Spillovers of Program Benefits with Mismeasured Networks

Lina Zhang

Job Market Paper
(Click here for the latest version)

September 22, 2020

Abstract

In studies of program evaluation under network interference, correctly measuring spillovers of the intervention is crucial for making appropriate policy recommendations. However, increasing empirical evidence has shown that network links are often measured with errors. This paper explores the identification and estimation of treatment and spillover effects when the network is mismeasured. I propose a novel method to nonparametrically point-identify the treatment and spillover effects, when two network observations are available. The method can deal with a large network with missing or misreported links and possesses several attractive features: (i) it allows heterogeneous treatment and spillover effects; (ii) it does not rely on modelling network formation or its misclassification probabilities; and (iii) it accommodates samples that are correlated in overlapping ways. A semiparametric estimation approach is proposed, and the analysis is applied to study the spillover effects of an insurance information program on the insurance adoption decisions.

Keywords: Causal Inference; Spillover; Heterogeneity; Random Experiments; Mismeasured Networks.

¹Department of Econometrics and Business Statistics, Monash University (lina.zhang@monash.edu). I am grateful to David Frazier, Donald Poskitt and Xueyan Zhao for their invaluable guidance and support. I would like to thank Jiti Gao, Didier Nibbering, Bing Peng, Denni Tommasi, Takuya Ura and Benjamin Wong for helpful comments. All errors are mine.
1 Introduction

Measuring correctly the spillovers of a program intervention is incredibly relevant to understand whether and how the intervention influences individuals’ outcome through their social interactions, and provide meaningful policy advices aiming at effective treatment allocation (Angelucci and Di Maro, 2016; Viviano, 2019). Existing methods studying spillover effects typically assume that accurate network links of all sampled units are available (see Leung, 2020b; Ma, Wang, and Tresp, 2020; Vazquez-Bare, 2019; Viviano, 2019, for example). However, such an assumption is hard to verify in practice and questionable in many settings (Sävje, 2019). For example, Angelucci et al. (2010) point out that when constructing the generational family network via respondents’ surnames using the PROGRESA data, false connections exist between unrelated families sharing the same surnames. In the study of technology diffusion among pineapple farmers in Ghana, Conley and Udry (2010) notice the potential misclassification of the information neighbors, due to the lack of precise definition of the information network and the existence of multi-contextual social connections. Comola and Fafchamps (2017) also document a massive discrepancy (about 73%) between the inter-household transfers reported by givers and receivers from a village in Tanzanian.1 Ignoring or mis-connecting either side of the responses may lead to misclassified risk-sharing networks.2

This paper investigates the identification and estimation of the treatment and spillover effects of a program intervention with mismeasured network data. There are several attractive features of the method proposed in this paper. First, it allows flexible forms of heterogeneity in the treatment and spillover effects, which is important to inform how treatment response varies across population (Manski, 2001). In addition, the analysis can be applied to settings with a large network that need not be block-diagonal and that contains missing or misreported links. Moreover, modeling of the network formation or its misclassification probability is not required to implement the proposed method.

I focus on a randomized program intervention and a superpopulation model studied in Leung (2020b). If the network is correctly measured, the direct treatment effect can be identified from the variation of the ego unit’s own treatment status, and the spillover effect can be identified via the variation of a statistic summarizing the exposure to the treated peers. However, the network measurement errors introduced in this paper sophisticate the identification by contaminating the true channels of the network interference. Therefore, ignoring those errors will lead to biased estimation. The measurement errors considered in this paper are nonclassical; that is, they depend on the network interactions. In addition, the measurement errors are assumed to be independent of

---

1The ratio 73% is computed as the number of reported transfers coming from only giver or only receiver (1250) over the number of total reported transfers (1721), see Comola and Fafchamps (2017) page 560-561.
2Similar non-reciprocal problem has also been found in other survey data, e.g. 40% of risk-sharing network links from rural Philippines (Fafchamps and Lund, 2003) and more than 10% of the friendship among adolescents in Add-Health dataset (Calvó-Armengol, Patacchini, and Zenou, 2009; Patacchini, Rainone, and Zenou, 2017) are non-reciprocal.
the potential outcomes and the treatment, conditional on the statistic of network and exogenous covariates. Such an independence assumption is referred to as “nondifferential”, and is often invoked in the literature studying measurement error models (see Bound, Brown, and Mathiowetz, 2001, for example).

In this paper, I propose a novel strategy to nonparametrically point-identify the treatment and spillover effects with a mismeasured network proxy, when an instrumental variable for the latent network (or equivalently, an additional network proxy) is available. The identification consists of two steps. Firstly, I adopt the matrix diagonalization method proposed by Hu (2008) to identify several distributions of the true number of network neighbors (hereafter degree), under the help of the instrument network proxy. Secondly, the distribution involving the true number of treated network neighbors, which measures the exposure to the treated peers, is identified. The identification in the second step relies on the observation that, network proxies in some studies might satisfy one assumption: there exists only one type of measurement error. It means that the network proxy either includes no false links while allowing missing ones (“no false positive”), or includes no missing links but allowing false ones (“no false negative”). The one type of measurement error assumption dramatically simplifies the interdependence of the observed network-based variables with their latent counterparts, which is the main difficulty of the identification. Testable implication of such an assumption is also available.

Inference in network settings are nonstandard due to the data correlation induced by the network interaction. In particular, outcomes of two units are correlated if they are connected or they share common network neighbors (Leung, 2020b). In this paper, the mismeasured network adds to the complication by introducing extra source of correlation, through the spillover of the measurement errors.3 Such spillover occurs, because a false network connection of two units will alter both of their observable exposures to the treated neighbors. In addition to the above network-induced correlation, this paper also considers the data dependency due to general forms of heteroscedasticity, autocorrelation and clustering, so that units who are not friends nor share common friends may also correlate with each other. Such correlation may caused by, for instance, family background, school culture, or community diversity. All sources of data correlation described above generate distinct technical issues for the causal inference.

I propose a semiparametric estimation approach, which overcomes the difficulty caused by the spillover of the measurement errors, and the resulting estimator is shown to be consistent and asymptotically normal. To derive limit theorems, I extend the univariate central limit theorem (CLT) of Chandrasekhar and Jackson (2016) to multivariate settings. The estimation approach in this paper possesses several advantages: (i) it fits the situations where there may exist no clear spatial or ordered structure; (ii) it does not require a large number of independent subnetworks.

3Similar to this paper, Sävje (2019) also finds that misspecification of the treatment exposures will cause extra data dependence.
and allows general forms of data dependence; and (iii) it allows sufficiently large number of units have nonzero correlation with an increasing number of other units.

In the simulation exercises, I verify the advantage of the proposed methodology over the naive estimation ignoring the network measurement errors. The bias reduction provided by the semiparametric approach is substantial and its causal inference is more reliable than that of the naive estimation. Moreover, it is confirmed that the semiparametric method still outperforms the naive estimation, even if its key identification assumption, for example, the one type of measurement error, is mildly violated.

The rest of this paper is organized as follows. Section 2 reviews related literature. Section 3 introduces the model setup and the definition of treatment and spillover effects. The nonparametric identification is established in Section 4. Section 5 presents the semiparametric estimation and its asymptotic properties. Section 6 shows the simulation results and Section 7 presents an empirical application to study the spillover effect of an information program on insurance take-up decisions of rice farmers. Section 8 concludes. All proofs are relegated to the Appendix and the supplementary materials.

2 Literature Review

Spillover effects of the treatment via network interactions have been documented in many applications, e.g., cash transfer programs (Barrera-Osorio et al., 2011), health programs (Dupas, 2014), public policy programs (Kremer and Miguel, 2007), education programs (Opper, 2019), and information diffusion (Banerjee et al., 2013). Misclassification is a pervasive problem of network data and has been noticed in e.g. Advani and Malde (2018), Chandrasekhar and Lewis (2011), De Paula (2017) and Kossinets (2006).

This paper is one among few papers that have studied the spillover effect of a program intervention with mismeasured network. Hardy, Heath, Lee, and McCormick (2019) consider a parametric model for the potential outcomes and for the network misclassifications, and use a likelihood-based approach to estimate the spillover effect. In a nonparametric setting, when only a network proxy is available, He and Song (2018) provide a lower bound for the spillover effect under the restriction that the spillover is nonnegative. This paper is substantially different from the papers above, because it does not reply on modeling the network misclassifications, and more importantly, it provides a formal solution for the nonparametric point-identification of the spillover effect, when two network proxies are available.4

4Sävje, Aronow, and Hudgens (2017) find that when there is limited or moderate degree of network interactions, ignoring the network interference would not impact the asymptotic properties of the average treatment effect estimators. Chin (2018) studies the average treatment effects under unmodeled network interference. However, neither of them explore the spillover effect.
One of the papers studying the structural model of social interactions with mismeasured network is related to this paper.\(^5\) Gao and Li (2019) explore the endogenous and exogenous peer effects via the linear-in-means model with two mismeasured network proxies. Their identification result depends on three key assumptions. First, there exist two different latent network structures for the same group of individuals. Second, the error contaminated network-based variables are assumed to be independent conditional on their latent counterparts, which implicitly requires the networks to be non-stochastic. At last, a copula is used to capture the dependence between the mismeasured network effects. Like this paper, Gao and Li (2019) exploit the matrix diagonalization method, however my analysis focuses on the reduce-form treatment response function which is modeled nonparametrically, enabling flexible forms of heterogeneous treatment and spillover effects.\(^6\) In addition, my identification strategy does not require different network structures for the same set of individuals, the non-stochastic network, or a copula structure for the network-based variables. Instead, the identification in this paper is achieved via restricting the network measurement error.

Consequences and solutions of misclassified network on estimating network statistics or network formation are discussed by, e.g. Breza, Chandrasekhar, McCormick, and Pan (2020), Candelaria and Ura (2020), Comola and Fafchamps (2017), Kossinets (2006), Liu (2013) and Thirkettle (2019). However, it is not clear how to apply these methods to identify treatment and spillover effects in a causal setting.

Literature exploring limit theorems using network dependent data is developing rapidly. Some papers assume that the social network can be partitioned into a large number of disjoint and independent subnetworks (see Lewbel et al., 2019; Vazquez-Bare, 2019, for example). However, this independence assumption may not be plausible in practice, because it ignores the links across subnetworks. Chandrasekhar and Lewis (2011) adopt mixing conditions to restrict the dependence of network links, while, in many contexts, there is no underlying metric space to define the standard “mixing” forms of dependence. Leung (2020b) introduces the notion of “dependence graph” to capture the network-correlated effects, and derives limit theorems under the conditional local dependence; that is, the outcomes of two units are independent if they are not friends nor share common friends. However, in the setting considered in this paper, the measurement errors disrupt the true network dependence structure, so that some seemingly uncorrelated units may actually correlate with each other due to the latent network connections, and vice versa. Therefore, an alternative data dependence structure is needed. I adopt the “dependence neighborhoods” structure proposed by Chandrasekhar and Jackson (2016) to control the data correlation, which does not require to observe the correct network links and employs less restrictions on the dependence

---

\(^5\)Other related papers are, e.g. Chandrasekhar and Lewis (2011), De Paula, Rasul, and Souza (2018a), Goldsmith-Pinkham and Imbens (2013), Lewbel, Qu, and Tang (2019) and Sojourner (2013).

\(^6\)See Hardy et al. (2019), Leung (2019a) and Manski (2013) which also emphasis the difference between the structural model of social interactions and the reduced-form model focusing on the treatment response function.
structure. The dependency neighborhood used in this paper is similar to the dependency graph of Leung (2020b) in the sense that they both aim to control the data dependence. Nonetheless, they are different, since the dependency neighborhoods can capture more general forms of correlation induced by network measurement error and by the unobservables.7

3 Model Setup

Let $D = \{D_i\}_{i \in \mathcal{P}}$ and $Z = \{Z_i\}_{i \in \mathcal{P}}$ denote vectors consisting of units’ (or individuals, nodes, agents) treatment status and observable characteristics for the super-population $\mathcal{P}$, respectively. Denote $A^*$ as the true, latent and binary adjacency matrix, corresponding to an unweighted and undirected random network over the super-population $\mathcal{P}$. Each row of $A^*$, denoted by $A^*_i$, represents unit $i$’s network connection with unit $j$.8 Let $A^*_ij = 1$ if $i$ and $j$ are linked (or equivalently, network neighbors9), and $A^*_ij = 0$ otherwise. As a convention, self links are ruled out, i.e. $A^*_ii = 0$ for $\forall i \in \mathcal{P}$. Given the adjacency matrix $A^*$, we define the set of unit $i$’s first-degree network neighbors by $N^*_i = \{j \in \mathcal{P} : A^*_ij = 1\}$. Denote $|N^*_i| = \sum_{j \in \mathcal{P}} A^*_ij$ as the cardinality of $N^*_i$, and $|N^*_i|$ is usually referred to as the “degree” of unit $i$. For each $i \in \mathcal{P}$, define the outcome $Y_i$ as

$$Y_i = \tilde{r}(i, D, A^*, Z, \varepsilon_i),$$

where $\tilde{r}$ is a unknown real-valued function and $\varepsilon_i$ is an unobservable error term. The $Y_i$ in (1) acknowledges that one units outcome depends on not only his or her own treatment status, but also the treatments assigned to other units, i.e. the spillover effect. I impose the assumption below to restrict the dependence of the outcome $Y_i$ on $(i, D, A^*, Z, \varepsilon_i)$.

Assumption 3.1 (Network Interference) For $\forall i, k \in \mathcal{P}$, $\forall (D, A^*, Z)$ and $\forall (\tilde{D}, \tilde{A}^*, \tilde{Z})$,

$$\tilde{r}(i, D, A^*, Z, e) = \tilde{r}(k, \tilde{D}, \tilde{A}^*, \tilde{Z}, e),$$

for all $e \in \Omega_{\varepsilon_i} \cup \Omega_{\varepsilon_k}$, if the following conditions hold simultaneously: (i) $D_i = \tilde{D}_k$; (ii) $\sum_{j \in \mathcal{P}} A^*_ij = \sum_{j \in \mathcal{P}} \tilde{A}^*_kj$; (iii) $\sum_{j \in \mathcal{P}} A^*_ij D_j = \sum_{j \in \mathcal{P}} \tilde{A}^*_kj \tilde{D}_j$; (iv) $Z_i = \tilde{Z}_k$.

Assumption 3.1 states that the outcome is fully determined by (i) unit’s own treatment status; (ii)

7Other papers study limit theorems of network dependent data include Chin (2018), Kojevnikov, Marmer, and Song (2019), Kuersteiner (2019), Lee and Ogburn (2020), Leung and Moon (2019), Leung (2019b, 2020a), Liu and Hudgens (2014), Song (2018), van der Laan (2014) and references therein.

8The vectors of treatment status and observable characteristics, and the adjacency matrix are infinite-dimensional. We follow Leung (2020b) and obviate further details to ease the illustration.

9It is worthy to notice that there are two different definitions of neighbors utilized in this paper. The first one, which is referred to as “network neighbors”, is defined by the network links $D$. The second one, which is referred to as “dependent neighbors”, is defined via the dependency neighborhoods and is used to characterize correlations of random variables of interest. See Section 5.1 for more details.
the degree; (iii) the number of the first-order treated network neighbors $S^*_i := \sum_{j \in P} A^*_{ij} D_j$; and (iv) unit’s own covariates. The same assumption is used in Leung (2020b).

Assumption 3.1 substantially reduces the dimensionality of the outcome and reveals two crucial features of the network interactions. Firstly, the interference occurs locally, only among the first-order network neighbors. Thus, $(D_i, S^*_i)$ can be viewed as the “effective treatment” of Manski (2013). Secondly, the outcome is invariant to any permutations of the treatments received by the first-order network neighbors, meaning that the interactions are anonymous. The anonymous interaction is also referred to as “stratified interference”, see Baird et al. (2018), Basse and Feller (2018) and Hudgens and Halloran (2008) among others. Under Assumption 3.1, equation (1) can be simplified to

$$Y_i = r(D_i, S^*_i, Z_i, |N^*_i|, \varepsilon_i), \quad \forall i \in P$$

(2)

where $r$ represents a real-valued unknown function. Such an outcome structure permits adequate controls for the observable and unobservable heterogeneity of the treatment response. Given (2), it is easy to see that unit $i$’s outcome $Y_i$ is directly affected by his or her own treatment status $D_i$ (treatment effect), and is also affected by $S^*_i$ because of the exposure to the treated peers (spillover effect). The network $N^*_i$ affects the outcome via two pathways: the degree $|N^*_i|$ and the treated network neighbors incorporated in $S^*_i$. The degree $|N^*_i|$ is a critical attribute because it quantifies the influence of each unit in the social network. Controlling $|N^*_i|$ in (2) enables us to target subpopulation based on different levels of influence, and it also acts as a control variable as it allows the correlation between the degree and the unobservables.

Throughout the paper, the following notations are used. For any generic random variables $X$ and $Y$, denote $f_X$ and $f_X|Y$ as the probability function of $X$ and the conditional probability function of $X$ given $Y$, respectively. Let $\Omega_X$ denote the support of the random variable $X$. By notation abuse, let $|B|$ denote the cardinality of any set $B$, or the absolute value for any scalar $B$. For any vector $a \in \mathbb{R}^p$, let $\|a\|_1 = \sum_{i=1}^{p} |a_i|$ be its $L^1$ norm, $\|a\| = (a'a)^{1/2}$ be its Euclidean norm and $\|a\|_\infty = \max_{1 \leq i \leq p} |a_i|$. Given a matrix $A = (a_{ij})$, we set $\|A\| = [tr(A'A)]^{1/2}$ and $\|A\|_\infty = \max_{1 \leq i,j \leq p} |a_{ij}|$. More generally, for an array (or a vector) of functions, say $a = \{a_i\}$ with $a_i : \Omega_X \mapsto \mathbb{R}$, denote $\|a\|_\infty = \sup_{x \in \Omega_X} \sup_{i} |a_i(x)|$, where $i$ could stand for a multiple index. For an arbitrary parameter $\beta$, denote $d_\beta = dim(\beta)$. \(\perp\) means statistical independence.

### 3.1 Treatment and Spillover Effects

To motivate the potential identification issues, let me begin by defining key concepts and introducing basic assumptions.

---

10 Aronow and Samii (2017), Leung (2019a) and Sävje (2019) consider the possible mis-specification of models similarly defined by Assumption 3.1, and tests for Assumption 3.1 are feasible in Athey et al. (2018).

11 Similar control variable method is used in e.g. Johnsson and Moon (2015).
Definition 1 (CASF) For $\forall (d, s, z, n) \in \{0, 1\} \times \Omega_{S^*, Z, |N^*|}$, the conditional average structural function (CASF) is defined as

$$m^*(d, s, z, n) = E[r(d, s, Z_i, |N^*_i|, \varepsilon_i) | Z_i = z, |N^*_i| = n].$$

In this paper, I focus on the treatment and spillover effects, measuring the average change in potential outcomes in response to the counterfactual manipulation of the treatment assigned to the ego unit and to the network peers, respectively.\(^{12}\)

Definition 2 (Treatment and Spillover Effects) For $\forall (s, z, n) \in \Omega_{S^*, Z, |N^*|}$, define

- treatment effect: $\tau_d(s, z, n) = m^*(1, s, z, n) - m^*(0, s, z, n)$,
- spillover effect: $\tau_s(s, z, n) = m^*(0, s, z, n) - m^*(0, 0, z, n)$.

The assumption below introduces the ignorability conditions accounting for the network interference, based on which the causal effects of interest can be recovered if the actual network data is available.

Assumption 3.2

(a) (Randomized Treatment) $\{D_i\}_{i \in \mathcal{P}}$ are i.i.d. and $\{D_i\}_{i \in \mathcal{P}} \perp \{\varepsilon_j, Z_j, N^*_j\}_{j \in \mathcal{P}}$.

(b) (Unconfounded Network) For $\forall i \in \mathcal{P}$, $\varepsilon_i \perp \{N^*_i, \{D_j\}_{j \in \mathcal{N}^*_i} \} | Z_i, |N^*_i|$.\(^{13}\)

Assumption 3.2 (a) states that the treatment is randomly assigned, does not impact the network, and is independent to the potential outcomes. The randomized intervention is relevant for a wide range of experimental contexts, including Miguel and Kremer (2004), Aral and Walker (2012), Oster and Thornton (2012), Cai et al. (2015b) to name a few, and see Athey and Imbens (2017) for a review. Therefore, Assumption 3.2 (a) is a straightforward starting point for the analysis.

Assumption 3.2 (b) requires the unconfounded network, which is weaker than the fully exogenous network, by allowing the correlation between the degree $|N^*_i|$ and the unobservable characteristics, for example, through the spillovers of unobservables. See Leung (2020b) for similar assumption and supportive examples. The network unconfoundedness to the treatment and the potential outcomes is likely to hold in randomized experiments where the network data is collected before the intervention.\(^{13}\)

\(^{12}\)Similar definitions measuring the direct effect and the spillover effect of treatment are also introduced in Hudgens and Halloran (2008) and Sobel (2006) to name a few. See Tchetgen and VanderWeele (2012) for a discussion about relationships between various notions of causal effects in the presence of network interference. The analysis in this paper can be straightforwardly extended to studies dealing with other notions of treatment effect estimands.

\(^{13}\)It is also worth to notice that although Assumption 3.2 (b) allows the dependence between the network $N^*_i$ and unobservable $\varepsilon_i$ through $|N^*_i|$ and observable exogenous characteristics, it does not allow the unobserved homophily in network formation where the unobservables are correlated to the potential outcomes.
Assumption 3.3 (Distribution)

(a) \( \{ Z_i \}_{i \in P} \) are i.i.d. and \(|N^*_i|\) given \( Z_i \) is identically distributed across \( i \in P \).

(b) For \( \forall i \in P \), \( \varepsilon_i \) given \( (Z_i, |N^*_i|) \) is identically distributed.

Assumption 3.3 (a) implies that the covariate \( Z_i \) is of randomly drawn samples, which is standard in the network effect models literature, e.g. Johnsson and Moon (2015) and Auerbach (2019).\(^\text{14}\)

Condition (a) also requires the conditional distribution of the degree to be invariant across units. An example of the dyadic network formation in Appendix A can be used to verify the existence of such an identical distribution. Also see a strategic network formation model in Leung (2020b) that satisfies (a). Moreover, the identical distribution of the error term \( \varepsilon_i \) given \( (Z_i, |N^*_i|) \) in condition (b) permits that the expressions of the CASF, the treatment effect \( \tau_d \) and the spillover effect \( \tau_s \) are all identical for any unit \( i \in P \).

If the actual network \( N^*_i \) is correctly observed, under the assumptions introduced so far, the CASF can be identified by\(^\text{15}\)

\[
m^*(d, s, z, n) = E \left[ Y_i \mid D_i = d, S^*_i = s, Z_i = z, |N^*_i| = n \right],
\]

which ensures that the treatment and spillover effects are also identifiable. However, it appears that in many applications we fail to obtain fully accurate network information. Ignoring the missing or misclassified network links may lead to biased estimation and misleading causal implications.

### 3.2 Bias of CASF with Mismeasured Network

In this subsection, I will present the potential bias of the CASF identified from the mismeasured network data. Assume that researchers randomly draw \( N \) units from the population \( P \), and collect their outcomes of interest, treatment status, covariates, network information and treatment assignments of their network neighbors. Thus, researchers can observe:

\[
(Y_i, D_i, Z_i, N_i, \{ D_j \}_{j \in N_i}), \text{ for } i = 1, 2, ..., N,
\]

\(^\text{14}\)In the analysis of this paper, it is feasible to relax the i.i.d. of \( Z_i \) and allow it to possess dependent structure under the framework described in Section 5.1. We maintain such an i.i.d. assumption only for illustration simplicity.

\(^\text{15}\)For \( \forall (d, s, z, n) \in \{0, 1\} \times \Omega_{S^*, Z, |N^*|} \), it can be shown that

\[
E \left[ Y_i \mid D_i = d, S^*_i = s, Z_i = z, |N^*_i| = n \right] = E \left[ r(D_i, S^*_i, Z_i, |N^*_i|, \varepsilon_i) \mid D_i = d, S^*_i = s, Z_i = z, |N^*_i| = n \right] = m^*(d, s, z, n),
\]

where the second equality is due to the unconfoundedness of \( (D_i, S^*_i) \) in Lemma B.1 and the last equality is by Definition 1.
where \( N_i \) denotes the observed identities of unit \( i \)'s network neighbors with cardinality \( |N_i| \), and the convention of no self connections is maintained, i.e. \( i \notin N_i \). Note that there are no restrictions on the sampling scheme of the network data. Namely, \( N_i \) can be obtained from a single and fully observed network, or from a (possibly partially observed) sampled network. In addition, \( N_i \) can be either self-reported, acquired from the administrative data, or constructed by researchers based on some specific rules. Throughout the paper, \( N_i \) is referred to as “network proxy”. Given \( N_i \), the number of the observed treated network neighbors is denoted by \( S_i = \sum_{j \in N_i} D_j \).

The assumption below extends Assumption 3.2 to accommodate the observable network proxy by restricting the misclassification of the network links.

