color	Description
Lingam with HSIC	red	% of choosing $X \to Y$
	blue	% of choosing $Y \to X$
Test-based approach	purple	% of rejecting in both directions
	green	% of failing to reject in both directions
	blue	% of reject only $X \to Y$
	red	% of reject only $Y \to X$

Talk Outline

- Motivation
- Methods
- Simulations
- Real Data Results
- Conclusion

Real Data Results

- The Food Consumption Data measures the average annual rate of change of population and the average annual rate of change of the total dietary consumption for total population
  - Known causal direction is that population change causes change in total dietary consumption



Talk Outline

- Motivation
- Methods
- Simulations
- Real Data Results
- Conclusions

Conclusions

Findings

- The Test-based approach assesses when there are **assumption violations** as well as estimate the causal direction at the same time
- Able to assess the uncertainty of the causal direction estimate through power-like metrics and p-value

Conclusions

Next Steps

- Want to assess the finite sample performance for more complicated causal discovery models and extend our results to the multivariate case

Thank you! Questions?

Appendix

Simulation Setup

Table 1: Simulation Settings for Varying Linearity

Linear	Polynomial = 1	$Y = sign(X-a)\|X-a\| * \beta + \epsilon$
Slightly Nonlinear	Polynomial = 1.5	$Y = sign(X-a)\|X-a\|^{1.5} * \beta + \epsilon$
Nonlinear	Polynomial = 5	$Y = sign(X-a)\|X-a\|^{5} * \beta + \epsilon$

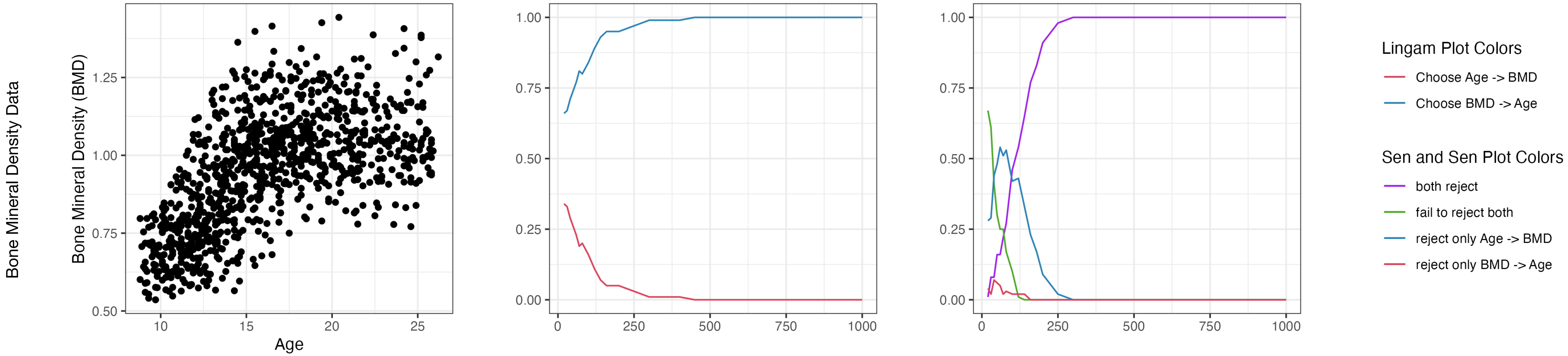
Table 2: Simulation Settings for Varying Levels of Gaussianity, k is number of mixtures

Gaussian	$X \sim N(0,1), \epsilon \sim N(\mu_1, \sigma_1)$
Slightly Non-Gaussian	$X \sim N(0,1), \epsilon \sim GMM(k=2)$
Non-Gaussian	$X \sim N(0,1), \epsilon \sim GMM(k=3)$

Test Approach	Line Color	Description	Interpretation
Lingam with HSIC	red	% of choosing $X \to Y$	The red and blue lines represent the chance of choosing the correct and incorrect direction for each sample size.
	blue	% of choosing $Y \to X$	
Test-based approach	purple	% of rejecting in both directions	Indication of linearity assumption violation.
	green	% of failing to reject in both directions	Indication of small sample size or Gaussianity assumption violation.
	blue	% of reject only $X \to Y$	Indication of favoring the incorrect direction.
	red	% of reject only $Y \to X$	Indication of favoring the correct direction.

Real Data: Bone Mineral Density Data Results

- Bone Mineral Density Data contains 1003 relative spinal bone mineral density measurements on 261 North American adolescents
  - Known causal direction is that age causes the spinal bone mineral density measurements for adolescents

