DSE4231: Topics in Data Science and the Digital Economy

Evaluating the GLiDeR Model: A Comparative Study on Simulated Datasets

Group Project

Name
Justin Cheong

Table of Contents

1. Executive Summary

This paper evaluates the paper titled "Covariate selection with group lasso and doubly robust estimation of causal effects" (Koch et al., 2018). Koch et al. (2018) introduced a method to estimate the average treatment effect (ATE) called the grouped lasso and doubly robust estimate (GLiDeR). This paper replicated the findings obtained by Koch et al. (2018) using 13 different simulated methods, with 3 sets of variables with varying numbers of variables. Moreover, this paper introduces some potential improvements to the GLiDeR model, which will be further evaluated in this paper.

2. Introduction

2.1 Motivation

Covariate selection is crucial in causal inference, as improper selection can bias treatment effect estimates, leading to misleading conclusions. Traditional methods like LASSO and stepwise selection focus on prediction accuracy rather than explicitly identifying confounders, often resulting in omitted variables or irrelevant predictors.

2.2 Literature Review

To address these limitations, GLiDeR was introduced, which jointly selects covariates for both treatment and outcome models using group penalties (Koch et al., 2018), improving confounder selection and enhancing causal effect estimation. Moreover, GLiDeR leverages doubly robust estimation, ensuring consistent treatment effect estimates even if one of the models is misspecified. Given the challenges of model misspecification and high-dimensional data, GLiDeR is a promising approach to improving robustness in causal inference, making it valuable for applied research in economics, healthcare, and policy evaluation.

In their original study, Koch et al. (2018) compared GLiDeR against several alternative methods in a simulation study to compare its effectiveness.

1. **Two-Stage Model Averaged Doubly Robust (MADR):** Simple average of ATE estimates from different models
2. **Backward Selection:** A conventional variable selection method.
3. **Adaptive LASSO:** A refinement of LASSO that adaptively penalizes coefficients based on their importance.
4. **Saturated Model:** A model that includes all covariates, serving as a benchmark.

Koch et al. (2018) found that GLiDeR consistently demonstrated superior robustness and selection stability, particularly in scenarios with model misspecification and highly correlated covariates. Its ability to achieve higher MSE ratios (indicates lower MSE compared to the saturated model) makes it a more reliable approach compared to alternative methods.

2.3 Objectives of study

Our project aims to replicate, critically analyse and extend the findings of Koch et al. (2018) through (1) an exact replication of the original study, (2) carry out further robustness analysis under the below scenarios.

1. **Trigonometric Formulation of outcomes and treatment Data Generating Process (DGP).** The paper uses a variety of treatment and outcome models, including linear, interaction terms and polynomials. Extending this, we want to assess whether GLiDeR remains effective when accommodating for periodic relationships.
2. **Inclusion of qualitative covariates.** Study uses only continuous covariates from a normal distribution. We want to assess if GLiDeR can adapt to categorical covariates.
3. **Alternative LASSO variations to select covariates.** In the study, GLiDeR only used LASSO to select covariates. We assess if various adaptations such as regularized LASSO (rLASSO) can be used in a similar grouped and doubly robust setting.
4. **Dimensionality Reduction via Prescreening.** We want to assess if prescreening is helpful to reduce dimensionality and allow GLiDeR to choose the most informative predictor instead of a correlated but uninformative covariate.

3. Methodology

3.1 Data Generation Process

To replicate the original study, we generate a simulated dataset consisting of covariates, treatment assignment and an outcome variable. Numerous simulation scenarios are considered to evaluate the effects of varying levels of confounding, model misspecification, covariate structure, number of irrelevant variables (i.e., covariates unrelated to outcome and treatment), and sample size. To compare the performances of the various models, **Table 1** summarises the types of data that has been produced, and the corresponding scenarios describing the types of cases used to test for the model robustness.

Table 1: Types of Scenarios Introduced

Scenarios	Specification of Scenarios
1, 2, 3, 4	Correct specifications of <u>both</u> treatment and outcome variables
5, 6, 7	Misspecification of outcome models
8, 9	Misspecification of treatment models
10	High dimensionality of the datasets
11 (New)	Trigonometric misspecification of outcome model
12 (New)	Trigonometric misspecification of treatment model
13 (New)	Inclusion of categorical variables

To generate each scenario, we generate potential confounders from a normal distribution $V \sim N(\mu_v, \sigma^2_v)$. Treatment A is drawn from $Bernoulli[expit\{f(V)\}]$, where $expit(x) = \frac{exp(x)}{exp(x)+1}$ and outcome variable as $N(A + g(V), \sigma^2_y)$. The various combinations of scenarios can be summarised in **Table 2**. It is to note that Scenarios 1 to 10 are the same as that of Koch (2018) since the project asks for replication, while 11 to 13 are for exploration. Scenario 11 applies sin transformation to treatment variable, scenario 12 applies cosine transformation to outcome variable while Scenario 13 introduces categorical variables in the outcome.

