TAKING UNCERTAINTY SERIOUSLY: BAYESIAN MARGINAL STRUCTURAL MODELS FOR CAUSAL INFERENCE

A. Jordan Nafa and Andrew Heiss

University of North Texas and Georgia State University

September 16th, 2022

INTRODUCTION

INTRODUCTION

- In many fields of political science baseline random assignment and experimental design are impossible

INTRODUCTION

- In many fields of political science baseline random assignment and experimental design are impossible
  - For many research questions, observational data is our only option but this makes causal inference difficult

INTRODUCTION

- In many fields of political science baseline random assignment and experimental design are impossible
    - For many research questions, observational data is our only option but this makes causal inference difficult
    - Assumption of *strict exogeneity* is usually unrealistic and valid instruments are a rare beast (Liu, Wang, and Xu 2022; Mogstad and Torgovitsky 2018; Swamy, Tavlas, and Hall 2015)

INTRODUCTION

- In many fields of political science baseline random assignment and experimental design are impossible
  - For many research questions, observational data is our only option but this makes causal inference difficult
  - Assumption of *strict exogeneity* is usually unrealistic and valid instruments are a rare beast (Liu, Wang, and Xu 2022; Mogstad and Torgovitsky 2018; Swamy, Tavlas, and Hall 2015)
  - Need to rely on weaker assumptions for causal identification (Acharya, Blackwell, and Sen 2016; Blackwell and Glynn 2018; Forastiere, Mattei, and Ding 2018)

INTRODUCTION

- In many fields of political science baseline random assignment and experimental design are impossible
  - For many research questions, observational data is our only option but this makes causal inference difficult
  - Assumption of *strict exogeneity* is usually unrealistic and valid instruments are a rare beast (Liu, Wang, and Xu 2022; Mogstad and Torgovitsky 2018; Swamy, Tavlas, and Hall 2015)
  - Need to rely on weaker assumptions for causal identification (Acharya, Blackwell, and Sen 2016; Blackwell and Glynn 2018; Forastiere, Mattei, and Ding 2018)
- How can we estimate causal effects and describe their uncertainty using observational data?

CAUSAL INFERENCE IN POLITICAL SCIENCE

CAUSAL INFERENCE IN POLITICAL SCIENCE

- Cross-sectional time series data and causal inference

CAUSAL INFERENCE IN POLITICAL SCIENCE

- Cross-sectional time series data and causal inference
  - Causal inference literature in political science focuses largely on frequentist econometrics (i.e., Imai and Kim 2019, 2020)

CAUSAL INFERENCE IN POLITICAL SCIENCE

- Cross-sectional time series data and causal inference
  - Causal inference literature in political science focuses largely on frequentist econometrics (i.e., Imai and Kim 2019, 2020)
    - Still relies on strict exogeneity assumption

CAUSAL INFERENCE IN POLITICAL SCIENCE

- Cross-sectional time series data and causal inference
  - Causal inference literature in political science focuses largely on frequentist econometrics (i.e., Imai and Kim 2019, 2020)
    - Still relies on strict exogeneity assumption
  - Recent works drawing on approaches developed in biostatistics outline a framework for estimating causal effects under the relatively weaker assumption of *sequential ignorability* (Acharya, Blackwell, and Sen 2016; Blackwell and Glynn 2018)

CAUSAL INFERENCE IN POLITICAL SCIENCE

- Cross-sectional time series data and causal inference
  - Causal inference literature in political science focuses largely on frequentist econometrics (i.e., Imai and Kim 2019, 2020)
    - Still relies on strict exogeneity assumption
  - Recent works drawing on approaches developed in biostatistics outline a framework for estimating causal effects under the relatively weaker assumption of *sequential ignorability* (Acharya, Blackwell, and Sen 2016; Blackwell and Glynn 2018)
- Our goal in this research note is to extend the approach to causal inference under selection on observables introduced by Blackwell and Glynn (2018) to a Bayesian framework

MARGINAL STRUCTURAL MODELS

MARGINAL STRUCTURAL MODELS

- Marginal structural models (MSMs) are a multi-stage approach to estimating causal effects where baseline random assignment is not possible (Robins 1997; Robins, Hernán, and Brumback 2000)

MARGINAL STRUCTURAL MODELS

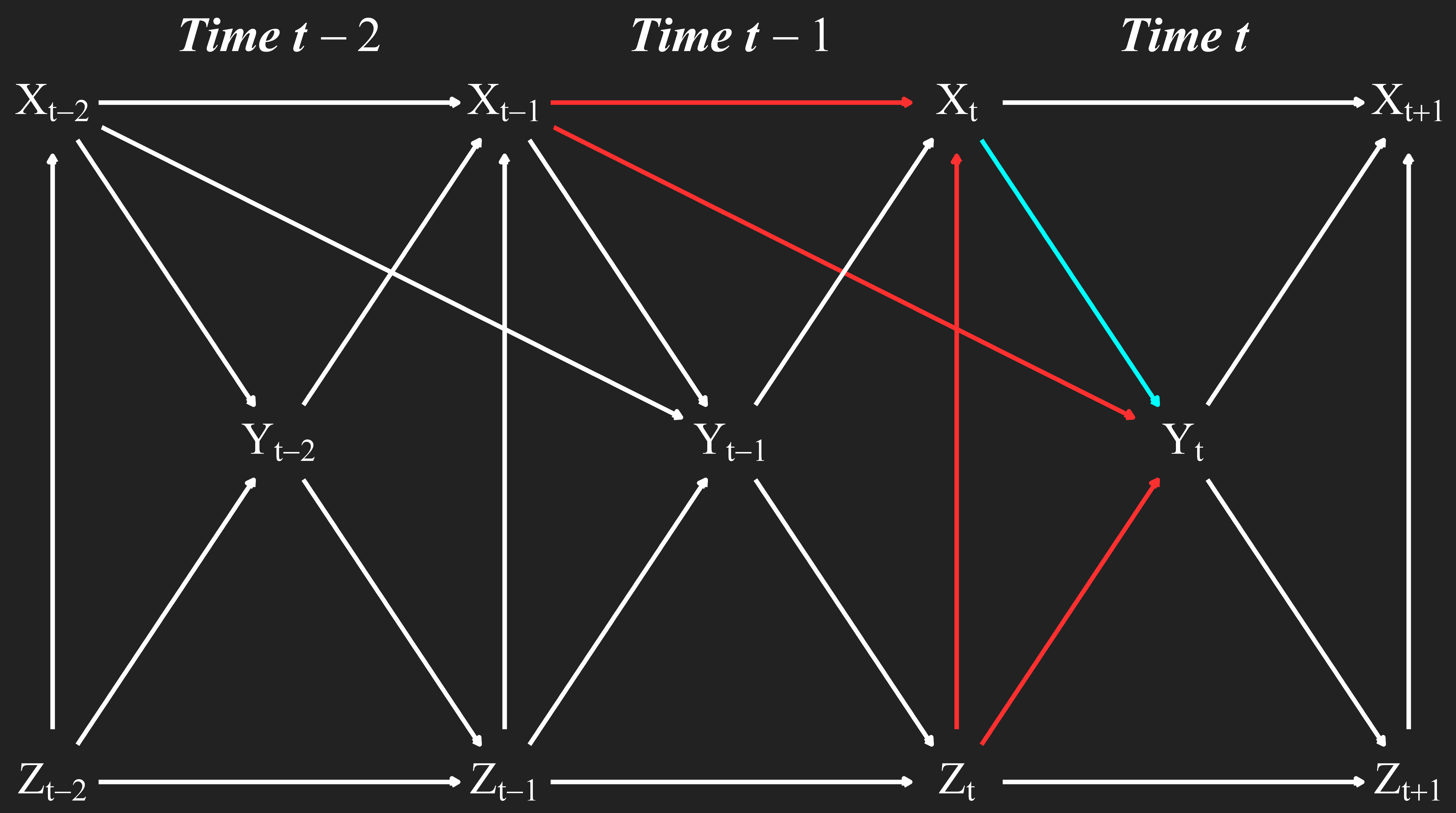
- Marginal structural models (MSMs) are a multi-stage approach to estimating causal effects where baseline random assignment is not possible (Robins 1997; Robins, Hernán, and Brumback 2000)
    - Relies on inverse probability of treatment weighting to achieve covariate balance by constructing pseudo-populations (Cole and Hernán 2008; Imai and Ratkovic 2015; Saarela et al. 2015)