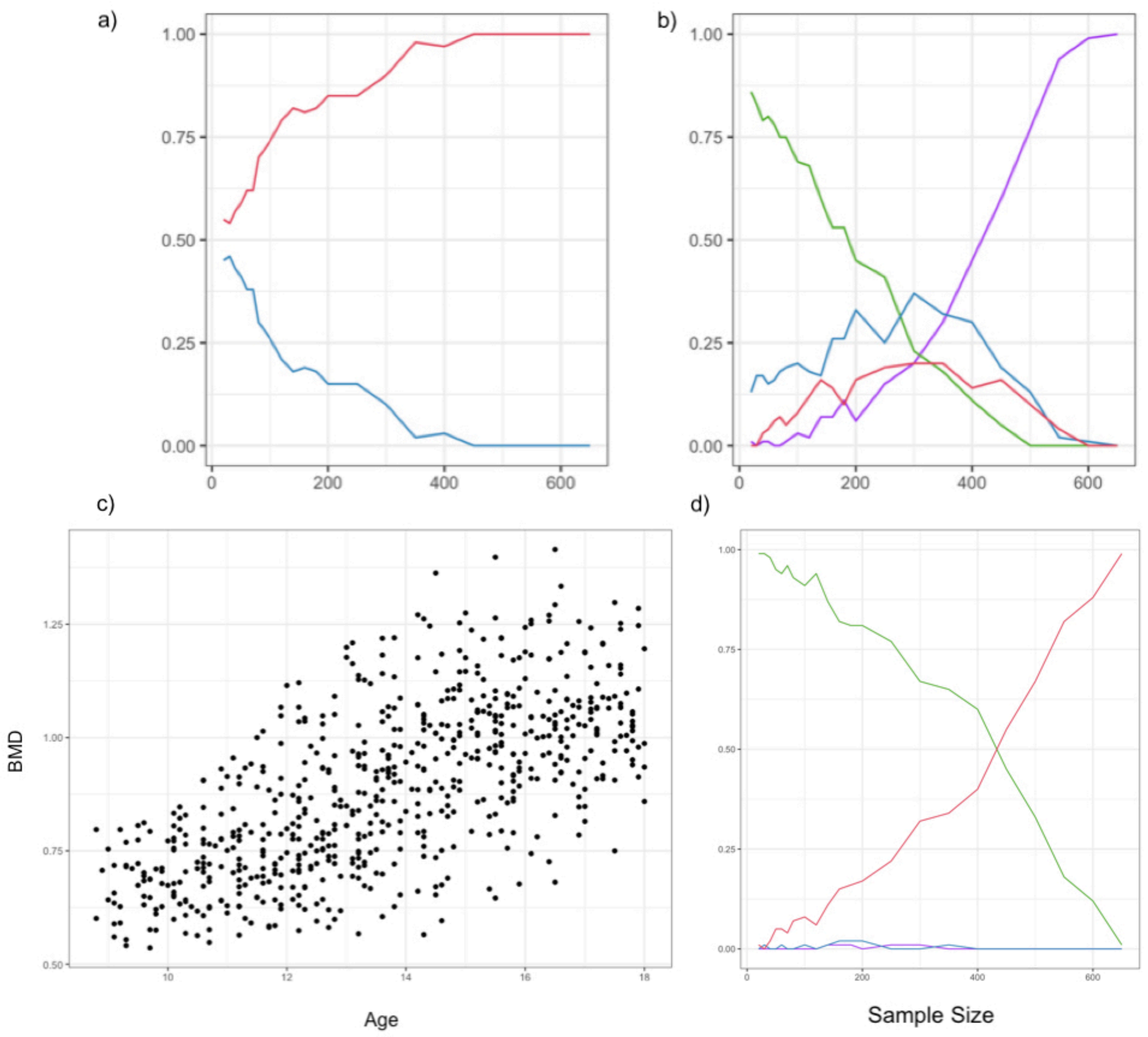


Scatterplot | LiNGAM Results | Test-based Results

Truncated Bone Mineral Density Data

Lingam Results for both transformed and untransformed data

Test-based Results for untransformed data



Scatterplot of truncated Bone Mineral Density data

Test-based Results for transformed data

- Linearity assumption is violated so using additive noise models to instead infer the causal direction
- Instead of checking for linearity, our method will be testing for the goodness-of-fit of the estimated non-linear models
- Instead of checking non-gaussianity, our method checks for non-identifiability
- Fit splines in both directions for Truncated BMD Data and able to detect the correct direction