Table 2: Equations used to generate relevant treatment and outcome variables

Scenario Number	$f(V)$ **(Treatment)**	$g(V)$ **(Outcome)**
1	$0.4V_1 + 0.3V_2 + 0.2V_3 + 0.1V_4$	0
2	$0.5V_1 + 0.5V_2 + 0.5V_3 + 0.1V_4$	$0.5V_1 + V_3 + 0.5V_4$
3	$0.1V_1 + 0.1V_2 + V_3 + V_4 + V_5$	$2V_1 + 2V_2$
4	$0.5V_1 + 0.4V_2 + 0.3V_3 + 0.2V_4 + 0.1V_5$	$0.5V_1 + V_2 + 1.5V_3 + 2V_4 + 2.5V_5$
5	$0.5V_1 + 0.5V_2 + 0.1V_3$	$V_3 + V_4 + V_5 + \sum\sum V_i V_j$
6	$V_1 + V_2 + V_5$	$\sum\sum 0.5V_i V_j$
7	$0.2V_1 + 0.2V_2 + 0.2V_5$	$0.25V_3 + (V_1 + V_2)^2 - (V_1^2 - V_3)^2 + (V_4^2 - 0.5V_5)(V_3 - 0.5V_4)$
8	$V_3 + V_4 + V_5 + \sum\sum V_i V_j$	$0.5V_1 + 0.5V_2 + 0.1V_3$
9	$(V_1 + V_2 + 0.5V_3)^2$	$0.5V_1 + 0.5V_2 + 0.1V_3$
10	$0.2V_1 - 2V_2 + V_5 - V_6 + V_7 - V_8$	$2V_1 + 0.2V_2 + 5V_3 + 0.5V_4$
11	$2V_1 + V_2 + 5V_3$	$sin(0.5V_1 + 0.5V_3 + 0.5V_4)$
12	$cos(2V_1 + V_2 + 5V_3)$	$V_1 + V_3 + V_4 - 5V_5$
13	$2V_1 + V_2 + 5V_3 + V_{categorical,1} + V_{categorical,2}$	$V_1 + V_3 + V_4 + V_5^2$

3.2 Replication Study Model Implementation

To estimate the treatment effect, $\hat{\tau}$, the below methods were implemented in R.

3.2.1 Group LASSO with Doubly Robust Estimation (GLiDeR)

GLiDeR uses Group LASSO to simultaneously select covariates for the treatment and outcome models, ensuring doubly robust estimation to improve results under potential model misspecification.

1. Perform variable selection using Group LASSO by minimizing objective function:

   $$\hat{\beta} = argmin_{\beta} \frac{1}{n}\sum_{1}^{n} \phi_{sum}(Y_{std}, A_i, V_i; \beta) + \lambda \sum_{k=1}^{p} W_k \|\beta_k\|_2,$$

   Where ϕ_{sum} is the sum of loss functions of both the outcome and treatment models, Y_{std} is standardized outcome variable and λ is the chosen tuning parameter.
2. Perform Doubly Robust Estimation

   $$\hat{\tau}_{DR} = \frac{1}{n}\sum_{1}^{n}[\frac{A_i Y_i}{\pi(X_i; \hat{\gamma})} - \frac{(1-A_i)Y_i}{1-\pi(X_i; \hat{\gamma})} - [\frac{A_i - \pi(X_i; \hat{\gamma})}{\pi(X_i; \hat{\gamma})}]\mu(1, X_i; \hat{\alpha}) - [\frac{A_i - \pi(X_i; \hat{\gamma})}{\pi(X_i; \hat{\gamma})}]\mu(0, X_i; \hat{\alpha})] \quad (1)$$

   Where $\hat{\tau}_{DR}$ is the doubly robust estimator. This combines both the outcome and propensity score model, while only requiring one of them to be correctly specified to attain a consistent estimate.

More details on the specific steps, such as groupings, weights, tuning parameter selection and algorithm for step 1, variable selection using Group LASSO can be found in appendix A.

3.2.2 Backward Selection

Backward selection was applied, removing one covariate at a time based on p-value. This was repeated until all covariates remaining had p-value below 0.05, and were statistically significant predictors of outcome. Next, the remaining set of covariates was then used in the estimation of $\hat{\tau}_{DR}$ in the same manner as for GLiDeR, using equation (1).