MARGINAL STRUCTURAL MODELS

- Marginal structural models (MSMs) are a multi-stage approach to estimating causal effects where baseline random assignment is not possible (Robins 1997; Robins, Hernán, and Brumback 2000)
  - Relies on inverse probability of treatment weighting to achieve covariate balance by constructing pseudo-populations (Cole and Hernán 2008; Imai and Ratkovic 2015; Saarela et al. 2015)
  - Adjusting for biasing paths in the propensity model allows for identification of causal effects that are impossible to estimate in a single model due to post-treatment bias

MARGINAL STRUCTURAL MODELS

- Marginal structural models (MSMs) are a multi-stage approach to estimating causal effects where baseline random assignment is not possible (Robins 1997; Robins, Hernán, and Brumback 2000)
  - Relies on inverse probability of treatment weighting to achieve covariate balance by constructing pseudo-populations (Cole and Hernán 2008; Imai and Ratkovic 2015; Saarela et al. 2015)
  - Adjusting for biasing paths in the propensity model allows for identification of causal effects that are impossible to estimate in a single model due to post-treatment bias
  - Possible to estimate lagged effects and “treatment histories” in cross-sectional time series data under complex temporal dependence (Blackwell and Glynn 2018)



Time t – 2
Time t – 1
Time t
X_{t-2}
X_{t-1}
X_t
X_{t+1}
Y_{t-2}
Y_{t-1}
Y_t
Z_{t-2}
Z_{t-1}
Z_t
Z_{t+1}

WHY BAYESIAN ESTIMATION?

WHY BAYESIAN ESTIMATION?

- Frequentist uncertainty estimates are based on assumptions about *sampling distributions*

WHY BAYESIAN ESTIMATION?

- Frequentist uncertainty estimates are based on assumptions about *sampling distributions*
  - Yet, in many areas of political science our data comprise an *apparent population* rather than a random sample from a larger population of interest (Berk, Western, and Weiss 1995; Gill 2001)

WHY BAYESIAN ESTIMATION?

- Frequentist uncertainty estimates are based on assumptions about *sampling distributions*
  - Yet, in many areas of political science our data comprise an *apparent population* rather than a random sample from a larger population of interest (Berk, Western, and Weiss 1995; Gill 2001)
    - For example, all sovereign countries between 1945 and 2020 or all states in U.S. over some time period

WHY BAYESIAN ESTIMATION?

- Frequentist uncertainty estimates are based on assumptions about *sampling distributions*
  - Yet, in many areas of political science our data comprise an *apparent population* rather than a random sample from a larger population of interest (Berk, Western, and Weiss 1995; Gill 2001)
    - For example, all sovereign countries between 1945 and 2020 or all states in U.S. over some time period
  - It doesn't make sense to think in terms of random samples from a population if your observed data *is the population* (Gill and Heuberger 2020; Western and Jackman 1994)

WHY BAYESIAN ESTIMATION?

WHY BAYESIAN ESTIMATION?

- A Bayesian framework provides a straightforward approach to accounting for and propagating uncertainty in the specification of the propensity model

WHY BAYESIAN ESTIMATION?

- A Bayesian framework provides a straightforward approach to accounting for and propagating uncertainty in the specification of the propensity model
  - Bayesian Model Averaging (BMA) and cross-validation based stacking approaches allow us to avoid choosing a single specification for the propensity model (Kaplan and Chen 2014; Corwin Matthew Zigler and Dominici 2014)

WHY BAYESIAN ESTIMATION?

- A Bayesian framework provides a straightforward approach to accounting for and propagating uncertainty in the specification of the propensity model
    - Bayesian Model Averaging (BMA) and cross-validation based stacking approaches allow us to avoid choosing a single specification for the propensity model (Kaplan and Chen 2014; Corwin Matthew Zigler and Dominici 2014)
    - Acknowledges that we are virtually always uncertain about what the true set of confounders is

WHY BAYESIAN ESTIMATION?

- A Bayesian framework provides a straightforward approach to accounting for and propagating uncertainty in the specification of the propensity model
  - Bayesian Model Averaging (BMA) and cross-validation based stacking approaches allow us to avoid choosing a single specification for the propensity model (Kaplan and Chen 2014; Corwin Matthew Zigler and Dominici 2014)
  - Acknowledges that we are virtually always uncertain about what the true set of confounders is
  - May help reduce the degree to which our results depend on the propensity model being correctly specified (Hahn, Murray, and Carvalho 2020)

WHY BAYESIAN ESTIMATION?

WHY BAYESIAN ESTIMATION?

- Inverse probability of treatment weights are an estimated quantity with associated uncertainty (Liao and Zigler 2020)

WHY BAYESIAN ESTIMATION?

- Inverse probability of treatment weights are an estimated quantity with associated uncertainty (Liao and Zigler 2020)
  - Most applications of IPTW methods in political science assume the weights are deterministic (i.e., Blackwell and Glynn 2018; Kurtz and Lauretig 2021; Ladam, Harden, and Windett 2018)

WHY BAYESIAN ESTIMATION?

- Inverse probability of treatment weights are an estimated quantity with associated uncertainty (Liao and Zigler 2020)
  - Most applications of IPTW methods in political science assume the weights are deterministic (i.e., Blackwell and Glynn 2018; Kurtz and Lauretig 2021; Ladam, Harden, and Windett 2018)
  - Need to propagate uncertainty in the design stage weights to outcome stage (Liao and Zigler 2020)

WHY BAYESIAN ESTIMATION?

- Inverse probability of treatment weights are an estimated quantity with associated uncertainty (Liao and Zigler 2020)
  - Most applications of IPTW methods in political science assume the weights are deterministic (i.e., Blackwell and Glynn 2018; Kurtz and Lauretig 2021; Ladam, Harden, and Windett 2018)
  - Need to propagate uncertainty in the design stage weights to outcome stage (Liao and Zigler 2020)
- Highlights a problem with fully Bayesian estimation of MSMs which requires the models be estimated jointly (Robins, Hernán, and Wasserman 2015; Corwin M. Zigler et al. 2013)

A BAYESIAN PSEDUO-LIKELIHOOD APPROACH

BAYESIAN DESIGN STAGE ESTIMATION

For some binary treatment X_{it}, the posterior expectation of the stabilized inverse probability of treatment weights for each unit i at time t is

$$\mathrm{IPW}_{it} = \prod_{t=1}^{t} \frac{\int \Pr[X_{it} \mid X_{it-1},\ C_i]\pi(\theta)d\theta}{\int \Pr[X_{it} \mid Z_{it},\ X_{it-1},\ Y_{it-1},\ C_i]\pi(\theta)d\theta}$$

BAYESIAN DESIGN STAGE ESTIMATION

For some binary treatment X_{it}, the posterior expectation of the stabilized inverse probability of treatment weights for each unit i at time t is

$$\text{IPW}_{it} = \prod_{t=1}^{t} \frac{\int \Pr[X_{it} \mid X_{it-1},\ C_i]\pi(\theta)d\theta}{\int \Pr[X_{it} \mid Z_{it},\ X_{it-1},\ Y_{it-1},\ C_i]\pi(\theta)d\theta}$$