Lingam Plot Colors

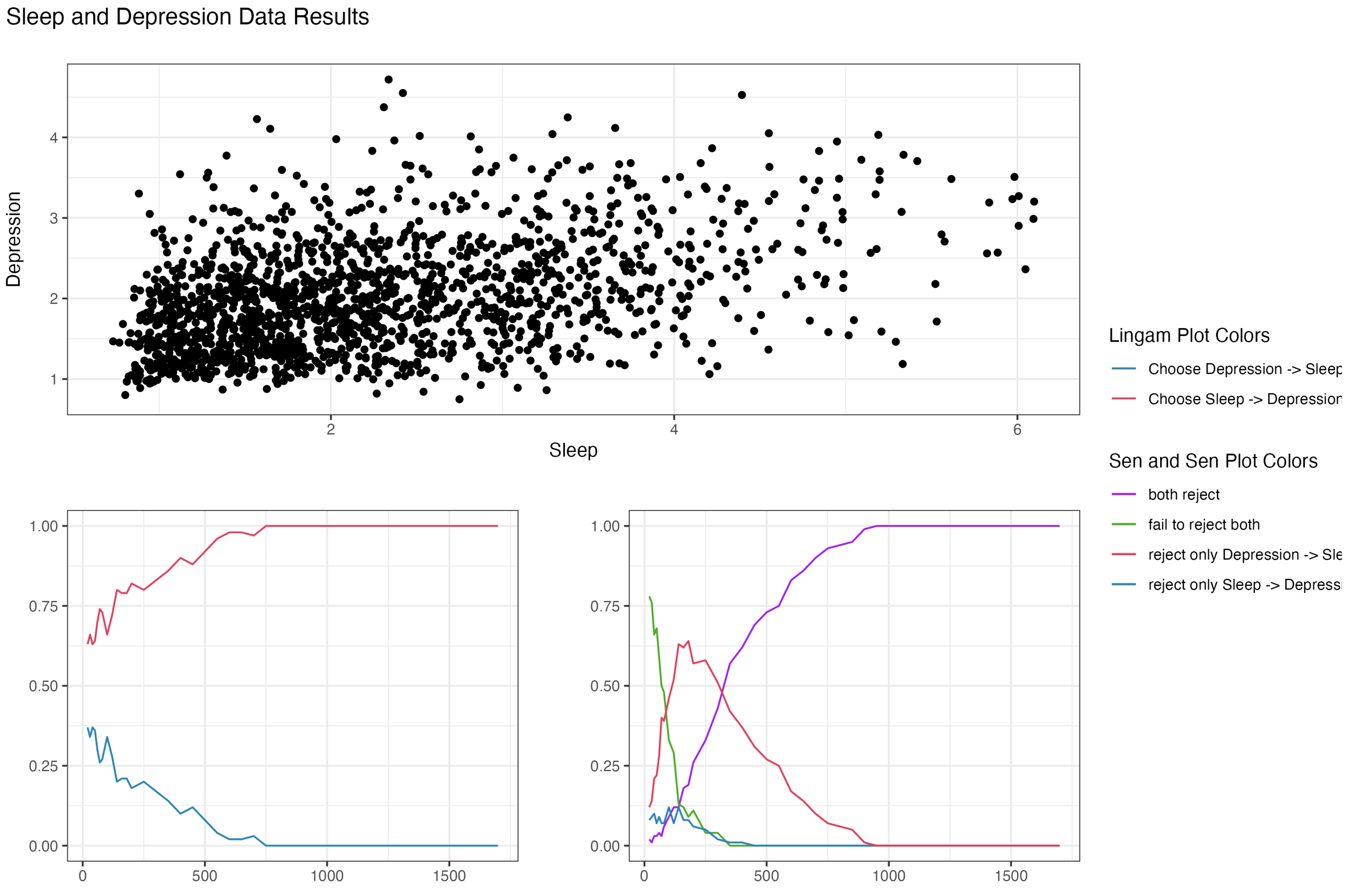
- Choose X->Y
- Choose Y->X

Sen and Sen Plot Colors

- both reject
- fail to reject both
- reject only X -> Y
- reject only Y -> X

Test Approach	Line Color	Description	Interpretation
Lingam with HSIC	red	% of choosing $X \rightarrow Y$	The red and blue lines represent the chance of choosing the correct and incorrect direction for each sample size.
	blue	% of choosing $Y \rightarrow X$	
Test-based approach	purple	% of rejecting in both directions	Indication of linearity assumption violation.
	green	% of failing to reject in both directions	Indication of small sample size or Gaussianity assumption violation.
	blue	% of reject only $X \rightarrow Y$	Indication of favoring the incorrect direction.
	red	% of reject only $Y \rightarrow X$	Indication of favoring the correct direction.

Sleep and Depression Data



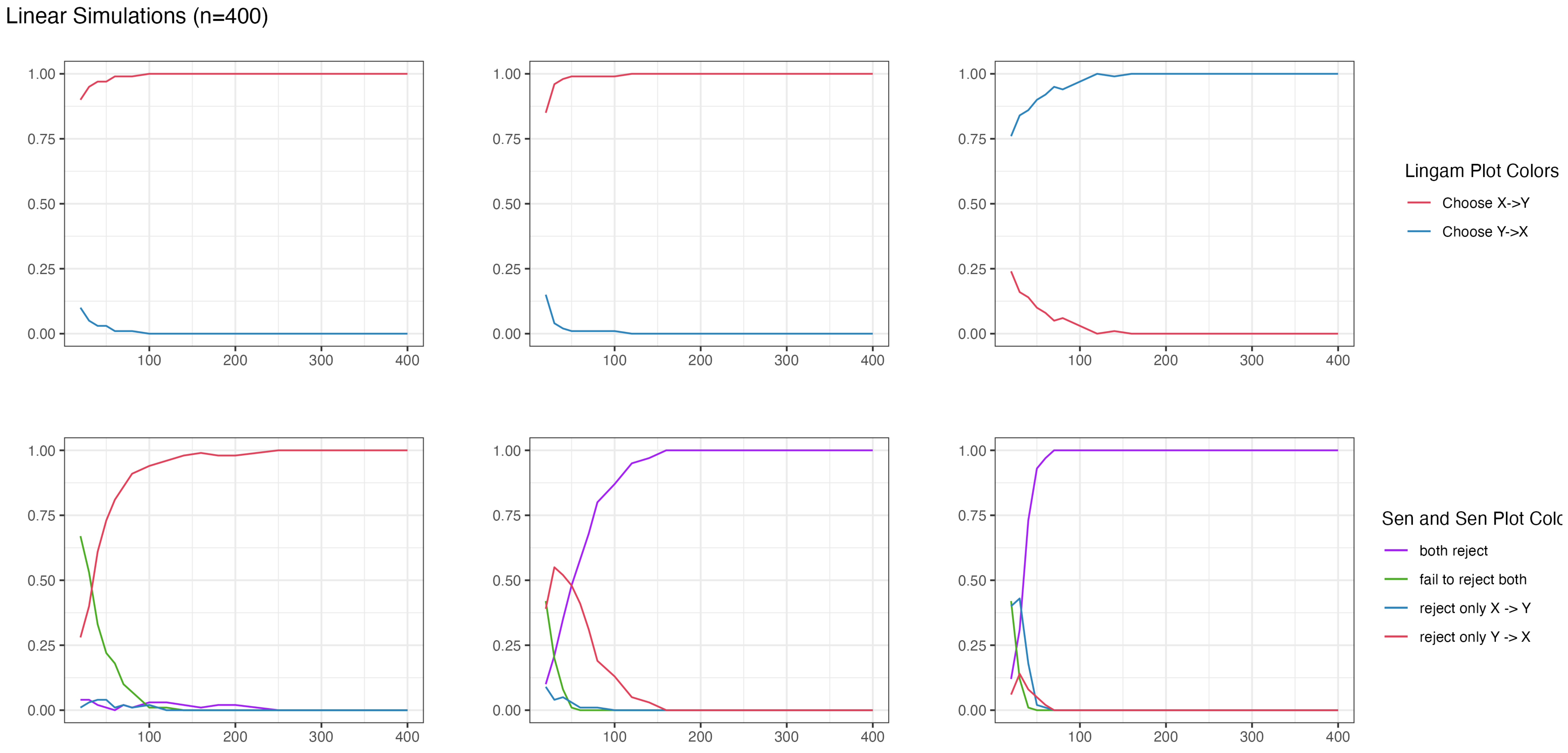
Lingam Results

Sen & Sen Results

Test Approach	Line Color	Description	Interpretation
Lingam with HSIC	red	% of choosing $X \to Y$	The red and blue lines represent the chance of choosing the correct and incorrect direction for each sample size.
	blue	% of choosing $Y \to X$	
Test-based approach	purple	% of rejecting in both directions	Indication of linearity assumption violation.
	green	% of failing to reject in both directions	Indication of small sample size or Gaussianity assumption violation.
	blue	% of reject only $X \to Y$	Indication of favoring the incorrect direction.
	red	% of reject only $Y \to X$	Indication of favoring the correct direction.