**Assumption 3.4 (Nondifferential Misclassification)**

(a) \( \{D_i\}_{i \in P} \perp \{\varepsilon_j, Z_j, N_j^*, N_j\}_{j \in P} \);

(b) For \( \forall i \in P, \varepsilon_i \perp (N_i^*, \{D_j\}_{j \in N_i^*}, N_i, \{D_j\}_{j \in N_i}) | Z_i, |N_i^*| \).

(c) For \( \forall i \in P, |N_i| \) given \( (Z_i, |N_i^*|) \) is identically distributed.

Assumption 3.4 (a) and (b) indicates that given the actual network information and individual’s characteristics, the observed proxy \( N_i \) does not contain relevant information to predict the outcome, which is often referred to as “nondifferential misclassification” in the measurement error models literature, e.g. Battistin and Sianesi (2011), Hu (2008) and Lewbel (2007). In addition, Assumption 3.4 (c) holds in many contexts, for example, when units fail to respond with probability proportional to their actual degrees (“the load effect”), or inversely proportional to their actual degrees (“the periphery effect”), see Kossinets (2006). I also provide one set of sufficient conditions for Assumption 3.4 (c) in Appendix A.

Now, let us denote the conditional mean function of the outcome given the observables as

\[
m_i(d, s, z, n) = E[Y_i|D_i = d, S_i = s, Z_i = z, |N_i| = n],
\]

where the subscript \( i \) of \( m_i \) represents the possibly non-identical conditional mean of the outcome given the observables, which is caused by the unknown dependence between the error contaminated network-variables \((S_i, |N_i|)\) and their latent counterparts \((S_i^*, |N_i^*|)\). The relationship between \( m_i \) and \( m^* \) can be obtained by the proposition below.

**Proposition 3.1** Under Assumptions 3.1-3.4, for \( \forall i \in P \) and \( \forall (d, s, z, n) \in \{0, 1\} \times \Omega_{S,Z,|N|}, \)

\[
m_i(d, s, z, n) = \sum_{(s^*, n^*) \in \Omega_{S^*,|N^*|}} m^*(d, s^*, z, n^*) f_{S_i^*,|N_i^*|} | D_i = d, S_i = s, Z_i = z, |N_i| = n (s^*, n^*).$

Proposition 3.1 characterizes the bias in the CASF estimand if ignoring the measurement errors of the network links. The expression of \( m_i \) makes it clear that the bias of \( m_i \) is governed by the latent
distribution of the actual network-based variable \((S^*_i, |N^*_i|)\) given its observed proxies \((S_i, |N_i|)\). The bias will be larger, if the misclassification probability of \((S_i, |N_i|)\) is higher. Importantly, due to the nonparametric setting of \(m^*\), simply differencing \(m_i(1, s, z, n)\) and \(m_i(0, s, z, n)\) in general cannot give the treatment effect \(\tau_d(s, z, n)\), even though the treatment is randomized and correctly-observed. While, it will be true if the response to the variation of ego unit’s own treatment status is homogeneity in both the observables and unobservables, relying on strong structural restriction. Similar weighted average expressions of the identifiable parameter are founded in Gao and Li (2019) for the endogenous peer effects and in Hardy et al. (2019) for the treatment spillover effects.\(^\text{16}\)

### 4 Identification

Suppose that two network proxies are available for each sampled individual \(i \in \{1, 2, ..., N\}\), denoted by \(N_i\) and \(\tilde{N}_i\). These two proxies may come from repeated observations of a sampled network over time, different dimensions of connections (e.g. kinship and borrowing-lending), multi-contextual interactions (e.g. various social events or afflictions), or self-reported and administrative network data. Intuitively, the additional network proxy \(\tilde{N}_i\) can be understood as an instrument for the true latent network, as in Hu (2008) and Hu and Schennach (2008), that is conceptually similar to the ones utilized in conventional instrumental variable methods. Following the same construction as before, for \(\tilde{N}_i\), denote its degree as \(|\tilde{N}_i|\) and the number of treated network neighbors as \(\tilde{S}_i = \sum_{j \in \tilde{N}_i} D_j\), respectively.

Let me first introduce an useful lemma which plays a key role in decomposing the latent distribution \(f_{S^*_i, |N^*_i|} | Z_i, |N_i|\) into identifiable components. Without loss of generality, I state the result for one network proxy \(N_i\), the same result holds for \(\tilde{N}_i\).

**Lemma 4.1** Under Assumption 3.2 (a) and 3.4 (a), we have that

(a) \(N^*_i \perp S^*_i | Z_i, |N^*_i|\) and \(|N_i| \perp S^*_i | Z_i, |N^*_i|\);

(b) \(N_i \perp S_i | Z_i, |N_i|\) and \(|N^*_i| \perp S_i | Z_i, |N_i|\);

(c) for \(\forall (s, n) \in \Omega_{S, |N_i|}\), \(f_{S^*_i | Z_i, |N^*_i| = n} (s) = f_{S_i | Z_i, |N_i| = n} (s) = C_s^nf_D(1)^sf_D(0)^{n-s}\), for \(f_D(d) := Pr(D_i = d)\) with \(d \in \{0, 1\}\).

Lemma 4.1 (a) delivers two implications. The distribution of the exposure to treated peers, \(S^*_i\), is fully determined by the degree and the exogenous covariates, instead of the identity of interacted

---

\(^\text{16}\)In order to consistently estimate the CASF \(m^*\), practitioners should either be aware of and able to restrict the degree of mismeasurement in the observed network, or identify and consistently estimate the intermediate latent distribution \(f_{S^*_i, |N^*_i|} | D_i, S_i, Z_i, |N_i|\). The first solution can be accomplished utilizing a single network proxy if the probability of misclassification decreases as sample size increases. Rigorous study along this line is provided in the supplemental material.
peers or the observable network degree. It further restricts the anonymous interactions. Lemma 4.1 (b) states the similar properties of $S_i$. In addition, the identifiability of $f_{S_i^*|Z_i,|\mathcal{N}_i^*|}$ in (c) is intuitive, because the summation of any given $n$ i.i.d. treatment status follows a binomial distribution.

Given Proposition 3.1 and Lemma 4.1, to identify the CASF $m^*$, we can first decompose the latent distribution function $f_{S_i^*|\mathcal{N}_i^*||D_i,S_i,Z_i,|\mathcal{N}_i|}$, for one of the two network proxies, as follows.

**Proposition 4.2** Under Assumptions 3.2 and 3.4, we have that

$$f_{S_i^*|\mathcal{N}_i^*||D_i,S_i,Z_i,|\mathcal{N}_i|} = \frac{f_{S_i|S_i^*,Z_i,|\mathcal{N}_i^*|,|\mathcal{N}_i|} \times f_{S_i^*|Z_i,|\mathcal{N}_i^*|} \times f_{\mathcal{N}_i||Z_i,|\mathcal{N}_i|} \times f_{|\mathcal{N}_i^*||Z_i|}}{f_{S_i|Z_i,|\mathcal{N}_i|} \times f_{|\mathcal{N}_i||Z_i|}}. \tag{3}$$

From Proposition 4.2, it is easy to see that the latent distribution of interest can be expressed as a product of six distributions, where $f_{|\mathcal{N}_i||Z_i}$, $f_{S_i^*|Z_i,|\mathcal{N}_i^*|}$ and $f_{S_i|Z_i,|\mathcal{N}_i|}$ can be identified directly from the observables under the assumptions exploited in the previous sections.

In what follows, I will deal with the identification of the rest distributions in the decomposition above in two steps. First, given the two observed networks $\mathcal{N}_i$ and $\tilde{\mathcal{N}}_i$, apply the method of matrix diagonalization of Hu (2008) to achieve the identification of $f_{|\mathcal{N}_i||Z_i,|\mathcal{N}_i^*|}$ and $f_{|\mathcal{N}_i^*||Z_i|}$. Due to the complex and unconstrained interdependence between the observed $(S_i,|\mathcal{N}_i|)$, $(\tilde{S}_i,|\tilde{\mathcal{N}}_i|)$ and their latent counterpart $(S_i^*,|\mathcal{N}_i^*|)$ through the underlying network $\mathcal{N}_i^*$, even with two network proxies, it is not feasible to identify the latent distribution $f_{S_i|S_i^*,Z_i,|\mathcal{N}_i^*|,|\mathcal{N}_i|}$ by simply repeating the matrix diagonalization approach. Therefore, in the second step, I introduce a crucial assumption on the network measurement error, which dramatically simplifies the interdependence and ensures the identification of $f_{S_i|S_i^*,Z_i,|\mathcal{N}_i^*|,|\mathcal{N}_i|}$.

### 4.1 Identification via Matrix Diagonalization

Assumptions 4.1 to 4.4 below are crucial when establishing the identification results via the matrix diagonalization technique similar to that used in Hu (2008). Modifications to the assumptions and to the method are made accordingly, to fit the network setting considered in this paper.

**Assumption 4.1 (Exclusion Restriction)** $|\mathcal{N}_i| \perp |\tilde{\mathcal{N}}_i||Z_i,|\mathcal{N}_i^*|$.  

Assumption 4.1 can be interpreted as standard exclusion restriction that $|\tilde{\mathcal{N}}_i|$ does not provide extra information about $|\mathcal{N}_i|$ than the actual degree $|\mathcal{N}_i^*|$ already provides. It can also be understood as that the instrumental variable $\tilde{\mathcal{N}}_i$ is conditionally independent to the measurement errors contained in the proxy $\mathcal{N}_i$. It rules out the situations where both network proxies are mismasured due to random omission of the same group of units when constructing the networks. One set of sufficient conditions for Assumption 4.1 is given in Appendix A. The exclusion restriction is the key to implement the matrix diagonalization method.

**Assumption 4.2 (Sparsity)** $\Omega_{|\mathcal{N}_i|} = \Omega_{|\mathcal{N}|} = \Omega_{|\mathcal{N}_i^*|}$ with finite cardinality $K_{|\mathcal{N}|}$. 

11
Assumption 4.2 requires the network to be sparse, i.e. each individual has finite friends, and the number of friends does not increase with the sample size. The sparse network is commonly observed in many empirical applications (Chandrasekhar, 2016), and is a standard assumption in network effects literature, e.g. De Paula, Richards-Shubik, and Tamer (2018b), Qu and Lee (2015) and Viviano (2019). By the i.i.d. of the treatment assignment, it is clear that $\Omega_{S} = \Omega_{S'} = \Omega_{S''}$.

To illustrate the basic idea of the matrix diagonalization technique, let me introduce the following notations. Without loss of generality, set $\Omega_{|\mathcal{N}|} = \Omega_{|\mathcal{N}|} = \Omega_{|\mathcal{N}|} = \{0, \ldots, K_{|\mathcal{N}|} - 1\}$. Denote the $K_{|\mathcal{N}|} \times K_{|\mathcal{N}|}$ matrix $F_{|\mathcal{N}|,|\mathcal{N}^*|}$ as

$$F_{|\mathcal{N}|,|\mathcal{N}^*|} = \begin{bmatrix} f_{|\mathcal{N}|,|\mathcal{N}^*| = 0} (0) & \cdots & f_{|\mathcal{N}|,|\mathcal{N}^*| = K_{|\mathcal{N}|} - 1} (0) \\ \vdots & \ddots & \vdots \\ f_{|\mathcal{N}|,|\mathcal{N}^*| = 0} (K_{|\mathcal{N}|} - 1) & \cdots & f_{|\mathcal{N}|,|\mathcal{N}^*| = K_{|\mathcal{N}|} - 1} (K_{|\mathcal{N}|} - 1) \end{bmatrix}.$$  \hfill (4)

In a similar vein as $F_{|\mathcal{N}|,|\mathcal{N}^*|}$, define two observable $F_{|\mathcal{N}|,|\mathcal{N}^*|}$ via replacing $f_{|\mathcal{N}|,|\mathcal{N}^*|}$ by $f_{|\mathcal{N}|,|\mathcal{N}^*|}$. In addition, define a $K_{|\mathcal{N}|} \times K_{|\mathcal{N}|}$ diagonal matrix

$$T_{|\mathcal{N}|,|\mathcal{N}^*|} = \text{diag} \left( E[Y_i | Z_i, \mathcal{N}^*_i = 0], E[Y_i | Z_i, \mathcal{N}^*_i = 1], \ldots, E[Y_i | Z_i, \mathcal{N}^*_i = K_{|\mathcal{N}|} - 1] \right).$$

The main idea of the matrix diagonalization method is to identify the latent distributions of interest via diagonalizing the directly observable distribution $E_{|\mathcal{N}|,|\mathcal{N}^*|} \times F_{|\mathcal{N}|,|\mathcal{N}^*|}^{-1}$

$$T_{|\mathcal{N}|,|\mathcal{N}^*|} = F_{|\mathcal{N}|,|\mathcal{N}^*|}^{-1} \times \left( E_{|\mathcal{N}|,|\mathcal{N}^*|} \times F_{|\mathcal{N}|,|\mathcal{N}^*|}^{-1} \right) \times F_{|\mathcal{N}|,|\mathcal{N}^*|}.$$

Then, recover the latent distributions in $F_{|\mathcal{N}|,|\mathcal{N}^*|}$ and $F_{|\mathcal{N}|,|\mathcal{N}^*|}$ via the eigen-decomposition approach: columns of $F_{|\mathcal{N}|,|\mathcal{N}^*|}$ are the eigenvectors of matrix $E_{|\mathcal{N}|,|\mathcal{N}^*|} \times F_{|\mathcal{N}|,|\mathcal{N}^*|}^{-1}$ and the diagonal elements of $T_{|\mathcal{N}|,|\mathcal{N}^*|}$ are the corresponding eigenvalues. Note that the discussion above is based on the presumption about the invertibility of $F_{|\mathcal{N}|,|\mathcal{N}^*|}$ and $F_{|\mathcal{N}|,|\mathcal{N}^*|}$, which is formalized by Assumption 4.3 below.

**Assumption 4.3 (Rank Condition)** The ranks of $F_{|\mathcal{N}|,|\mathcal{N}^*|}$ and $F_{|\mathcal{N}|,|\mathcal{N}^*|}$ are both $K_{|\mathcal{N}|}$.

The next assumption is the key to identifying latent probabilities via eigen-decomposition.

**Assumption 4.4 (Eigen-decomposition)**

(a) For $\forall n, n' \in \Omega_{|\mathcal{N}|}$ such that $n \neq n'$, we have $E[Y_i | Z_i, \mathcal{N}^*_i = n] \neq E[Y_i | Z_i, \mathcal{N}^*_i = n']$. 

12
(b) For \( \forall n^* \in \Omega_{|\mathcal{N}_i^*|} \) and any \( n \neq n^* \), we have
\[
 f_{|\mathcal{N}_i||Z_i||\mathcal{N}_i^*| = n^*}(n^*) > f_{|\mathcal{N}_i||Z_i||\mathcal{N}_i^*| = n^*}(n), \quad f_{|\hat{\mathcal{N}}_i||Z_i||\mathcal{N}_i^*| = n^*}(n^*) > f_{|\hat{\mathcal{N}}_i||Z_i||\mathcal{N}_i^*| = n^*}(n).
\]

Assumption 4.4 (a) is a sufficient condition to avoid duplicate eigenvalues so that the eigen-decomposition is unique. It is automatically satisfied if \( E[Y_i|Z_i, |\mathcal{N}_i^*|] \) is monotone in \( |\mathcal{N}_i^*| \) and it also holds for more general scenarios. Noticing that the condition (a) is a special case of a more general condition \( E[\varpi(Y_i)|Z_i, |\mathcal{N}_i^*| = n] \neq E[\varpi(Y_i)|Z_i, |\mathcal{N}_i^*| = n'] \), where the transformation function \( \varpi(\cdot) \) can be user-specified, such as \( \varpi(y) = (y - E[Y_i])^2 \) (variance) or \( \varpi(y) = 1[y \leq y_0] \) (quantile) for some given \( y_0 \). Assumption 4.4 (b) permits that the order of the eigenvectors is identifiable. It indicates that the observable degrees are informative proxies for the latent degree, which implicitly assumes that the probability of correctly reporting is higher than that of misreporting. Similar restrictions are widely invoked in the literature of measurement error models. See e.g. Battistin and Sianesi (2011), Battistin, De Nadai, and Sianesi (2014), Chen, Hong, and Nekipelov (2011), Hu and Schennach (2008), Lewbel (2007) and Mahajan (2006).

**Theorem 4.3** Suppose Assumption 3.4 is satisfied by \( \mathcal{N}_i \) and \( \mathcal{N}_i^* \). Under Assumptions 3.1-3.3 and 4.1,

(a) \( f_{|\mathcal{N}_i||Z_i|, |\mathcal{N}_i^*|} \) and \( f_{|\hat{\mathcal{N}}_i||\mathcal{N}_i||Z_i|, |\mathcal{N}_i^*|} \) are identical across \( i \in \mathcal{P} \).

(b) If further assume Assumptions 4.2-4.4 hold, then \( f_{|\mathcal{N}_i^*||Z_i|, |\mathcal{N}_i^*|, |\mathcal{N}_i^*|} \) and \( f_{|\hat{\mathcal{N}}_i||Z_i|, |\mathcal{N}_i^*|} \) are non-parametrically identified.

### 4.2 Identification via One Type of Measurement Error

Next, let me proceed with the identification of \( f_{S_i|S_i^*, Z_i, |\mathcal{N}_i^*|} \). The matrix diagonalization method is infeasible in this step, because of the violation of the exclusion restriction analogue to Assumption 4.1. In other words, the conditional independence \( S_i \perp \tilde{S}_i \) given \( (S_i^*, Z_i) \) with \( Z_i = (Z_i, |\mathcal{N}_i^*|) \) does not hold.\(^{17}\) To be more specific, consider the expression of \( S_i \) in terms of \( S_i^* \) below

\[
 S_i = S_i^* - \sum_{j \in \mathcal{N}_i^*/\mathcal{N}_i} D_j + \sum_{j \in \mathcal{N}_i/\mathcal{N}_i^*} D_j, \tag{5}
\]

where for any sets \( A \) and \( B \), let \( A/B := A \cap B^c \) with \( B^c \) being the complement of \( B \). The set \( \mathcal{N}_i^*/\mathcal{N}_i \) contains all the missing network links of \( i \) (false negative), and the set \( \mathcal{N}_i/\mathcal{N}_i^* \) includes all the false network links (false positive). Similarly, \( \tilde{S}_i = S_i^* - \sum_{j \in \mathcal{N}_i^*/\mathcal{N}_i} D_j + \sum_{j \in \mathcal{N}_i/\mathcal{N}_i^*} D_j \). Consider an extreme case, where the super-population \( \mathcal{P} \) is a finite set and unit \( i \) connects to all other

\(^{17}\)Intuitively, \( Z_i = (Z_i, |\mathcal{N}_i^*|) \) as a whole can be regarded as a covariate, because its distribution is identifiable based on Theorem 4.3.
units, i.e. \( N^*_i = \mathcal{P} \). Then the only possible misclassification for \( N_i \) and \( \tilde{N}_i \) will be underreporting, leading to both \( S_i \) and \( \tilde{S}_i \) smaller than \( S^*_i \). Apparently, \( S_i \) and \( \tilde{S}_i \) are positively correlated in this case, contradicting the exclusion restriction.

Based on the discussion above, the main issue of identifying \( f_{S_i|S^*_i,z_i,N_i,|N^*_i|} \) arises from the dependence between \((S_i, |N_i|)\) and \((S^*_i, |N^*_i|)\). Such dependence is not easy to characterize, because \((S_i, |N_i|)\) and \((S^*_i, |N^*_i|)\) relate to each other via the underlying network \( N^*_i \) which is unobservable, and the arbitrary measurement error further complicates their relationship. The latter is because, without imposing any constraint on the measurement error, given \((S^*_i, |N^*_i|) = n^*, |N_i| = n\), there will be various realizations of \( N_i \) and \( N^*_i \), each of which may lead to substantially different \( S_i \). For example, when \( n = n^* \), it is possible that all network links are classified correctly, therefore \( N_i = N^*_i \). If so, \( S_i \) would be entirely determined by its latent counterpart \( S^*_i \). While, it is also possible that not even a single element in \( N_i \) and \( N^*_i \) is the same, although they have the same cardinality. If that be the case, then \( S_i \) would be solely governed by the treatment status of the misreported false network neighbors \( \sum_{j \in N_i} D_{ij} \), and would not depend on \((S^*_i, |N^*_i|)\) anymore. Therefore, without further restricting the measurement error, there will be too little information and too much uncertainty to pin down \( f_{S_i|S^*_i,z_i,N_i,|N^*_i|} \).

For any given \( n \in \Omega_{|N_i|} \) and \( n^* \in \Omega_{|N^*_i|} \), the \((n+1) \times (n^*+1)\) unknown conditional probabilities of \( S_i \) which characterize the dependence structure between \((S_i, |N_i|)\) and \((S^*_i, |N^*_i|)\), can be formalized by the \((n+1) \times (n^*+1)\) matrix below:

\[
F_{S_i|S^*_i,N_i,|N^*_i|} = \begin{pmatrix}
    f_{S_i|S^*_i=0,z_i,|N_i|=n,|N^*_i|=n^*(0) & \cdots & f_{S_i|S^*_i=n^*,z_i,|N_i|=n,|N^*_i|=n^*(0)} \\
    \vdots & \ddots & \vdots \\
    f_{S_i|S^*_i=0,z_i,|N_i|=n,|N^*_i|=n^*(n)} & \cdots & f_{S_i|S^*_i=n^*,z_i,|N_i|=n,|N^*_i|=n^*(n)}
\end{pmatrix}
\]

Denote a \((n+1) \times 1\) vector \( F_{S_i|Z_i,|N_i|} \) and a \((n^*+1) \times 1\) vector \( F_{S^*_i|Z_i,|N^*_i|} \) by

\[
F_{S_i|Z_i,|N_i|} = [f_{S_i|Z_i,|N_i|=n(0)}, \ldots, f_{S_i|Z_i,|N_i|=n(n)}]', \quad F_{S^*_i|Z_i,|N^*_i|} = [f_{S^*_i|Z_i,|N^*_i|=n^*(0)}, \ldots, f_{S^*_i|Z_i,|N^*_i|=n^*(n^*)}].
\]

where both of the vectors are identifiable. It then yields a system of \((n+1)\) linear equations with \((n+1) \times (n^*+1)\) unknowns from Lemma 4.1 (b) and the law of total probability:\footnote{Equation (7) is because \( f_{S_i|Z_i,|N_i|=n(s)} = f_{S_i|Z_i,|N_i|=n,|N^*_i|=n^*(s)} = \sum_{s^* \in \Omega_{|N_i|}} f_{S_i|Z_i,|N^*_i|=n^*(s^*)} = \sum_{s^* \in \Omega_{|N_i|}} f_{S^*_i|Z_i,|N^*_i|=n^*(s^*)} f_{S_i|S^*_i=s^*,|N_i|=n,|N^*_i|=n^*} \times f_{S_i|S^*_i=s^*,|N^*_i|=n^*(s^*)} = \sum_{s^* \in \Omega_{|N_i|}} f_{S_i,|N_i|=n,|N^*_i|=n^*(s^*)} f_{S^*_i|Z_i,|N^*_i|=n^*(s^*)}.}

\[
F_{S_i|Z_i,|N_i|} = F_{S_i|Z_i,|N_i|,|N^*_i|,Z} \times F_{S^*_i|Z_i,|N^*_i|}, \quad (7)
\]

which, however, is underdetermined because there are fewer equations than unknowns. Therefore, it is necessary to impose restrictions to reduce the number of unknown parameters in order to get
enforces a sparsity constraint on the unknowns: given \( S_i^* = s^* \), the probability of \( S_i = s \) with \( s < s^* \) should be zero, as the only source of misclassification in \( S_i \) is from those false connections. Therefore, the elements above the diagonal of matrix \( F \) are all zero. Secondly, \( N_i^* \subset N_i \) also dramatically simplifies the dependence structure between \( (S_i, |N_i|) \) and \( (S_i^*, |N_i^*|) \) via limiting the possible realizations of \( N_i \) and \( N_i^* \): the elements in each \( k \)-diagonal \( (k = -1, -2, \ldots, -n) \) of matrix \( F_{S_i^* |N_i^*| |N_i^*|, Z} \) will be the same. It is because, under no false negative, no matter what the number of actual treated friends \( S_i^* \) is, the conditional distribution of \( S_i \) will be the same as long as its increase relative to the truth, \( S_i - S_i^* \), is the same. Intuitively, \( f_{S_i |S_i^* = s^*, Z_i, |N_i| = n, |N_i^*| = n^*} (s^*) \) is equal to the probability of randomly choosing \( s - s^* \) units out of \( n - n^* \) units, which does not vary with the realizations of \( S_i^*, |N_i| \) and \( |N_i^*| \).

