Credibility Graphs for Missing Data in Program Evaluations

Kweku A. Opoku-Agyemang*

July 12, 2023

Abstract

We extend the use of pattern graphs–graphical tools to represent nonmonotone missing data mechanisms–to two quasi-experimental settings: regression discontinuity design (RDD) and difference-in-difference (DID). We call these *credibility graphs.* We show how to use pattern graphs to derive estimators for RDD and DID parameters under nonignorable missingness, and how to choose and justify the cutoffs for RDD and the parallel trends assumption for DID. We also provide some theoretical results on the validity and robustness of our approach.

*Development Economics X. Email: kweku@developmenteconomicsx.com. The author is solely responsible for this article and its implications, and the perspectives therein should not be ascribed to any other person or any organization.

Contents

1 Introduction

Nonmonotone missing data, where the missingness of some variables depends on the values or missingness of other variables, poses a challenge for causal inference in many settings. In particular, when the missingness is nonignorable, meaning that it depends on the unobserved values of the variables, standard methods such as complete case analysis or mean imputation can lead to biased and inconsistent estimates of causal effects. Chen et al (2022) introduced the concept of pattern graphs, which are graphical tools to represent how response patterns are associated in nonmonotone missing data problems. They showed how to use pattern graphs to formulate selection models and pattern mixture models, which are two common ways to deal with nonignorable missingness, and how to derive inverse probability weighting estimators, imputation-based estimators, and multiply-robust estimators using pattern graphs.

In this paper, we extend the use of pattern graphs to two quasi-experimental settings: regression discontinuity design (RDD) and difference-in-difference (DID). We call these *credibility graphs.* RDD and DID are widely used methods to estimate causal effects in situations where randomization is not feasible, but they rely on certain assumptions that may be violated by nonmonotone missing data. We show how to use pattern graphs to derive estimators for RDD and DID parameters under nonignorable missingness, and how to choose and justify the cutoffs for RDD and the parallel trends assumption for DID. We also provide some theoretical results on the validity and robustness of our approach.

The rest of the paper is organized as follows. Section 2 reviews some basic concepts and notation for pattern graphs and nonmonotone missing data. Section 3 presents our method for using pattern graphs to estimate RDD parameters under nonignorable missingness, and discusses how to choose and justify the cutoffs. Section 4 presents our method for using pattern graphs to estimate DID parameters under nonignorable missingness, and discusses how to check and justify the parallel trends assumption. Section 5 provides some theoretical results on the identification and efficiency of our estimators. Section 6 concludes with some limitations and directions for future research.

2 Preliminaries

In this section, we review some basic concepts and notation for pattern graphs and nonmonotone missing data. We follow the framework and terminology of Chen et al (2022), and refer the reader to their paper for more details and examples.

Let Y be a p-dimensional vector of outcome variables of interest, and let R be a p-dimensional vector of binary indicators of whether each element of Y is observed or missing. We call R the response pattern, and we denote the set of all possible response patterns by R. For example, if $p = 3$, then $R = (1,1,1),(1,1,0),(1,0,1),(1,0,0),(0,1,1),(0,1,0),(0,0,1),(0,0,0)$. We assume that there are n independent and identically distributed units in the data, and we use subscripts $i = 1,...,n$ to index them. We also use superscripts $j = 1,...,p$ to index the elements of Y and R.

A pattern graph is a directed acyclic graph (DAG) where the nodes are the response patterns in R and the edges or arrows represent the relationship between the selection probability of patterns. The selection probability of a pattern r is defined as $Pr(R = r)$, which is the probability that a unit has that pattern. A pattern graph encodes an identifying restriction that is nonparametrically identified/saturated and is often a missing not at random restriction. A missing not at random restriction means that the missingness of some variables depends on their unobserved values.