- X_{it-1} and Y_{it-1} denote the treatment status and outcome for the i^{th} unit in the previous period respectively

BAYESIAN DESIGN STAGE ESTIMATION

For some binary treatment X_{it}, the posterior expectation of the stabilized inverse probability of treatment weights for each unit i at time t is

$$\mathrm{IPW}_{it} = \prod_{t=1}^{t} \frac{\int \Pr[X_{it} \mid X_{it-1},\ C_i]\pi(\theta)d\theta}{\int \Pr[X_{it} \mid Z_{it},\ X_{it-1},\ Y_{it-1},\ C_i]\pi(\theta)d\theta}$$

- X_{it-1} and Y_{it-1} denote the treatment status and outcome for the i^{th} unit in the previous period respectively
- C_i is a set of time-invariant baseline covariates

BAYESIAN DESIGN STAGE ESTIMATION

For some binary treatment X_{it}, the posterior expectation of the stabilized inverse probability of treatment weights for each unit i at time t is

$$\mathrm{IPW}_{it} = \prod_{t=1}^{t} \frac{\int \Pr[X_{it} \mid X_{it-1},\ C_i]\pi(\theta)d\theta}{\int \Pr[X_{it} \mid Z_{it},\ X_{it-1},\ Y_{it-1},\ C_i]\pi(\theta)d\theta}$$

- X_{it-1} and Y_{it-1} denote the treatment status and outcome for the i^{th} unit in the previous period respectively
- C_i is a set of time-invariant baseline covariates
- Z_{it} is a set of time-varying covariates that satisfies sequential ignorability

BAYESIAN DESIGN STAGE ESTIMATION

For some binary treatment X_{it}, the posterior expectation of the stabilized inverse probability of treatment weights for each unit i at time t is

$$\mathrm{IPW}_{it} = \prod_{t=1}^{t} \frac{\int \Pr[X_{it} \mid X_{it-1},\ C_i]\pi(\theta)d\theta}{\int \Pr[X_{it} \mid Z_{it},\ X_{it-1},\ Y_{it-1},\ C_i]\pi(\theta)d\theta}$$

- X_{it-1} and Y_{it-1} denote the treatment status and outcome for the i^{th} unit in the previous period respectively
- C_i is a set of time-invariant baseline covariates
- Z_{it} is a set of time-varying covariates that satisfies sequential ignorability
- Although we focus mainly on the average treatment effect at times t and $t-1$, it is possible to estimate longer lags and other estimands as well.

BAYESIAN DESIGN STAGE ESTIMATION

It is also possible to extend IPTW to cases in which X_{it} is continuous, in which case the stabilized weights are

$$\text{IPW}_{it} = \prod_{t=1}^{t} \frac{f_{X_{it} \mid X_{it-1}, C_i}\left[\left(X_{it} \mid X_{it-1},\ C_i\right);\ \mu,\ \sigma^2\right]}{f_{X_{it} \mid Z_{it}, X_{it-1}, Y_{it-1}, C_i}\left[\left(X_{it} \mid Z_{it},\ X_{it-1},\ Y_{it-1},\ C_i\right);\ \mu,\ \sigma^2\right]}$$

BAYESIAN DESIGN STAGE ESTIMATION

It is also possible to extend IPTW to cases in which X_{it} is continuous, in which case the stabilized weights are

$$\text{IPW}_{it} = \prod_{t=1}^{t} \frac{f_{X_{it}|X_{it-1},C_i}[(X_{it} \mid X_{it-1},\ C_i);\ \mu,\ \sigma^2]}{f_{X_{it}|Z_{it},X_{it-1},Y_{it-1},C_i}[(X_{it} \mid Z_{it},\ X_{it-1},\ Y_{it-1},\ C_i);\ \mu,\ \sigma^2]}$$

- Each of the parameters X, Y, Z, and C in the numerator and denominator are the same as in the binary version

BAYESIAN DESIGN STAGE ESTIMATION

It is also possible to extend IPTW to cases in which X_{it} is continuous, in which case the stabilized weights are

$$\text{IPW}_{it} = \prod_{t=1}^{t} \frac{f_{X_{it}|X_{it-1},C_i}[(X_{it} \mid X_{it-1},\ C_i);\ \mu,\ \sigma^2]}{f_{X_{it}|Z_{it},X_{it-1},Y_{it-1},C_i}[(X_{it} \mid Z_{it},\ X_{it-1},\ Y_{it-1},\ C_i);\ \mu,\ \sigma^2]}$$

- Each of the parameters X, Y, Z, and C in the numerator and denominator are the same as in the binary version
- The $f_{\ldots}(\cdot)$ expressions represent a probability density function with mean μ and variance σ^2

BAYESIAN DESIGN STAGE ESTIMATION

It is also possible to extend IPTW to cases in which X_{it} is continuous, in which case the stabilized weights are

$$\mathrm{IPW}_{it} = \prod_{t=1}^{t} \frac{f_{X_{it} \mid X_{it-1}, C_i}[(X_{it} \mid X_{it-1},\ C_i);\ \mu,\ \sigma^2]}{f_{X_{it} \mid Z_{it}, X_{it-1}, Y_{it-1}, C_i}[(X_{it} \mid Z_{it},\ X_{it-1},\ Y_{it-1},\ C_i);\ \mu,\ \sigma^2]}$$

- Each of the parameters X, Y, Z, and C in the numerator and denominator are the same as in the binary version
- The $f_{\dots}(\cdot)$ expressions represent a probability density function with mean μ and variance σ^2
- We'll focus mainly on binary treatment regimes, though this particular method tends to behave better for a continuous X in some cases

THE BAYESIAN PSEUDO-LIKELIHOOD

To propagate uncertainty in the distribution of weights from the design stage while avoiding the problem of feedback inherent in joint estimation, we develop a Bayesian Pseudo-Likelihood estimator (Savitsky and Toth 2016; Williams and Savitsky 2020a, 2020b)

$$\hat{\pi}(\theta \mid y, \tilde{w}) \;\propto\; \left[\prod_{i=1}^{n} \Pr(y_i \mid \theta)^{\tilde{w}_i}\right] \pi(\theta)$$

THE BAYESIAN PSEUDO-LIKELIHOOD

To propagate uncertainty in the distribution of weights from the design stage while avoiding the problem of feedback inherent in joint estimation, we develop a Bayesian Pseudo-Likelihood estimator (Savitsky and Toth 2016; Williams and Savitsky 2020a, 2020b)

$$\hat{\pi}(\theta \mid y, \tilde{w}) \;\propto\; \left[\prod_{i=1}^{n} \Pr(y_i \mid \theta)^{\tilde{w}_i}\right] \pi(\theta)$$

- $\tilde{w}_i$ is the realized IPT weight for the i^{th} observation

THE BAYESIAN PSEUDO-LIKELIHOOD

To propagate uncertainty in the distribution of weights from the design stage while avoiding the problem of feedback inherent in joint estimation, we develop a Bayesian Pseudo-Likelihood estimator (Savitsky and Toth 2016; Williams and Savitsky 2020a, 2020b)

$$\hat{\pi}(\theta \mid y, \tilde{w}) \;\propto\; \left[\prod_{i=1}^{n} \Pr(y_i \mid \theta)^{\tilde{w}_i}\right] \pi(\theta)$$

- $\tilde{w}_i$ is the realized IPT weight for the i^{th} observation
- $\prod_{i=1}^{n} \Pr(y_i \mid \theta)^{\tilde{w}_i}$ is the pseudo-likelihood and π denotes the prior probability for a parameter θ