Now, under no false negative, for any \( n^* \leq n \), the matrix \( F_{S_i |S_i^* = s^*, Z_i, |N_i| = n, |N_i^*| = n^*} \) can be simplified to

\[
\begin{bmatrix}
  f_{\Delta S_i |Z_i, |N_i| = |N_i^*| = \Delta n (0)} & 0 & \cdots & 0 \\
  f_{\Delta S_i |Z_i, |N_i| = |N_i^*| = \Delta n (1)} & f_{\Delta S_i |Z_i, |N_i| = |N_i^*| = \Delta n (0)} & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  f_{\Delta S_i |Z_i, |N_i| = |N_i^*| = \Delta n (n)} & \cdots & f_{\Delta S_i |Z_i, |N_i| = |N_i^*| = \Delta n (n-1)} & f_{\Delta S_i |Z_i, |N_i| = |N_i^*| = \Delta n (n)} \\
\end{bmatrix},
\]

with \((n + 1)\) unknowns, which is the same as the number of equations, ensuring a unique solution for (7) and the identification of \( f_{S_i |S_i^*, Z_i, |N_i| = |N_i^*|} \). The discussion above only requires one of the two network proxies satisfying the desired property, while does not impose any restriction on the measurement error of the other proxy expect for those having been assumed previously. Without loss of generality, hereafter we use \( N_i \) to denote the one that satisfies the requirement.

**Assumption 4.5 (One Type of Measurement Error)** For each unit \( i \in \mathcal{P} \), the proxy \( N_i \) sat-

\[\text{Under no false negative, we do not consider the case } n < n^*, \text{ because } N_i^* \subset N_i \text{ implies that the event } (|N_i^*|, |N_i|) = (n^*, n) \text{ with } n < n^* \text{ is a zero probability even, and a conditional probability conditional on a zero probability even is undefined. Similarly, under no false positive, we do not consider the case } n > n^*.

\[\text{It is worth to note the equivalence between } f_{S_i |S_i^* = s^*, Z_i, |N_i| = |N_i^*|} \text{ and } f_{S_i^* |S_i = s, Z_i, |N_i| = |N_i^*|} \text{ via re-scaling:}
\]

\[f_{S_i |S_i^* = s^*, Z_i, |N_i| = n, |N_i^*| = n^*} (s) = f_{S_i^* |S_i = s, Z_i, |N_i| = n, |N_i^*| = n^*} (s^*) f_{S_i |Z_i, |N_i| = n} (s) / f_{S_i^* |Z_i, |N_i^*| = n^*} (s^*),\]

where the equality is based on Lemma 4.1. Therefore, under no false positive, Similar arguments can be applied to the case when no false positive holds.
isfies either no false positive, i.e. \( N_i \subset N_i^* \), or no false negative, i.e. \( N_i^* \in N_i \).

Borrowing the terminology from Calvi et al. (2018), Assumption 4.5 is referred to as “one type of measurement error”. As can be seen from the next lemma, exploiting Assumption 4.5 benefits us the significant simplicity of the interdependence between the observable \((S_i, |N_i|)\) and the latent \((S_i^*, |N_i^*|)\), which dramatically reduces the number of unknown probabilities.

**Lemma 4.4** Suppose Assumptions 3.2, 3.4 and 4.5 hold. Let \( \Delta s = |s - s^*| \) and \( \Delta n = |n - n^*| \). For \( \forall (s^*, n^*) \in \Omega_{S^*|N^*|} \) and \( \forall (s, n) \in \Omega_{S_i|N_i|} \), \( f_{S_i|s^*,|N_i^*|=n^*,|N_i|=n,z_i}(s) \) is identical across \( i \in \mathcal{P} \).

(a) If no false negative \( N_i^* \subset N_i \) holds, then for \( n^* \leq n \),

\[
f_{s^*|s^*,|N_i^*|=n^*,|N_i|=n,z_i}(s) = \begin{cases} 
C_{\Delta n}^{\Delta s} f_D(1)^{\Delta s} f_D(0)^{\Delta n-\Delta s}, & \text{if } s^* \leq s \text{ and } \Delta s \leq \Delta n \\
0, & \text{otherwise.}
\end{cases}
\]

(b) If no false positive \( N_i \subset N_i^* \) holds, then for \( n \leq n^* \)

\[
f_{S_i^*|s_i=|N_i^*|=n^*,|N_i|=n,z_i}(s) = \begin{cases} 
C_{\Delta n}^{\Delta s} f_D(1)^{\Delta s} f_D(0)^{\Delta n-\Delta s}, & \text{if } s \leq s^* \text{ and } \Delta s \leq \Delta n \\
0, & \text{otherwise.}
\end{cases}
\]

It perhaps not surprising that \( S_i \) conditional on \((S_i^*, |N_i^*|, |N_i|, Z_i)\) follows a binomial distribution, given the equivalence of \( f_{S_i|s^*,|N_i^*|=n^*,|N_i|=n,z_i}(s^*)\) to the probability of randomly assigning treatment to \( \Delta s \) out of \( \Delta n \) units. The result in Lemma 4.4 enables a faster and easier way to identifying \( f_{S_i^*|s_i,|N_i^*|=n^*,|N_i|=n,z_i}(s) \) without relying on solving the linear system. Nevertheless, the linear system greatly facilitates the identification analysis, determines the identification status of \( f_{S_i^*|s_i,|N_i^*|=n^*,|N_i|=n,z_i} \), and the solution of the system produces the same result to that obtained by simply exploiting the binomial distribution.

**Theorem 4.5** Under Assumptions 3.2-3.4 and 4.5, \( f_{S_i^*||N_i^*|=|D_i,s_i,z_i,|N_i|} \) is identical across \( i \in \mathcal{P} \) and nonparametrically identified.

“No false positive” is satisfied in many situations, for instance, when the mismeasurement is caused by sampling-induced error, such as missing links (“induced subgraph” in Kossinets, 2006); restricting the network within a village (Angelucci et al., 2010); limiting the maximum number of nominated friends (Cai et al., 2015b). It is also satisfied when non-sampling-induced error arises, for example, when a survey respondent becomes uninterested in naming the full list of his or her friends due to survey fatigue; the lack of measurability of abstract but meaningful connections, e.g. esteem or authority; constructing network by intersecting repeated network observations assuming

\[\text{The conditions } (s^*, n^*) \in \Omega_{S^*|N^*|} \text{ and } (s, n) \in \Omega_{S_i|N_i} \text{ implicitly imply that } 0 \leq s \leq n \text{ and } 0 \leq s^* \leq n^*.\]
that the overlap includes those effectual interactions; keeping only reciprocated network links when non-reciprocated or undirected network links exist (Comola and Fafchamps, 2017); collecting data in certain contexts where participants are unwilling to cooperate, like criminals’ connections or adolescent sexual network (Kossinets, 2006); constructing the network based on a particular dimension of the social connections, while ignoring other relevant interactions (Conley and Udry, 2010).

“No false negative” is also a reasonable assumption, for instance, when observing a large network including ineffectual interactions, like social media friends, email connections and virtual communities; simply assuming all units within a certain geographical boundary are linked; when the observed network is formed as a union of multiple-dimensional networks, e.g. kinship, borrower-lender relationship and advice-giving (Banerjee et al., 2013); assuming a link exists if either side of the two nodes reports a interaction; or when constructing a network based on participation in multiple social events or affiliations (“multicontextual approach” in Kossinets, 2006).

If the network proxy $N_i$ satisfies the one type of measurement error assumption, the matrix $F_{\{N\}|Z|N^*}$ in (4) should be upper triangular if no false positive, and lower triangular if no false negative. Based on Theorem 4.3, since $F_{\{N\}|Z|N^*}$ is identifiable, it is possible to test the one type of measurement error assumption via the null hypothesis that all elements in either the upper or the lower triangular of matrix $F_{\{N\}|Z|N^*}$ are zero. One possible testing approach is the subsampling or bootstrap method proposed by Romano and Shaikh (2012) with proper adjustments to accommodate the network data. A formal test if left for future research.

Given the results in Theorem 4.3 and Theorem 4.5, the identification of the CASF, the treatment and spillover effects estimands can be achieved.

**Theorem 4.6 (Identification)** Suppose Assumption 3.4 is satisfied by $\tilde{N_i}$ and $N_i$. Let Assumptions 3.1-3.3, and 4.1-4.5 hold.

(a) For $\forall (d, s, z, n) \in \{0, 1\} \times \Omega_{S,Z,|N|}$ such that $f_{\{N\}|Z|=z}(n) > 0$, $m_i(d, s, z, n) = E[Y_i|D_i = d, S_i = s, Z_i = z, |N_i| = n]$ is identical for all $i \in \mathcal{P}$.

(b) The CASF $m^*$, the treatment effect estimand $\tau_d$ and the spillover effect estimand $\tau_s$ are nonparametrically identifiable wherever they are well-defined.

---

22Other possible testing approaches may be established following Leung (2020a) if the $\sqrt{N}$ convergence rate of estimator for $F_{\{N\}|Z|N^*}$ is satisfied. It might be the case if the outcome $Y_i$ and covariates in $Z_i$ in this paper are discrete, then a smooth kernel estimation is not needed and the $\sqrt{N}$ convergence rate can be achieved based on the proof of Theorem 5.2 in Section 5.
4.3 Discussion and Extension

4.3.1 Anonymous Interactions

As implied by Lemma 4.1, the anonymous interactions $S_i^* \perp \mathcal{N}_i^* | Z_i, |\mathcal{N}_i^*|$ is critical for accomplishing the identification of $m^*$. The key factor to ensure the anonymous interactions is that, for any given unit $i$, the treatment assignments to units other than $i$, $\{D_j\}_{j \in \mathcal{P}, j \neq i}$ conditional on $(Z_i, |\mathcal{N}_i^*|)$ are i.i.d. across $j$. It might be violated if there exist some exogenous covariates that enter the network formation process and also influence the treatment assignment. It is because, if one would like to believe the homophily effects in the network formation, i.e. individuals are more likely to establish a link if they are similar, then unit $i$’s characteristic and her peers’ identity together will reveal relevant information on the characteristics of her peers and non-peers. Then, conditioning on the covariate $Z_i$, the i.i.d. of $\{D_j\}_{j \in \mathcal{P}, j \neq i}$ would fail to hold.

4.3.2 Unconfounded Treatment

Given the discussion in Section 4.3.1, it is apparent that there exist two settings where the fully randomized treatment can be relaxed to allow the treatment being randomized based on individuals’ characteristics. The first setting accounts for the homophily effects, while requires the existence of a subset of individual’s characteristics $Z_{1,i} \subset Z_i$ such that $Z_{1,i}$ does not affect the network formation. Then, the treatments can be randomly assigned based on $Z_{1,i}$. For example, in the microfinance program, interventions can be allocated randomly give participants’ social status (like occupation), which is unlikely to determine their network measured by “go to pray together”, as people with whom a individual goes to pray should rely closely on their religion, gender and caste, instead of the social status. The second setting suits situations where it is reasonable to believe that the network is formed following the random graph model of Erdös and Rényi (1959), i.e. each link is formed independently with the same probability. Then, the treatments can be randomly assigned based on $Z_i$. Studying the consequences of relaxing this condition to adopt more general unconfounded treatment assignment is an interesting area for future exploration.

4.3.3 Directed or Weighed Links

Readers may observe that the analysis so far does not require the network $\mathcal{N}_i^*$ to be undirected, and the generalization to directed network is straightforward. If the unweighted restriction is also relaxed, then the spillover effects can be captured by $S_i^* = \sum_{j \in \mathcal{N}_i^*} \pi(Z_{1,j}) D_j$ with $Z_{1,j}$ being a subset of $Z_i$ and $\pi(\cdot)$ being a known weighting function. For the same reason discussed in Section 4.3.1, it is required that $Z_{1,j}$ should not impact the network formation. For example, in the microfinance program, a unit with a higher degree of financial literacy might be assigned a higher weight. While, the financial literacy is unlikely to have direct impacts on the network of women.
from South India, which is collected before the microfinance program is implemented.

5 Asymptotic Properties

This section is organized as below. Subsection 5.1 introduces the notion of dependency neighborhood, which is the stepstone to establish the asymptotic properties of the proposed estimation approach in this paper. Subsection 5.2 presents the nonparametric kernel estimation and subsection 5.3 discusses the semiparametric estimation procedure.

5.1 Dependency Neighborhoods

Let \( W_i \) be some observable random variable or vector. For each unit \( i \) and sample size \( N \), the dependency neighborhood of unit \( i \), denoted by \( \Delta(i, N) \), is such that \( \Delta(i, N) \subset \{1, 2, ..., N\} \), \( i \in \Delta(i, N) \) and satisfies conditions in Assumption 5.1 below. Any unit \( j \) such that \( j \in \Delta(i, N) \) is referred to as unit \( i \)' dependent neighbor, while the dependent neighbor is not necessary a network neighbor. Following Chandrasekhar and Jackson (2016), I define the dependency neighborhood by restricting the relative correlation of \( \{W_i\}_{i=1}^N \) inside and outside of \( \{\Delta(i, N)\}_{i=1}^N \). For any integrable function \( b \), denote the sum of covariance of all pairs of units in each other’s dependency neighborhood as

\[
\Sigma^b_N = \sum_{i=1}^N \sum_{j \in \Delta(i, N)} Cov(b(W_i), b(W_j)),
\]

which captures the variation of \( b(W_i) \) of all \( N \) units and the dependence across all pairs \( (b(W_i), b(W_j)) \) with \( j \) being dependent neighbor of \( i \). The assumption below characterizes two principle properties of the dependency neighborhood.

Assumption 5.1 (Dependency Neighborhood) For any integrable function \( b : \Omega_W \mapsto \mathbb{R}^d_b \),

(a) \( \Sigma^b_N \to \infty \) as \( N \to \infty \);

(b) \( \sum_{i=1}^N \sum_{j \notin \Delta(i, N)} Cov(b(W_i), b(W_j)) = o(\Sigma^b_N) \).

Condition (a) ensures that the dependence among units in each other’s dependent neighborhood does not vanish and contains sufficient information which is necessary when deriving asymptotic properties for statistics that is constructed using these dependent variables. Intuitively, condition (b) requires that \( \Delta(i, N) \) is a collection of units that have relatively high correlation with the ego unit \( i \) compared to those in its complement. The objects in \( \Delta(i, N) \) may not be unique, because it is defined asymptotically. In addition, the size of each \( \Delta(i, N) \) may change (generally expand) as sample size increases.
As mentioned in Chandrasekhar and Jackson (2016), there is substantial freedom in constructing these sets in different studies. For example, the dependency neighborhoods can be defined based on individuals’ participation in common actions, affiliation, and social events regardless of their network interactions; individuals’ identities (groups) that lead to strong social norms and clear barriers across groups, such as caste, tribe or race (Currarini et al., 2009, 2010); or simply defined via social or geographical location, like occupation, class, school, village or community. Essentially, the dependency neighborhoods \( \{ \Delta(i, N) \}_{i=1}^{N} \) can be understood as defined by individual’s exogenous attributes and the analysis in this paper is conducted conditional on these attributes: that is, the dependent neighborhoods are treated as non-stochastic.

5.2 First Step Kernel Estimation

The nonparametric kernel estimation of density function has been extensively studied, see Newey and MacFadden (1994), Newey (1994) and Li and Racine (2007) among others. To ease illustration, we denote the observable variable by \( W_i = (W_i^c, W_i^d)' \) where \( W_i^c \) represents the vector containing continuous variables and \( W_i^d \) is the vector containing discrete variables. Denote the support of \( W_i^c \) and \( W_i^d \) as \( \Omega_{W^c} \) and \( \Omega_{W^d} \), respectively. Note that \( W_i \) may be used to denote different observable variables at different places. For a bandwidth \( h > 0 \) and \( \forall w = (w^c, w^d)' \in \Omega_{W^c,W^d} \), denote

\[
K(W_i^c, w^c) = \frac{1}{h^Q} \prod_{q=1}^{Q} \kappa \left( \frac{W_{i,q}^c - w_{q}^c}{h} \right),
\]

with \( \kappa(\cdot) \) being a univariate kernel function and \( Q \) is the dimension of vector \( W_i^c \). Denote the nonparametric kernel estimator of the probability function of interest as

\[
\hat{f}_{W_i}(w) = \frac{1}{N} \sum_{i=1}^{N} K(W_i^c, w^c) 1 \left[ W_i^d = w^d \right]. \tag{10}
\]

Given (10), the estimators for a nuisance parameter \( \gamma_N \) is

\[
\hat{\gamma}_N = \left[ \hat{f}_{N_i|N_i}, Y_i, Z_i^c, \hat{f}_{N_i|N_i}, Z_i^d, \hat{f}_{S_i|N_i}, Z_i, \hat{f}_{S_i|N_i}, Z_i, \hat{f}_{Z_i}, Z_i \right]'.
\]

Let \( \gamma^0 \) be the true value of \( \hat{\gamma}_N \). Assumption below provides sufficient conditions for the uniform convergence of the nonparametric kernel estimation.

**Assumption 5.2** Let \( W_i^c = (Y_i, Z_i^c)' \) and \( W_i^d = (D_i, Z_i^d, S_i, |N_i|, \tilde{S}_i, |\tilde{N}_i|)' \).

(a) \( \Omega_{W^c} \subset \mathbb{R}^Q \) is a compact and convex set and the cardinality of \( \Omega_{W^d} \) is finite.

\(^{23}\)For expositional simplicity, we restrict the bandwidth for all continuous variables to be the same. In practice, our method also allows for different bandwidths, while a data-driven method for bandwidth selection is not the focus of this paper.
(b) Each element in $\gamma^0$ is bounded and continuously differentiable in $w^c$ to order two with bounded derivatives on an open set containing $\Omega_{w^c}$.

(c) $\kappa(\cdot)$ is nonnegative kernel function and is differentiable with uniformly bounded first derivative. In addition, for some constant $K_1, K_2 > 0$

$$\int \kappa(v)dv = 1, \quad \kappa(v) = \kappa(-v), \quad \int v^2 \kappa(v)dv = K_1, \quad \int \kappa(v)^2 dv = K_2.$$ 

(d) $h \to 0$, $Nh^Q \to \infty$, $\ln(N)/(Nh^Q) \to 0$, as $N \to \infty$.

(e) Let $\tilde{r}_N = \sup_{1 \leq i \leq N} |\Delta(i, N)|$. The cardinalities of dependency neighborhoods satisfy

$$\tilde{r}_N \left[\ln(N)/(Nh^Q)\right]^{1/2} = O(1), \quad \frac{1}{N} \sum_{i=1}^{N} |\Delta(i, N)|^2 = O(1).$$

Conditions (a) and (b) state the regularity conditions of the support and of the data distribution. Conditions (c) and (d) describe features of the kernel function and the bandwidth, which are standard for nonparametric kernel estimation. In addition, to accommodate the dependence across units, we need to impose restrictions on the dependency neighborhood. Condition (e) allows the situation where sufficiently large number of units possess increasing number of dependent neighbors, say $O(\ln(N)/Nh^Q)^{1/2}$ units having $O([Nh^Q/\ln(N)]^{1/4})$ dependent neighbors and the rest having bounded number of dependent neighbors. Note that although we require the sparse network, the number of dependent neighbors may increase with sample size.

To address issues arising from the dependence between observations, we accommodate the method of Masry (1996), which is based on the approximation theorems developed by Bradley et al. (1983), to approximate dependent random variables by independent ones. For any given sample size $N$, partition the index set $\{1, 2, ..., N\}$ into $q_N$ mutually exclusive subsets $S_1, ..., S_{q_N}$ with $\bigcup_{1 \leq i \leq q_N} S_i = \{1, 2, ..., N\}$. The subscript of $q_N$ means that it may go to infinity as $N \to \infty$.

Let $i_1 = 1$ and $S_1 = \Delta(i_1, N)$. For any given sample size $N$, find $i_2$ to be the unit that is not a dependent neighbor of $i_1$ but has the largest number of common dependent neighbors with $i_1$, i.e. $i_2$ satisfying

$$\left\{ i_2 \notin \Delta(i_1, N), \left|\Delta(i_1, N) \cap \Delta(i_2, N)\right| \geq \left|\Delta(i_1, N) \cap \Delta(j, N)\right| \text{ for } \forall j \notin \Delta(i_1, N) \right\}.$$ 

Then, $i_2$ can be understood as the most correlated non-dependent neighbor of $i_1$. Given $i_2$ we can set $S_2 = \Delta(i_2, N)/\Delta(i_1, N)$. If there are more than one units, say two units, satisfy the above

\footnote{For any given sample size $N$, the partition exists and every sampled unit is included in exactly one set of $S_1, ..., S_{q_N}$. The largest possible value of $q_N$ will be $N$.}
requirements simultaneously, we can combine them as one set. Similarly, we can repeat the process and find $i_k \ (k \leq q_N)$ as the unit that satisfies

$$
\begin{align*}
\left\{ i_k & \not\in \bigcup_{1 \leq l \leq k-1} \Delta(i_l, N), \quad \bigg| \bigcup_{1 \leq l \leq k-1} \Delta(i_l, N) \bigcap \Delta(i_k, N) \bigg| \\
& \geq \bigg| \bigcup_{1 \leq l \leq k-1} \Delta(i_l, N) \bigcap \Delta(j, N) \bigg| \\
& \text{for } \forall j \not\in \bigcup_{1 \leq l \leq k-1} \Delta(i_l, N) \right\},
\end{align*}
$$

which implies that $i_k$ is the most correlated non-dependent neighbor of \{\textit{i}_1, \textit{i}_2, ..., \textit{i}_{k-1}\}. Then, let $S_k = \Delta(i_k, N) / \bigcup_{1 \leq l \leq k-1} S_l$. Given the above partition, define the dependence coefficient

$$
\alpha_k = \sup_{A \in \mathcal{F}^{k-2}, B \in \mathcal{F}_k} |Pr(A, B) - Pr(A)Pr(B)|,
$$

where $\mathcal{F}^{k-2} = \sigma (\{W_i, \ i \in \bigcup_{1 \leq l \leq k-2} S_l\})$ and $\mathcal{F}_k = \sigma (\{W_i, \ i \in S_k\})$ for $k = 1, 2, ..., q_N$. The coefficient $\alpha_k$ measures the dependence strength between observations in the two sets $\bigcup_{1 \leq l \leq k-2} S_l$ and $S_k$. Without loss of generality, suppose $q_N$ is an even integer.

**Assumption 5.3** Let $L_N = [N/(\ln(N)h^{Q+2})]^{Q/2}$. The dependence coefficient $\alpha_k$ satisfies

$$
\Psi_N := L_N \left( \frac{N}{\ln(N)} \right)^{1/2} \sum_{k=1}^{q_N} \alpha_k, \quad \sum_{N=1}^{\infty} \Psi_N < \infty.
$$

Assumption 5.3 controls the asymptotic dependence among observables and is akin to the mixing coefficient decaying condition in a setting with network-induced data dependence. It ensures that uniform convergence still holds even when the large scale of dependency among samples exists and it allows for nonzero dependence outside the dependency neighborhood. Similar assumption is exploited in Masry (1996) to restrict the time series data, and in Sävje (2019) to control the dependence between network measurement errors.

**Lemma 5.1** Assumption 5.3 is satisfied, if either one of the following two conditions holds.

(i) For any $i \in \mathcal{N}$, $W_j \perp W_i$ if $j \in \Delta(i, N)$ and $l \not\in \Delta(i, N)$;

(ii) For $i_1, i_2, ..., i_k$ constructed as described above, $W_j \perp W_i$ if $j \in \bigcup_{r=1}^k \Delta(i_r, N)$ and $l \not\in \left\{ \bigcup_{r=1}^k \Delta(i_r, N) \bigcup \Delta(i_{k+1}, N) \right\}$, where $i_{k+1}$ is the unit having the largest number of common dependent-neighbors with units $i_1, i_2, ..., i_k$.

The proof of Lemma 5.1 is trivial therefore omitted. Condition (i) indicates that Assumption 5.3 holds if dependence neighborhoods are independent clusters. Suppose $j \in \Delta(i_k, N)$, then
condition (ii) requires that \( W_j \) and \( W_l \) are independent, if unit \( l \) is not a dependent neighbor; firstly, of unit \( i_k \); secondly, of the units \( i_1, \ldots, i_{k-1} \) with whom unit \( i_k \) has the largest number of common dependent neighbors; lastly, of the unit \( i_{k+1} \) who has the largest number of common dependent neighbors with units \( i_1, \ldots, i_k \). Condition (ii) is much weaker than condition (i) and it does not contradict Assumption 5.2 (e). Because such an independence in (ii) is required across dependent neighborhoods, and it imposes no restrictions on the number of units in each dependent neighborhood.

**Theorem 5.2** Let Assumptions 5.2 and 5.3 hold, then \( \left\| \hat{\gamma}_N - \gamma^0 \right\|_\infty = O_p \left( \left[ \ln(N)/(Nh^Q) \right]^{1/2} + h^2 \right) \).

The uniform convergence rate of the kernel estimation in Theorem 5.2 is consistent with that of the conventional kernel estimation under i.i.d. or strong mixing data settings. See e.g. Newey (1994), Li and Racine (2007) and Masry (1996).

Let \( \hat{\phi}_N := \phi(\hat{\gamma}_N) \) represent the estimator of the latent distribution function \( f_{\mathcal{S}^* \mid \mathcal{N}^*_i \mid \mathcal{D}_i \mid \mathcal{S}_i \mid \mathcal{Z}_i \mid \mathcal{N}_i} \). According to Proposition 4.2, we can obtain a plug-in estimator \( \hat{\phi}_N \) via replacing the distributions on the right hand side of (3) by their kernel estimators based on \( \hat{\gamma}_N \) in (10). Denote \( \phi^0 = \phi(\gamma^0) \) as the true latent distribution function. Given the uniform convergence of \( \hat{\gamma}_N \) in Theorem 5.2, we only need to consider the convergence of \( \hat{\phi}_N \) in a small neighborhood of \( \gamma^0 \).