Linearity Simulation Results (n = 400)

Linearity Simulations



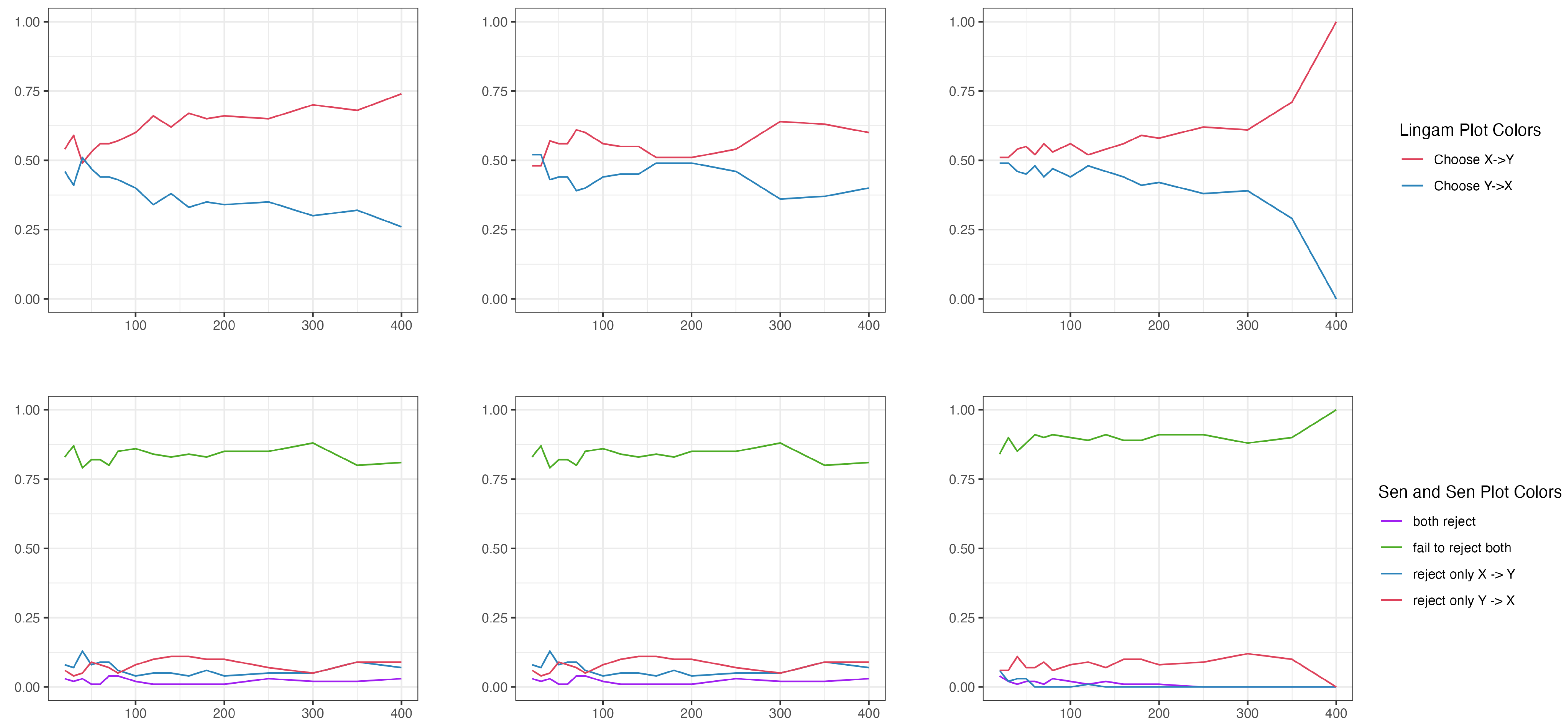
Polynomial = 1 Polynomial = 1.5 Polynomial = 5

Test Approach	Line Color	Description	Interpretation
Lingam with HSIC	red	% of choosing $X \to Y$	The red and blue lines represent the chance of choosing the correct and incorrect direction for each sample size.
	blue	% of choosing $Y \to X$	
Test-based approach	purple	% of rejecting in both directions	Indication of linearity assumption violation.
	green	% of failing to reject in both directions	Indication of small sample size or Gaussianity assumption violation.
	blue	% of reject only $X \to Y$	Indication of favoring the incorrect direction.
	red	% of reject only $Y \to X$	Indication of favoring the correct direction.

Gaussianity Simulation Results (n = 400)

Gaussianity Simulations

Gaussian Simulations (n=400)