3.2.3 Two-Stage Model Averaged Doubly Robust (MADR)

The Two-Stage MADR is a doubly robust estimation method designed to address uncertainty in model selection (Cefalu et al., 2017). MADR aggregates estimates across multiple candidate models, providing robustness against model misspecification. MADR consists of two main stages (1) MADR constructs multiple candidate models using propensity score and the outcome (2) MADR then calculates the ATE of each combination of outcome and propensity score model. Subsequently, the individual ATE are combined into a single averaged estimate of treatment effect.

In our project, we selected the linear model and random forest models as the candidate models to not only produce an estimate of the treatment effect under linear conditions, but also under conditions where nonlinearity between predictors and the outcome variables may exist. This ensures that a more robust model is obtained with greater consideration of different types of relationships between the covariates and outcome variables.

3.2.4 Adaptive LASSO

Adaptive LASSO was also considered, implementing a weight, based on initial OLS estimates of coefficients, to the standard LASSO. In our project, we generated adaptive weights as the inverse of the absolute OLS estimates, hence assigning smaller penalties to larger coefficients, and larger penalties to smaller coefficients (Zou, 2006). This reduces bias in coefficient estimates relative to the standard LASSO method, and potentially improves the variable selection process as the model is more likely to select the correct covariates. Instead of simultaneously selecting covariates for both the treatment and outcome models, adaptive LASSO only selects covariate for the outcome model. To select the tuning parameter for this method, 10-fold cross validation is used.

3.3 Proposed Extension Models Implementation

3.3.1 Trigonometric Formulation of Outcome and Treatment DGP

We extend the DGP used in the original study by using periodic relationships. We included a *sin* term for the outcome model (Scenario 11) and a *cos* term for the treatment model (Scenario 12) to capture non-linear relationships that traditional polynomial terms might miss. This is because in empirical studies, outcomes are frequently periodic or cyclical, such as the business cycle or temperature variations. Through this, we assess the robustness of GLiDeR in handling cyclical dependencies.

3.3.2 Inclusion of qualitative covariates

The original study used only continuous covariates. However, in real-world applications, many datasets contain categorical variables, such as gender, education level, or socioeconomic status. To evaluate GLiDeR's ability to select categorical variables alongside continuous ones, we introduce categorical covariates and encode them using one-hot encoding. Through this, we aim to identify treatment effect heterogeneity. We generate two sets of multi-class categorical variables. We use 3 classes for each variable for simplicity.

3.3.3 Alternative regularization techniques like LASSO with data-driven penalty (rLASSO)

The GLiDeR model utilises Grouped LASSO as a regularizer. This method groups coefficients of the same covariate from both models, forcing covariates to be included/excluded in both models simultaneously, therefore addressing the challenge of identifying confounders and outcome predictors in high-dimensional problems. However, in the case where the causal effect is driven by individual variables rather than covariate groups, normal LASSO may outperform the Grouped LASSO (Huang et al., 2009).

With the rLASSO performing variable selection independently for treatment and outcome models, flexible confounder adjustment is allowed. Unlike Grouped LASSO, which enforces joint selection, rLASSO applies adaptive shrinkage, selecting covariates based on individual contributions rather than group presence. Its data-driven penalty adapts to model sparsity and variance, making it robust in high-dimensional settings (Belloni et al., 2012). While GLiDeR excels when covariates cluster into meaningful groups, rLASSO is preferable for individual-level selection, making it effective for heterogeneous treatment effects. Therefore, we decided to include this method on top of adaptive LASSO to better determine the performance of the GLiDeR method.

3.3.4 Dimensionality Reduction via Prescreening

In high-dimensional settings, where p >>> n, incorporating all covariates will make the GLIDER model ineffective by incorporating more noise than signal, making computational cost very expensive. To mitigate this issue, we intend to introduce prescreening, which reduces the dimensionality by filtering out covariates that exhibit little to no marginal correlation with the treatment and outcome.

One method of prescreening is the Sure Independence Screening (SIS) method. SIS ranks each covariate based on its marginal correlation with the outcome, retaining only those that exhibit strong associations (Fan & Lv, 2008). The key advantage of SIS is that it preserves relevant predictors with high probability under mild conditions, making it an effective way to perform variable selection in ultra-high-dimensional settings.

However, SIS does not account for joint effects or interactions among covariates, as it considers variables independently. To address this limitation, we will apply SIS followed by normal LASSO. The SIS step acts as an initial screening filter, reducing the number of variables to a manageable subset, after which LASSO will be applied to perform joint selection and shrinkage. This two-step approach ensures a parsimonious model while capturing important predictive signals while also balancing computational efficiency. To test if prescreening enhances GLiDeR, an additional GLiDeR-SIS model was also done.