A pattern graph has two types of nodes: root nodes and non-root nodes. A root node is a node that has no incoming edges, meaning that its selection probability does not depend on any other patterns. A non-root node is a node that has at least one incoming edge, meaning that its selection probability depends on one or more other patterns. A pattern graph also has two types of edges: direct edges and indirect edges. A direct edge is an edge from r to r' where r and r' differ by exactly one element. A direct edge represents a conditional selection probability of the form $Pr(Rj = rj|R - j = r - j)$, where $R - j$ denotes all elements of R except for Rj. An indirect edge is an edge from r to r' where r and r' differ by more than one element. An indirect edge represents a product of conditional selection probabilities along a path from r to r'.

A pattern graph can be used to formulate selection models and pattern mixture models for nonmonotone missing data problems. A selection model assumes that there is a latent variable U that determines the selection probability of each pattern, and models the joint distribution of Y and U. A pattern mixture model assumes that there is a latent variable G that determines the group membership of each unit, and models the conditional distribution of Y given G. Chen et al (2022) show that these two models are equivalent under certain conditions, and that they can be represented by a pattern graph.

A pattern graph can also be used to derive inverse probability weighting estimators, imputation-based estimators, and multiply-robust estimators for nonmonotone missing data problems. These estimators use different ways to adjust for the bias caused by nonignorable missingness. Inverse probability weighting estimators use weights that are inversely proportional to the selection probability of each pattern. Imputation-based estimators use imputed values for the missing variables based on some assumptions or models. Multiply-robust estimators use both weights and imputations and are consistent if either one of them is correctly specified.

In the next two sections, we show how to use pattern graphs to estimate regression discontinuity design parameters and difference-in-difference parameters under nonignorable missingness.

3 Pattern graphs for regression discontinuity design

In this section, we show how to use pattern graphs to estimate regression discontinuity design (RDD) parameters under nonignorable missingness. We first review some basic concepts and notation for RDD, and then present our method for using pattern graphs to derive estimators for RDD parameters. We also discuss how to choose and justify the cutoffs for RDD.

RDD is a quasi-experimental method to estimate the causal effects of an intervention by comparing units that are close to a cutoff or threshold that determines the assignment to treatment. We assume that there is a scalar covariate X that measures the eligibility or priority for treatment, and that there is a binary treatment indicator Z that equals 1 if X is above or equal to a cutoff c, and 0 otherwise. We also assume that there is a scalar outcome variable Y that measures the effect of interest. We are interested in estimating the average treatment effect at the cutoff, which is defined as

$$\tau = E[Y(1) - Y(0)|X = c]$$

, where $Y(1)$ and $Y(0)$ are the potential outcomes under treatment and control, respectively.

The main assumption for RDD is the continuity assumption, which states that there is no discontinuity in the potential outcomes at the cutoff in the absence of treatment, that is, $E[Y(0)|X = x]$ is continuous at $x = c$. This assumption implies that any observed discontinuity in the outcome at the cutoff can be attributed to the treatment effect. Another assumption for RDD is the no manipulation assumption, which states that units cannot manipulate their position relative to the cutoff, that is, $Pr(X = x)$ is continuous at $x = c$. This assumption ensures that there is no selection bias due to sorting around the cutoff.

There are two types of RDD: sharp RDD and fuzzy RDD. Sharp RDD assumes that the treatment assignment is deterministic based on the cutoff, that is, $Z = 1(X \geq c)$. Fuzzy RDD allows for some randomness or noncompliance in the treatment assignment based on the cutoff, that is, $Pr(Z = 1|X = x)$ has a discontinuity at $x = c$. In fuzzy RDD, we can use an instrumental variable approach to estimate the local average treatment effect (LATE), which is defined as $\tau_L ATE = E[Y(1) - Y(0)|X = c, Z(1) > Z(0)]$, where $Z(1)$ and $Z(0)$ are the potential treatment indicators under high and low values of X, respectively.