THE BAYESIAN PSEUDO-LIKELIHOOD

To propagate uncertainty in the distribution of weights from the design stage while avoiding the problem of feedback inherent in joint estimation, we develop a Bayesian Pseudo-Likelihood estimator (Savitsky and Toth 2016; Williams and Savitsky 2020a, 2020b)

$$\hat{\pi}(\theta \mid y, \tilde{w}) \ \propto \ \left[\prod_{i=1}^{n} \Pr(y_i \mid \theta)^{\tilde{w}_i}\right] \pi(\theta)$$

- $\tilde{w}_i$ is the realized IPT weight for the i^{th} observation
- $\prod_{i=1}^{n} \Pr(y_i \mid \theta)^{\tilde{w}_i}$ is the pseudo-likelihood and π denotes the prior probability for a parameter θ
- $\hat{\pi}(\theta \mid y, \tilde{w})$ represents the Bayesian pseudo-posterior for θ

PARAMETERIZATION OF THE WEIGHTS

PARAMETERIZATION OF THE WEIGHTS

- We decompose the matrix of weights from the design stage into a location component λ and a scale component δ

PARAMETERIZATION OF THE WEIGHTS

- We decompose the matrix of weights from the design stage into a location component λ and a scale component δ
- The weight for each observation is sampled as

$$\tilde{w}_{it} \sim \lambda_{it} + \delta_{it} \cdot \pi(\delta_{it})$$

where $\pi(\delta_{it})$ is a regularizing prior on the scale of the weights such as an exponential distribution with rate $\lambda > 3.5$ or Beta distribution with shape parameters $\alpha = 2$ and $\beta \geq 2$

PARAMETERIZATION OF THE WEIGHTS

- We decompose the matrix of weights from the design stage into a location component λ and a scale component δ
- The weight for each observation is sampled as

$$\tilde{w}_{it} \sim \lambda_{it} + \delta_{it} \cdot \pi(\delta_{it})$$

  where $\pi(\delta_{it})$ is a regularizing prior on the scale of the weights such as an exponential distribution with rate $\lambda > 3.5$ or Beta distribution with shape parameters $\alpha = 2$ and $\beta \geq 2$
- Provides computational stability and shuts down extreme values when the IPT weights have high variance

PARAMETERIZATION OF THE WEIGHTS

- We decompose the matrix of weights from the design stage into a location component λ and a scale component δ
- The weight for each observation is sampled as

$$\tilde{w}_{it} \sim \lambda_{it} + \delta_{it} \cdot \pi(\delta_{it})$$

  where $\pi(\delta_{it})$ is a regularizing prior on the scale of the weights such as an exponential distribution with rate $\lambda > 3.5$ or Beta distribution with shape parameters $\alpha = 2$ and $\beta \geq 2$
- Provides computational stability and shuts down extreme values when the IPT weights have high variance
- Straightforward extensions for nested data structures via double-weighted estimation (Savitsky and Williams 2021)

SIMULATION STUDY DESIGN

SIMULATION STUDY OVERVIEW

SIMULATION STUDY OVERVIEW

- To assess parameter recovery and bias, we adapt the original simulation design from Blackwell and Glynn (2018)

SIMULATION STUDY OVERVIEW

- To assess parameter recovery and bias, we adapt the original simulation design from Blackwell and Glynn (2018)
- We simulate 2000 data sets of varying dimensions, manipulating the path $X_{it-1} \longrightarrow Z_{it}$

SIMULATION STUDY OVERVIEW

- To assess parameter recovery and bias, we adapt the original simulation design from Blackwell and Glynn (2018)
- We simulate 2000 data sets of varying dimensions, manipulating the path $X_{it-1} \longrightarrow Z_{it}$
  - Periods $\in \{20, 50\}$

SIMULATION STUDY OVERVIEW

- To assess parameter recovery and bias, we adapt the original simulation design from Blackwell and Glynn (2018)
- We simulate 2000 data sets of varying dimensions, manipulating the path $X_{it-1} \longrightarrow Z_{it}$
  - Periods $\in \{20, 50\}$
  - Groups $\in \{25, 45, 65, 85, 100\}$

SIMULATION STUDY OVERVIEW

- To assess parameter recovery and bias, we adapt the original simulation design from Blackwell and Glynn (2018)
- We simulate 2000 data sets of varying dimensions, manipulating the path $X_{it-1} \longrightarrow Z_{it}$
  - Periods $\in \{20, 50\}$
  - Groups $\in \{25, 45, 65, 85, 100\}$
- Objectives