**Corollary 5.3** Let Assumption 3.1-3.4 and 4.1-4.5 hold. Under assumptions in Theorem 5.2, suppose that there exists a constant \( \epsilon > 0 \) such that \( f_{\mathcal{N}_i \mid \mathcal{Z}_i} > \epsilon \). Then, for \( \eta \to 0 \) as \( N \to \infty \),

\[
\sup_{\|\hat{\gamma}_N - \gamma^0\|_\infty \leq \eta} \left\| \hat{\phi}_N - \phi^0 \right\|_\infty = O_p(\|\hat{\gamma}_N - \gamma^0\|_\infty).
\]

### 5.3 Semiparametric Estimation

In this subsection, we study the estimation of CASF \( m^* \) by simplifying \( m^* = m^*(\cdot; \theta) \) as known function up to unknown parameter \( \theta \in \Theta \subset \mathbb{R}^{d_\theta} \). Consequently, \( m_i(\cdot) = m(\cdot) = m(\cdot; \theta, \phi) \) is also known up to \((\theta, \phi)\).\(^{25}\) Note that the identification of \( m^* \) studied in Section 4 does not rely on such an simplification. More importantly, imposing such a parametric structure on \( m^* \) still allows flexible heterogeneity of the treatment and spillover effects, which can be captured by interactions of \( D_i \) and \( S_i \), with covariate \( Z_i \) and network degree \( |\mathcal{N}_i| \), as well as using their polynomials.

#### 5.3.1 Consistency

For notational simplicity, let \( X^*_i := (D_i, S^*_i, Z_i, |\mathcal{N}^*_i|)' \) and \( X_i := (D_i, S_i, Z_i, |\mathcal{N}_i|)' \) with support \( \Omega_{X^*} \) and \( \Omega_X \), respectively. In addition, denote \( T^*_i = (S^*_i, |\mathcal{N}^*_i|)' \). Let \( x^*_i := (d, s^*_i, z, n^*_i) \) with

\(^{25}\)Based on Theorem 4.5, we know that \( m_i(\cdot) \) is identical across all \( i \). Thus, we can suppress the subscript \( i \), i.e. \( m_i(\cdot) = m(\cdot) \). In addition, \( m(\cdot) = m(\cdot; \theta, \phi) \) is because of \( m(\cdot) \) being a function of the CASF \( m^*(\cdot; \theta) \) and nuisance parameter \( \phi \).
\[ t_j^* = (s_j^*, n_j^*) \in \Omega_{S^*,|N^*|}, \text{ and } j \in \{1, 2, ..., K_T\} \] represents the lexicographical ordering of the possible values of \( T_k \) as described in (B.29). Similarly, let \( x_j := (d, s_j, z, n_j) \) with \( t_j := (s_j, n_j) \in \Omega_{S,|N|} \).

By definition of \( m(\cdot; \theta, \phi) \), the following moment condition holds:

\[ E \left[ Y_i - m(X_i; \theta, \phi) | X_i \right] = 0. \]

From Proposition 3.1, \( m(\cdot; \theta, \phi) \) and the CASF \( m^*(\cdot; \theta) \) are linked through the formula \( m(x; \theta, \phi) = \sum_{j=1}^{K_T} m^*(x_j^*; \theta) f_{T_j^* | X_i=x}(t_j^*) \). Recall that \( X_i \) is identically distributed for all \( i \) under assumptions in Section 3. Denote the objective function and its sample analogue as

\[ \mathcal{L}(\theta, \phi) = E \left\{ \tau_i [Y_i - m(X_i; \theta, \phi)]^2 \right\}, \quad \text{and} \quad \mathcal{L}_N(\theta, \phi) = \frac{1}{N} \sum_{i=1}^{N} \tau_i [Y_i - m(X_i; \theta, \phi)]^2, \]

where \( \tau_i := \tau(X_i) \) is non-negative weight. Following Newey (1994), we use the weight function \( \tau \) to focus the optimization problem on regions where the kernel estimation is relatively reliable.\(^{26}\)

Then, \( \theta \) can be estimated by minimizing \( \mathcal{L}_N(\theta, \phi_N) \) given the estimator \( \hat{\phi}_N \) from Theorem 5.2:

\[ \hat{\theta}_N = \arg \min_{\theta \in \Theta} \mathcal{L}_N(\theta, \hat{\phi}_N). \] (11)

Let \( W_i = (Y_i, X_i')' \) be the vector containing all the observed variables and \( w = (y, x')' \in \Omega_W \).

**Assumption 5.4**

\( (i) \) \( \Theta \subset \mathbb{R}^{d_\Theta} \) is compact, \( \theta^0 \in \text{int}(\Theta) \) and \( \theta^0 \) is identifiable from the weighted conditional moment function \( \mathcal{L}(\theta, \phi^0) = 0. \)

\( (ii) \) \( \tau(\cdot) \) is nonnegative and \( \sup_{x \in \Omega_X} |\tau(x)| < C \) for some constant \( C > 0. \)

\( (iii) \) \( m^*(x; \theta) \) is continuous in \( \theta \) for all \( x \in \Omega_X \), and is an integrable function of \( X_i \) for all \( \theta \in \Theta. \)

\( (iv) \) Denote the random variable \( x_{i,j}^* = (D_i, s_j^*, Z_i, n_j^*) \) with \( t_j^* = (s_j^*, n_j^*) \in \Omega_{T^*} \) and \( j = 1, 2, ..., K_T. \)

There exists a function \( h_1(x) \) such that \( |m^*(x; \theta)|^2 \leq h_1(x) \) for all \( \theta \in \Theta, \) and \( E[h_1(x_{i,j}^*)] < \infty \) for all \( j = 1, 2, ..., K_T. \)

\( (v) \) Let \( e(w, \theta) := \tau(x)[y - m(x; \theta, \phi^0)]^2 \) and \( e_i(\theta) := e(W_i, \theta). \) For any given constant \( \eta > 0, \) denote \( U_i(\theta, \eta) = \sup_{\theta' \in \Theta, \|\theta - \theta'\| < \eta} |e_i(\theta') - e_i(\theta)|. \) There exists a function \( h_2(w) \) such that \( |e(w, \theta)| \leq h_2(w) \) for all \( \theta \in \Theta \) and \( E[h_2(W_i)] < \infty. \) In addition, \( \sup_{\theta \in \Theta} E[|e_i(\theta)|^{2+\delta}] < C \) for some constants \( \delta > 0 \) and \( C > 0. \)

\(^{26}\)Hu (2008) also adopts the weight function and set it to be a fixed trimming \( \tau(x) = 1_{[x \in X]} \) with \( X \subset \Omega_X \) a fixed set. In this paper, we follow Hu (2008) to use the fixed trimming weight function. Other types of weight functions such as data-driven weight functions or methods for selection of weight functions are out of the scope of this paper.
Theorem 5.4 (Consistency) Let assumptions in Theorem 4.6 hold. Under Assumptions 5.1-5.4, we have $\|\hat{\theta}_N - \theta^0\| = o_p(1)$.

5.3.2 Asymptotic Normality

To show asymptotic normality of the estimator $\hat{\theta}_N$, we need to account for the presence of the nuisance parameter $\phi$ and the various forms of dependence arising from the mismeasured network data, which requires a significant generalization of the classical CLT. In particular, the oft-used CLT developed for mixing processes does not work for our purpose, as they rely on some particular ordering structure to measure the “distance” between units. Therefore, I adopt and extend the univariate CLT for network data proposed by Chandrasekhar and Jackson (2016) to multivariate setting, see Lemma E.6 in the Appendix, which will be applied in this section to derive the asymptotic normality for $\hat{\theta}_N$.

Let $g(W_i; \theta, \phi) = \tau_i[Y_i - m(X_i; \theta, \phi)]\frac{\partial m(X_i; \theta, \phi)}{\partial \theta}$. Then, from the first order condition of the optimization problem (11), $\hat{\theta}_N$ solves $\frac{1}{N} \sum_{i=1}^{N} g(W_i; \hat{\theta}_N, \hat{\phi}_N) = 0$. Then, by the mean value theorem we can obtain

$$0 = \frac{1}{N} \sum_{i=1}^{N} g(W_i; \hat{\theta}_N, \hat{\phi}_N) = \frac{1}{N} \sum_{i=1}^{N} g(W_i; \theta^0, \hat{\phi}_N) + \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta'} (\hat{\theta}_N - \theta^0),$$  \hspace{1cm} (12)

where $\bar{\theta}_N$ is between $\hat{\theta}_N$ and $\theta^0$. If $\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta'}$ is invertible, rearranging (12) leads to

$$\sqrt{N}(\hat{\theta}_N - \theta^0) = \left[ \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta'} \right]^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(W_i; \theta^0, \hat{\phi}_N).$$

Let us introduce some useful notations. Recall that $\phi(\cdot) = \phi(\cdot; \gamma)$. We set $t := (t_1, \ldots, t_{K_T})'$ and $\phi(t; \gamma) = [f_{T_1'|X_i}(t_1), \ldots, f_{T_{K_T}'|X_i}(t_{K_T})]'. $ Let $1_{d_w}$ be a $d_w \times 1$ vector of ones. Denote $\nu(w; \theta, \gamma) = E\left[ \tau(X_i) \frac{\partial}{\partial \theta} \mathcal{R}(W_i; \theta, \phi) \frac{\partial \phi(t; \gamma)}{\partial \gamma} \right]_{\gamma = \gamma(w)} 1_{d_w} \left| w \right]$ and $\delta(W_i; \theta, \gamma) := \nu(W_i; \theta, \gamma) - E[\nu(W_i; \theta, \gamma)]$, where

$$\mathcal{R}(W_i; \theta, \phi) = \begin{bmatrix} [Y_i - m(X_i; \theta, \phi)] m^*(x_{i,1}^*; \theta) \\ \vdots \\ [Y_i - m(X_i; \theta, \phi)] m^*(x_{i,K_T}^*; \theta) \end{bmatrix}'.$$

To simplify notation, denote $\nu(W_i) := \nu(W_i; \theta^0, \gamma^0)$ and $\delta(W_i) := \delta(W_i; \theta^0, \gamma^0)$.

Assumption 5.5

(i) $m^*(x; \theta)$ is continuously differentiable in $\theta$ up to order three with bounded third order deriv-
Assumption 5.6

Condition (iii) ensures that the limit of the Hessian matrix exists and is invertible.

Denote the dependence neighborhoods covariance matrix

Lemma 5.5 (i) and (ii) introduce regularity conditions on the smoothness of the CASF $m^*(\cdot, \theta)$. Condition (iii) ensures that the limit of the Hessian matrix exists and is invertible.

Assumption 5.6

(i) $\frac{N^{1/2} \log(N)}{(Nh^d)} \rightarrow 0$ and $Nh^d \rightarrow 0$ as $N \rightarrow \infty$.

(ii) $\nu(w; \theta, \gamma) = \nu(w^e, w^d; \theta, \gamma)$ is continuously differentiable in $w^e$ almost everywhere and satisfies

$\sum_{w^d} \int \nu(w) dw^c < \infty$. In addition, $||Var[\nu(W_i)]|| < \infty$.

Assumption 5.6 implies that the convergence rate of $\hat{\gamma}_N$ is faster than $N^{1/4}$. It is a typical restriction on the bandwidth to guarantee the asymptotic normality for semiparametric two-step estimators that depend on kernel density, e.g. Newey and MacFadden (1994).

We first show that the $d_\theta \times d_\theta$ Hessian matrix $\frac{1}{N} \sum_{i=1}^N \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta^r}$ converges in probability uniformly.

Lemma 5.5 Let the assumptions in Theorem 5.4 hold.

(a) Under Assumption 5.5, for a small enough $\eta \rightarrow 0$ as $N \rightarrow \infty$, we have

$$\sup_{\|\hat{\gamma}_N - \gamma^0\| < \eta} \left\| \frac{1}{N} \sum_{i=1}^N \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta^r} - E \left[ \frac{\partial g(W_i; \theta^0, \phi^0)}{\partial \theta^r} \right] \right\| = o_p(1).$$

(b) Under Assumption 5.6, we can get

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N g(W_i; \theta^0, \phi^0) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left[ g(W_i; \theta^0, \phi^0) + \delta(W_i) \right] + o_p(1).$$

Denote the dependence neighborhoods covariance matrix

$$\Sigma_N = \sum_{i=1}^N \sum_{j \in \Delta(i,N)} E \left\{ \left[ g(W_i; \theta^0, \phi^0) + \delta(W_i) \right] \left[ g(W_j; \theta^0, \phi^0) + \delta(W_j) \right] \right\}.$$
To ease the notations, denote the $d_a \times 1$ vector $\tilde{g}_i = g(W_i; \theta^0, \phi^0) + \delta(W_i)$ with $\tilde{g}_i = (\tilde{g}_{i1}, ..., \tilde{g}_{id_a})'$. Then, $\Sigma_N^\# = \sum_{i=1}^{N} \sum_{j \in \Delta(i, N)} E[\tilde{g}_i \tilde{g}_j']$. In addition, by notation abuse, let $S_i^c = \sum_{j \in \Delta(i, N)} \tilde{g}_j$ be the covariance outside the dependency neighborhoods. For any vector $a$, let $a \geq 0$ mean that each of its entry is nonnegative. For any matrix $A = \{a_{ij}\}$, vec$(A)$ denotes the vectorization of $A$ and $|A| = \{|a_{ij}|\}$.

**Assumption 5.7**

(i) For all $i \in \mathcal{P}$, $\Delta(i, N)$ is symmetric such that $j \in \Delta(i, N)$ if and only if $i \in \Delta(j, N)$.

(ii) There exists a finite, strictly positive-definite and symmetric matrix $\Omega \in \mathbb{R}^{d_a} \times \mathbb{R}^{d_a}$ such that $\frac{\|\Sigma_N^\# - \Omega\|}{\Sigma_N^\#} \to 0$ as $N \to \infty$.

(iii) The following conditions hold for $\{g_i\}_{i=1}^{N}$.

(a) $\left\| \sum_{i=1}^{N} \sum_{j,k \in \Delta(i, N)} E\left[|\text{vec}(\tilde{g}_i \tilde{g}_j') \tilde{g}_k'\right] \right\|_{\infty} = o\left(\left\|\Sigma_N^\#\right\|_{3/2}\right)$;

(b) $\left\| \sum_{i,k=1}^{N} \sum_{j \in \Delta(i, N)} \sum_{l \in \Delta(k, N)} E\left[(\tilde{g}_i \tilde{g}_j' - E[\tilde{g}_i \tilde{g}_j'])' (\tilde{g}_k \tilde{g}_l' - E[\tilde{g}_k \tilde{g}_l'])\right] \right\|_{\infty} = o\left(\left\|\Sigma_N^\#\right\|_{2}\right)$;

(c) $\left\| \sum_{i=1}^{N} \sum_{j \in \Delta(i, N)} \text{Cov} (\tilde{g}_i, \tilde{g}_j) \right\|_{\infty} = o\left(\left\|\Sigma_N^\#\right\|_{\infty}\right)$;

(d) $E \left[\tilde{g}_i S_i^c | S_i^c\right] \geq 0$ for all $i \in \mathcal{P}$.

Assumption 5.7 (i) guarantees that the covariance matrix $\Sigma_N^\#$ is symmetric. Condition (ii) ensures that the samples possess sufficiently large variation so that the CLT holds. Meanwhile, it requires the limit of $\Sigma_N^\#/N$ being a constant matrix $\Omega$, instead of varying with sample size, which imposes restriction on the allowable divergence rate of $\Sigma_N^\#$ to some degree. Similar assumptions are used to study asymptotic properties of covariance matrix estimator, e.g. in White and Domowitz (1984).

Moreover, Assumption 5.7 (iii) is crucial for multivariate normal approximation under the dependency neighborhood structure, which further guarantees the CLT. Similar assumption is used in Chandrasekhar and Jackson (2016) to establish the asymptotic normality. We extend the assumption to accommodate general multivariate random vectors without imposing any restrictions on their support. In particular, conditions (iii) (a) and (b) restrict the rate of dependency between the dependency sets, while (c) limits the rate of dependency outside the dependence sets. Besides, condition (d) states that on average, units outside each other’s dependency neighborhood do not tend to interact negatively.\footnote{Chandrasekhar and Jackson (2016) also use condition that is similar to Assumption 5.7 (iii) (d) to ease their proof. We note that the condition (iii) (d) is not necessary for the asymptotic normality in this paper and can be replaced by more primitive assumptions.}
Theorem 5.6 (Asymptotic Normality) Suppose assumptions in Theorem 5.4, Assumptions 5.5-5.7 hold. Then
\[ \sqrt{N}(\hat{\theta}_N - \theta^0) \xrightarrow{d} \mathcal{N}(0, H^{-1}\Omega H^{-1}), \]
where \( H = E[\partial g(W_i; \theta^0, \phi^0)/\partial \theta'] \) and \( \mathcal{N} \) represents normal distribution.

Notably, the consistency and asymptotic normality of \( \hat{\theta}_N \) only require the existence of dependency neighborhoods, rather than the accurate knowledge of them. Nevertheless, if the knowledge of dependency neighborhoods \( \{\Delta(i, N)\}_{i=1}^N \) is available, it suffices a consistent variance estimator. Given that the form of \( \delta(w) \) is known, following Newey and MacFadden (1994), we construct the estimator of \( \delta(W_i) \) by substituting \( (\hat{\theta}_N, \hat{\gamma}_N) \) for \( (\theta^0, \phi^0) \), i.e. \( \hat{\delta}(W_i) := \delta(W_i; \hat{\theta}_N, \hat{\gamma}_N) \). The corollary below provides a consistent estimator of the variance-covariance matrix \( H^{-1}\Omega H^{-1} \), which is essential when constructing asymptotic confidence intervals and conducting hypothesis tests.

Corollary 5.7 (Variance Estimator) Under assumptions in Theorem 5.6, we can get
\[ \|H_N^{-1}\hat{\Omega}_N H_N^{-1} - H^{-1}\Omega H^{-1}\| \xrightarrow{P} 0 \text{ as } N \to \infty, \]
where
\[ \hat{H}_N = \frac{1}{N} \sum_{i=1}^N \partial g(W_i; \hat{\theta}_N, \hat{\phi}_N), \]
\[ \hat{\Omega}_N = \frac{1}{N} \sum_{i=1}^N \sum_{j \in \Delta(i, N)} \left[ g(W_i; \hat{\theta}_N, \hat{\phi}_N) + \hat{\delta}(W_i) \right] \left[ g(W_j; \hat{\theta}_N, \hat{\phi}_N) + \hat{\delta}(W_j) \right]',. \]

It is worth to note that the consistent variance estimator \( H_N^{-1}\hat{\Omega}_N H_N^{-1} \) is robust to mild degree of misspecification of the dependency neighborhoods. For example, if there is only finite units whose dependency neighborhoods are misspecified, the variance estimator is still consistent due to the consistency of \( (\hat{\theta}_N, \hat{\phi}_N, \hat{\gamma}_N) \) and Assumption 5.7 (iii)(c). Moreover, if the knowledge of dependency neighborhoods is not available at all, one may resort to the resampling method proposed by Leung (2020a) to conduct inference for the parameter of interest. Rigorous study is left for future research.\(^{28}\)

6 Simulation

In this section, I illustrate the estimation performance by Monte Carlo simulations. A similar data generating process (DGP) and network formation design to those used by Leung (2020b) is

\(^{28}\)The resampling method of Leung (2020a) requires that the first-stage kernel estimation to be \( \sqrt{N} \)-consistent, which is satisfied if outcome and covariates \( (Y_i, Z_i) \) are discrete random variables.
adopted, where I introduce measurement errors to the observable networks. Consider the following DGP of the outcome variable:

$$Y_i = \theta_0 + \theta_1 D_i + \theta_2 D_i | N_i^r | Z_i + \theta_3 S_i^r + \theta_4 S_i^{r^2} + \theta_5 S_i^r Z_i + \theta_6 S_i^r | N_i^r | + \varepsilon_i,$$  \hspace{1cm} (13)

where the i.i.d. treatment $D_i$ is independently generated with the probability of being treated equal 0.3, and the covariate $Z_i \overset{i.i.d.}{\sim} \text{Bernoulli}(0.5)$. In addition, the error term $\varepsilon_i = \varepsilon_i^{idio} + \varepsilon_i^{peer}$ with $\varepsilon_i^{idio}$ being the idiosyncratic disturbance and $\varepsilon_i^{peer} = \sum_{j \in P} A_{ij}^* v_j$ capturing the unobservable peer effects, where $v_j \overset{i.i.d.}{\sim} N(0, 0.5)$. We set parameters $\theta = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)' = (0, 1, 1/3, 1, -1, -1/2, 1)'$.

Throughout this section, we aim to estimate the treatment effects $\tau_d(0, 0, 3)$ and $\tau_d(0, 1, 3)$ (with true value 1 and 2, respectively), and the spillover effects $\tau_s(1, 0, 3)$ and $\tau_s(1, 1, 3)$ (with true values 3 and 2.5, respectively).

Suppose the actual network is formed as following. We randomly allocate units on a $[0, 1] \times [0, 1]$ space and assume their exogenous geographic locations $\rho_i \in [0, 1] \times [0, 1]$ are correctly observed. Assume the links satisfying

$$A_{ij}^* = 1[\beta_1 + \beta_2 (Z_i + Z_j) + \beta_3 d(\rho_i, \rho_j) + \zeta_{ij} > 0] \times 1[i \neq j],$$

where $\zeta_{ij} = \zeta_{ji}$ is a random shock that is i.i.d. across dyads with distribution $N(0, 1)$ and independent of $Z_i$ and $\rho_i$ for all $i$ and $j$. $d(\rho_i, \rho_j)$ indicates the distance between two units

$$d(\rho_i, \rho_j) = \begin{cases} 0, & \text{if } r^{-1}\|\rho_i - \rho_j\|_1 \leq 1, \\ \infty, & \text{otherwise} \end{cases},$$

with the scaling constant $r = (r_{deg}/N)^{1/2}$ to guarantee the network sparsity and the parameter $r_{deg}$ controls the average degree: the larger value of $r_{deg}$ is, the larger average degree $E[|N_i^r|]$ will be. We consider two levels of sparsity $r_{deg} = 5$ and 8. Set $\beta = (\beta_1, \beta_2, \beta_3)' = (-0.25, 0.5, -1)$. Key statistics of the latent network $A^* = \{A_{ij}^*\}_{i,j=1}^N$ are summarized in Table 1. Given the true network formed as described above, suppose two self-reported and mismeasured network proxies are available for all units: for $i, j = 1, 2, ..., N$,

$$A_{ij} = \omega_i [U_{ij} A_{ij}^* + V_{ij} (1 - A_{ij}^*)] + (1 - \omega_i) A_{ij}^*,$$

$$\tilde{A}_{ij} = \bar{\omega}_i [\tilde{U}_{ij} A_{ij}^* + \tilde{V}_{ij} (1 - A_{ij}^*)] + (1 - \bar{\omega}_i) A_{ij}^*,$$

where $\omega_i, U_{ij}, V_{ij}, \bar{\omega}_i, \tilde{U}_{ij}$ and $\tilde{V}_{ij}$ are mutually independent and randomly generated binary indicators, taking value one with probabilities $p\omega, p'U, p'V, p\bar{\omega}, p\tilde{U}, p\tilde{V}$, respectively. In particular, taking $A_{ij}$ as an example, $\omega_i$ indicates whether unit $i$ ever misreports his or her links, and $p\omega$ captures the overall level of misreporting. If unit $i$ misreports, there are two types of classification
errors: if \( U_{ij} = 0 \) then a actually linked pair \((i, j)\) with \( A_{ij}^* = 1 \) is misclassified as unlinked (i.e. false negative), and if \( V_{ij} = 1 \) then a actually unlinked pair \((i, j)\) with \( A_{ij}^* = 0 \) is treated as linked (false positive). Therefore, \( 1 - p^U \) and \( p^V \) are the probability of false negative and false positive, respectively.

Following the design of Leung (2020b), assume the full network is collected for both proxies when they are available, meaning that \( \mathcal{P} = \{1, 2, \ldots, N\} \) and \( |\mathcal{N}| = \sum_{i=1}^{N} A_{ij}, \ S_i = \sum_{i=1}^{N} A_{ij}D_j, |\tilde{\mathcal{N}}| = \sum_{i=1}^{N} \tilde{A}_{ij} \) and \( \tilde{S}_i = \sum_{i=1}^{N} \tilde{A}_{ij}D_j \). Given the DGP design, the dependency neighborhood of each unit \( i \) can be set as a collection of units that are located close to unit \( i \) with distance less than \( r \), equivalently, \( \Delta(i, N) = \{ j \in \{1, 2, \ldots, N\}, \| \rho_i - \rho_j \|_1 \leq r \} \).

We generate data using sample size \( N \in \{1000, 2000, 5000\} \) with replications \( M = 1000 \). In the first-step kernel estimation, bandwidth is set to be \( h = N^{-3/8} \). In what follows, we estimate the treatment effects of interest under two scenarios based on the availability of network information.