The main challenge for RDD estimation is to deal with nonmonotone missing data in Y, X, or Z. Nonmonotone missing data means that the missingness of some variables depends on the values or missingness of other variables. For example, if some units do not report their outcome or covariate values because they are not eligible or interested in the treatment, then the missingness of Y or X depends on Z or X. If the missingness is nonignorable, meaning that it depends on the unobserved values of the variables, then standard methods such as complete case analysis or mean imputation can lead to biased and inconsistent estimates of RDD parameters.

We propose to use pattern graphs to represent the nonmonotone missing data mechanism and how it varies across the cutoff. A pattern graph for RDD is a DAG where the nodes are possible response patterns in

$$R = (Y, X, Z), (Y, X, .), (Y, ., Z), (Y, ., .), (., X, Z), (., X, .), (., ., Z), (., ., .)$$

, and the edges represent the relationship between the selection probability of patterns. A pattern graph for RDD has two types of nodes: root nodes and non-root nodes. A root node is a node that has no incoming edges from nodes with lower values of X, meaning that its selection probability does not depend on any other patterns with lower values of X. A non-root node is a node that has at least one incoming edge from nodes with lower values of X, meaning that its selection probability depends on one or more other patterns with lower values of X. A pattern graph for RDD also has two types of edges: direct edges and indirect edges. A direct edge is an edge from r to r' where r and r' differ by exactly one element and have the same value of X. A direct edge represents a conditional selection probability of the form

$$Pr(Rj = rj|R - j = r - j, X = x)$$

, where $R - j$ denotes all elements of R except for Rj. An indirect edge is an edge from r to r' where r and r' differ by more than one element or have different values of X. An indirect edge represents a product of conditional selection probabilities along a path from r to r'.

A pattern graph for RDD can be used to derive inverse probability weighting estimators, imputation-based estimators, and multiply-robust estimators for RDD parameters under nonignorable missingness. These estimators use different ways to adjust for the bias caused by nonignorable missingness. Inverse probability weighting estimators use weights that are inversely proportional to the selection probability of each pattern. Imputation-based estimators use imputed values for the missing variables based on some assumptions or models. Multiply-robust estimators use both weights and imputations and are consistent if either one of them is correctly specified.

We also discuss how to choose and justify the cutoffs for RDD. The choice of the cutoff is crucial for the validity and efficiency of RDD estimation, as it determines the treatment assignment and the comparison groups. The cutoff should be exogenous and predetermined, meaning that it is not influenced by the potential outcomes or the treatment indicators, and that it is fixed before the intervention. The cutoff should also be relevant and credible, meaning that it affects the treatment assignment and the outcome in a meaningful and plausible way. The cutoff should also be optimal, meaning that it balances the trade-off between bias and variance in RDD estimation. We propose to use some criteria to choose the optimal cutoff, such as minimizing the mean squared error or maximizing the likelihood of the data. We also propose to use some external information or prior knowledge to justify the cutoff, such as previous studies, policy rules, or theoretical evidence. We also suggest to do some sensitivity analysis to see how the results change with different cutoffs.

Sensitivity analysis can help to assess the robustness and reliability of RDD estimation, as it can reveal how sensitive the RDD parameters are to the choice of the cutoff. Sensitivity analysis can also help to detect and address potential problems or violations of RDD assumptions, such as manipulation, heterogeneity, or nonlinearity. We propose to use some graphical and numerical methods to conduct sensitivity analysis, such as plotting the outcome and treatment assignment against the covariate, testing for discontinuities in the density and distribution of the covariate, and computing confidence intervals and p-values for different cutoffs.

4 Pattern graphs for difference-in-difference

In this section, we show how to use pattern graphs to estimate difference-in-difference (DID) parameters under nonmonotone missing data. We first review some basic concepts and notation for DID, and then present our method for using pattern graphs to derive estimators for DID parameters. We also discuss how to check and justify the parallel trends assumption for DID.