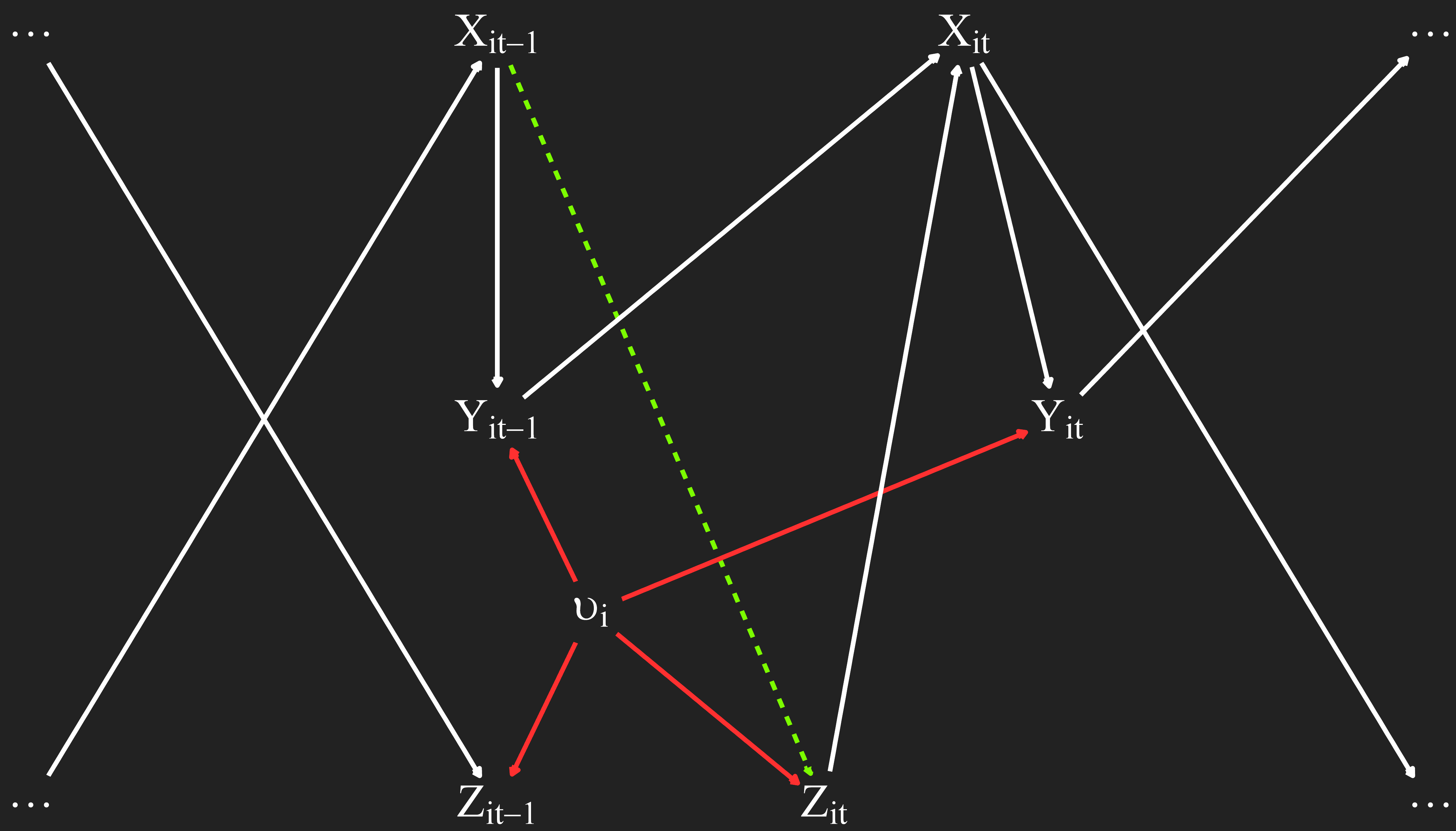
SIMULATION STUDY OVERVIEW

- To assess parameter recovery and bias, we adapt the original simulation design from Blackwell and Glynn (2018)
- We simulate 2000 data sets of varying dimensions, manipulating the path $X_{it-1} \longrightarrow Z_{it}$
  - Periods $\in \{20, 50\}$
  - Groups $\in \{25, 45, 65, 85, 100\}$
- Objectives
  - Identify both $X_{it} \longrightarrow Y_{it}$ and $X_{it-1} \longrightarrow Y_{it}$

SIMULATION STUDY OVERVIEW

- To assess parameter recovery and bias, we adapt the original simulation design from Blackwell and Glynn (2018)
- We simulate 2000 data sets of varying dimensions, manipulating the path $X_{it-1} \longrightarrow Z_{it}$
  - Periods $\in \{20, 50\}$
  - Groups $\in \{25, 45, 65, 85, 100\}$
- Objectives
  - Identify both $X_{it} \longrightarrow Y_{it}$ and $X_{it-1} \longrightarrow Y_{it}$
  - Compare our Bayesian Pseduo-Likelihood approach against the more common auto-regressive distributed lag (ARDL) specification

DAG FOR THE SIMULATED DATA



ARDL MODEL SPECIFICATION

$$y_{it} \sim Normal(\mu_{it}, \epsilon^2)$$

$$\mu_{it} = \alpha + \beta_1 X_{it} + \beta_2 X_{it-1} + \beta_3 Y_{it-1} + \beta_4 Y_{it-2} + \beta_5 Z_{it} + \beta_6 Z_{it-1} + \epsilon$$

with priors

$$\alpha \sim Normal(\bar{y},\ 2 \cdot \sigma_y) \qquad \beta_k \sim Normal\left(0,\ 1.5 \cdot \frac{\sigma_y}{\sigma_x}\right)$$

$$\epsilon \sim Exponential\left(\frac{1}{\sigma_y}\right)$$

MSM DESIGN STAGE SPECIFICATION

As illustrated in the equation for the stabilized weights, we specify two separate models for the numerator and denominator with weakly informative independent normal priors on α and β

$$\Pr(X_{it} = 1 \mid \theta_{it}) \sim Bernoulli(\theta_{it})$$

$$\theta_{it} = \text{logit}^{-1}(\alpha + X_n\beta_k)$$

$$\text{with priors}$$

$$\alpha \sim Normal(0,\ 2) \qquad \beta_k \sim Normal(0,\ 1)$$

MSM DESIGN STAGE SPECIFICATION

As illustrated in the equation for the stabilized weights, we specify two separate models for the numerator and denominator with weakly informative independent normal priors on α and β

$$\begin{aligned}\Pr(X_{it} = 1 \mid \theta_{it}) &\sim \mathit{Bernoulli}(\theta_{it}) \\ \theta_{it} &= \text{logit}^{-1}(\alpha + X_n\beta_k) \\ \text{with priors} & \\ \alpha \sim \mathit{Normal}(0,\ 2) & \qquad \beta_k \sim \mathit{Normal}(0,\ 1)\end{aligned}$$

- For the numerator model, the matrix X_n is simply X_{it-1}

MSM DESIGN STAGE SPECIFICATION

As illustrated in the equation for the stabilized weights, we specify two separate models for the numerator and denominator with weakly informative independent normal priors on α and β

$$\Pr(X_{it} = 1 \mid \theta_{it}) \sim Bernoulli(\theta_{it})$$

$$\theta_{it} = \text{logit}^{-1}(\alpha + X_n \beta_k)$$

$$\text{with priors}$$

$$\alpha \sim Normal(0,\ 2) \qquad \beta_k \sim Normal(0,\ 1)$$

- For the numerator model, the matrix X_n is simply X_{it-1}
- For the denominator model, $X_n = \{Z_{it},\ X_{it-1},\ Y_{it-1}\}$