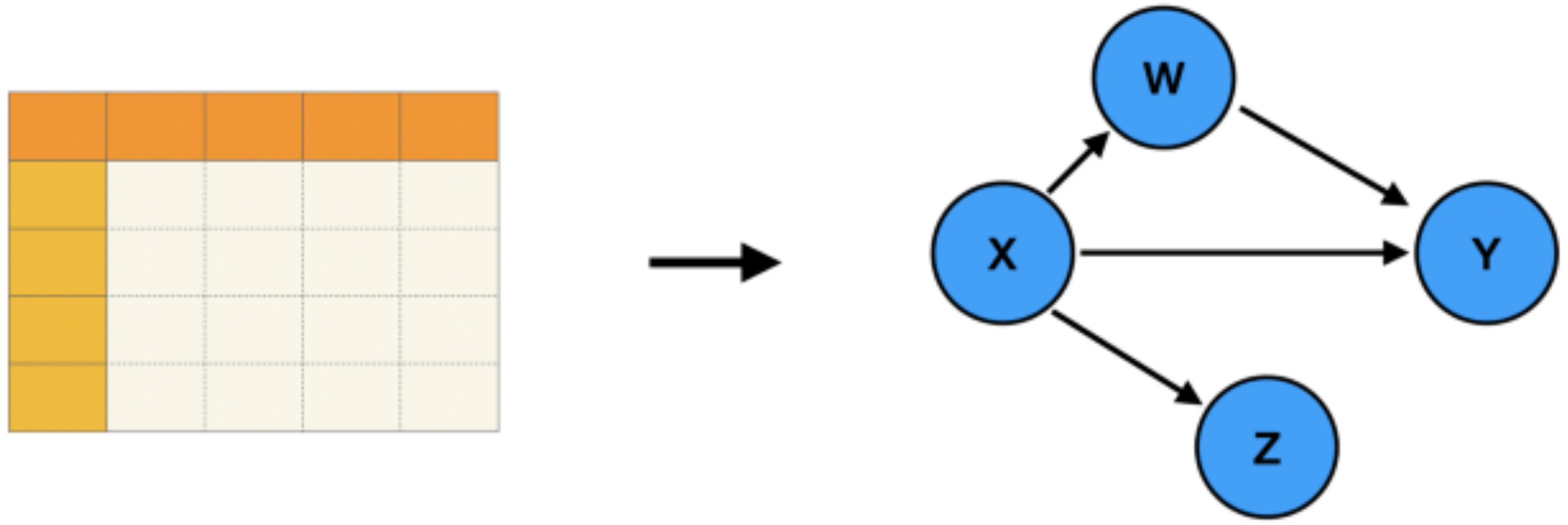


GMM with 3 mixtures | GMM with 2 mixtures | Gaussian

Test Approach	Line Color	Description	Interpretation
Lingam with HSIC	red	% of choosing $X \to Y$	The red and blue lines represent the chance of choosing the correct and incorrect direction for each sample size.
	blue	% of choosing $Y \to X$	
Test-based approach	purple	% of rejecting in both directions	Indication of linearity assumption violation.
	green	% of failing to reject in both directions	Indication of small sample size or Gaussianity assumption violation.
	blue	% of reject only $X \to Y$	Indication of favoring the incorrect direction.
	red	% of reject only $Y \to X$	Indication of favoring the correct direction.

Causal Discovery

- Causal discovery methods aim to infer a causal directionality structure from the data
- Generally two directions:
    1. Functional-based (e.g LiNGAM)
    2. Constraint-based (PC Algorithm)
- Interested in bivariate case, so we cannot use conditional independence based algorithms like the PC algorithm



Bivariate LiNGAM

- Shimizu et al. (2006) proposed the LiNGAM model
- Assumptions:
    1. Linearity
    2. Non-gaussian error terms
    3. Acyclicity
    4. No unobserved confounders

Bivariate LiNGAM

- In the bivariate case the goal is to decide between 2 possible linear causal models:

    1. $X \to Y$ $\quad$ $(Y = \beta X + \eta_Y, X \perp \eta_Y)$
    2. $Y \to X$ $\quad$ $(X = \rho Y + \eta_X, Y \perp \eta_X)$

- But LiNGAM only outputs the causal direction ***without any statistical guarantees***, have no idea if the output is right or wrong due to assumption violations

Sen and Sen Test

- Sen and Sen (2014) proposes a goodness of fit and independence test based on the Hilbert-Schmidt independence criterion (HSIC)
- Similarly to LiNGAM, this method makes the following assumptions
  1. Linearity
  2. Non-gaussian error terms
  3. Acyclicity
  4. No unobserved confounders

Sen and Sen Causal Discovery

- The Sen and Sen test tests the following null hypothesis:

$$H_0 : X \perp \eta, \text{relationship between } X \text{ and } Y \text{ is linear}$$

- In the bivariate case, interested in testing the following set of hypothesis:

$$H_1 = \begin{cases} H_Y^0 : X \to Y, H_Y^1 : Y \to X \\ H_X^0 : Y \to X, H_X^1 : X \to Y \end{cases}$$

Sen and Sen Causal Discovery

- With assumptions have that:

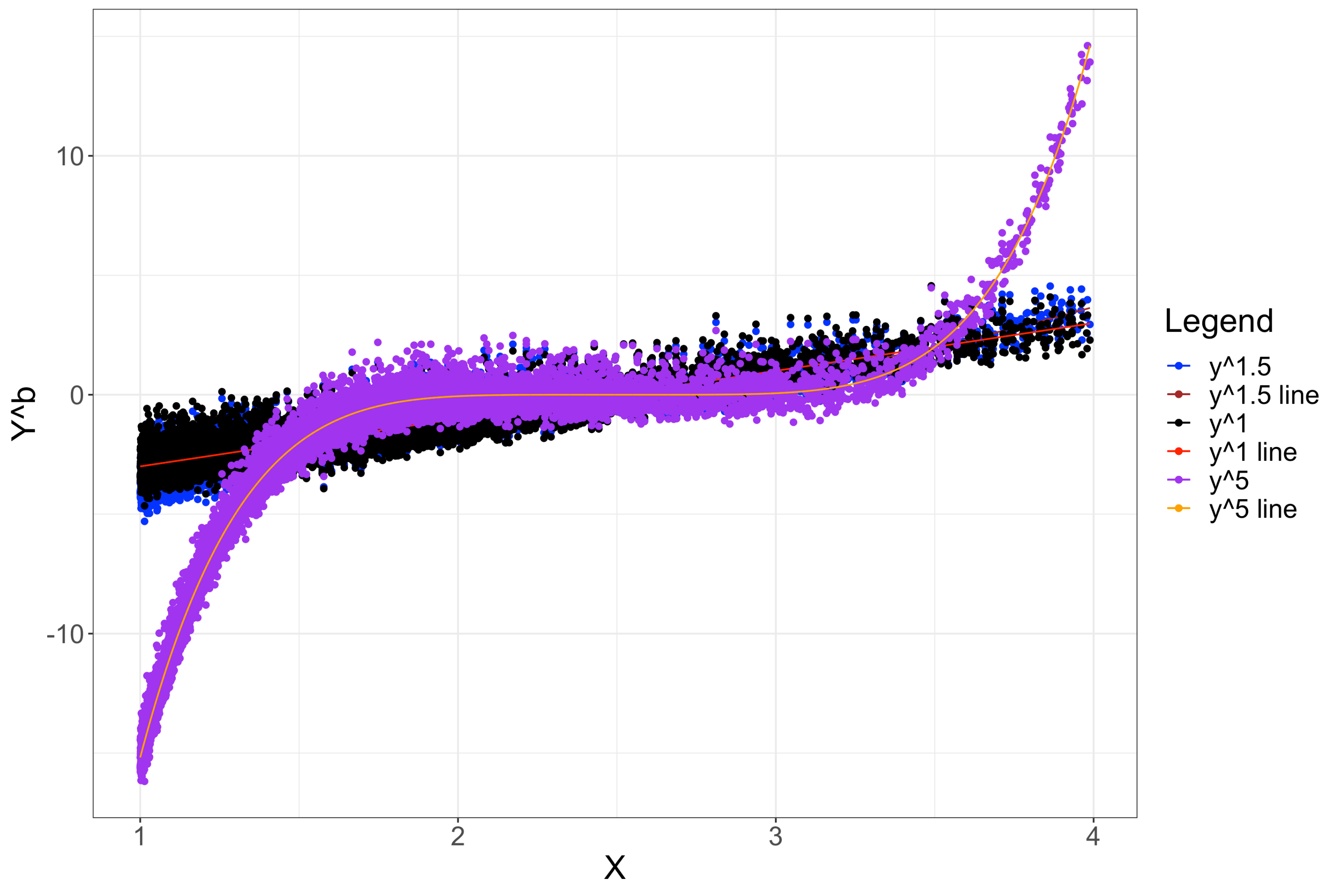
$$X \rightarrow Y \quad \Rightarrow \quad Y = X + \epsilon$$