3.4 Evaluation Metrics

This section outlines the metrics used to assess the performance of our models. The evaluation is based on Monte Carlo (MC) simulations, with 1000 MC datasets generated for each scenario. The metrics are computed as follows:

1. **Monte Carlo Bias:** The difference between the ATE across the 1000 MC simulations and the true effect. This is the systematic error in the model's estimates.
2. **Monte Carlo Standard Deviation:** The variability in the estimated effect across the 1000 MC simulations. This is the precision of the model's estimates.
3. **Mean Squared Error (MSE) Ratio:**

$$MSE\ Ratio\ =\ \frac{Saturated\ Model\ MSE}{Method\ MSE}$$

Where MSE is computed as the average squared difference between estimated ATE and true ATE across all MC simulations.

$$MSE\ =\ \frac{1}{N}\sum_{i=1}^{N}(\hat{\tau_i}-\tau)^2$$

A better model should have larger MSE ratio, smaller MC bias and smaller MC SE.

4. Results and Discussion

This section describes the overall results for the models in terms of MC bias, MSE Ratio (MSE Ratio), and MC standard deviation.

4.1 Results

MC Bias Results. Our findings given in **Figure 1** show that across 5, 10, and 25 variables used, GLiDeR consistently performed better than competing models. The highest bias across most models was observed at scenario 6, which refers to the outcome variable being fully dependent on a correlated product between variables. Any model using GLiDeR is observed to perform well and have a bias close to 0. Another key observation is MADR being observed to have consistently higher bias relative to other models other than scenario 11 for 5 variables and 10 variables, where its bias lies close to 0. These results align closely with the findings in Koch et al. (2018), where GLiDeR demonstrated low bias across all scenarios, particularly under correct model specification in scenario 1 to 4. With the addition of trigonometric functions in scenarios 11 and 12, all models performed well in scenario 12, while only SIS LASSO performed well in scenario 11. Addition of the categorical variables in scenario 13 also saw strong performances across most models where the average bias lies around 0. Only SIS LASSO and MADR struggled to predict ATE accurately for scenario 13, where they showed a positive and negative bias respectively (**Figure 1**).

Handling of nonlinear transformations (Scenario 11 and 12). As shown by **Figure 1**, SIS LASSO tends to do well for nonlinear transformations with their ability to identify independent and marginal relationships between individual features and the outcomes. On the other hand, models such as MADR and GLiDeR may struggle to capture the nonlinearity as MADR (by definition) is dependent on the model-averaged doubly robust value, while GLiDeR uses group LASSO which is largely only designed for linear relationships. The poorly performing LASSO and rLASSO are also typically designed for linear relationships, and backward selection may struggle with the identification of features as it does not consider key interactions as they solely select the features based on their p-values or AIC.

Weakness at scenario 6. Most of the models fail at scenario 6, where the outcome model has no linear structure, but rather is simply the product of two columns. This is where the group LASSO trait in GLiDeR was effective where it is able to group both columns together to select them as a single group, rather than the other models which select the variables separately to determine the ATE. Therefore, it can be observed that GLiDeR and SIS GLiDeR emerged as the best models to predict the ATE and the lowest MC bias (**Figure 1**).