MSM OUTCOME MODEL SPECIFICATION

$$y_{it} \sim Normal(\mu_{it}, \epsilon^2)^{\tilde{w}_{it}}$$

$$\mu_{it} = \alpha + \beta_1 X_{it} + \beta_2 X_{it-1} + \epsilon$$

where

$$\tilde{w}_{it} \sim \lambda_{it} + \delta_{it} \cdot \pi(\delta)$$

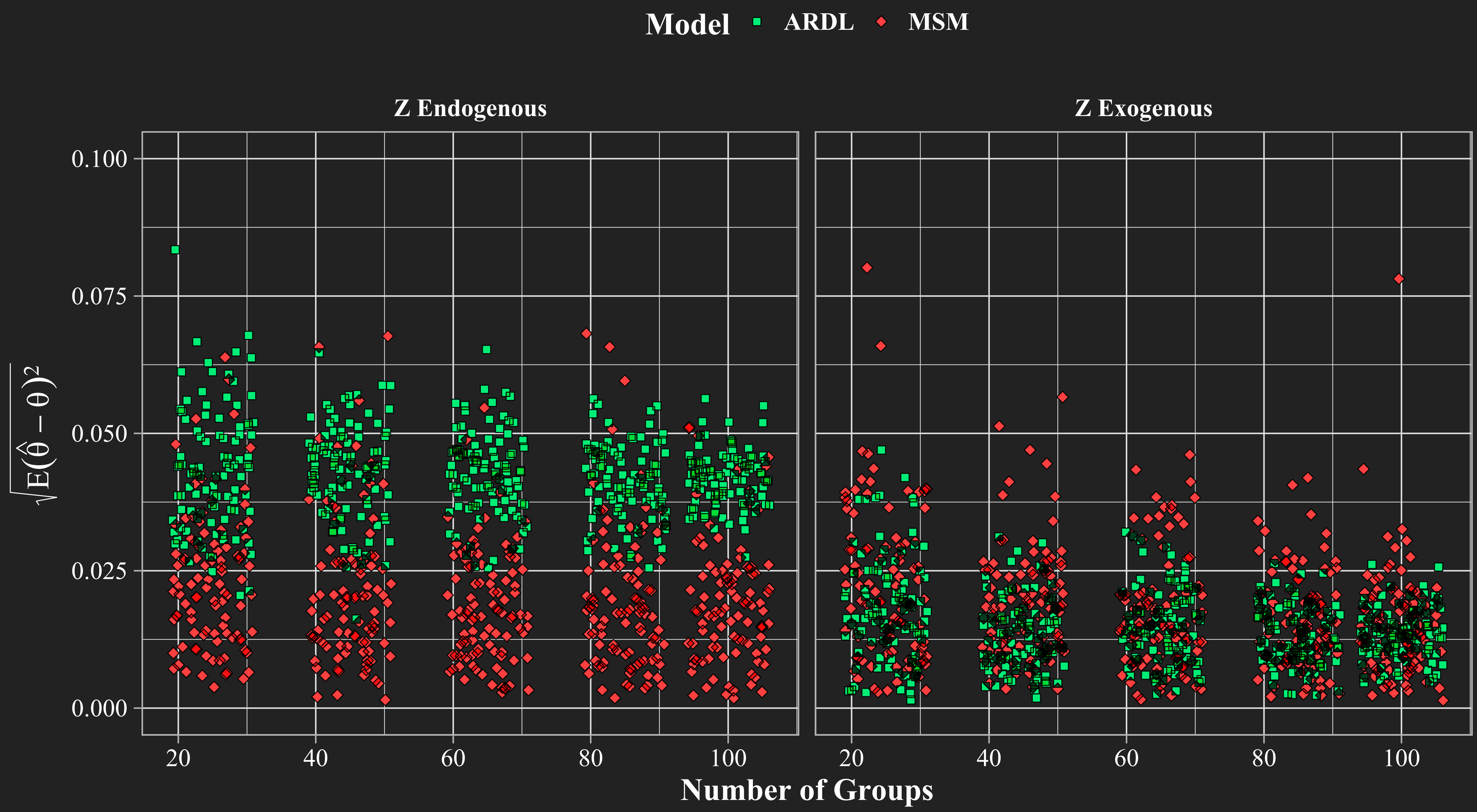
with priors

$$\alpha \sim Normal(\bar{y},\ 2 \cdot \sigma_y) \qquad \beta_k \sim Normal\left(0,\ 1.5 \cdot \frac{\sigma_y}{\sigma_x}\right)$$

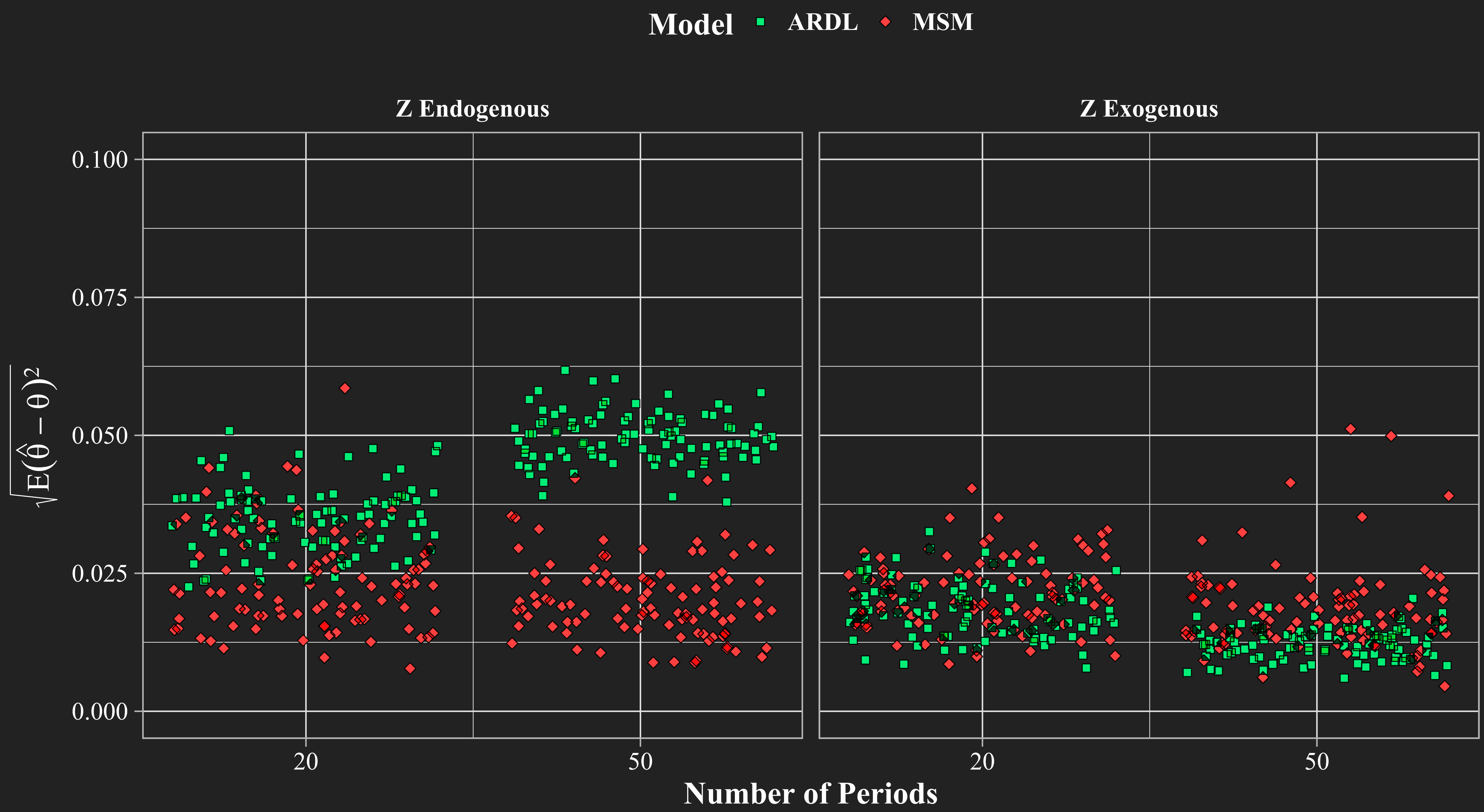
$$\epsilon \sim Exponential\left(\frac{1}{\sigma_y}\right) \qquad \delta_{it} \sim Beta(2,\ 5)$$