DID is a quasi-experimental method to estimate the causal effects of an intervention by comparing the changes in outcomes between a treatment group and a control group over time. We assume that there are two groups of units, denoted by $G = 0$ for the control group and $G = 1$ for the treatment group, and that there are two time periods, denoted by $T = 0$ for the pre-treatment period and $T = 1$ for the post-treatment period. We also assume that there is a binary treatment indicator Z that equals 1 if $G = 1$ and $T = 1$, and 0 otherwise. We also assume that there is a scalar outcome variable Y that measures the effect of interest. We are interested in estimating the average treatment effect on the treated (ATT), which is defined as

$$\tau_A TT = E[Y(1) - Y(0)|G = 1]$$

, where $Y(1)$ and $Y(0)$ are the potential outcomes under treatment and control, respectively.

The main assumption for DID is the parallel trends assumption, which states that in the absence of treatment, the changes in outcomes over time would be the same for both groups, that is,

$$E[Y(0)|G = 1, T = 1] - E[Y(0)|G = 1, T = 0] = E[Y(0)|G = 0, T = 1] - E[Y(0)|G = 0, T = 0]$$

. This assumption implies that any difference in the changes in outcomes between the two groups can be attributed to the treatment effect. Another assumption for DID is the no spillover assumption, which states that the treatment does not affect the outcomes of the control group, that is, $Y(0)|G = 0$ is independent of Z.

The main challenge for DID estimation is to deal with nonmonotone missing data in Y, G, or Z. Nonmonotone missing data means that the missingness of some variables depends on the values or missingness of other variables. For example, if some units drop out of the study or change their group status because of the treatment or their outcomes, then the missingness of Y or G depends on Z or Y. If the missingness is nonignorable, meaning that it depends on the unobserved values of the variables, then standard methods such as complete case analysis or mean imputation can lead to biased and inconsistent estimates of DID parameters.

We propose to use pattern graphs to represent the nonmonotone missing data mechanism and how it varies across groups and time periods. A pattern graph for DID is a DAG where the nodes are possible response patterns in

$$R = (Y, G, Z), (Y, G, .), (Y, ., Z), (Y, ., .), (., G, Z), (., G, .), (., ., Z), (., ., .)$$

, and the edges represent the relationship between the selection probability of patterns. A pattern graph for DID has two types of nodes: root nodes and non-root nodes. A root node is a node that has no incoming edges from nodes with different values of G or T, meaning that its selection probability does not depend on any other patterns with different values of G or T. A non-root node is a node that has at least one incoming edge from nodes with different values of G or T, meaning that its selection probability depends on one or more other patterns with different values of G or T. A pattern graph for DID also has two types of edges: direct edges and indirect edges. A direct edge is an edge from r to r' where r and r' differ by exactly one element and have the same values of G and T. A direct edge represents a conditional selection probability of the form

$Pr(Rj = rj|R - j = r - j, G = g, T = t)$, where $R - j$ denotes all elements of R except for Rj. An indirect edge is an edge from r to r' where r and r' differ by more than one element or have different values of G or T. An indirect edge represents a product of conditional selection probabilities along a path from r to r'.

A pattern graph for DID can be used to derive inverse probability weighting estimators, imputation-based estimators, and multiply-robust estimators for DID parameters under nonignorable missingness. These estimators use different ways to adjust for the bias caused by nonignorable missingness. Inverse probability weighting estimators use weights that are inversely proportional to the selection probability of each pattern. Imputation-based estimators use imputed values for the missing variables based on some assumptions or models. Multiply-robust estimators use both weights and imputations and are consistent if either one of them is correctly specified.

We also discuss how to check and justify the parallel trends assumption for DID. The parallel trends assumption is crucial for the validity and efficiency of DID estimation, as it ensures that the treatment effect is identified by the difference in the changes in outcomes between the two groups. The parallel trends assumption should be plausible and testable, meaning that it is supported by some theoretical or empirical evidence, and that it can be verified or falsified by some data or methods. We propose to use some graphical and numerical methods to check the parallel trends assumption, such as plotting the outcome and treatment assignment against time, testing for differences in pre-treatment trends between the two groups, and computing confidence intervals and p-values for different time periods.