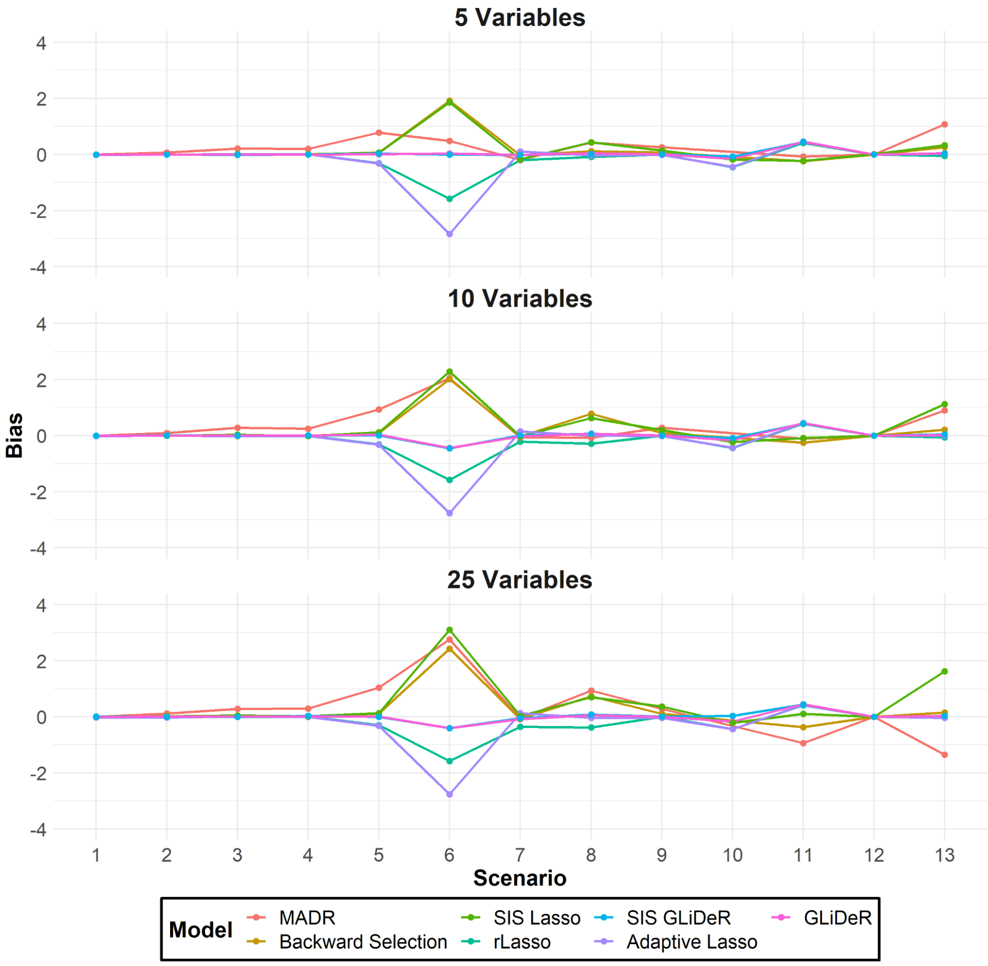

***Figure 1 : MC Bias** across models for different numbers of variables*

<u>MSE Ratio Results</u>. As explained above, the higher the MSE ratio, the greater the performance of the model tested. Models having a ratio above 1 indicates that it performs better than the saturated model. As shown in **Figure 2** below, GLiDeR consistently achieves a higher MSE ratio compared to the saturated model across all scenarios, indicating better model performance. This outperformance is especially notable in scenarios 5 to 13, under partially misspecified models where other methods, including the saturated model, perform poorly while GLiDeR remains robust. Following the incorporation of SIS, poorer performance is observed with a lighter shade relative to GLiDeR across all scenarios, and this method even results in its relatively poorer performance compared to other LASSO methods such as rLASSO, SIS LASSO, and adaptive LASSO. Backward selection and MADR is observed to also consistently perform poorly as shown by its lighter shade across all scenarios. However,

backward selection had been observed to perform equivalently to LASSO models for scenario 8, particularly when dimensionality is higher at p = 10 and p = 25.

Incorporation of SIS (All scenarios). With SiS having the ability to minimise dimensionality through the screening out of irrelevant features early in the process, this allows the estimation to be more effective, and in our project, we applied it to two methods - GLiDeR, and vanilla LASSO. Firstly, when SIS was incorporated into normal LASSO, it showed greater accuracy in the prediction of the ATE based on the data relative to GLiDeR at higher dimensionalities given its removal of more variables prior to analysis. This is relative to the GLiDeR model which applies group LASSO and may have slightly overfitted the model given that many of the variables may be weakly correlated, thereby resulting in a greater ATE estimation and hence bias. However, when SIS was incorporated into GLiDeR, the estimation of ATE appears to be the same as the GLiDeR without the additional step of SIS, with the bias of the SIS GLiDeR model following closely to the GLiDeR model, and performing either equivalently or worse than it across all scenarios.