Table 1: Statistics of Latent Links

| \( r_{deg} \) | \( \mathcal{N}_i \) | \( S_i \) | total | \( \tilde{\mathcal{N}}_i \) | \( \tilde{S}_i \) | total |
|---|---|---|---|---|---|---|
| \( N \) avg. | max | avg. | max | avg. | max | avg. | max |
| 1k | 5.65 | 15.52 | 1.69 | 7.31 | 5649 | 8.92 | 21.45 | 2.68 | 9.53 | 8919 |
| 2k | 5.73 | 16.36 | 1.72 | 7.90 | 11458 | 9.08 | 22.54 | 2.72 | 10.26 | 18167 |
| 5k | 5.80 | 17.39 | 1.74 | 8.55 | 29018 | 9.23 | 23.78 | 2.77 | 11.07 | 46165 |

Note: statistics reported in this table are the average over 1000 replications.

6.1 Semiparametric Estimation with Two Network Proxies

Set the overall misclassification rates \( p^\omega = \tilde{p}^\omega = 0.6 \). For the first proxy, let \( 1 - p^U \in \{0.2, 0.4\} \) and \( p^V = \delta^V / N \) with \( \delta^V \in \{0.1, 0.5\} \) to permit the network sparsity. For the second proxy, set \( 1 - p^\tilde{U} \in \{0.2, 0.4\} \) and \( p^\tilde{V} = 0 \). Then, the first proxy possesses both false negative and false positive classification errors, and the second one contains no false positive. Table 3 reports the statistics of the two mismeasured network proxies for different misclassification rates. We can see that when \( p^U \) or \( p^\tilde{U} \) is 0.2, the misclassification rates is relatively low, varying around 12% to 17%. While when \( p^U \) or \( p^\tilde{U} \) is set to be 0.4, the misclassification rates become quite high, varying around 24% to 29%. We conduct and compare three estimations:

1. **SPE**: the semiparametric estimation studied in Section 5.3 using two proxies;

and two naive parametric estimations via ordinary least square (OLS) and ignoring potential misclassification errors:

1. **Naive 1**: regression of \( Y_i \) on \((1, D_i, D_i|\mathcal{N}_i|Z_i, S_i, S_i^2, S_iZ_i, S_i|\mathcal{N}_i|)\);
(3) **Naive 2**: regression of $Y_i$ on $(1, D_i, D_i|\tilde{N}_i|Z_i, \tilde{S}_i, \tilde{S}_i^2, \tilde{S}_iZ_i, \tilde{S}_i|\tilde{N}_i|)$.

Table 4 to Table 7 display the estimation results for the treatment and the spillover effects via the above three approaches: SPE, Naive 1 and Naive 2. The bias, the standard deviation (sd), the mean squared error (mse), and the coverage rate of the 95% confidence interval for the true value of the causal parameter (cr) are reported.

For the treatment effect $\tau_d(0,0,3)$ (Table 4), the estimation of three methods are comparable in terms of the mse and the coverage rate (cr). It is reasonable because the treatment status of each ego unit is correctly observed and the network measurement error does not impact the estimation of $\tau_d(0,0,3)$ via the naive methods for units with $Z_i = 0$.

For the treatment effect $\tau_d(0,1,3)$ (Table 5), the spillover effects $\tau_s(1,0,3)$ (Table 6) and $\tau_s(1,1,3)$ (Table 7), their finite sample estimation reveals several patterns. First and most importantly, the bias of the SPE is remarkably lower than the bias of the two naive estimations in most cases, especially when the network degree is relatively small ($r_{deg} = 5$), the misclassification rate is relatively low $(1 - p^U = 1 - p^{\tilde{U}} = 0.2)$, or the sample size is sufficiently large $(N = 5000)$.

In addition, as expected, the bias of the SPE is decreasing as sample size increases for most cases. While, the two naive estimations are biased in all settings, and the bias is quite severe when the misclassification rate is relatively high $(1 - p^U = 1 - p^{\tilde{U}} = 0.4)$ or the network degree is relatively large ($r_{deg} = 8$). Increasing the sample size fails to mitigate the bias of the two naive estimations. For instance, consider the estimation of the spillover $\tau_s(1,0,3)$ under $r_{deg} = 8$ in panel (b) of Table 6. Under low misclassification rate $1 - p^U = 1 - p^{\tilde{U}} = 0.2$, $\delta^V = 0.1$ and $N = 1000$, the bias of SPE (0.076) is 11.6% of the bias of Naive 1 (0.653), and is 9.7% of the bias of Naive 2 (0.780). When sample size increases to $N = 5000$, the bias of SPE (-0.034) becomes to 5.2% of the bias of Naive 1 (0.650) and 4.5% of the bias of Naive 2 (0.753). While, under high misclassification rate $1 - p^U = 1 - p^{\tilde{U}} = 0.4$ and $\delta^V = 0.1$, the naive estimations suffer even severer bias: the biases of both Naive 1 and Naive 2 become doubled compared to those under the low misclassification rate. Although the bias of SPE also increases in cases with high misclassification rate compared to that in cases with low misclassification rate, it diminishes with sample size. Hence, the simulations verify that ignoring the network classification errors would result in non-negligible bias which cannot be eliminated via increasing sample size.

In addition, we can see that the sd and the mse of SPE decrease with sample size. The mse of SPE outperforms the those of Naive 1 and Naive 2 in most cases. Moreover, it is apparent that the coverage rate of the SPE performs better, because it is not only closer to the nominal rate 95% compared to the rates of two naive methods, but also approaching to the nominal rate as sample sizes becomes larger. However, for the two naive approaches, their coverage rates drops rapidly when samples size increases or when the misclassification become worse. Take the spillover effect $\tau_s(1,1,3)$ under $r_{deg} = 8$ (panel (b) in Table 7) as an example, when the misclassification rate is low $1 - p^U = 1 - p^{\tilde{U}} = 0.2$ and $\delta^V = 0.1$, the coverage rate is 11.9% for Naive 1 and is 6.2% for
Naive 2, while it is 93.1% for SPE. When $N = 5000$, it goes down to 0% for both Naive 1 and Naive 2, but increases to 93.8% for SPE.

At last, the accuracy of the SPE decreases as $r_{deg}$ increases, or as the misclassification rate increases. To sum up, the SPE works significantly better than the naive estimations neglecting the network misclassification, especially when the sample size is relatively large. In addition, the asymptotic properties in Section 5 are verified by the simulation results.

### 6.2 Robustness of the Semiparametric Estimation

Two key identification assumptions, i.e. the exclusion restriction and the one type of measurement error, may be violated in some applications. In this section, I consider more empirically important questions: Is SPE robust to the violation of these two assumptions? Does SPE still perform better than the naive estimation if any violation is present? To answer these questions, consider the following two scenarios where the observable networks are generated for sensible departures from either of the two identification conditions:

1. **violating “exclusion restriction”:** generate random error $(U_{ij}^*, V_{ij}^*, \tilde{U}_{ij}^*, \tilde{V}_{ij}^*)$ from a joint normal distribution for all $i, j = 1, 2, ..., N$,

$$
\begin{pmatrix}
U_{ij}^* \\
V_{ij}^* \\
\tilde{U}_{ij}^* \\
\tilde{V}_{ij}^*
\end{pmatrix} = \mathbb{N}\left( \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \varrho & 0 & 0 \\ 0 & 1 & 0 & \varrho \\ \varrho & 0 & 1 & 0 \\ 0 & \varrho & 0 & 1 \end{pmatrix} \right) \right), \quad U_{ij} = 1[\Phi(U_{ij}^*) < 1 - p^U], \quad V_{ij} = 1[\Phi(V_{ij}^*) < p^V], \\
\tilde{U}_{ij} = 1[\Phi(\tilde{U}_{ij}^*) < 1 - \tilde{p}^U], \quad \tilde{V}_{ij} = 1[\Phi(\tilde{V}_{ij}^*) < \tilde{p}^V].
$$

where $\varrho \in \{0.05, 0.1\}$ controls the correlation between the misclassification errors;

2. **violating “one type of measurement error”:** generate $\tilde{V}_{ij}$ via $p_{\tilde{V}} = \delta^V/N$ with $\delta^V \in \{0.05, 0.1\}$; while keeping anything else the same with the design in Section 6.1. Results for the three approaches are reported in Table 8 and Table 9.\textsuperscript{29}

To check the robustness of the SPE method when the exclusion restriction is violated, let us compare the results in Table 4 to Table 7 with their counterparts in Table 8 and Table 9. We can see that the violation of either assumptions aggravates the performance of SPE in most of the cases, but only at a limited degree.

Take the spillover $\tau_s(1, 0, 3)$ as an example. When $r_{deg} = 5$, $N = 5000$ and misclassification rate is relatively low $(1 - p^V = 1 - p_{\tilde{V}} = 0.2, \delta^V = 0.1)$, the bias and the mse of SPE under the point identification condition are 0.073 and 0.049, respectively, with the coverage rate 94.5%. These numbers become to 0.101 (bias), 0.097 (mse) and 92.0% (coverage rate); when the exclusion

\textsuperscript{29}Due to the space limitation, Table 9 only displays the results for cases with relatively large sample size ($N = 5000$), which is sufficient to illustrate the asymptotic performance of the SPE relative to the naive approaches.
restriction fails to hold ($\rho = 0.1$), and become to 0.104 (bias), 0.103 (mse) and 93.3% (coverage rate) when the one type of measurement error condition is moderately violated ($\delta \hat{V} = 0.05$).

The answer of the question that whether SPE still outperforms the naive estimation can be obtained by comparing the results of the SPE and the two naive approaches in Table 8 and Table 9. For the treatment effect $\tau_d(0, 0, 3)$, the bias and the mse of SPE are smaller than those of the two naive methods when the misclassification rate is relatively low ($1 - p^U = 1 - p^{\hat{U}} = 0.2$); while the SPE produces slightly larger bias compared to that of the two naive methods when the misclassification rate is relatively high ($1 - p^U = 1 - p^{\hat{U}} = 0.4$). For the treatment effect $\tau_d(0, 1, 3)$, the spillover effects $\tau_s(1, 0, 3)$ and $\tau_s(1, 1, 3)$, in almost all cases the bias and the mse of the SPE are better than those of the two naive estimation. Notably, the coverage rate of the SPE significantly surpasses that of the naive estimations. For example, consider the case where $r_{deg} = 8$ with low misclassification rate ($1 - p^U = 1 - p^{\hat{U}} = 0.2$). If the exclusion restriction is violated, the coverage rate of the spillover effects $\tau_s(1, 0, 3)$ and $\tau_s(1, 1, 3)$ obtained by the SPE method lies in the range of 93.0% to 94.3%, while it drops down dramatically to less than 6% for $\tau_s(1, 0, 3)$ and even becomes to 0% for $\tau_s(1, 1, 3)$ when using native estimations. If the one type of measurement error assumption fails, the coverage rate of $\tau_s(1, 0, 3)$ and $\tau_s(1, 1, 3)$ computed via the SPE varies from 94.9% to 95.5%, while it varies from 0% to a little below 4% for the naive estimations.

The results in this section show that (i) the SPE approach is robust to mild violation of the one type of measurement error assumption; and (ii) the SPE is still superior to the naive methods except in rare cases, in the sense that the bias reduction provided by the SPE is substantial and its causal inference is much more reliable.

7 Empirical Application: Diffusion of Insurance Information among Rice Farmers

This section applies the proposed SPE method to data on social network of rice farmers from 185 villages of rural China. The data is collected by Cai, De Janvry, and Sadoulet (2015b) to investigate the take-up decisions of weather insurance, which is typically adopted with low rates even when the government provides heavy subsidies. The primary interest of Cai et al. (2015b) is to study whether and how the diffusion of insurance knowledge through social network affects the insurance take-up rate.\(^3\) Thus, two rounds of sessions are offered with a three-days gap to allow information sharing by the first round participants. In each round, there are two types of sessions held simultaneously: the 20 minutes simple session where only contract is discussed, and the 45 minutes intensive session where details of how the insurance operates and the expected benefits are explained. About 5000 rice-producing households from those 185 villages are randomly

\(^3\)Data is available at Cai, De Janvry, and Sadoulet (2015a) https://doi.org/10.3886/E113593V1.
assigned to one of the two information sessions aiming at generating household-level variation in insurance knowledge. The authors are particularly interested in the spillover effects: whether the second round participants’ take-up decisions are affected by their friends who are invited to the first round intensive session. Hence, the baseline model for the treatment and spillover effects is:

\[ \text{Takeup}_{ig} = \theta_0 + \theta_1 \text{Intensive}_{ig} + \theta_2 \text{Network}_{ig} + \theta_3 \text{Cov}_{ig} + \theta_4 \text{NetSize}_{ig} + \eta_g + \varepsilon_{ig}, \]  

(14)

where \( \text{Takeup}_{ig} \) is a binary indicator of whether the household \( i \) in village \( g \) decide to buy the insurance, \( \text{Intensive}_{ig} \) is a dummy variable taking value one if the household is invited to intensive session, \( \text{Network}_{ig} \) is the fraction of household \( i \)’s friends who have been invited to the first round intensive session, \( \text{NetSize}_{ig} \) is a set of dummies indicating network degree, \( \text{Cov}_{ig} \) includes household characteristics and \( \eta_g \) represents village fixed effect.\(^{31}\) Household characteristics in \( \text{Cov}_{ig} \) include gender, age and education of household head, rice production area, risk aversion and perceived probability of future disasters. Five Dummies in \( \text{NetSize}_{ig} \) are indicators of the number of nominated friends equal to one to five, where the dummy of zero nominated friends is dropped to avoid collinearity. Instead of the baseline model (14), I also consider an alternative model specification where the interaction term \( \text{Intensive}_{ig} \times \text{Network}_{ig} \) is included.\(^{32}\)

Data from the social network survey is used to construct the household-level network measures. The social network survey requires the sampled household heads to nominate five friends with whom they discuss rice production or financial issues, while not all the respondents list five friends.\(^{33}\) No geographical restriction is imposed, which means the nominated friends can either live in the same village with the respondent or outside the village. This network measure is non-reciprocal and is referred to as “general measure” in Cai et al. (2015b). The general measure may contain two types of measurement error: those with less than five friends are likely to report false friends (false positive) and those with more than five friends may censor the number of network links (false negative). Another household-level network measure used in Cai et al. (2015b), referred to as “strong measure”, is defined as the bilaterally linked friends (reciprocal) using the same information from the social network survey. The social network survey is conducted before the experiment, therefore the network formation should not be affected by the treatment assignments nor the take-up decisions.

The analysis in this section utilizes both these two measures, and assumes that the strong measure includes only false negative links. It is worth noting that although the two network measures are probably correlated even conditional on the true network information, according to

\(^{31}\)If household \( i \) nominates zero friends, then \( \text{Network}_{ig} \) is set to be zero.

\(^{32}\)In the same spirit of Cai et al. (2015b), because the treatment \( \text{Intensive}_{ig} \) is whether household is invited to an intensive session or not, the treatment and spillover effects are studied from an intention-to-treat perspective. Nevertheless, almost 90% of households who are invited to one of the sessions actually attend. Therefore, the dropout is not a main concern.

\(^{33}\)About 95% of respondents report five friends, the rest report less than five friends.
the simulation results in Section 6.2, the SPE analysis can still be viewed as a bias-reduction method in the presence of network measurement error. The first step estimation is implemented as described in Section 5.2 by setting the dependence neighborhoods as villages, meaning that the dependent neighbors of an respondent \( i \) are those from the same village with \( i \).\(^{34}\)

Two further remarks are worth noticing. Firstly, the second round participants are not impacted by the take-up decisions made by the first round participants if this information is not revealed to them (see Table 6 column 7 and Table 7 column 6 of Cai et al., 2015b). In addition, according to the survey, there is only 9% of the households who are not informed of any first round take-up information know at least one of their friends’ decision. Thus, the endogenous peer effects, i.e. the spillovers of friends’ take-up decisions, should not be of major concern in this application. Secondly, the first round simple session also exhibits no significant spillover effects to the second round participants (see Table 2 column 3 of Cai et al., 2015b).

Table 2: Effect of Social Networks on Insurance Take-up

|                  | Naive |           | SPE  | Naive |           |
|------------------|-------|-----------|------|-------|-----------|
|                  | General | Strong |      | General | Strong |
| Intensive        | 0.0298 | 0.0228   | 0.0265 | 0.0809** | 0.0409   | 0.0556   |
|                  | (0.0332)| (0.0334) | (0.0462) | (0.0397) | (0.0341) | (0.0735) |
| Network          | 0.291*** | 0.113* | 0.196 | 0.444*** | 0.231*** | 0.244 |
|                  | (0.0820) | (0.0606) | (0.2492) | (0.1089) | (0.0859) | (0.2472) |
| Intensive*Network|       |         |      |        |         |
|                  | -0.329** | -0.221** | -0.106 |       |
|                  | (0.161) | (0.111) | (0.189) |

\( \eta_g \) | Yes | Yes | Yes | Yes | Yes | Yes |
\( Cov_{ig} \) | Yes | Yes | Yes | Yes | Yes | Yes |

Note: Samples are from the second round sessions “Simple2-NoInfo” and “Intensive2-NoInfo” as defined and used by Cai et al. (2015b). Number of observations is 1255. Standard error (se) is reported in the parenthesis. For the naive method, column “General” shows the result using the general measure of the network and column “Strong” display the result using the strong measure of the network. The SPE method is implemented by assuming the classification error is correlated to literacy. The se of the naive method is computed using clustered standard error with villages as clusters. The se of the SPE method is calculated based on Corollary 5.7 with villages as dependency neighborhoods.

Estimation results are summarized in Table 2. The baseline model (columns (1) to (3)) and the alternative model with interaction term of the treatment and the network exposure (columns (4) to (6)) are estimated using the household-level samples from the second round sessions, where no overall attendance/take-up rate nor individual insurance purchase results at the first round sessions in their village are revealed to the participants. Results for the Naive method using general measure of the network data in columns (1) and (4) in Table 2 are the same to those

\(^{34}\)To mitigate estimation error arising from small sample size, the first step estimation uses samples from both the first and the second rounds and their network data based on the social network survey, with sample size 4588.
in Table 2 columns (2) and (4) of Cai et al. (2015b), based on which they draw two conclusions. First, the spillover effect on insurance take-up is significantly positive. For example, column (1) (or column (2)) reveals that having additional 20% increase in the ratio of friends attending the first round intensive session will lead to a $29.1\% \times 20\% = 5.82\%$ (or $11.3\% \times 20\% = 2.26\%$) increase in farmer’s own take-up probability. Second, people are less likely to be affected by their friends if they attend the intensive session themselves. Column (4) (or column (5)) in Table 2 reveals that for farmers who have been directly educated about the insurance details, if the ratio of friends attending the first round intensive session increases 20%, their own take-up probability will increase $(44.4\%-32.9\%) \times 20\% = 2.3\%$ (or $(23.1\%-22.1\%) \times 20\% = 0.2\%$); while this probability increase will be $44.4\% \times 20\% = 8.88\%$ (or $23.1\% \times 20\% = 4.62\%$) for farmers who haven’t attended the intensive session.

If the general measure possesses network misclassification, then the estimates of the naive approach are biased. The SPE method can then be used to provide some guidance of the degree and direction of the potential bias. The SPE estimates in Table 2 are obtained by assuming that the measurement errors of the two network measures (both general and strong) are dependent on the household-head’s literacy. By comparing the results in columns (1) and (2) to those in column (3), we can see that the SPE estimate of the spillover effect induced by additional 20% treated friends is $19.6\% \times 20\% = 3.92\%$. Thus, the naive method using the general measure may overestimate the spillover effect, while the naive method using the strong measure is likely to underestimate the spillover effect. In addition, based on the SPE method columns (6), people who attend the intensive session themselves facing a $(24.4\%-10.6\%) \times 20\% = 2.76\%$ increase in their take-up probability when extra 20% of their friends are exposed to the intensive insurance-information education. While this change arises to $24.4\% \times 20\% = 4.88\%$ for people who are not attend the intensive session. Hence, the comparison between column (5) and column (7) indicates that results for the Naive method using the general measure underestimate the spillover effect for the treated individuals, and overestimate the spillover effect for the untreated ones. Besides, the Naive method using the strong measure dramatically underestimates the spillover effect for the treated individuals, but only slightly underestimates the spillover effect for the controlled individuals.

8 Conclusion

Motivated by applications of program evaluation under network interference, this paper studies the identification and estimation of treatment and spillover effects when the network is mismeasured. The novel identification strategy proposed in this paper utilizes two network proxies, where one of them is used as an instrumental variable for the latent network and the other is assumed to contain only one type of measurement error. A semiparametric estimation approach for the
effects of interest is also provided. The simulation results confirm that the proposed estimation (i) outperforms the naive estimation neglecting the network misclassification, and (ii) remains to be a preferred alternative to the naive estimation, even if its key assumption is mildly violated. Therefore, the proposed estimation serves as an effective way to reduce the bias caused by the network measurement errors, and provide reliable causal inference.

The proposed semiparametric estimation approach exploits a parametric structural assumption of the outcome variable to avoid the curse of dimensionality, which opens new questions on the trade-off between the potential model misspecification and the network mismeasurement-robust estimation. It is also meaningful and feasible to investigate the estimation in a more flexible semiparametric setup, including partially linear model, index model and random-coefficient model.

This paper is particularly suited for studies where the treatment is randomly assigned with perfect compliance. While, for some empirical studies, it is reasonable to allow for non-compliance (Vazquez-Bare, 2020). Future research could further explore the impacts of relaxing the perfect compliance, and develop methods for the identification and estimation of spillover effects to accommodate the non-compliance.

Finally, this paper assumes that the specification of how the exposure to the treated peers affecting the outcome is correct, meaning that the spillover effect is local through the first-order peers. The literature on network effects often stresses the existence of the higher-order interference, i.e. the interference with friends of friends. It sophisticates the analysis in this paper by introducing the spillover of the treatment and of the measurement errors from higher-order interactions. Moreover, it also makes the dependence structure among the observable and latent network-based variables more complicated. It is nontrivial how the analysis of this paper can be extended to deal with higher-order interference. However, for studies where the treatment response is primarily governed by the first-order spillover, it is possible to apply the analysis of this paper via assuming the higher-order interference effects are omitted to the unobservables. The rationale is that, based on the study of Leung (2019a) and Sävje (2019), the exposure misspecification ignoring higher-order interference does not alter the estimation results, if the specification errors are well counterbalanced by the deceasing data correlation as the order of the interference increases. Rigorous exploration along this direction is worthwhile in the future research.

References

Advani, A. and B. Malde (2018): “Credibly identifying social effects: Accounting for network formation and measurement error,” Journal of Economic Surveys, 32, 1016–1044.

Andrew, A. L., K.-W. E. Chu, and P. Lancaster (1993): “Derivatives of eigenvalues and eigenvectors of matrix functions,” SIAM journal on matrix analysis and applications, 14, 903–926.
Angelucci, M., G. De Giorgi, M. A. Rangel, and I. Rasul (2010): “Family networks and school enrolment: Evidence from a randomized social experiment,” *Journal of public Economics*, 94, 197–221.

Angelucci, M. and V. Di Maro (2016): “Programme evaluation and spillover effects,” *Journal of Development Effectiveness*, 8, 22–43.

Aral, S. and D. Walker (2012): “Identifying influential and susceptible members of social networks,” *Science*, 337, 337–341.

Aronow, P. M. and C. Samii (2017): “Estimating average causal effects under general interference, with application to a social network experiment,” *The Annals of Applied Statistics*, 11, 1912–1947.

Athey, S., D. Eckles, and G. W. Imbens (2018): “Exact p-values for network interference,” *Journal of the American Statistical Association*, 113, 230–240.

Athey, S. and G. W. Imbens (2017): “The econometrics of randomized experiments,” in *Handbook of economic field experiments*, Elsevier, vol. 1, 73–140.

Auerbach, E. (2019): “Identification and estimation of a partially linear regression model using network data,” *arXiv preprint arXiv:1903.09679*.

Baird, S., J. A. Bohren, C. McIntosh, and B. Özler (2018): “Optimal design of experiments in the presence of interference,” *Review of Economics and Statistics*, 100, 844–860.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): “The diffusion of microfinance,” *Science*, 341, 1236498.

Barrera-Osorio, F., M. Bertrand, L. L. Linden, and F. Perez-Calle (2011): “Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia,” *American Economic Journal: Applied Economics*, 3, 167–95.

Basse, G. and A. Feller (2018): “Analyzing two-stage experiments in the presence of interference,” *Journal of the American Statistical Association*, 113, 41–55.

Battistin, E., M. De Nadai, and B. Sianesi (2014): “Misreported schooling, multiple measures and returns to educational qualifications,” *Journal of Econometrics*, 181, 136–150.

Battistin, E. and B. Sianesi (2011): “Misclassified treatment status and treatment effects: an application to returns to education in the United Kingdom,” *Review of Economics and Statistics*, 93, 495–509.

Bound, J., C. Brown, and N. Mathiowetz (2001): “Measurement error in survey data,” in *Handbook of econometrics*, Elsevier, vol. 5, 3705–3843.