$$Y \rightarrow X \quad \Rightarrow \quad X = Y + \delta$$

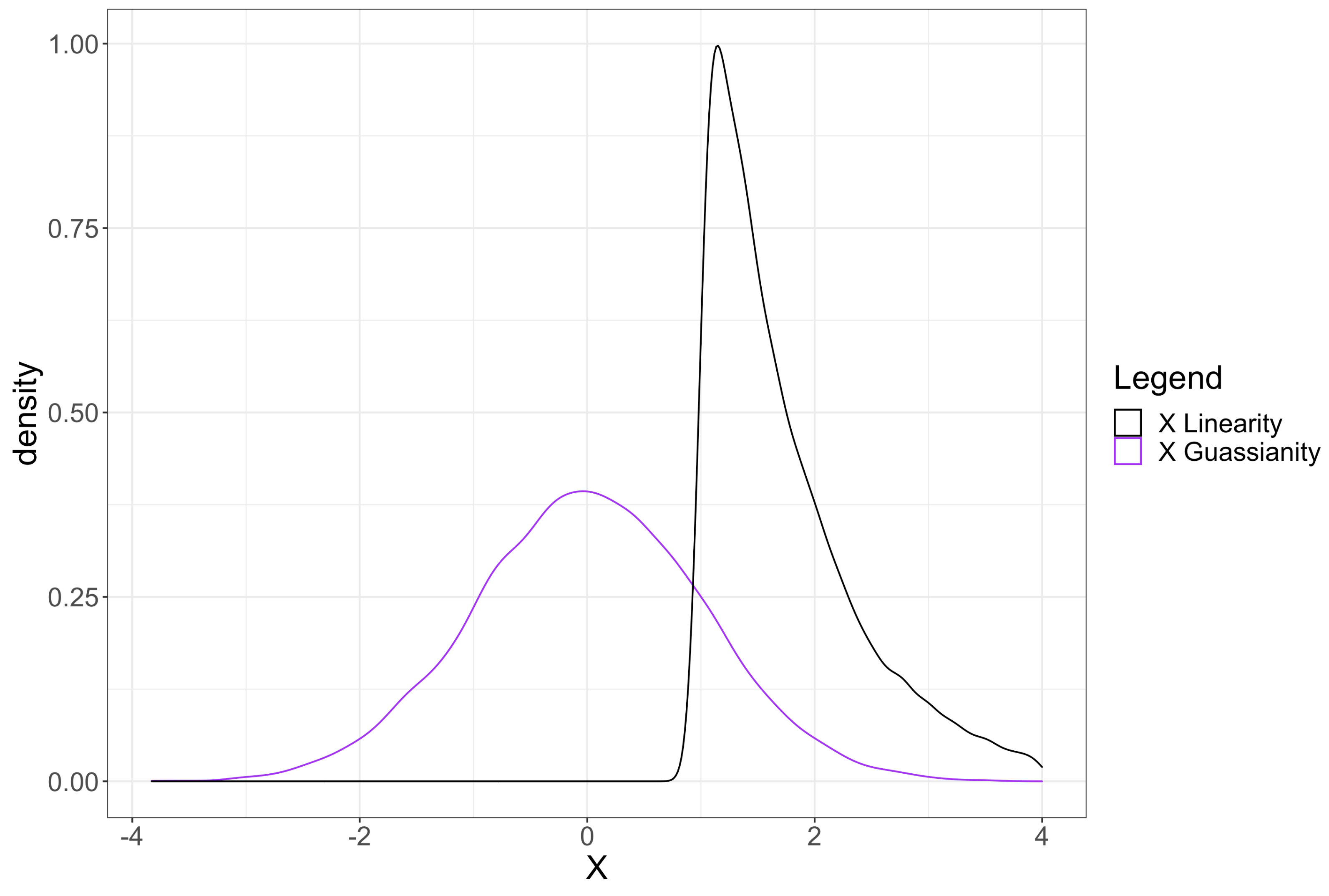
- So can translate H_1 to

$$H_1 * = \begin{cases} H_Y^0 : X \perp \epsilon, H_Y^1 : X, \epsilon \text{ dependent} \\ H_X^0 : Y \perp \delta, H_X^1 : Y, \delta \text{ dependent} \end{cases}$$