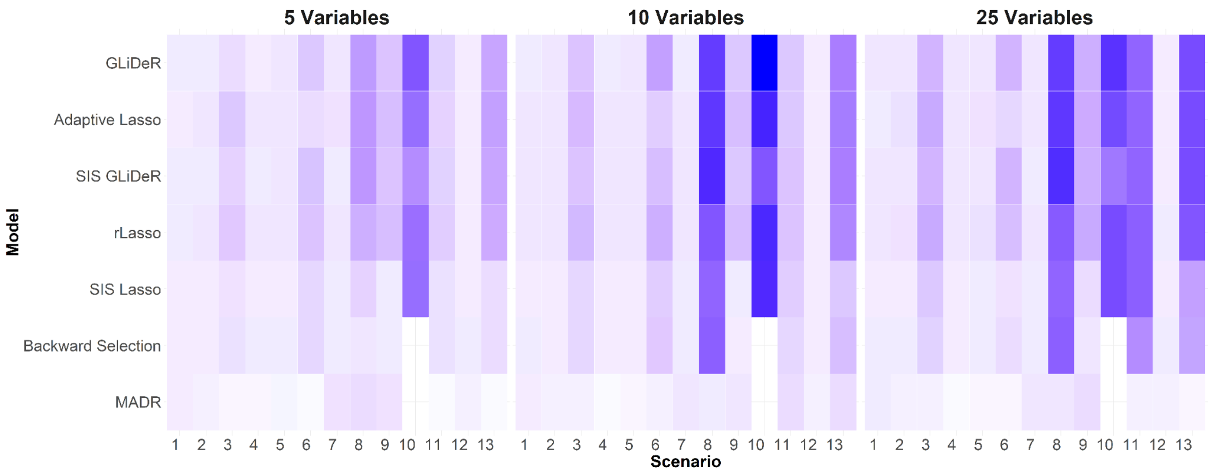

Figure 2 : Heatmap of log-transformed MSE ratios across Scenarios and Models. *Log-transformation with (log(1 + MSE ratio)) applied for clarity. Darker shades indicate higher MSE ratio and better performance. Performance is similar across 5, 10 and 25 variables.*

Poor performance of saturated models. As the saturated model includes all covariates regardless of relevance, it often suffers from variance inflation because inclusion of irrelevant or treatment-only-related variables. In contrast, GLiDeR selectively includes covariates associated with both treatment and outcome, or just the outcome, and excludes variables that can introduce unnecessary noise. From this insight, we then examine MC SD to see if other models are also similarly affected by variability as the saturated model is.

MC Standard Deviation (SD) Results. In **Figure 3**, we find that all LASSO variants and backward selection consistently attain short and centered violins, showing that they are precise and stable across different scenarios. However, MADR consistently produces tall violins, demonstrating high variability, which is observed to increase with dimensionality.

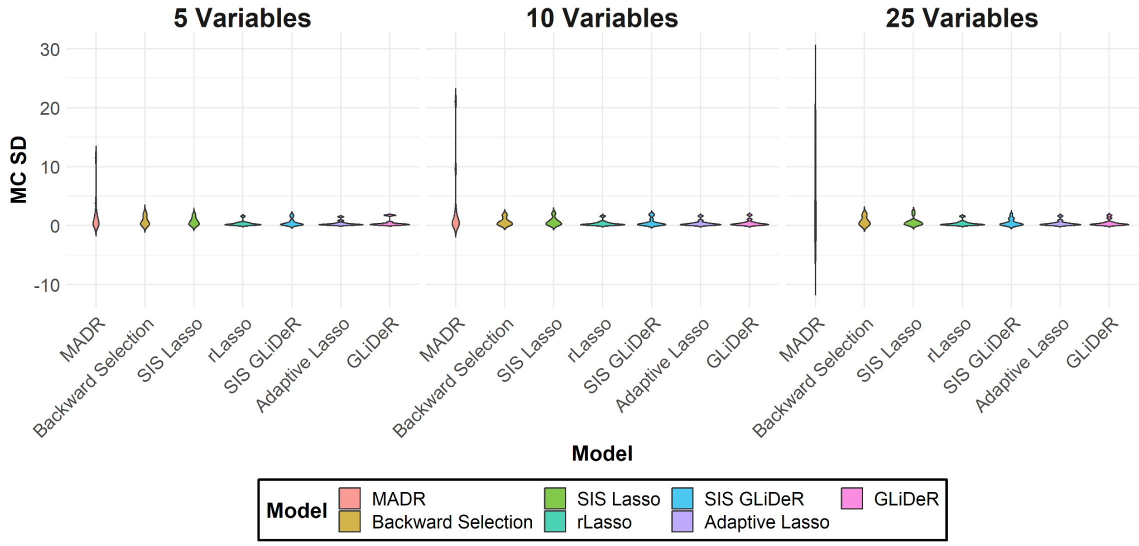

Figure 3 : Distribution of MC SD across models and variable counts. *A short and flat violin at 0 represents the least variability amongst scenarios.*

4.2 Discussion

<u>Overall comments about the GLiDeR model</u>. It is observed that the GLiDeR model performs the best in nearly all scenarios, and coupled with its use of the doubly robust formula to compute ATE, it is a method worth considering in the accurate computation of ATE. While it performed equivalently well in simpler models where both models were correctly specified, its true predictive power is seen in Scenario 11 and 12 where we introduced periodic relationships aimed at replicating seasonal or cyclic real-world patterns. Under such scenarios, we observed that GLiDeR remained robust under sine transformation in the outcome variable compared to other models which did not perform as well. As expected, cosine transformation to treatment variable did not prove to hinder the performance of other models given that the treatment variable was binary, and hence the use of the *cosine* function was not expected to hinder the process of ATE computation. Additionally, GLiDeR also excelled in when categorical covariates were incorporated in Scenario 13, since many datasets now contain factors like gender or region.