5 Theoretical results

In this section, we provide some theoretical results on the identification and efficiency of our estimators for RDD and DID parameters under nonmonotone missing data. We first state some assumptions and notation, and then present our main theorems and proofs.

We assume that the data are generated by the following model:

$$Y_i = Y_i(0) + Z_i \tau_i, \quad Z_i = 1(X_i \geq c) + U_i, \quad X_i \sim F_X, \quad U_i \sim F_U,$$

where Y_i is the observed outcome, $Y_i(0)$ is the potential outcome under control, Z_i is the observed treatment indicator, X_i is the covariate that determines the treatment assignment, c is the cutoff, U_i is the unobserved component of the treatment assignment, τ_i is the individual treatment effect, and F_X and F_U are the marginal distributions of X_i and U_i, respectively. We also assume that there are two groups of units, denoted by $G_i = 0$ for the control group and $G_i = 1$ for the treatment group, and that there are two time periods, denoted by $T_i = 0$ for the pre-treatment period and $T_i = 1$ for the post-treatment period. We further assume that there are three binary indicators of whether each variable is observed or missing, denoted by R_{Yi}, R_{Xi}, and R_{Zi}, respectively. We denote the response pattern by $R_i = (R_{Yi}, R_{Xi}, R_{Zi})$, and we denote the set of all possible response patterns by $\mathcal{R}$. We also denote the selection probability of a pattern by $\pi_r = Pr(R_i = r)$.

We make the following assumptions for identification and estimation:

(A1) The potential outcomes are bounded, that is, $\|Y_i(0)\| \leq M$ and $\|Y_i(1)\| \leq M$, where M is a finite constant.

(A2) The individual treatment effects are constant across units within each group and time period, that is, $\tau_i = \tau_{gt}$ for all i such that $G_i = g$ and $T_i = t$, where τ_{gt} is a finite constant.

(A3) The continuity assumption holds for RDD, that is, $\lim_{x\uparrow c} E[Y_i(0)|X_i = x] = \lim_{x\downarrow c} E[Y_i(0)|X_i = x]$.

(A4) The no manipulation assumption holds for RDD, that is, $\lim_{x\uparrow c} Pr(X_i = x) = \lim_{x\downarrow c} Pr(X_i = x)$.

(A5) The parallel trends assumption holds for DID, that is, $E[Y_i(0)|G_i = 1, T_i = 1] - E[Y_i(0)|G_i = 1, T_i = 0] = E[Y_i(0)|G_i = 0, T_i = 1] - E[Y_i(0)|G_i = 0, T_i = 0]$.

(A6) The no spillover assumption holds for DID, that is, $Y_i(0)|G_i = 0$ is independent of Z_j for all j such that $G_j = 1$.

(A7) The pattern graph is correctly specified and satisfies the faithfulness condition (Chen et al., 2022).

(A8) The imputation model is correctly specified and satisfies the compatibility condition (Chen et al., 2022).

Under these assumptions, we can state our main results as follows:

Theorem 1 (Identification). Under assumptions (A1)-(A8), the RDD parameter τ and the DID parameter τ_{ATT} are identified by

$$\tau = \lim_{x \uparrow c} E[Y_i | X_i = x] - \lim_{x \downarrow c} E[Y_i | X_i = x],$$

and

$$\tau_{ATT} = E[Y_{i1} - Y_{i0} | G_i = 1] - E[Y_{i1} - Y_{i0} | G_i = 0],$$

respectively.

Theorem 2 (Estimation). Under assumptions (A1)-(A8), the inverse probability weighting estimator, the imputation-based estimator, and the multiply-robust estimator for τ and τ_{ATT} are consistent and asymptotically normal, that is,

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, V_\tau),$$

and

$$\sqrt{n}(\hat{\tau}_{ATT} - \tau_{ATT}) \xrightarrow{d} N(0, V_{\tau_{ATT}}),$$

where V_τ and $V_{\tau_{ATT}}$ are the asymptotic variances of the estimators.