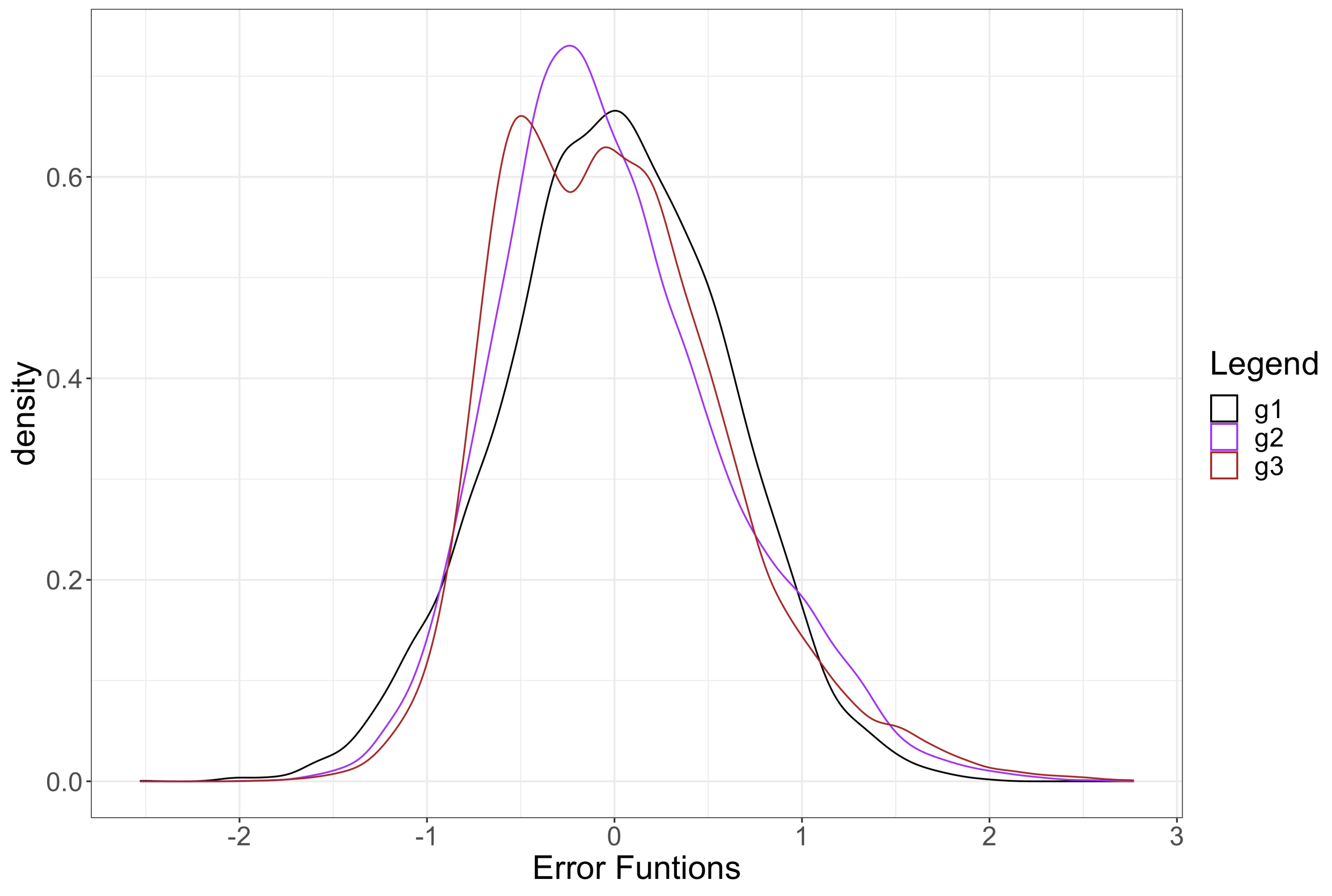
Simulation Setup



Settings of Linearity



Simulation of X Distribution



Settings of Gaussianity Errors

Discussion

- Still have room to explore how algorithms behave when there is a weak signal between X and Y
- If there is both a weak signal as well as assumption violations, we hypothesize that there might not be enough "delay in detection of assumption violation" for the Sen and Sen algorithm to determine the correct causal direction
- Can see an example with the truncated version of the Bone Mineral Density data

What is Power Analysis

- $Power = P(reject\ H_0 | H_1\ true)$
- Statistical power is one piece of a puzzle of 3 other related parts:
    1. Effect size (es): size of magnitude of a result present in the population

What is Power Analysis

- $Power = P(reject\ H_0 | H_1\ true)$
- Statistical power is one piece of a puzzle of 3 other related parts:
    1. Effect size (es): size of magnitude of a result present in the population
    2. Sample Size

What is Power Analysis

- $Power = P(reject\ H_0 | H_1\ true)$
- Statistical power is one piece of a puzzle of 3 other related parts:

  1. Effect size (es): size of magnitude of a result present in the population
  2. Sample Size
  3. Significance

What is Power Analysis

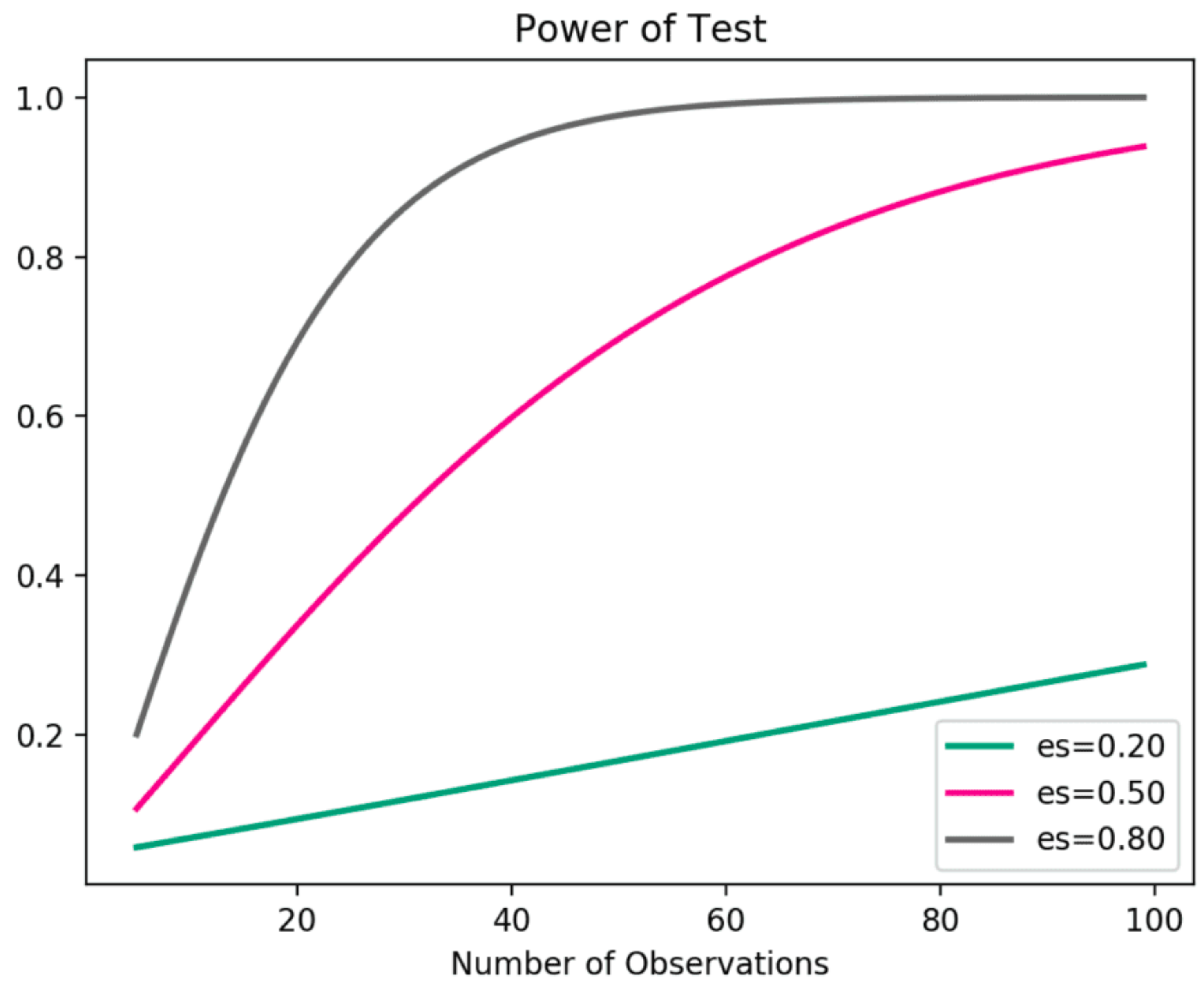
- $Power = P(reject\ H_0 \mid H_1\ true)$