SIMULATION RESULTS

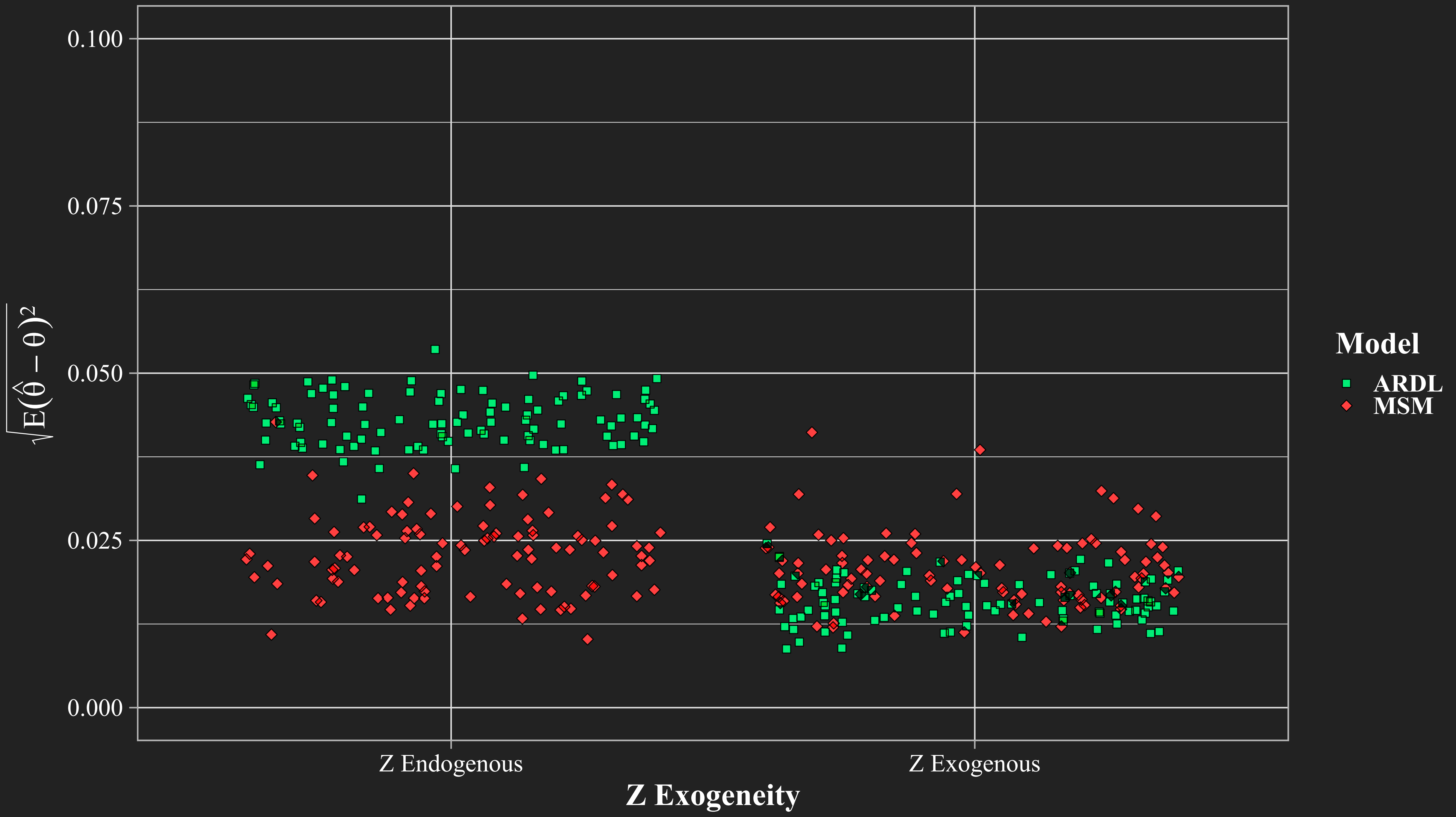
SIMULATION RESULTS



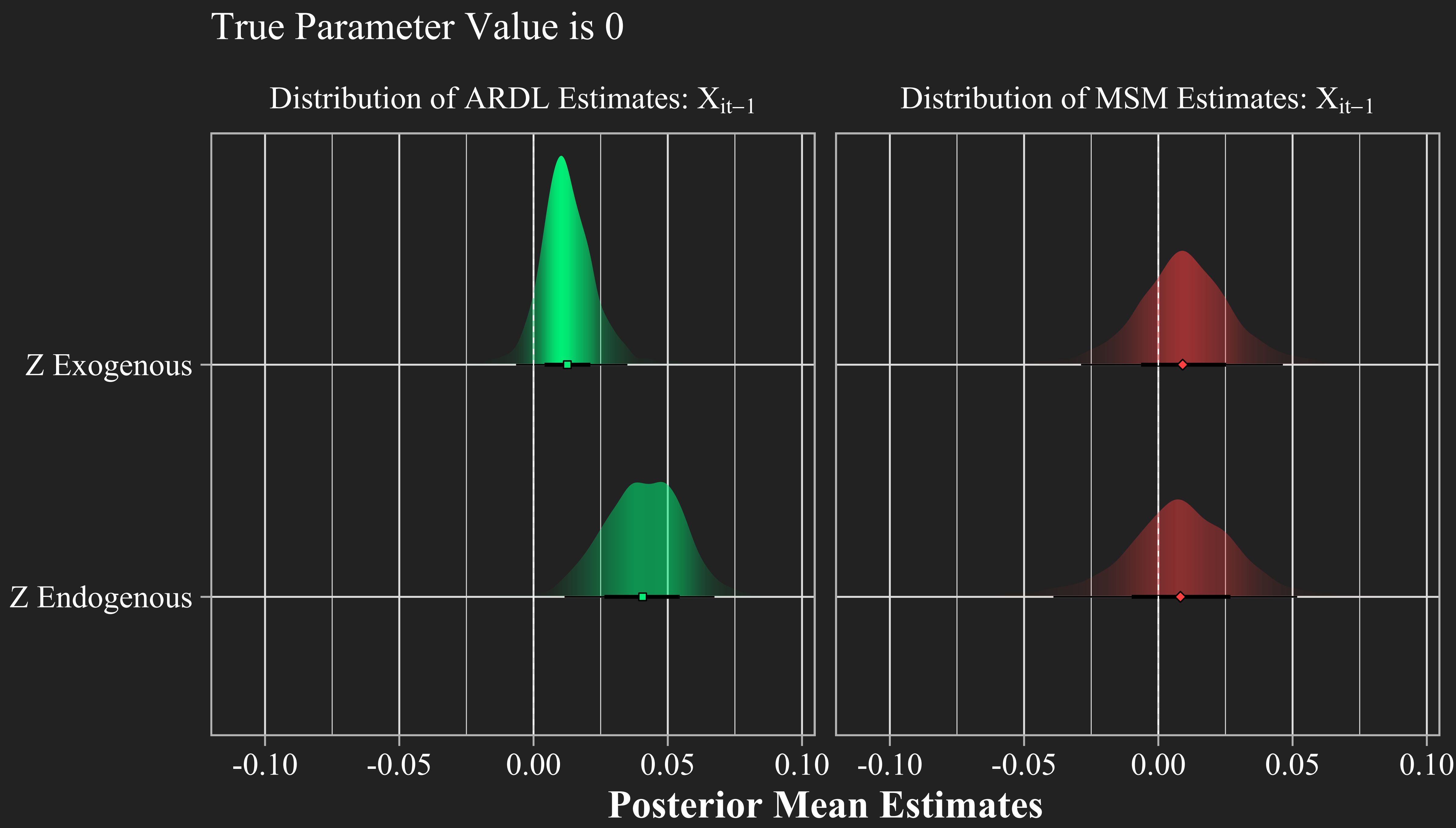
SIMULATION RESULTS



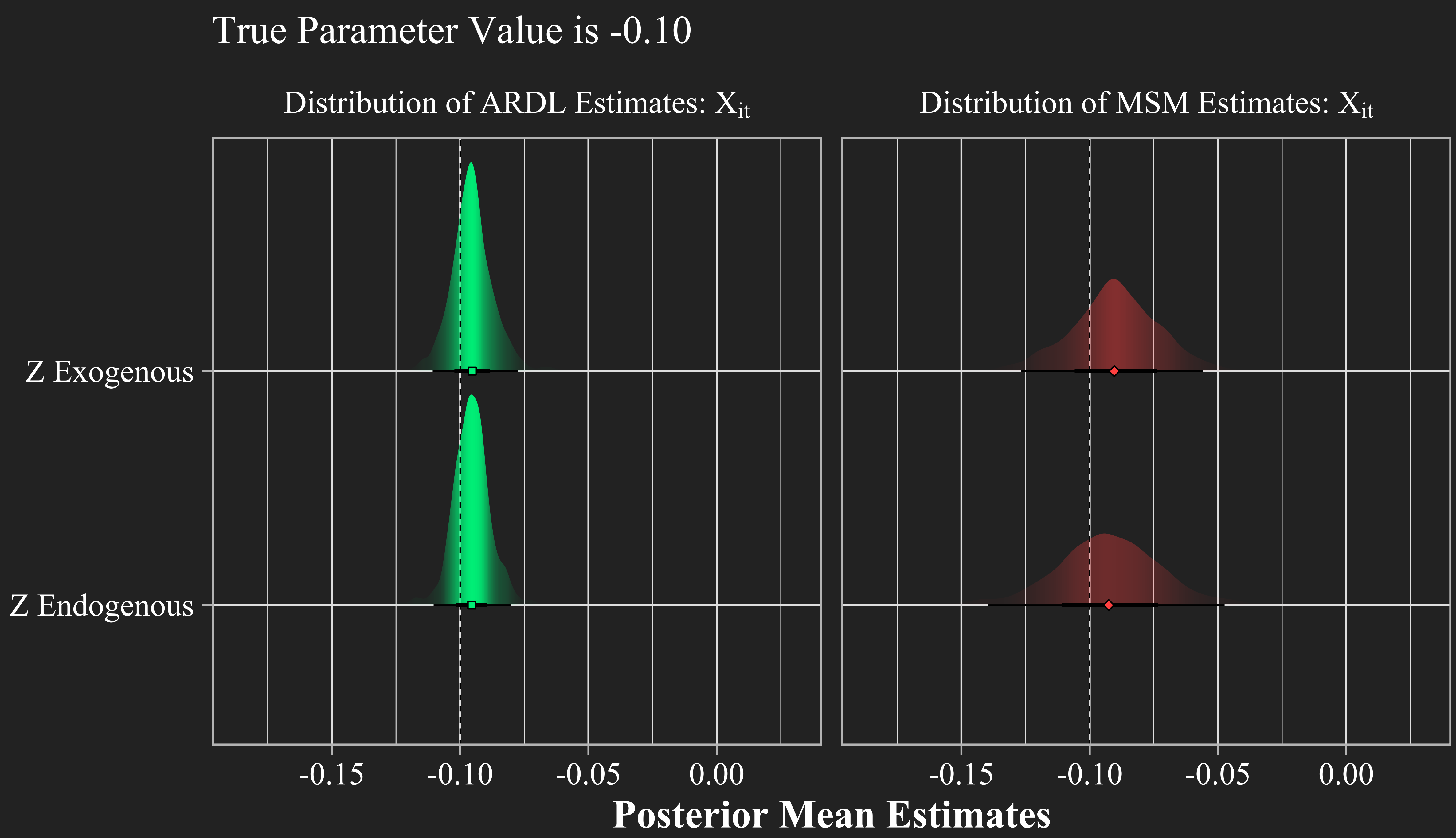
SIMULATION RESULTS



SIMULATION RESULTS



SIMULATION RESULTS



CONCLUSIONS

CONCLUSION

CONCLUSION

- Overall, our proposed procedure performs well in terms of parameter recovery under fairly general conditions

CONCLUSION

- Overall, our proposed procedure performs well in terms of parameter recovery under fairly general conditions
- Going forward, we need to apply this to some real world political science examples

CONCLUSION

- Overall, our proposed procedure performs well in terms of parameter recovery under fairly general conditions
- Going forward, we need to apply this to some real world political science examples
- Planned R package implementing our procedure by building on the brms package as a backend

CONCLUSION

- Overall, our proposed procedure performs well in terms of parameter recovery under fairly general conditions
- Going forward, we need to apply this to some real world political science examples
- Planned R package implementing our procedure by building on the brms package as a backend
  - Makes it super easy for anyone who knows standard R model syntax to use

REFERENCES

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110(3): 512–29.

Berk, Richard A., Bruce Western, and Robert E. Weiss. 1995. "Statistical Inference for Apparent Populations." *Sociological Methodology* 25: 421.

Blackwell, Matthew, and Adam N. Glynn. 2018. "How to Make Causal Inferences with Time-Series Cross-Sectional Data Under Selection on Observables." *American Political Science Review* 112: 1067–82.

Cole, Stephen R., and Miguel A. Hernán. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology* 168(6): 656–64.

Forastiere, Laura, Alessandra Mattei, and Peng Ding. 2018. "Principal Ignorability in Mediation Analysis: Through and Beyond Sequential Ignorability." *Biometrika* 105(4): 979–86.

Gill, Jeff. 2001. "Whose Variance Is It Anyway? Interpreting Empirical Models with State-Level Data." *State Politics & Policy Quarterly* 1(3): 318–38.

Gill, Jeff, and Simon Heuberger. 2020. "Bayesian Modeling and Inference: A Post-Modern Perspective." In *The SAGE Handbook of Research Methods in Political Science and International Relations*, eds. Luigi Curini and Robert Franzese. London, UK: SAGE, 961–84.

Hahn, P. Richard, Jared S. Murray, and Carlos M. Carvalho. 2020. "Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects (with Discussion)." *Bayesian Analysis*

15(3).

Imai, Kosuke, and In Song Kim. 2019. "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63(2): 467–90.

———. 2020. "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis* 29(3): 405–15.

Imai, Kosuke, and Marc Ratkovic. 2015. "Robust Estimation of Inverse Probability Weights for Marginal Structural Models." *Journal of the American Statistical Association* 110(511): 1013–23.