<u>Handling of categorical variables (Scenario 13)</u>. GLiDeR tends to perform very well as it is able to group categorical variables and select them for ATE computation, resulting in a higher MSE ratio relative to other models across the various dimensions. Covariate selection algorithms such as adaptive and rLASSO models also appear to perform better with their ability to shrink the coefficients, and adaptive LASSO being able to vary the penalty on correlated variables. Through the assumption of independence between covariates, SIS LASSO struggles as it is unable to pick up the correlation between the categorical variables. This leads to a positive bias being incurred in the ATE prediction. Therefore, most model selection algorithms are able to easily handle categorical variables in variable selection.

Incorporation of SIS to GLiDeR. As an intended enhancement to GLiDeR to provide further guidance towards improving the process of variable selection, the SIS stage aimed to exclude variables in advance to allow GLiDeR to better discern the key variables from the large number of variables available. However, it was observed that this enhancement did not significantly improve the results in terms of MSE bias, ratio, and standard deviation, but rather worsened the results. This could be due to overselection taking place, where wrong variables (which would have been selected by GLiDeR) were excluded from SIS, thereby resulting in underselection bias (Fan & Lv, 2008). To solve this issue, iterative SIS (ISIS) could be considered instead to recover relevant predictors initially excluded from analysis due to SIS (Saldana & Feng, 2018).

Evaluation of MADR. It can be observed that while the ATE predictions of MADR is generally comparable to that of GLiDeR under typical scenarios, it exhibits high variability over multiple scenarios given its instability in model averaging. Being a model reliant on a combination of multiple models to produce predictions of ATE, instability can occur if the models disagree significantly in different scenarios, thereby resulting in model uncertainty. Therefore, this resulted in the high variance in MADR results, particularly when the dimensionality of the dataset is high.

Evaluation of backward selection. Similar to MADR, backward selection's ATE predictions are similar to GLiDeR, having low bias across most scenarios, but higher variance in high-dimensional datasets. This is likely due to its reliance on iterative model refinement, which may introduce instability if the selected variables differ significantly across scenarios. While MADR combines multiple models to produce ATE predictions, backward selection focuses on refining a single model iteratively. This distinction explains why MADR might exhibit higher variability due to disagreements among models but potentially lower bias due to averaging effects. Backward selection's MSE ratio is competitive, with lower ratios compared to other models, suggesting less efficient performance across varying dimensions.

Evaluation of LASSO variant models. Through our extension, we found that adaptive LASSO and rLASSO also display decent performance in both variability and bias, while SIS LASSO shows higher variability. The MSE ratio for rLASSO is consistently high and most competitive among these models, whereas adaptive LASSO's MSE ratio decreases significantly in high-dimensional settings. When there are many variables, the process of estimating the adaptive weights in adaptive LASSO can become unstable, particularly if some variables are highly correlated, therefore resulting in poor variable selections and subsequently poorer performance due to incorrect assignment of weights to variables. This is compared to the data-driven rLASSO which is able to automatically adapt to the sparsity and noise levels in the data which reduces overfitting. For SIS LASSO, its reliance on marginal correlation could result in unstable variable selection in cases where variables with weaker marginal correlations are actually important when considered jointly.

5. Conclusion and Further work

This paper did a comprehensive evaluation on the model proposed by Koch et al. (2018), with findings being closely aligned to the original paper by Koch et al. (2018). GLiDeR stood out amongst the other methods in both the replication and extension study. It is able to balance variable relevance across both treatment and outcomes, allowing it to outperform our competing models across most metrics.

This paper evaluates the models using 13 simulated examples, ensuring that the models are sufficiently robust when handling different types of data. However, it is noted that our resources were unable to evaluate models with variables more than 25 due to the lack of computational power. Therefore, further work can be done to analyse the performances of the various models at higher dimensionalities. Apart from that, alternative regularization methods such as Ridge and Elastic Net regression can also be explored as benchmarks for GLiDeR to justify the use of group LASSO in the process of variable selection.

*Project was done with my mates, Tze Wyne, Wilfred Woo and Young Zhan Heng. Contact for more details.

6. Appendix

Appendix A: Coding Files Directory

The github link for the project and the respective files can be found here. The data files are also included in the zip file alongside the report.

Table A-1 describes the different coding folders and files in the github.