- Statistical power is one piece of a puzzle of 3 other related parts:

  1. Effect size (es): size of magnitude of a result present in the population

  2. Sample Size

  3. Significance



Causal Discovery and Power

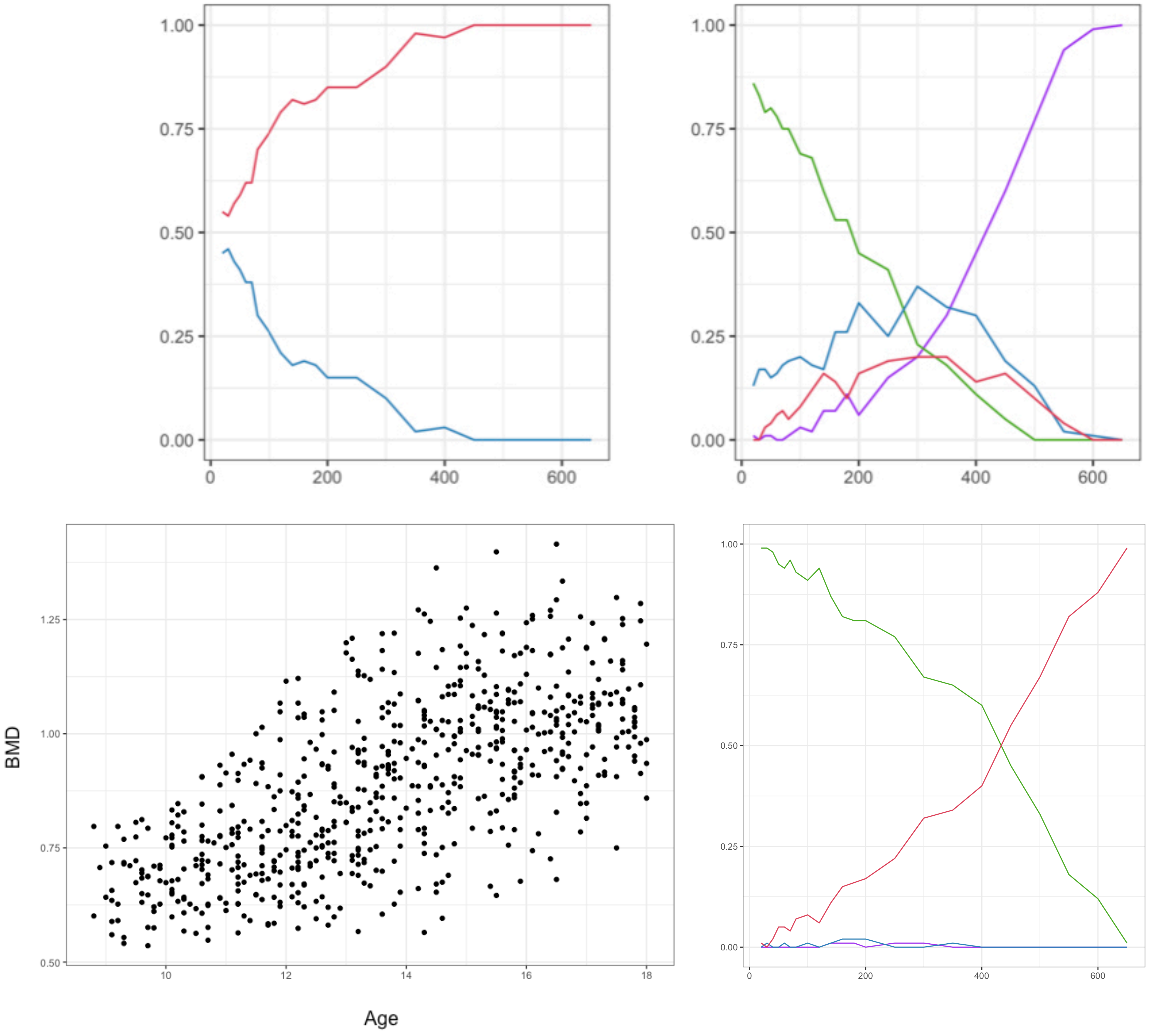
- In causal discovery, power translates to the probability of correctly identifying the causal direction

$$P(\text{reject } (X \rightarrow Y) \,|\, Y \rightarrow X) \text{ and } P(\text{reject } (Y \rightarrow X) \,|\, X \rightarrow Y)$$

- Example: if lack of sleep does in fact cause depression, low power would mean we would not be able to determine the causal direction between sleep and depression

- Will use power analysis to understand if sample size would affect the results of causal discovery and how this differs across methods

- Specifically will analyze power under linearity assumption violations for both LiNGAM and Sen and Sen



Lingam Plot Colors

- Choose X->Y
- Choose Y->X

Sen and Sen Plot Colors

- both reject
- fail to reject both
- reject only X -> Y
- reject only Y -> X