Bradley, R. C. et al. (1983): “Approximation theorems for strongly mixing random variables.” *The Michigan Mathematical Journal*, 30, 69–81.

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2020): “Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data,” *American Economic Review*, 110, 2454–2484.
Cai, J., A. De Janvry, and E. Sadoulet (2015a): “Replication data for: Social networks and the decision to insure,” Nashville, TN: American Economic Association, Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2019-10-12. https://doi.org/10.3886/E113593V1.

——— (2015b): “Social networks and the decision to insure,” American Economic Journal: Applied Economics, 7, 81–108.

Calvi, R., A. Lewbel, and D. Tommasi (2018): “Women’s Empowerment and Family Health: Estimating LATE with Mismeasured Treatment,” Available at SSRN 2980250.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): “Peer effects and social networks in education,” The Review of Economic Studies, 76, 1239–1267.

Candelaria, L. E. and T. Ura (2020): “Identification and inference of network formation games with misclassified links,” Tech. rep., University of Warwick, Department of Economics.

Chandrasekhar, A. (2016): “Econometrics of network formation,” The Oxford Handbook of the economics of networks, 303–357.

Chandrasekhar, A. and R. Lewis (2011): “Econometrics of sampled networks,” Working Paper.

Chandrasekhar, A. G. and M. O. Jackson (2016): “A network formation model based on subgraphs,” Available at SSRN 2660381.

Chen, L. H., L. Goldstein, and Q.-M. Shao (2010): Normal approximation by Steins method, Springer Science & Business Media.

Chen, X., H. Hong, and D. Nekipelov (2011): “Nonlinear models of measurement errors,” Journal of Economic Literature, 49, 901–37.

Chin, A. (2018): “Central limit theorems via Stein’s method for randomized experiments under interference,” arXiv preprint arXiv:1804.03105.

Comola, M. and M. Fafchamps (2017): “The missing transfers: Estimating misreporting in dyadic data,” Economic Development and Cultural Change, 65, 549–582.

Conley, T. G. and C. R. Udry (2010): “Learning about a new technology: Pineapple in Ghana,” American Economic Review, 100, 35–69.

CurraSlide, S., M. O. Jackson, and P. Pin (2009): “An economic model of friendship: Homophily, minorities, and segregation,” Econometrica, 77, 1003–1045.

——— (2010): “Identifying the roles of race-based choice and chance in high school friendship network formation,” Proceedings of the National Academy of Sciences, 107, 4857–4861.

De Paula, A. (2017): “Econometrics of network models,” in Advances in economics and econometrics: Theory and applications, eleventh world congress, Cambridge University Press Cambridge, 268–323.
De Paula, A., I. Rasul, and P. Souza (2018a): “Recovering social networks from panel data: identification, simulations and an application,” LACEA Working Paper Series No. 0001, Available at SSRN: http://dx.doi.org/10.2139/ssrn.3322049.

De Paula, Á., S. Richards-Shubik, and E. Tamer (2018b): “Identifying preferences in networks with bounded degree,” *Econometrica*, 86, 263–288.

Dupas, P. (2014): “Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment,” *Econometrica*, 82, 197–228.

Erdős, P. and A. Rényi (1959): “On random graphs,” *Mathematicae Debrecen*, 6, 290–297.

Fafchamps, M. and S. Lund (2003): “Risk-sharing networks in rural Philippines,” *Journal of Development Economics*, 71, 261–287.

Gao, W. and S. Li (2019): “Identification and Estimation of Peer Effects in Latent Networks,” *Working Paper*.

Goldsmith-Pinkham, P. and G. W. Imbens (2013): “Social networks and the identification of peer effects,” *Journal of Business & Economic Statistics*, 31, 253–264.

Goldstein, L. and Y. Rinott (1996): “Multivariate normal approximations by Stein’s method and size bias couplings,” *Journal of Applied Probability*, 33, 1–17.

Hardy, M., R. M. Heath, W. Lee, and T. H. McCormick (2019): “Estimating spillovers using imprecisely measured networks,” *arXiv preprint arXiv:1904.00136*.

He, X. and K. Song (2018): “Measuring diffusion over a large network,” *arXiv preprint arXiv:1812.04195*.

Hu, Y. (2008): “Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution,” *Journal of Econometrics*, 144, 27–61.

Hu, Y. and S. M. Schennach (2008): “Instrumental variable treatment of nonclassical measurement error models,” *Econometrica*, 76, 195–216.

Hudgens, M. G. and M. E. Halloran (2008): “Toward causal inference with interference,” *Journal of the American Statistical Association*, 103, 832–842.

Imbens, G. W. (2000): “The role of the propensity score in estimating dose-response functions,” *Biometrika*, 87, 706–710.

Johnsson, I. and H. R. Moon (2015): “Estimation of peer effects in endogenous social networks: control function approach,” *Review of Economics and Statistics*, 1–51.

Kojevnikov, D., V. Marmer, and K. Song (2019): “Limit theorems for network dependent random variables,” *arXiv preprint arXiv:1903.01059*.

Kossinets, G. (2006): “Effects of missing data in social networks,” *Social networks*, 28, 247–268.

Kremer, M. and E. Miguel (2007): “The illusion of sustainability,” *The Quarterly Journal of Economics*, 122, 1007–1065.
Kuersteiner, G. M. (2019): “Limit Theorems for Data with Network Structure,” arXiv preprint arXiv:1908.02375.

Lee, Y. and E. L. Ogburn (2020): “Network Dependence Can Lead to Spurious Associations and Invalid Inference,” Journal of the American Statistical Association, 1–31.

Leung, M. P. (2019a): “Causal Inference Under Approximate Neighborhood Interference,” Available at SSRN 3479902.

——— (2019b): “A weak law for moments of pairwise stable networks,” Journal of Econometrics, 210, 310–326.

——— (2020a): “Dependence-Robust Inference Using Resampled Statistics,” arXiv preprint arXiv:2002.02097.

——— (2020b): “Treatment and spillover effects under network interference,” Review of Economics and Statistics, 102, 368–380.

Leung, M. P. and H. R. Moon (2019): “Normal Approximation in Large Network Models,” arXiv preprint arXiv:1904.11060.

Lewbel, A. (2007): “Estimation of average treatment effects with misclassification,” Econometrica, 75, 537–551.

Lewbel, A., X. Qu, and X. Tang (2019): “Social networks with misclassified or unobserved links,” Tech. rep., Boston College Department of Economics.

Li, Q. and J. S. Racine (2007): Nonparametric econometrics: theory and practice, Princeton University Press.

Liu, L. and M. G. Hudgens (2014): “Large sample randomization inference of causal effects in the presence of interference,” Journal of the American Statistical Association, 109, 288–301.

Liu, X. (2013): “Estimation of a local-aggregate network model with sampled networks,” Economics Letters, 118, 243–246.

Ma, Y., Y. Wang, and V. Tresp (2020): “Causal Inference under Networked Interference,” arXiv preprint arXiv:2002.08506.

Mahajan, A. (2006): “Identification and estimation of regression models with misclassification,” Econometrica, 74, 631–665.

Manski, C. F. (2001): “Designing programs for heterogeneous populations: The value of covariate information,” American Economic Review, 91, 103–106.

——— (2013): “Identification of treatment response with social interactions,” The Econometrics Journal, 16, S1–S23.

Masry, E. (1996): “Multivariate local polynomial regression for time series: uniform strong consistency and rates,” Journal of Time Series Analysis, 17, 571–599.
Meckes, E. et al. (2009): “On Steins method for multivariate normal approximation,” in High dimensional probability V: the Luminy volume, Institute of Mathematical Statistics, 153–178.

Miguel, E. and M. Kremer (2004): “Worms: identifying impacts on education and health in the presence of treatment externalities,” Econometrica, 72, 159–217.

Newey, W. and D. MacFadden (1994): “Large Sample Estimation and Hypothesis Testing, Chapter 36,” Handbook of Econometrics Vol. 4.

Newey, W. K. (1994): “Kernel estimation of partial means and a general variance estimator,” Econometric Theory, 10, 1–21.

Opper, I. M. (2019): “Does helping john help sue? evidence of spillovers in education,” American Economic Review, 109, 1080–1115.

Oster, E. and R. Thornton (2012): “Determinants of technology adoption: Peer effects in menstrual cup take-up,” Journal of the European Economic Association, 10, 1263–1293.

Patacchini, E., E. Rainone, and Y. Zenou (2017): “Heterogeneous peer effects in education,” Journal of Economic Behavior & Organization, 134, 190–227.

Qu, X. and L.-F. Lee (2015): “Estimating a spatial autoregressive model with an endogenous spatial weight matrix,” Journal of Econometrics, 184, 209–232.

Romano, J. P. and A. M. Shaikh (2012): “On the uniform asymptotic validity of subsampling and the bootstrap,” The Annals of Statistics, 40, 2798–2822.

Ross, N. (2011): “Fundamentals of Stein’s method,” Probability Surveys, 8, 210–293.

Sävje, F. (2019): “Causal inference with misspecified exposure mappings,” Tech. rep., Yale University.

Sävje, F., P. M. Aronow, and M. G. Hudgens (2017): “Average treatment effects in the presence of unknown interference,” arXiv preprint arXiv:1711.06399.

Sobel, M. E. (2006): “What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference,” Journal of the American Statistical Association, 101, 1398–1407.

Sojourner, A. (2013): “Identification of peer effects with missing peer data: Evidence from Project STAR,” The Economic Journal, 123, 574–605.

Song, K. (2018): “Measuring the graph concordance of locally dependent observations,” Review of Economics and Statistics, 100, 535–549.

Stein, C. (1986): “Approximate computation of expectations,” Lecture Notes-Monograph Series, 7, i–164.

Tauchen, G. (1985): “Diagnostic testing and evaluation of maximum likelihood models,” Journal of Econometrics, 30, 415–443.
Tchetgen, E. J. T. and T. J. VanderWeele (2012): “On causal inference in the presence of interference,” *Statistical Methods in Medical Research*, 21, 55–75.

Thirkettle, M. (2019): “Identification and Estimation of Network Statistics with Missing Link Data,” Tech. rep., Working Paper.

van der Laan, M. J. (2014): “Causal inference for a population of causally connected units,” *Journal of Causal Inference*, 2, 13–74.

Vazquez-Bare, G. (2019): “Identification and estimation of spillover effects in randomized experiments,” *arXiv preprint arXiv:1711.02745*.

——— (2020): “Causal spillover effects using instrumental variables,” *arXiv preprint arXiv:2003.06023*.

Viviano, D. (2019): “Policy targeting under network interference,” *arXiv preprint arXiv:1906.10258*.

White, H. and I. Domowitz (1984): “Nonlinear regression with dependent observations,” *Econometrica: Journal of the Econometric Society*, 143–161.
Appendix

We first introduce notations used in the Appendix. $I_K$ is the $K \times K$ identity matrix. $\lambda_{\text{max}}(A)$ and $\lambda_{\text{min}}(A)$ denote the largest and the smallest eigenvalues of a square matrix $A$, respectively. We use $C$ to represent some positive constant and its value may be different at different uses. s.o. denotes the terms of smaller order. Appendix Section E contains some useful lemmas and is relegated to the supplementary material.

A Examples

This section provides sufficient conditions or examples for the assumptions in the main text.

Example 1 (Assumption 3.3 (a)) Suppose the network links follow the dyadic formation below:

$$A_{ij}^* = 1[\omega(Z_i, Z_j) > \eta_{ij}] \cdot 1[i \neq j], \text{ with } i, j \in \mathcal{P}$$

where $\omega : \Omega_2^2 \mapsto \mathbb{R}$ and the unobserved link specific error term $\eta_{ij}$ is independent to $\{Z_i\}_{i \in \mathcal{P}}$ and is i.i.d. across $(i, j)$. Then, $A_{ij}^*$ given $Z_i$ is a function of $(Z_j, \eta_{ij})$, which is i.i.d. across $j$ and $|\mathcal{N}_i^*| = \sum_{j \in \mathcal{P}} A_{ij}^*$ is identically distributed following the binomial distribution. Such a network formation is considered in e.g. Johnsson and Moon (2015).

Example 2 (Assumption 3.4 (c)) For any given latent $A_{ij}^*$, consider the following data-generating process of the observable $A_{ij}$

$$A_{ij} = U_{ij} A_{ij}^* + V_{ij}(1 - A_{ij}^*), \text{ with } i, j \in \mathcal{P} \quad (A.1)$$

where $\mathcal{N}_i = \{j \in \mathcal{P} : A_{ij} = 1\}$ and the classification errors $(U_{ij}, V_{ij})$ are random indicators taking values from $\{0, 1\}$. From (A.1), we can obtain that

$$|\mathcal{N}_i| = \sum_{j \in \mathcal{P}} A_{ij} = \sum_{j \in \mathcal{P}} (U_{ij} - V_{ij}) A_{ij}^* + \sum_{j \in \mathcal{P}} V_{ij} = \sum_{j \in \mathcal{N}_i^*} (U_{ij} - V_{ij}) + \sum_{j \in \mathcal{P}} V_{ij}.$$  

Let two vectors $\mathbf{U}_i = \{U_{ij}\}_{j \in \mathcal{P}}$ and $\mathbf{V}_i = \{V_{ij}\}_{j \in \mathcal{P}}$. If the random vector $(\mathbf{U}_i, \mathbf{V}_i)$ is conditionally independent to $\mathcal{N}_i^*$ and identically distributed across $i \in \mathcal{P}$ given $(Z_i, |\mathcal{N}_i^*|)$, then the identical distribution of $|\mathcal{N}_i|$ given $(Z_i, |\mathcal{N}_i^*|)$ holds.

Example 3 (Assumption 4.1) For each $i \in \mathcal{P}$ and any given latent $A_{ij}^*$, suppose the observable links are generated as

$$A_{ij} = \omega_j [U_{ij} A_{ij}^* + V_{ij}(1 - A_{ij}^*)], \quad \tilde{A}_{ij} = \tilde{\omega}_j [\tilde{U}_{ij} A_{ij}^* + \tilde{V}_{ij}(1 - A_{ij}^*)], \quad \text{with } j \in \mathcal{P} \quad (A.2)$$

with $U_{ij}$, $V_{ij}$, $\tilde{U}_{ij}$, $\tilde{V}_{ij}$, $\omega_j$ and $\tilde{\omega}_j$ are all binary random variables taking values from $\{0, 1\}$. $\omega_j$ and $\tilde{\omega}_j$ can be understood as indicators of sampling-induced errors, e.g. $\omega_j = 0$ means unit $j$ is not sampled when constructing $\mathcal{N}_i$, while only links among pairs of sampled units are accounted for. $(U_{ij}, V_{ij})$ and $(\tilde{U}_{ij}, \tilde{V}_{ij})$ can be understood as indicators of non-sampling-induced errors, e.g.
$\tilde{U}_{ij} = 0$ represents unit $i$'s misreporting of her link with unit $j$ when constructing $\tilde{N}_i$. Then, the observed sets of links are $\mathcal{N}_i = \{j \in \mathcal{P} : A_{ij} = 1\}$ and $\tilde{\mathcal{N}}_i = \{j \in \mathcal{P} : \tilde{A}_{ij} = 1\}$. Therefore,

$$
|\mathcal{N}_i| = \sum_{j \in \mathcal{P}} A_{ij} = \sum_{j \in \mathcal{P}} (U_{ij} - V_{ij}) \omega_j A_{ij}^* + \sum_{j \in \mathcal{N}_i^*} (U_{ij} - V_{ij}) \omega_j + \sum_{j \in \mathcal{P}} V_{ij} \omega_j,

|\tilde{\mathcal{N}}_i| = \sum_{j \in \mathcal{P}} \tilde{A}_{ij} = \sum_{j \in \mathcal{P}} (\tilde{U}_{ij} - \tilde{V}_{ij}) \omega_j A_{ij}^* + \sum_{j \in \mathcal{N}_i^*} (\tilde{U}_{ij} - \tilde{V}_{ij}) \omega_j + \sum_{j \in \mathcal{P}} \tilde{V}_{ij} \omega_j.
$$

(A.3)

Then, one set of sufficient conditions for Assumption 4.1 is provided by the lemma below.

**Lemma A.1** Let Assumption 3.4 (a) holds for both $\mathcal{N}_i$ and $\tilde{\mathcal{N}}_i$. Suppose the random vector $(U_{ij}, V_{ij}, \tilde{U}_{ij}, \tilde{V}_{ij}, \omega_j, \tilde{\omega}_j)$ given $(Z_i, |\mathcal{N}_i^*|)$ is i.i.d. across $j$ for all $i \in \mathcal{P}$. If

(a) $\{U_{ij}, V_{ij}, \omega_j\}_{j \in \mathcal{P}} \perp \{\tilde{U}_{ik}, \tilde{V}_{ik}, \tilde{\omega}_k\}_{k \in \mathcal{P}} | Z_i, |\mathcal{N}_i^*|$

(b) $\{U_{ij}, V_{ij}, \tilde{U}_{ij}, \tilde{V}_{ij}, \omega_j, \tilde{\omega}_j\}_{j \in \mathcal{P}} \perp |\mathcal{N}_i^*| | Z_i, |\mathcal{N}_i^*|$

then Assumption 4.1 is satisfied by $|\mathcal{N}_i|$ and $|\tilde{\mathcal{N}}_i|$ given in (A.3).

**Proof of Lemma A.1.** (i) From condition (a) that $\{U_{ij}, V_{ij}, \omega_j\}_{j \in \mathcal{P}} \perp \{\tilde{U}_{ik}, \tilde{V}_{ik}, \tilde{\omega}_k\}_{k \in \mathcal{P}} | Z_i, |\mathcal{N}_i^*|$, we have $\mathcal{N}_i \perp \tilde{\mathcal{N}}_i | Z_i, |\mathcal{N}_i^*|$, which implies $|\mathcal{N}_i| \perp |\tilde{\mathcal{N}}_i| | Z_i, |\mathcal{N}_i^*|$. If we can further show that $|\mathcal{N}_i| \perp |\mathcal{N}_i^*| | Z_i, |\mathcal{N}_i^*|$ and $|\tilde{\mathcal{N}}_i| \perp |\mathcal{N}_i^*| | Z_i, |\mathcal{N}_i^*|$ hold, then the desired result follows, because

$$
f_{|\mathcal{N}_i||Z_i,|\mathcal{N}_i^*|,|\tilde{\mathcal{N}}_i|}(n) = \sum_{\mathcal{J} \in \Omega_{\mathcal{N}_i^*}} f_{|\mathcal{N}_i||Z_i,|\mathcal{N}_i^*||\tilde{\mathcal{N}}_i|,|\mathcal{N}_i^*|}(\mathcal{J}) \times f_{\mathcal{N}_i^*||Z_i,|\mathcal{N}_i^*|}(n) \times f_{|\mathcal{N}_i^*||Z_i,|\mathcal{N}_i^*|}(n),
$$

which indicates $|\mathcal{N}_i| \perp |\tilde{\mathcal{N}}_i| | Z_i, |\mathcal{N}_i^*|$. Given the expressions in (A.3), based on the i.i.d. of $(U_{ij}, V_{ij}, \omega_j)$ across $j$, applying the same arguments used to prove Lemma 4.1 (a), we can show that given $(Z_i, |\mathcal{N}_i^*|)$, the distribution of $\sum_{j \in \mathcal{N}_i^*} (U_{ij} - V_{ij}) \omega_j$ does not depend on $\mathcal{N}_i^*$, i.e., $\sum_{j \in \mathcal{N}_i^*} (U_{ij} - V_{ij}) \omega_j \perp |\mathcal{N}_i^*| | Z_i, |\mathcal{N}_i^*|$. In addition, from condition (b) we can obtain the independence of $\sum_{j \in \mathcal{P}} V_{ij} \omega_j$ to $\mathcal{N}_i^*$ given $Z_i, |\mathcal{N}_i^*|$. Thus, it follows from the above results and (A.3) that $|\mathcal{N}_i| \perp |\mathcal{N}_i^*| | Z_i, |\mathcal{N}_i^*|$. Similarly, $|\tilde{\mathcal{N}}_i| \perp |\mathcal{N}_i^*| | Z_i, |\mathcal{N}_i^*|$ also holds. ■

**B Proofs**

**B.1 Proofs of Section 3**

**Lemma B.1** Under Assumption 3.2, we have that for $\forall i \in \mathcal{P}$, $\varepsilon_i \perp (D_i, S_i^*) | Z_i, |\mathcal{N}_i^*|$. 

**Proof of Lemma B.1.** If we can show that $Pr(\varepsilon_i < e | D_i, S_i^*, Z_i, |\mathcal{N}_i^*|) = Pr(\varepsilon_i < e | Z_i, |\mathcal{N}_i^*|)$, then the stated result follows. By the law of total probability, we have for $\forall e \in \Omega_\varepsilon$,

$$
Pr(\varepsilon_i < e | D_i, S_i^*, Z_i, |\mathcal{N}_i^*|)
$$

45
\[ E \left[ Pr \left( \varepsilon_i < e \middle| D_i, S_i^*, Z_i, |N_i^*|, \{D_j\}_{j \in N_i^*} \right) \right] = E \left[ Pr \left( \varepsilon_i < e \middle| Z_i, |N_i^*|, \{D_j\}_{j \in N_i^*} \right) \right] = E \left[ Pr \left( \varepsilon_i < e \middle| Z_i, |N_i^*| \right) \right], \tag{B.1} \]

where the expectation is with respect to \( f_{N_i^*}(D_i) \). By definition, \( S_i^* = \sum_{j \in N_i^*} D_j \), therefore, \( S_i^* \) becomes fixed when given \( (N_i^*, \{D_j\}_{j \in N_i^*}) \). In addition, since Assumption 3.2 (a) implies that \( D_i \) is independent to \( (\varepsilon_i, Z_i, |N_i^*|, S_i^*, \{D_j\}_{j \in N_i^*}) \). Then, we know that \( (D_i, S_i^*) \) can be eliminated from the conditional probability of \( \varepsilon_i < e \) in (B.1), i.e.

\[ Pr(\varepsilon_i < e|D_i, S_i^*, Z_i, |N_i^*|) = E \left[ Pr \left( \varepsilon_i < e \middle| Z_i, |N_i^*|, \{D_j\}_{j \in N_i^*} \right) \right] \]

where the second line is from Assumption 3.2 (b) and the last line implies \( \varepsilon_i \perp (D_i, S_i^*)|Z_i, |N_i^*| \).

**Lemma B.2** Under Assumptions 3.2 and 3.4, \( \varepsilon_i \perp (D_i, S_i^*, S_i, |N_i|)|Z_i, |N_i^*| \) for \( \forall i \in \mathcal{P} \).

**Proof of Lemma B.2.** Denote \( P_i^* = (N_i^*, \{D_j\}_{j \in N_i^*}) \) and \( P_i = (N_i, \{D_j\}_{j \in N_i}) \). Then, we know from Assumptions 3.2 (a) and 3.4 (a) that \( D_i \perp (P_i^*, P_i) \), because of the facts that \( i \notin N_i^* \), \( i \notin N_i \), \( D_i \perp \{\{D_j\}_{j \in N_i^*}\} |N_i^*, N_i \) and \( \{D_j\}_{j \in \mathcal{P}} \perp (N_i^*, N_i) \). Moreover, since \( S_i^* \) and \( S_i \) are functions of \( P_i^* \) and \( P_i \), respectively, we have \( D_i \perp (S_i^*, S_i) \). By the law of total probability,

\[ Pr \left( \varepsilon_i < e \middle| D_i, S_i^*, S_i, |N_i|, Z_i, |N_i^*| \right) = Pr \left( \varepsilon_i < e \middle| S_i^*, S_i, |N_i|, Z_i, |N_i^*| \right) = E \left[ Pr \left( \varepsilon_i < e \middle| S_i^*, S_i, |N_i|, Z_i, |N_i^*|, P_i^*, P_i \right) \right] \]

for \( \forall e \in \Omega_e \), where the expectation is with respect to \( f_{P_i^*, P_i|S_i^*, S_i, |N_i|, Z_i, |N_i^*|} \). We know that \( S_i^*, S_i, |N_i| \) are fixed given \( (P_i^*, P_i) \). Thus, equation (B.3) becomes to

\[ Pr \left( \varepsilon_i < e \middle| D_i, S_i^*, S_i, |N_i|, Z_i, |N_i^*| \right) = E \left[ Pr \left( \varepsilon_i < e \middle| Z_i, |N_i^*|, P_i^*, P_i \right) \right] \]

where the second equality above is due to Assumption 3.4 (b).