Table A-1: Description of coding files

Type of File	Folder / File Name	Description
Folder	*old_plots/*	Old plots
	individual_model_results/	Results for individual models
	plots/	Plots
	processed_results/	Processed results
File	*data_sim_fn.R*	Generation of simulated data scenarios
	glider_fn.R	GLiDeR functions
	models_fn.R	All functions for models, and also includes the code to process the results to obtain MC ratios
	visualisations.Rmd	Code to generate plots

To Note:

Naming Convention of Files. For a given file *overall_5var_scene10_50_simulation.rdata*, it is to note that Scenario 10 has a different number of predictors from the other scenarios, hence there is a separate number specifying the number of predictors used in Scenario 10. *scene10_50* indicates that the number of predictors in scenario 10 is 50. *5var* indicates that there are only 5 predictors for all scenarios (other than Scenario 10).

Method of saving files. Given that each element has its own dataframe, we saved the dataframes as a list for easy access by the '$' command to call the respective model's data. The MC ratio list object is of the same convention as the file mentioned in the previous point, except it is called *mcratio*. The components refer to the specific results including bias, spread, and MC ratio. Files that start with "*overall_*" refer to the raw combined simulation results from the 1000 MC simulations for the models, and are list objects.

MADR and Backward Selection Models. MADR and backward selection has only 12 scenarios (missing scenario 10) instead of 13 given that Koch et al. (2018) did not do it as well

Appendix B: GLiDeR Procedure in depth

1. Define covariate groups and weights
    a. Group features: Standardize all covariates and form groups where each group contains coefficients for covariate in both the outcome and treatment models. Additionally, define an intercept term group for both the outcome and treatment model, as well as a treatment effect group. These are defined separately to ensure that they are not penalized through the group weights.
    b. Compute adaptive group weights: For each covariate group, calculate weight as $W_k = \frac{\sqrt{2}}{|v_k|}$, where numerator is the default group penalty weight given by a general group LASSO formulation and denominator is the absolute value of our regression coefficient estimate from a full model (eg. OLS). For the intercept groups and treatment effect group, the weights are set to zero, meaning they are not penalized. If covariate transformations are included, numerator is square root of corresponding predictors and denominator is the absolute value of the l2 norm of initial coefficient estimates. This encourages outcome-related (rather than treatment) covariates since the denominator is the outcome from a full model. If a covariate has strong association with outcome, denominator will be large and penalty across the group will be small.
2. Apply modified group LASSO
    a. Typical Group LASSO is modified to group together coefficients corresponding to the same covariate across both the outcome model and the propensity score model. This forces covariates to be either included or excluded in both the models simultaneously.
    b. Loss function is taken to be the sum of loss functions for both outcome model and treatment model.
    c. Define tuning parameters to control degree of shrinkage by performing Groupwise Majorization Descent (GMD) as described by Yang and Zou (2015). Next apply generalized cross validation (GCV) on the outcome model for all tuning parameters using the below GCV statistic and select optimal tuning parameter.

    $$\mathrm{GCV}(\lambda) = \frac{\sum_{i=1}^{n} (Y_i - \hat{\alpha}_1(\lambda)V_{1i} - \cdots - \hat{\alpha}_p(\lambda)V_{pi} - \hat{\alpha}_{p+1}(\lambda) - \hat{\alpha}_{p+2}(\lambda)A_i)^2}{\left(1 - \left(2 + \sum_{j=1}^{K} I\left(||\hat{\alpha}_j(\lambda)|| > 0\right) + \sum_{j=1}^{K} \frac{|\hat{\alpha}_j(\lambda)|}{|v_j|}\right)/n\right)^2}$$

    d. Plug the estimated coefficients and corresponding optimal tuning parameter into both the outcome and propensity score models. Then use models in doubly robust estimator formula to attain τ.

7. References

Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(5), 849–911. https://doi.org/10.1111/j.1467-9868.2008.00674.x

Huang, J., Ma, S., Xie, H., & Zhang, C.-H. (2009). A group bridge approach for Variable selection. *Biometrika*, *96*(2), 339–355. https://doi.org/10.1093/biomet/asp020

Koch, B., Vock, D. M., & Wolfson, J. (2018). Covariate Selection with Group LASSO and Doubly Robust Estimation of Causal Effects. *Biometrics*, *74*(1), 8–17. https://doi.org/10.1111/biom.12736

Tibshirani, R. (1996). *Regression Shrinkage and Selection via the LASSO on JSTOR*. (n.d.). Retrieved 19 March 2025, from https://www.jstor.org/stable/2346178

SPSS. (2024, January 18). Backward Regression in SPSS - Explained, Performing, Reported. *Statistical Analysis Services For Academic Researches*. https://spssanalysis.com/backward-regression-in-spss/

Zou, H. (2006). The Adaptive LASSO and Its Oracle Properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. https://doi.org/10.1198/016214506000000735

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x