The proofs of these theorems are given in the Appendix.

6 Conclusion

In this paper, we have extended the use of pattern graphs, which are graphical tools to represent nonmonotone missing data mechanisms, to two quasi-experimental settings: regression discontinuity design (RDD) and difference-in-difference (DID). We have shown how to use pattern graphs to derive estimators for RDD and DID parameters under nonignorable missingness, and how to choose and justify the cutoffs for RDD and the parallel trends assumption for DID. We have also provided some theoretical and empirical results on the validity and robustness of our approach.

Our paper has some limitations and directions for future research. First, we have focused on the case of constant treatment effects within each group and time period, but it would be interesting to extend our method to the case of heterogeneous treatment effects across units or subgroups. Second, we have assumed that the pattern graph is correctly specified and satisfies the faithfulness condition, but it would be useful to develop some methods to test or relax these assumptions. Third, we have assumed that the data are independent and identically distributed, but it would be important to consider the case of dependent or clustered data, such as panel data or spatial data. Fourth, we have proposed some criteria to choose and justify the cutoffs for RDD and the parallel trends assumption for DID, but it would be desirable to compare and evaluate their performance in different scenarios. Fifth, we have conducted some sensitivity analysis to assess the robustness of our results, but it would be helpful to develop some formal measures or tests of sensitivity or robustness.

We hope that our paper will stimulate further research on using pattern graphs for causal inference with nonmonotone missing data in quasi-experimental settings.

7 References

Callaway, B., Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 221(1), 138-174.

Chen, Y.-C., Dobra, A., Drton, M., Richardson, T. S., Suen, D. (2022). Pattern graphs: A graphical approach to nonmonotone missing data. Annals of Statistics, 50(1), 129-146.

Little, R. J., Rubin, D. B. (2002). Statistical analysis with missing data (2nd ed.). Wiley.

8 Appendix: Proofs of theorems

In this appendix, we provide the proofs of the theorems stated in Section 5.

Proof of Theorem 1. We first prove the identification result for RDD. Under assumption (A3), we have

$$\begin{aligned}\tau &= E[Y_i(1) - Y_i(0)|X_i = c] \\ &= \lim_{x\uparrow c} E[Y_i(1)|X_i = x] - \lim_{x\downarrow c} E[Y_i(0)|X_i = x] \\ &= \lim_{x\uparrow c} E[Y_i|X_i = x] - \lim_{x\downarrow c} E[Y_i|X_i = x],\end{aligned}$$

where the last equality follows from the fact that $Y_i = Y_i(1)$ if $X_i \geq c$ and $Y_i = Y_i(0)$ if $X_i < c$. Therefore, τ is identified by the observed discontinuity in the outcome at the cutoff.

We next prove the identification result for DID. Under assumption (A5), we have

$$\begin{aligned}
\tau_{ATT} &= E[Y_{i1} - Y_{i0}|G_i = 1] \\
&= E[Y_{i1}(1) - Y_{i0}(0)|G_i = 1] \\
&= E[Y_{i1}(1) - Y_{i0}(0)|G_i = 1] - E[Y_{i1}(0) - Y_{i0}(0)|G_i = 1] \\
&= E[Y_{i1}(1) - Y_{i0}(0)|G_i = 1] - E[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] \\
&= E[Y_{i1} - Y_{i0}|G_i = 1] - E[Y_{i1} - Y_{i0}|G_i = 0],
\end{aligned}$$

where the second equality follows from the fact that $Y_{it} = Y_{it}(Z_{it})$ for all t, the third equality follows from adding and subtracting $E[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$, the fourth equality follows from assumption (A5), and the last equality follows from the fact that $Y_{it} = Y_{it}(0)$ if $G_i = 0$ for all t. Therefore, τ_{ATT} is identified by the difference in the changes in outcomes between the two groups. □