**Proof of Proposition 3.1.** By Assumption 3.1 and the law of iterated expectation,

\[ m_i(d, s, z, n) = E \left[ r(D_i, S_i^*, Z_i, |N_i^*|, \varepsilon_i) \left| D_i = d, S_i = s, Z_i = z, |N_i| = n \right. \right] \]

\[ = \sum_{(s^*, n^*) \in \Omega_{S_i^*, |N_i^*|}} \left[ E \left[ r(d, s^*, z, n^*, \varepsilon_i) \left| D_i = d, S_i = s, Z_i = z, |N_i| = n, S_i^* = s^*, |N_i^*| = n^* \right. \right. \right] \]

\[ \times f_{S_i^*, |N_i^*| |D_i = d, S_i = s, Z_i = z, |N_i| = n}(s^*, n^*). \tag{B.4} \]
Based on Lemma B.2 that \( \epsilon_i \perp (D_i, S_i^*, S_i, |N_i|)|Z_i, |N_i^*| \), we have that (B.4) becomes to

\[
m_i(d, s, z, n) = \sum_{(s^*, n^*) \in \Omega_{S^*, |N^*|}} E\left[r(d, s^*, z, n^*, \epsilon_i)|Z_i = z, |N_i^*| = n^*\right] f_{S_i^*, |N_i^*|}|D_i = d, S_i = s, Z_i = z, |N_i| = n(s^*, n^*)
\]

where the last equality is by Definition 1. \( \blacksquare \)

### B.2 Proofs of Section 4

**Lemma B.3** Under Assumption 3.2 (a), suppose Assumption 3.4 (a) and (b) are satisfied by both \( N_i \) and \( \tilde{N}_i \). Then, \( Y_i \perp (|N_i|, |\tilde{N}_i|)|Z_i, |N_i^*| \) holds.

**Proof of Lemma B.3.** First, same arguments used in the proof of Lemma 4.1 (a) can be applied to show that \( S_i^* \perp (|N_i|, |\tilde{N}_i|)|Z_i, |N_i^*| \). Second, rewrite \( Y_i \) in terms of the potential outcomes:

\[
Y_i = \sum_{(d, s) \in \{0,1\} \times \Omega_{S^*}} 1[D_i = d, S_i^* = s] r(d, s, Z_i, |N_i^*|, \epsilon_i),
\]

where by the randomness of the treatment assignment and \( S_i^* \perp (|N_i|, |\tilde{N}_i|)|Z_i, |N_i^*| \), we know that \( 1[D_i = d, S_i^* = s] \perp (|N_i|, |\tilde{N}_i|)|Z_i, |N_i^*| \). Then, because Assumption 3.4 (b) implies that \( (|N_i|, |\tilde{N}_i|) \) is independent to the potential outcome \( r(d, s, Z_i, |N_i^*|, \epsilon_i) \) given \( (Z_i, |N_i^*|) \), we can conclude that \( Y_i \perp (|N_i|, |\tilde{N}_i|)|Z_i, |N_i^*| \). \( \blacksquare \)

**Proof of Lemma 4.1.** (a) If we can show that for any \((s, J) \in \Omega_{S^*, N^*}\) the equation below holds,

\[
f_{S_i^*, |N_i^*|=n}|Z_i, |N_i^*|=n(s, J) = f_{S_i^*, |N_i^*|=n}(s) f_{|N_i^*|=n}|Z_i, |N_i^*|=n(J), \tag{B.6}
\]

then the desired result follows. First of all, if either \( s > n \) or \(|J| \neq n\), (B.6) holds trivially. Therefore, we consider \((s, J)\) such that \( s \leq n \) and \(|J| = n\). Because for any fixed \( J \), \( \{D_j\}_{j \in J} \) is independent to \((Z_i, |N_i^*|, |\tilde{N}_i|)\) by Assumption 3.2 (a), then by i.i.d. of \( D_i \)

\[
f_{S_i^*, |N_i^*|=n, N_i^*=J}(s) = f_{\sum_{j \in J} D_j |Z_i, |N_i^*|=n, N_i^*=J}(s) = f_{\sum_{j \in J} D_j}(s) = C_n f_D(1) f_D(0)^{(n-s)}. \tag{B.7}
\]

where \( f_D(d) = Pr(D_i = d) \) with \( d \in \{0,1\} \). On the other hand, by the law of total probability,

\[
f_{S_i^*, |N_i^*|=n}|Z_i, |N_i^*|=n(s) = \sum_{J \in \Omega_{N^*}, s.t. |J|=n} f_{\sum_{j \in J} D_j |Z_i, |N_i^*|=n, N_i^*=J}(s) f_{N_i^*}|Z_i, |N_i^*|=n(J)
\]

\[= \sum_{J \in \Omega_{N^*}, s.t. |J|=n} f_{\sum_{j \in J} D_j}(s) f_{N_i^*}|Z_i, |N_i^*|=n(J)
\]

\[= C_n f_D(1) f_D(0)^{(n-s)}. \tag{B.8}
\]

Therefore, (B.7) and (B.8) lead to

\[
f_{S_i^*, |N_i^*|=n}|Z_i, |N_i^*|=n(s, J) = f_{S_i^*, |N_i^*|=n, N_i^*=J}(s) f_{N_i^*}|Z_i, |N_i^*|=n(J) = f_{S_i^*, |N_i^*|=n}|Z_i, |N_i^*|=n(s) f_{N_i^*}|Z_i, |N_i^*|=n(J).
\]
In addition, due to $S_i^* = \sum_{j \in \mathcal{N}_i^*} D_j$ and Assumption 3.4 (a), it is easy to see that $|\mathcal{N}_i| \perp S_i^* | Z_i, \mathcal{N}_i^*$ Thus, similar arguments used to show (B.10) give us

$$f_{S_i^* | Z_i, | \mathcal{N}_i| = n, | \mathcal{N}_i^*| = n^*}(s) = \sum_{J \in \Omega_{\mathcal{N}_i^*}, \text{ s.t. } |J|=n^*} f_{S_i^* | Z_i, | \mathcal{N}_i| = n, | \mathcal{N}_i^*| = n^*}(s) \times f_{\mathcal{N}_i^* | Z_i, | \mathcal{N}_i| = n, | \mathcal{N}_i^*| = n^*}(J)$$

$$= \sum_{J \in \Omega_{\mathcal{N}_i^*}, \text{ s.t. } |J|=n^*} f_{S_i^* | Z_i, | \mathcal{N}_i^*| = n^*}(s) \times f_{\mathcal{N}_i^* | Z_i, | \mathcal{N}_i| = n, | \mathcal{N}_i^*| = n^*}(J)$$

$$= f_{S_i^* | Z_i, | \mathcal{N}_i^*| = n^*}(s), \quad (B.9)$$

where the second equality is due to $|\mathcal{N}_i| \perp S_i^* | Z_i, \mathcal{N}_i^*$, the third equality is because of $\mathcal{N}_i^* \perp S_i^* | Z_i, | \mathcal{N}_i|$, in proof (a). Hence, (B.9) permits that $|\mathcal{N}_i| \perp S_i^* | Z_i, | \mathcal{N}_i^*$.

(b) Given $S_i = \sum_{j \in \mathcal{N}_i} D_j$, according to Assumptions 3.2 (a) and 3.4 (a), $\{D_j\}_{j \in \mathcal{P}}$ are i.i.d. and independent to $(Z_i, \mathcal{N}_i)$. Thus, applying the same arguments used to show (a), we can obtain for $s \leq n$, $f_{S_i | Z_i, | \mathcal{N}_i| = n, \mathcal{N}_i}(s) = f_{S_i | Z_i, | \mathcal{N}_i| = n}(s) = C_n^s f_D(s) f_D(0)^{(n-s)}$, leading to $\mathcal{N}_i \perp S_i | Z_i, | \mathcal{N}_i|$.

Moreover, because $S_i = \sum_{j \in \mathcal{N}_i} D_j$ is a function of $(\mathcal{N}_i, \{D_j\}_{j \in \mathcal{N}_i})$, the randomness of $S_i$ given $\mathcal{N}_i$ only comes from $D_j$ for $j \in \mathcal{N}_i$. In addition, since $\{D_j\}_{j \in \mathcal{P}}$ are independent to $(Z_i, \mathcal{N}_i^*, \mathcal{N}_i)$ as in Assumption 3.4 (a), it implies that $|\mathcal{N}_i^*| \perp S_i | Z_i, \mathcal{N}_i$. Hence, for $\forall (s, n, n^*) \in \Omega_{S_i | \mathcal{N}_i| | \mathcal{N}_i^*|}$,

$$f_{S_i | Z_i, | \mathcal{N}_i| = n, | \mathcal{N}_i^*| = n^*}(s) = \sum_{J \in \Omega_{\mathcal{N}_i^*}, \text{ s.t. } |J|=n} f_{S_i | Z_i, | \mathcal{N}_i| = n, | \mathcal{N}_i^*| = n^*}(s) \times f_{\mathcal{N}_i^* | Z_i, | \mathcal{N}_i| = n, | \mathcal{N}_i^*| = n^*}(J)$$

$$= \sum_{J \in \Omega_{\mathcal{N}_i^*}, \text{ s.t. } |J|=n} f_{S_i | Z_i, | \mathcal{N}_i| = n}(s) \times f_{\mathcal{N}_i^* | Z_i, | \mathcal{N}_i| = n, | \mathcal{N}_i^*| = n^*}(J)$$

$$= f_{S_i | Z_i, | \mathcal{N}_i| = n}(s), \quad (B.10)$$

where the second equality is because $|\mathcal{N}_i^*| \perp S_i | Z_i, \mathcal{N}_i$, the third equality is due to $\mathcal{N}_i \perp S_i | Z_i, | \mathcal{N}_i|$, as shown at the beginning of this proof. Therefore, $S_i \perp |\mathcal{N}_i^*| | Z_i, | \mathcal{N}_i|$ from (B.10).

(c) The proof in this step follows directly from the proofs in (a) and (b).

**Proof of Proposition 4.2.** Recall that by Bayes’ Theorem, we have

$$f_{S_i^* | \mathcal{N}_i| | D_i, S_i, Z_i, | \mathcal{N}_i|} = \frac{f_{S_i, | \mathcal{N}_i| | D_i, S_i^*, Z_i, | \mathcal{N}_i^*|} \times f_{S_i^* | \mathcal{N}_i^*| | D_i, Z_i}}{f_{S_i, | \mathcal{N}_i| | D_i, Z_i}}. \quad (B.11)$$

In what follows, we further rewrite the distributions in the numerator and the denominator to achieve the desired result. Based on Assumptions 3.2 and 3.4, we know that $\{D_i\}_{i \in \mathcal{P}}$ is i.i.d. and independent to $(Z_i, \mathcal{N}_i^*, \mathcal{N}_i)_{i \in \mathcal{P}}$. Thus, from the fact that $i \not\in \mathcal{N}_i^*$ and $i \not\in \mathcal{N}_i$, we can conclude that $D_i \perp (S_i^*, S_i, Z_i, \mathcal{N}_i^*, \mathcal{N}_i)$, for $\forall i \in \mathcal{P}$. \quad (B.12)

It further yields that $D_i \perp S_i | (S_i^*, Z_i, | \mathcal{N}_i^*|, | \mathcal{N}_i|)$ and $D_i \perp | \mathcal{N}_i| | (S_i^*, Z_i, | \mathcal{N}_i^*|)$. Therefore, consider
the first term in the numerator, for any \((s, n) \in \Omega_{s,|N|}\)

\[
f_{s_i,|N_i||D_i, S_i^*, Z_i, |N_i^*|}(s, n) = f_{S_i,|N_i||D_i, S_i^*, Z_i, |N_i^*|}(s) \times f_{|N_i||D_i, S_i^*, Z_i, |N_i^*|}(n)
= f_{S_i,|N_i^*|, |N_i^*|}(s) \times f_{|N_i||S_i^*, Z_i, |N_i^*|}(n)
= f_{S_i,|N_i^*|, |N_i^*|}(s) \times f_{|N_i||Z_i, |N_i^*|}(n),
\]

(B.13)

where the last equality is because of \(|N_i^*| \perp S_i^* | Z_i, |N_i^*|\) in Lemma 4.1. Besides, again by (B.12), we have \(D_i \perp S_i^* | Z_i, |N_i^*|\) and \(D_i \perp |N_i^*| | Z_i\). For the second term in the numerator,

\[
f_{S_i,|N_i||D_i, Z_i}(s, n) = f_{S_i,|N_i||D_i, Z_i}(s) \times f_{|N_i||D_i, Z_i}(n)
= f_{S_i,|N_i||D_i, Z_i}(s) \times f_{|N_i||Z_i}(n).
\]

(B.14)

Similarly, by (B.12), we can rewrite the denominator

\[
f_{s_i,|N_i||D_i, Z_i}(s, n) = f_{S_i,|N_i||D_i, Z_i}(s) \times f_{|N_i||Z_i}(n).
\]

(B.15)

Now, substituting (B.13), (B.14) and (B.15) into (B.11) leads to the stated result. ■

**Proof of Theorem 4.3.** (a) Due to the Assumption 3.3, it is clear that

\[
f_{Z_i, f_{|N_i^*||Z_i}|, f_{|N_i||N_i^*||Z_i|}, f_{|N_i||Z_i||N_i^*||Z_i|}, \text{ are all identical across } i \in \mathcal{P}. \tag{B.16}
\]

Now, according to \(|\tilde{N}_i| \perp |N_i||Z_i, |N_i^*|\) in Assumption 4.1, we can obtain

\[
f_{|N_i^*||\tilde{N}_i||N_i||Z_i|, z_i} = f_{|N_i||\tilde{N}_i||Z_i, |N_i^*|} \times f_{|N_i^*||Z_i|, z_i} = f_{|N_i||\tilde{N}_i||Z_i, |N_i^*|} \times f_{|N_i||Z_i, |N_i^*|} \times f_{|N_i||Z_i, |N_i^*|}.
\]

Because all the terms on the right hand side of the above equation are identical for all \(i\), then \(f_{|N_i^*||\tilde{N}_i||N_i||Z_i|, z_i}\) is identical for all \(i\), so as all the marginal and conditional distributions of \((|N_i^*|, |N_i||Z_i|, Z_i)\), which include \(f_{|N_i||Z_i|}\) and \(f_{|N_i||Z_i, |N_i^*|} \times f_{|N_i||Z_i, |N_i^*|}\).

In addition, recall that \(Y_i = r(D_i, S_i^*, Z_i, |N_i^*|, \varepsilon_i)\) as in Assumption 3.1. By Lemma B.2 and (B.12), we have \((\varepsilon_i, D_i) \perp (S_i^*, |N_i^*|, |N_i||Z_i|)\). Moreover, from Lemma 4.1 we know that \(S_i^* \perp (|N_i^*|, |N_i||Z_i|)\). Therefore,

\[
f_{|N_i^*||\tilde{N}_i||N_i||S_i^*, \varepsilon_i, D_i, Z_i, |N_i^*|} = f_{|N_i||\tilde{N}_i||N_i||S_i^*, \varepsilon_i, D_i, Z_i, |N_i^*|} \times f_{|N_i^*||Z_i|, z_i}
= f_{D} \times f_{\varepsilon_i,|N_i^*|} \times f_{|\tilde{N}_i||N_i||S_i^*, D_i, Z_i, |N_i^*|} \times f_{|N_i^*||Z_i|, z_i}
= f_{D} \times f_{\varepsilon_i,|N_i^*|} \times f_{S_i^*|Z_i, |N_i^*|} \times f_{\tilde{N}_i||Z_i, |N_i^*|} \times f_{|N_i||Z_i, |N_i^*|} \times f_{|N_i||Z_i, |N_i^*|} \times f_{|N_i||Z_i, |N_i^*|}.
\]

By the identical distribution of \(\varepsilon_i\) given \(Z_i, |N_i^*|\) in Assumption 3.3, and \(f_{S_i^*|Z_i, |N_i^*|} = C_n^{s*} f_{D}^{(1)}(1) f_{D}(0)^{(n_s - s^*)}\), together with (B.16), we can conclude that \((|N_i^*|, |N_i||Z_i), S_i^*, \varepsilon_i, D_i, Z_i)\) is identically distributed and so as \((|N_i^*|, |N_i||Z_i), Y_i, Z_i)\).

(b) In this proof, we first show that \(f_{|\tilde{N}_i||Z_i, |N_i^*|}\) and \(f_{|N_i||Z_i, |N_i^*|}\) are identified. We then verify the identification of \(f_{|N_i||Z_i, |N_i^*|}\). By the law of total probability, for any \((\tilde{n}, n, y) \in \Omega_{|\tilde{N}_i||N_i||Y_i}\),

\[
f_{|\tilde{N}_i||N_i||Y_i|, |Z_i, |N_i^*|}(\tilde{n}, n, y)
= \sum_{n^* \in \Omega_{|N_i^*|}} f_{|\tilde{N}_i||N_i||Y_i|, |Z_i, |N_i^*|, \tilde{n}^*}(\tilde{n}, n, y) \times f_{|N_i^*||Z_i|, n^*}(n^*)
\]

49
where the last equality is due to Assumption 4.1 and Lemma B.3. Integrate both sides of (B.17)

\[
\int_{y \in \Omega_Y} y f_{\tilde{X}_i|\tilde{N}, Y_i|Z_i}(\tilde{n}, n, y) dy
= \sum_{n^* \in \Omega_{|N^*|}} E[Y_i|Z_i, |N^*| = n^*] \times f_{\tilde{X}_i|\tilde{N}, |N^*| = n^*}(\tilde{n}) \times f_{|N^*||Z_i}(n^*) \times f_{|N^*|}(n^*).
\]

(B.18)

Besides, for any \((\tilde{n}, n) \in \Omega_{|\tilde{N}|, |N^*|}\), because of Assumption 4.1

\[
f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(\tilde{n}, n) = \sum_{n^* \in \Omega_{|N^*|}} f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(\tilde{n}) \times f_{|N^*||Z_i}(n^*)
= \sum_{n^* \in \Omega_{|N^*|}} f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(\tilde{n}) \times f_{|N^*||Z_i}(n^*) \times f_{|N^*|}(n^*).
\]

(B.19)

Recall that the notations below from the main text: for \(\forall y \in \Omega_Y\), the \(K_{|\tilde{N}||N^*|} \times K_{|\tilde{N}||N^*|}\) matrices

\[
E_{|\tilde{N}||N^*||Z_i} = \begin{bmatrix}
\int_{y \in \Omega_Y} y f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(0, 0, y) dy & \cdots & \int_{y \in \Omega_Y} y f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(0, K_{|\tilde{N}||N^*|} - 1, y) dy \\
\vdots & \ddots & \vdots \\
\int_{y \in \Omega_Y} y f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(K_{|\tilde{N}||N^*|} - 1, 0, y) dy & \cdots & \int_{y \in \Omega_Y} y f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(K_{|\tilde{N}||N^*|} - 1, K_{|\tilde{N}||N^*|} - 1, y) dy
\end{bmatrix},
\]

\[
F_{|\tilde{N}||N^*||Z_i} = \begin{bmatrix}
f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(0, 0) & \cdots & f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(0, K_{|\tilde{N}||N^*|} - 1) \\
\vdots & \ddots & \vdots \\
f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(K_{|\tilde{N}||N^*|} - 1, 0) & \cdots & f_{\tilde{X}_i|\tilde{N}, |N^*|}|Z_i(K_{|\tilde{N}||N^*|} - 1, K_{|\tilde{N}||N^*|} - 1)
\end{bmatrix}.
\]

In addition, recall and denote two \(K_{|\tilde{N}||N^*|}\) diagonal matrices

\[
T_{|\tilde{N}^*||Z_i} = \text{diag}(E[Y_i|Z_i, |\tilde{N}^*| = 0], E[Y_i|Z_i, |\tilde{N}^*| = 1], \cdots, E[Y_i|Z_i, |\tilde{N}^*| = K_{|\tilde{N}||N^*|} - 1]),
\]

\[
T_{|\tilde{N}^*||Z_i} = \text{diag}(f_{\tilde{N}^*||Z_i}(0), f_{\tilde{N}^*||Z_i}(1), \cdots, f_{\tilde{N}^*||Z_i}(K_{|\tilde{N}||N^*|} - 1)).
\]

Then, given the notations above, (B.17) and (B.19) can be rewritten in the following expressions:

\[
E_{|\tilde{N}||N^*||Y_i} = F_{|\tilde{N}||N^*||Z_i} \times T_{|\tilde{N}||Z_i}} \times T_{|\tilde{N}^*||Z_i} \times F'_{|\tilde{N}||N^*||Z_i},
\]

(B.20)

\[
F_{|\tilde{N}||N^*||Z_i} = F_{|\tilde{N}||N^*||Z_i} \times T_{|\tilde{N}^*||Z_i} \times F'_{|\tilde{N}||N^*||Z_i},
\]

(B.21)

where \(F_{|\tilde{N}||N^*||Z_i}\) and \(F_{|\tilde{N}||N^*||Z_i}\) are defined in the main text. Based on Assumption 4.3, we know that \(F_{|\tilde{N}||N^*||Z_i}\) and \(F_{|\tilde{N}||N^*||Z_i}\) are invertible. In addition, based on Assumption 4.4 (b), we have that \(f_{\tilde{N}^*||Z_i}(n) > \eta > 0\) for \(\forall n \in \Omega_{|\tilde{N}^*|}\) indicates the invertibility of \(T_{|\tilde{N}^*||Z_i}\). Hence, (B.21) implies that \(F_{|\tilde{N}||N^*||Z_i}\) is also invertible. It then yields from (B.20) and (B.21) that the square matrix
Proof of Lemma 4.4. Recall that \( \Delta E_{|\mathcal{N}|,|\mathcal{V}|,|y|Z} \) can be factorized as
\[
E_{|\mathcal{N}|,|\mathcal{V}|,|y|Z} \times F^{-1}_{|\mathcal{N}|,|\mathcal{V}|} = F_{|\mathcal{N}|,|\mathcal{V}|} \times T_{|Y|Z,|\mathcal{V}^*|} \times F^{-1}_{|\mathcal{N}|,|\mathcal{V}|},
\]
where the matrix \( E_{|\mathcal{N}|,|\mathcal{V}|,|y|Z} \times F^{-1}_{|\mathcal{N}|,|\mathcal{V}|} \) on the left hand side of the above equation is identifiable from the observed data, and the right hand side corresponds to its eigen-decomposition, whose eigenvalues are the diagonal entries of \( T_{|Y|Z,|\mathcal{V}^*|} \).

By Assumption 4.4 (a), all the \( K_{|\mathcal{V}|} \) eigenvalues in the diagonal matrix \( T_{|Y|Z,|\mathcal{V}^*|} \) are strictly positive and distinct. Thus, given the eigen-decomposition of matrix \( E_{|\mathcal{N}|,|\mathcal{V}|,|y|Z} \times F^{-1}_{|\mathcal{N}|,|\mathcal{V}|} \) in (B.22), its \( K_{|\mathcal{V}|} \) eigenvectors are linearly independent and are corresponding to the \( K_{|\mathcal{V}|} \) columns of \( F_{|\mathcal{N}|,|\mathcal{V}|,|y|Z} \). By simple algebra, we can solve the \( K_{|\mathcal{V}|} \) eigenvectors, meaning that the columns of \( F_{|\mathcal{N}|,|\mathcal{V}|,|y|Z} \) are identifiable. Moreover, Assumption 4.4 (b) ensures there is an unique maximum entry of each eigenvector, and its location reveals which eigenvalue it corresponds to. For example, if the largest value in some eigenvector appears in its first entry, then this eigenvector gives the latent probabilities \( f_i(\mathcal{N}^*_i|z_i,|\mathcal{V}^*_i|=1(1), f_i(\mathcal{N}^*_i|z_i,|\mathcal{V}^*_i|=1(2), ..., f_i(\mathcal{N}^*_i|z_i,|\mathcal{V}^*_i|=1(K_{|\mathcal{V}|}))' \) and corresponds to the eigenvalue \( E[Y_i|Z_i,|\mathcal{V}^*_i| = 1] \). Because the summation of each column in the matrix \( F_{|\mathcal{N}|,|\mathcal{V}|,|y|Z} \) is naturally normalized to be one, there is an unique solution for each eigenvector. The above discussions verify that \( F_{|\mathcal{N}|,|\mathcal{V}|,|y|Z} \) can be nonparametrically identified. Same arguments can be use to show the identifcation of \( F_{|\mathcal{V}|,|\mathcal{V}|,|y|Z} \).

Next, let us move on to \( f_i(\mathcal{N}^*_i|z_i) \). Define two \( K_{|\mathcal{V}|} \times 1 \) vectors as
\[
F_{|\mathcal{N}^*_i|Z_i} = \begin{bmatrix} f_i(\mathcal{N}^*_i|z_i(0)) & f_i(\mathcal{N}^*_i|z_i(1)) & \cdots & f_i(\mathcal{N}^*_i|z_i(K_{|\mathcal{V}|} - 1)) \end{bmatrix}',
F_{|\mathcal{V}|} = \begin{bmatrix} f_i(\mathcal{N}^*_i|z_i(0)) & f_i(\mathcal{N}^*_i|z_i(1)) & \cdots & f_i(\mathcal{N}^*_i|z_i(K_{|\mathcal{V}|} - 1)) \end{bmatrix}'.
\]
Based on the law of total probability, it is easy to get \( F_{|\mathcal{V}|} = F_{|\mathcal{V}|,|\mathcal{V}^*_i|,|y|Z} \times F_{|\mathcal{N}^*_i|Z_i} \). Since \( F_{|\mathcal{V}|,|\mathcal{V}^*_i|,|y|Z} \) is invertible, multiplying both sides of the above equation by \( F^{-1}_{|\mathcal{V}|,|\mathcal{V}^*_i|,|y|Z} \) gives us
\[
F_{|\mathcal{N}^*_i|Z_i} = F^{-1}_{|\mathcal{V}|,|\mathcal{V}^*_i|,|y|Z} \times F_{|\mathcal{V}|,|\mathcal{V}^*_i|,|y|Z},
\]
which indicates the identifiability of \( F_{|\mathcal{N}^*_i|Z_i} \).