Kaplan, David, and Jianshen Chen. 2014. "Bayesian Model Averaging for Propensity Score Analysis." *Multivariate Behavioral Research* 49(6): 505–17.

Kurtz, Marcus J., and Adam Lauretig. 2021. "Does Free-Market Reform Induce Protest? Selection, Post-Treatment Bias, and Depoliticization." *British Journal of Political Science* 52(2): 968–76.

Ladam, Christina, Jeffrey J. Harden, and Jason H. Windett. 2018. "Prominent Role Models: High-profile Female Politicians and the Emergence of Women as Candidates for Public Office." *American Journal of Political Science* 62: 369–81.

Liao, Shirley X., and Corwin M. Zigler. 2020. "Uncertainty in the Design Stage of Two-Stage Bayesian Propensity Score Analysis." *Statistics in Medicine* 39(17): 2265–90.

Liu, Licheng, Ye Wang, and Yiqing Xu. 2022. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science*.

Mogstad, Magne, and Alexander Torgovitsky. 2018. "Identification and Extrapolation of Causal Effects with Instrumental Variables." *Annual Review of Economics* 10(1): 577–613.

Robins, James M. 1997. "Causal Inference from Complex Longitudinal Data." In *Latent Variable Modeling and Applications to Causality*, Lecture notes in statistics, ed. Maia Berkane. New York: Springer-Verlag, 69–117.

Robins, James M., Miguel Ángel Hernán, and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11(5): 550–60.

Robins, James M., Miguel A. Hernán, and Larry Wasserman. 2015. "Discussion of 'on Bayesian Estimation of Marginal Structural Models'." *Biometrics* 71(2): 296–99.

Saarela, Olli, David A. Stephens, Erica E. M. Moodie, and Marina B. Klein. 2015. "On Bayesian Estimation of Marginal Structural Models." *Biometrics* 71(2): 279–88.

Savitsky, Terrance D., and Daniell Toth. 2016. "Bayesian Estimation Under Informative Sampling." *Electronic Journal of Statistics* 10(1).

Savitsky, Terrance D., and Matthew R. Williams. 2021. "Pseudo Bayesian Mixed Models Under Informative Sampling." *arXiv*.

Swamy, P. A. V. B., George Tavlas, and Stephen Hall. 2015. "On the Interpretation of Instrumental Variables in the Presence of Specification Errors." *Econometrics* 3(1): 55–64.

Western, Bruce, and Simon Jackman. 1994. "Bayesian Inference for Comparative Research." *American Political Science Review* 88: 412–23.

Williams, Matthew R., and Terrance D. Savitsky. 2020a. "Bayesian Estimation Under Informative Sampling with Unattenuated Dependence." *Bayesian Analysis* 15(1).

———. 2020b. "Uncertainty Estimation for Pseudo-Bayesian Inference Under Complex Sampling." *International Statistical Review* 89(1): 72–107.

Zigler, Corwin M. et al. 2013. "Model Feedback in Bayesian Propensity Score Estimation." *Biometrics* 69(1): 263–73.

Zigler, Corwin Matthew, and Francesca Dominici. 2014. "Uncertainty in Propensity Score Estimation: Bayesian Methods for Variable Selection and Model-Averaged Causal Effects." *Journal of the American Statistical Association* 109(505): 95–107.