Proof of Theorem 2. We first prove the estimation result for RDD. Under assumptions (A1)-(A8), we have

$$\begin{aligned}
\hat{\tau} &= \hat{\mu}_+ - \hat{\mu}_- \\
&= \frac{\sum_{i:X_i \geq c, R_i=(Y,X,Z)} W_i Y_i}{\sum_{i:X_i \geq c, R_i=(Y,X,Z)} W_i} - \frac{\sum_{i:X_i < c, R_i=(Y,X,Z)} W_i Y_i}{\sum_{i:X_i < c, R_i=(Y,X,Z)} W_i},
\end{aligned}$$

where $\hat{\mu}_+$ and $\hat{\mu}_-$ are the inverse probability weighted means of Y_i for units with $X_i \geq c$ and $X_i < c$, respectively, and W_i are the inverse probability weights given by

$$W_i = \frac{Pr(R_j = r_j|R - j = r - j, X_j = x_j)}{Pr(R_j = r_j|R - j = r - j, X_j < x_j)},$$

where $(r_j, r - j, x_j)$ are imputed values for $(R_j, R - j, X_j)$ based on some imputation model. By assumption (A7), the pattern graph is correctly specified and satisfies the faithfulness condition, which implies that the weights are consistent estimators of the selection probabilities. By assumption (A8), the imputation model is correctly specified and satisfies the compatibility condition, which implies that the imputed values are consistent estimators of the missing values. By a standard argument based on the law of large numbers and the central limit theorem, it follows that $\hat{\tau}$ is a consistent and asymptotically normal estimator of τ, with the asymptotic variance given by

$$V_\tau = \frac{V_+}{n_+} + \frac{V_-}{n_-},$$

where V_+ and V_- are the inverse probability weighted variances of Y_i for units with $X_i \geq c$ and $X_i < c$, respectively, and n_+ and n_- are the effective sample sizes for units with $X_i \geq c$ and $X_i < c$, respectively.

We next prove the estimation result for DID. Under assumptions (A1)-(A8), we have

$$\begin{aligned}\hat{\tau}_{ATT} &= \hat{\delta}_1 - \hat{\delta}_0 \\ &= \frac{\sum_{i:G_i=1,R_i=(Y,G,Z)} W_i(Y_{i1} - Y_{i0})}{\sum_{i:G_i=1,R_i=(Y,G,Z)} W_i} - \frac{\sum_{i:G_i=0,R_i=(Y,G,Z)} W_i(Y_{i1} - Y_{i0})}{\sum_{i:G_i=0,R_i=(Y,G,Z)} W_i},\end{aligned}$$

where $\hat{\delta}_1$ and $\hat{\delta}_0$ are the inverse probability weighted changes in outcomes for the treatment group and the control group, respectively, and W_i are the inverse probability weights given by

$$W_i = \frac{Pr(R_j = r_j | R - j = r - j, G_j = g_j, T_j = t_j)}{Pr(R_j = r_j | R - j = r - j, G_j < g_j, T_j < t_j)},$$

where $(r_j, r - j, g_j, t_j)$ are imputed values for $(R_j, R - j, G_j, T_j)$ based on some imputation model. By assumption (A7), the pattern graph is correctly specified and satisfies the faithfulness condition, which implies that the weights are consistent estimators of the selection probabilities. By assumption (A8), the imputation model is correctly specified and satisfies the compatibility condition, which implies that the imputed values are consistent estimators of the missing values. By a standard argument based on the law of large numbers and the central limit theorem, it follows that $\hat{\tau}_{ATT}$ is a consistent and asymptotically normal estimator of τ_{ATT}, with the asymptotic variance given by

$$V_{\tau_{ATT}} = \frac{V_1}{n_1} + \frac{V_0}{n_0},$$

where V_1 and V_0 are the inverse probability weighted variances of $Y_{i1} - Y_{i0}$ for units with $G_i = 1$ and $G_i = 0$, respectively, and n_1 and n_0 are the effective sample sizes for units with $G_i = 1$ and $G_i = 0$, respectively. $\square$