---

**Proof of Lemma 4.4.** Recall that \( \Delta S_i := S_i - S_i^* \). (a) Consider the case \( \mathcal{N}^*_i \subset \mathcal{N}_i \). For \( \forall (s, n) \in \Omega_{S,|\mathcal{V}^*_i|} \) and \( (s^*, n^*) \in \Omega_{S,|\mathcal{V}^*_i|} \) such that \( n^* \leq n \)
\[
f_{S_i|S_i^* = s^*, Z_i, |\mathcal{N}_i| = n, |\mathcal{V}^*_i| = n^*} (s) = \Delta f_{S_i|S_i^* = s^*, Z_i, |\mathcal{N}_i| = n, |\mathcal{V}^*_i| = n^*} (s - s^*)
= \sum_{(J^*, J)} f_{\Delta S_i|S_i^* = s^*, Z_i, |\mathcal{N}_i| = n, |\mathcal{V}^*_i| = n^*} (s - s^*) \times f_{\mathcal{N}^*_i|S_i^* = s^*, Z_i, |\mathcal{N}_i| = n, |\mathcal{V}^*_i| = n^*} (J^*, J),
\]
where the last line is based on the law of total probability. Because \( \mathcal{N}^*_i \subset \mathcal{N}_i \), we have that \( \mathcal{N}^*_i / \mathcal{N}_i \) is empty and \( \Delta S_i = \sum_{j \in \mathcal{N}^*_i / \mathcal{N}_i} D_j \). In addition, \( \mathcal{N}_i / \mathcal{N}^*_i \) and \( \mathcal{N}^*_i \) are mutually exclusive, i.e. if \( i \in \mathcal{N}_i / \mathcal{N}^*_i \) then \( i \notin \mathcal{N}^*_i \). Due to the i.i.d. of \( \{D_i\}_{i \in \mathcal{P}} \) (Assumption 3.2), and the independence between \( \{D_i\}_{i \in \mathcal{P}} \) and \( (Z_i, \mathcal{N}^*_i, \mathcal{N}_i) \) (Assumption 3.4), we have that \( \Delta S_i \perp S_i^* | Z_i, \mathcal{N}^*_i, \mathcal{N}_i \). Therefore,
\[
f_{\Delta S_i|S_i^* = s^*, Z_i, |\mathcal{N}_i| = n, |\mathcal{V}^*_i| = n^*} (s - s^*) = \Delta f_{S_i|Z_i, |\mathcal{N}_i| = n, |\mathcal{V}^*_i| = n^*} (s - s^*).
\]
Again by the independence of \( \{D_i\}_{i \in \mathcal{P}} \), once conditional on \( N_i/N_i^* = J/J^* \), we know that \( \Delta S_i = \sum_{j \in J/J^*} D_j \) follows a binomial distribution if \( s^* \leq s \) and \( \Delta s \leq \Delta n \), and is independent to the identity of network neighbors contained in \( (N_i^*, N_i) \). Then (B.25) becomes

\[
\begin{align*}
&f_{\Delta S_i | Z_i, |N_i| = n, |N_i^*| = n^*} = f_{\Delta S_i | Z_i, |N_i| = n} f_{|N_i^*| = n^*} = f_{\Delta S_i | Z_i} f_{|N_i| = n} f_{|N_i^*| = n^*} f_{J/J^*} (s - s^*) \\
&= f_{\Delta S_i | Z_i, |N_i| = n} f_{|N_i| = n} f_{|N_i^*| = n^*} f_{J/J^*} (s - s^*) \\
&= f_{\Delta S_i | Z_i} f_{|N_i| = n} f_{|N_i^*| = \Delta n} (\Delta s),
\end{align*}
\]

where the last equality follows the same arguments used to show Lemma 4.1 (a). Substituting (B.26) into (B.24) gives the desired result.

(b) Similar arguments used in proof for the case \( N_i^* \subset N_i \) can be applied to obtain the result for the case \( N_i \subset N_i^* \). Therefore, we omit the proof. \( \blacksquare \)

**Proof of Theorem 4.5.** (a) From Proposition 4.2, we know that

\[
f_{S_i^* | |N_i^*||D_i, S_i, Z_i, |N_i|} = \frac{f_{S_i^* | Z_i, |N_i^*||N_i|} f_{S_i^* | Z_i, |N_i^*|} f_{|N_i||N_i^*|} f_{Z_i | N_i^*|}}{f_{S_i^* | Z_i, |N_i|}}.
\]

Based on Lemma 4.1, we know that \( f_{S_i^* | Z_i, |N_i^*|=n^*} = C_{n^*} f_D(1) f_D(0)^{(n^*-s^*)} \) and \( f_{S_i | Z_i, |N_i|} = n(s) = C_n f_D(1) f_D(0)^{(n-s)} \). Similarly, from Lemma 4.4,

\[
f_{S_i | S_i^* = s^*, Z_i, |N_i| = n^*} = C_{\Delta n} f_D(1) f_D(0)^{(\Delta n - \Delta s)}.
\]

Because \( D_i \) is i.i.d., by Assumptions 3.3 and 3.4, \( f_{|N_i^*||Z_i|} f_{Z_i | |N_i^*||N_i|} \) and \( f_{|N_i||Z_i|} \) are identical for all \( i \). Therefore, all distributions on the right hand side of (B.27) are identical across \( i \), together with Theorem 4.3, \( f_{S_i | S_i^* = s^*, Z_i, |N_i| = n^*} \) can be nonparametrically identified. \( \blacksquare \)

**Proof of Theorem 4.6.** (a) From Theorem 4.5, we know that \( f_{S_i^* | |N_i^*||D_i, S_i, Z_i, |N_i|} \) is identical for all \( i \), together with Proposition 3.1, we know \( m_i \) is also identical for all \( i \in \mathcal{P} \).

(b) To ease the notation, denote \( T_i = (S_i, |N_i|)' \) and \( T_i^* = (S_i^*, |N_i^*|)' \). According to Assumption 4.2, the support of \( T_i \) and \( T_i^* \) are the same, and we denote it as \( \Omega_T = \{t_1, t_2, ..., t_{K_T}\} \) with \( t_k = (s_k, n_k) \in \Omega_{S_i|N_i} \). Let us rank the possible values in \( \Omega_T \) by the lexicographical ordering, according to the natural order of the integers in \( \Omega_{S_i|N_i} \), i.e.

\[
\begin{align*}
t_1 &= (0, 0), \\
t_2 &= (0, 1), \\
t_3 &= (1, 1), \\
t_4 &= (0, 1), \\
t_5 &= (1, 2), \\
t_6 &= (2, 2), \\
&\quad \ldots \\
t_{(K_{|N_i|-1})/2} &= (0, K_{|N_i|} - 1), \ldots, t_{(K_{|N_i|}-2)/2+1} = (K_{|N_i|} - 1, K_{|N_i|} - 1).
\end{align*}
\]

Because result in (a), \( m_i(\cdot) \) is identical for all \( i \), thus we suppress the subscript \( i \), i.e. \( m(d, s, z, n) := E[Y_i | D_i = d, S_i = s, Z_i = z, |N_i| = n] \). By notation abuse, we ignore the arguments \( (d, z) \) in functions \( m \) and \( m^* \), and introduce the following notations. For any \( (d, z) \in \{0, 1\} \times \Omega_{Z} \), denote
When \( N_f \) and \( M \) satisfy Assumption 4.5, we need to consider two cases. \( N_f \)

where \( m(t_k) \) represents the mean function \( m(d, s_k, z, n_k) = E[Y_i | D_i = d, S_i = s_k, Z_i = z, |N_i| = n_k] \).

Define the \( \mathbf{K}_T \times \mathbf{K}_T \) matrix

\[
F_{T^* | T, D = d, Z = z} = \begin{bmatrix}
  f_{T^* | T = t_1, D = d, Z = z}(t_1) & \cdots & f_{T^* | T = t_1, D = d, Z = z}(t_{K_T}) \\
  \vdots & \ddots & \vdots \\
  f_{T^* | T = t_K, D = d, Z = z}(t_1) & \cdots & f_{T^* | T = t_K, D = d, Z = z}(t_{K_T})
\end{bmatrix}.
\] (B.32)

From Proposition 3.1 and the notations in (B.30)-(B.32), we have for any \((d, z) \in \{0, 1\} \times \Omega_Z\)

\[
M_{Y | T, D = d, Z = z}(m) = F_{T^* | T, D = d, Z = z} \times M_{Y | T^*, D = d, Z = z}(m^*). \tag{B.33}
\]

Given Proposition 4.2, for \( \forall (d, z) \in \{0, 1\} \times \Omega_Z \), the elements in the main diagonal of \( F_{T^* | T, D = d, Z = z} \)

\[
\frac{f_{S^*_i | N^*_i \mid D_i = d, S_i = s, Z_i = z, |N_i| = n}(s, n)}{f_{S_i | Z_i = z, |N_i| = n}(s) \times f_{N^*_i | |N_i| = n}(n) \times f_{N^*_i | Z_i = z}(n)} \tag{B.34}
\]

where the second equality is because of Lemma 4.1 and Lemma 4.4. In addition, based on Assumption 4.4 (b), we know that \( f_{N^*_i | Z_i = z, |N_i| = n}(n) > 0 \), which also leads to \( f_{N^*_i | Z_i = z}(n) > 0 \). Therefore, by the preassumption that \( f_{N^*_i | Z_i = z}(n) > 0 \), we can conclude that

\[
\frac{f_{S^*_i | N^*_i \mid D_i = d, S_i = s, Z_i = z, |N_i| = n}(s, n)}{f_{S_i | Z_i = z, |N_i| = n}(s) \times f_{N^*_i | |N_i| = n}(n) \times f_{N^*_i | Z_i = z}(n)} > 0 \quad \forall (s, n) \in \Omega_{S^*_i \mid |N^*_i|}.
\]

In what follows, we prove the desired result in two steps. Firstly, we show that the square matrix \( F_{T^* | T, D = d, Z = z} \) is invertible. Secondly, we show that the CASF \( m^* \) is identifiable from (B.33).

**Step 1.** Consider any \( t^* = (s^*, n^*) \) and \( t = (s, n) \) such that \( 0 \leq s^* \leq n^* \) and \( 0 \leq s \leq n \). Under Assumption 4.5, we need to consider two cases.

Firstly, suppose \( N^*_i \subseteq N_i \) holds. Then, we know that \( S_i^* \leq S_i \) and \( |N^*_i| \leq |N_i| \). Thus, \( f_{T^* | T = t, D = d, Z = z}(t^*) = 0 \) if at least one of the restrictions \( s^* \leq s \) and \( n^* \leq n \) is violated. Similarly, when \( N^*_i \subseteq N^*_i \) holds, we have that \( S_i \leq S_i^* \) and \( |N_i| \leq |N^*_i| \). Then, \( f_{T^* | T = t, D = d, Z = z}(t^*) = 0 \) if at least one of the restrictions \( s \leq s^* \) and \( n \leq n^* \) is violated. Given the lexicographical ordering of the elements in \( \Omega_T \), it is easy to see that the matrix \( F_{T^* | T, D = d, Z = z} \) is lower triangular if \( N^*_i \subseteq N_i \), and is upper triangular if \( N_i \subseteq N^*_i \). Moreover, (B.34) implies that all the elements on the main diagonal of the triangular matrix \( F_{T^* | T, D = d, Z = z} \) are strictly positive. Since the eigenvalues of a triangular matrix are its diagonal entries, the matrix \( F_{T^* | T, D = d, Z = z} \) is therefore invertible.

**Step 2.** Next, we show that the CASF \( m^* \) is identifiable. Suppose \( m^* \) is not identifiable, then there exists \( \tilde{m}^* \neq m^* \) such that \( \tilde{m}^* \) is observationally equivalent to \( m^* \), in the sense that (B.33)
also holds for \( \tilde{m}^* \):

\[
M_Y|T,D=d,Z=z(m) = F_{T^*|T,D=d,Z=z}M_Y|T^*,D=d,Z=z(\tilde{m}^*). \tag{B.35}
\]

It then yields from (B.33) and (B.35) that

\[
0 = F_{T^*|T,D=d,Z=z} \left[ M_Y|T^*,D=d,Z=z(m^*) - M_Y|T^*,D=d,Z=z(\tilde{m}^*) \right]. \tag{B.36}
\]

Since \( F_{T^*|T,D=d,Z=z} \) is invertible, it follows from (B.36) that

\[
M_Y|T^*,D=d,Z=z(m^*) = M_Y|T^*,D=d,Z=z(\tilde{m}^*),
\]

meaning that \( \tilde{m}^*(t_k) = m^*(t_k) \) for all \( k = 1, 2, \ldots, K_T \), which contradicts \( \tilde{m}^* \neq m^* \). Therefore, we can conclude that \( m^* \) is identifiable. ■

### B.3 Proofs of Section 5

**Proof of Theorem 5.2.** For illustration simplicity, by notation abuse, we denote \( W_i \) as any generic vector of observable variables of interest, where \( W_i = (W_{i,d}^c, W_{i,d}^d)' \in \Omega_{W^c} \times \Omega_{W^d} \), with the \( Q \times 1 \) vector \( W_i^c := (W_{i1}^c, \ldots, W_{iQ}^c)' \) containing continuous variables and the vector \( W_i^d \) containing discrete variables. In this proof, we focus on the uniform convergence rate of the kernel estimation \( \hat{f}_{W_i}(w) \). Then, replacing \( W_i \) by the observable variables of interest gives the stated results.

Denote \( w = (w^c, w^d)' \) with \( w^c = (w_{1}^c, \ldots, w_{Q}^c)' \) and \( \hat{f}_{W_i}(w) = 1/N \sum_{i=1}^{N} \hat{f}_{ker}(w) \), where

\[
\hat{f}_{ker}(w) := K(W_i^c, w^c)1 \left[ W_i^d = w^d \right], \tag{B.37}
\]

with \( K(W_i^c, w^c) = h^{-Q} \prod_{q=1}^{Q} \kappa \left( (W_{iq}^c - w_q^c)/h \right) \). Let \( f_{W_i}(w) \) be the true distribution of \( W_i \). For any \( w \in \Omega_W \),

\[
\left| \hat{f}_{W_i}(w) - f_{W_i}(w) \right| \leq \left| \hat{f}_{W_i}(w) - E \left[ \hat{f}_{W_i}(w) \right] \right| + \left| E \left[ \hat{f}_{W_i}(w) \right] - f_{W_i}(w) \right|.
\]

Given the inequality above, we prove the uniform convergence of \( |\hat{f}_{W_i}(w) - f_{W_i}(w)| \) and its rate in two steps. In Step 1, we show that the bias of \( \hat{f}_{W_i}(w) \), i.e. \( |E[\hat{f}_{W_i}(w)] - f_{W_i}(w)| \), is \( O(h^2) \) uniformly. In Step 2, we show the uniform convergence of \( \hat{f}_{W_i}(w) \) to \( E[\hat{f}_{W_i}(w)] \) and establish its convergence rate.

**Step 1.** Firstly, let \( w^d \) and \( w^c \) be any generic element in \( \Omega_{W^d} \) and \( \Omega_{W^c} \), respectively. Then, for \( w = (w^c, w^d)' \)

\[
E \left[ \hat{f}_{ker}(w) \right] = h^{-Q} \sum_{w^d \in \Omega_{W^d}} \left[ 1[w^d = w^d] \int \prod_{q=1}^{Q} \kappa \left( \frac{w_q^c - w_q^c}{h} \right) f_{W_i^c, W_i^d}(w^c, w^d) dw^c \right],
\]

by changing of variables using \( v = (v_1, \ldots, v_Q)' \) with \( v_q = (w_q^c - w_q^c)/h \) and \( q = 1, \ldots, Q \),

\[
E \left[ \hat{f}_{ker}(w) \right] = \sum_{w^d \in \Omega_{W^d}} \left[ 1[w^d = w^d] \int \prod_{q=1}^{Q} \kappa(v_q) f_{W_i^c, W_i^d}(w^c + hv, w^d) dv \right].
\]
where we denote the shorthand notation \( w^c + hv := (w^c_1 + hv_1, ..., w^c_Q + hv_Q) \). Let the \( Q \times 1 \) vector \( f_c^{(1)}(w) := \partial f_{W_i}(w)/\partial w^c \) represent the first order derivative of \( f_{W_i}(w) \) with respect to \( w^c \), and let the \( Q \times Q \) matrix \( f_c^{(2)}(w) := \partial^2 f_{W_i}(w)/\partial w^c \partial w^c \) be the second order derivative of \( f_{W_i} \) with respect to \( w^c \). Consider the Taylor series expansion of \( f_{W_i, W_i}(w^c + hv, w^d) \) around \( w \):

\[
 f_{W_i, W_i}(w^c + hv, w^d) - f_{W_i, W_i}(w^c, w^d) = hf_c^{(1)}(w)'v + h^2v'f_c^{(2)}(\bar{w})v
\]  

(39)

where \( \bar{w} \) is between \((w^c + hv, w^d) \) and \((w^c, w^d) \). Since \( W_i \) is identically distributed based on Theorems 4.3 and 4.5, we have \( E[\hat{f}_{W_i}(w)] = E[f_{\ker}(w)] \). Plugging (39) into (38) gives

\[
 E \left[ \hat{f}_{W_i}(w) \right] - f_{W_i}(w) = \int \left[ hf_c^{(1)}(w)'v + h^2v'f_c^{(2)}(\bar{w})v \right] \prod_{q=1}^Q \kappa(v_q) dv
\]

\[
 = hf_c^{(1)}(w)' \int v \prod_{q=1}^Q \kappa(v_q) dv + h^2 \int v'f_c^{(2)}(\bar{w})v \prod_{q=1}^Q \kappa(v_q) dv
\]

\[
 \leq Ch^2 \int v'v \prod_{q=1}^Q \kappa(v_q) dv
\]

\[
 = Ch^2 \sum_{q=1}^Q \int v_q^2 \kappa(v_q) dv_q,
\]

(40)

where the inequality is because that each element in \( f_c^{(2)} \) is bounded uniformly in \( w^c \), and the symmetric kernel function \( \kappa(\cdot) \) in Assumption 5.2 (c) implies \( \int \kappa(v_q) v_q dv_q = 0 \), thus \( \int v \prod_{q=1}^Q \kappa(v_q) dv = (\int v_1 \kappa(v_1) dv_1, ..., \int v_Q \kappa(v_Q) dv_Q)' = (0, ..., 0)' \). From (40), we get

\[
 \sup_{w \in \Omega_W} \left| E \left[ \hat{f}_{W_i}(w) \right] - f_{W_i}(w) \right| \leq \sup_{w \in \Omega_W} \left| Ch^2 \sum_{q=1}^Q \int \kappa(v_q) v_q^2 dv_q \right| \leq CK_1 Q h^2 = O(h^2).
\]

(41)

**Step 2.** Next, we show the uniform convergence of \( |\hat{f}_{W_i}(w) - E[\hat{f}_{W_i}(w)]| \). Since \( \Omega_{W^c} \) is compact and \( \Omega_{W^d} \) has finite dimension as in Assumption 5.2 (a), for some constant \( C > 0 \), \( \Omega_W \) can be covered by less than \( L_N = C l_N^{-Q} \) open balls of radius \( l_N \), where for any \( w = (w^c, w^d)' \), \( \bar{w} = (\bar{w}^c, \bar{w}^d)' \) in the same ball, we let \( w^d = \bar{w}^d \). Denote the centers of these open balls as \( \bar{w}_j \) with \( j = 1, 2, ..., J(\epsilon) \) and \( J(\epsilon) \leq L_N \). For any \( w, \bar{w} \) in the same ball, the mean value theorem implies that

\[
 \sup_{||w - \bar{w}|| < \epsilon} \left| \hat{f}_{W_i}(w) - \hat{f}_{W_i}(\bar{w}) \right| \leq \sup_{||w - \bar{w}|| < \epsilon} \frac{1}{N} \sum_{i=1}^N \left| K_W(W_i^c, w^c) - K_W(W_i^c, \bar{w}^c) \right|
\]

\[
 = \sup_{||w - \bar{w}|| < \epsilon} \frac{1}{Nh^Q} \sum_{i=1}^N \left| \prod_{q=1}^Q \kappa \left( \frac{W_{iq}^c - w_{iq}^c}{h} \right) - \prod_{q=1}^Q \kappa \left( \frac{W_{iq}^c - \bar{w}_{iq}^c}{h} \right) \right|
\]

55
(B.42)

\[ \leq C l_N h^{-(Q+1)}, \]

where \( w^c \) denotes some intermediate value between \( (W_i^c - w^c)/h \) and \( (W_i^c - \bar{w}^c)/h \), and \( \bar{w}^c \) represents the first order derivative of \( \prod_{q=1}^{Q} \kappa(v_q) \) to \( v = (v_1, ..., v_Q)' \). The last line of (B.42) is because of the boundedness of \( \kappa(\cdot) \) and the uniform boundedness of its first order derivative (Assumption 5.2). Let \( \bar{w}_{je} \) denote the center of an open ball containing \( w \). Then,

\[
\sup_{w \in \Omega_w} \left| \hat{f}_{W_i}(w) - E[\hat{f}_{W_i}(w)] \right| \leq \max_{1 \leq j \leq L_N} \sup_{\|w - \bar{w}_{je}\| < \epsilon} \left| \hat{f}_{W_i}(w) - \hat{f}_{W_i}(\bar{w}_{je}) \right|
\]

\[
+ \max_{1 \leq j \leq L_N} \left| \hat{f}_{W_i}(\bar{w}_{je}) - E[\hat{f}_{W_i}(\bar{w}_{je})] \right|
\]

\[
+ \max_{1 \leq j \leq L_N} \sup_{\|w - \bar{w}_{je}\| < \epsilon} \left| E[\hat{f}_{W_i}(w)] - E[\hat{f}_{W_i}(\bar{w}_{je})] \right|
\]

\[ := R_1 + R_2 + R_3. \quad (B.43) \]

By (B.42), we find immediately that \( R_1 \) and \( R_3 \) can be bounded as below

\[ R_1 \leq C_1 l_N h^{-(Q+1)}, \text{ and } R_3 \leq C_3 l_N h^{-(Q+1)}, \]

for some constants \( C_1, C_3 \). The main task is then to find the convergence rate of \( R_2 \). Denote

\[ Q_{N,i} := Q_{N,i}(w) = (\hat{f}_{i}^{ker}(w) - E[\hat{f}_{i}^{ker}(w)])/N, \]

where to ease the notation, we suppress the argument \( w \) in \( Q_{N,i}(w) \). Then, \( \hat{f}_{W_i}(w) - E[\hat{f}_{W_i}(w)] = \sum_{i=1}^{N} Q_{N,i} \). Following the method of Masry (1996), which aims at approximating dependent random variables by independent ones, we further divide the proof for \( R_2 \) into two parts:

- Step 2.1 construct the approximation process;
- Step 2.2 shows that the independent random variable approximation converges uniformly and verifies the uniform convergence for the reminder term.

**Step 2.1.** Recall that \( S_1, ..., S_{q_N} \) are the mutually exclusive partitions of index set \( \{1, 2, ..., N\} \) with \( \bigcup_{i=1}^{q_N} S_i = \{1, 2, ..., N\} \). Define \( V_N(k) = \sum_{i \in S_k} Q_{N,i} \), for \( k = 1, ..., q_N \) and

\[
\begin{cases} 
W^{q_N/2}_N = \sum_{k=1}^{q_N/2} V_N(2k - 1), & \text{if } q_N \text{ is even} \\
W^{q_N/2}_N = \sum_{k=1}^{(q_N+1)/2} V_N(2k - 1), & \text{if } q_N \text{ is odd}
\end{cases}
\]

\[
W^{(q_N-1)/2}_N = \sum_{k=1}^{(q_N-1)/2} V_N(2k),
\]

so that \( \hat{f}_{W_i}(w) - E[\hat{f}_{W_i}(w)] = W'_N + W''_N \) with \( W'_N \) and \( W''_N \) are the sums of \( Q_{N,i} \) over the odd-numbered subsets \( \{S_{2k-1}\} \) and even-numbered subsets \( \{S_{2k}\} \), respectively. Then, for any \( \eta > 0 \),

\[ \Pr(R_2 > \eta) \leq \Pr \left( \max_{1 \leq j \leq L_N} |W'_N(\bar{w}_{je})| > \eta/2 \right) + \Pr \left( \max_{1 \leq j \leq L_N} |W''_N(\bar{w}_{je})| > \eta/2 \right) \]
Next, we bound \( \Pr(|W_N'(w)| > \eta/2) \) by applying Lemma E.3 and approximating the odd-numbered \( \{V_N(2k-1)\} \) series by independent random variables. Enlarging the probability space if necessary, let us introduce a random variable sequence \( \{U_1, U_2, \ldots\} \) of mutually independent uniform \([0,1]\) random variables, which is also independent to the odd-numbered sequence \( \{V_N(2k-1)\} \). Define \( V_N^*(0) = 0 \) and \( V_N^*(1) = V_N(1) \). Then by Lemma E.3, for each \( k \geq 2 \), there is a random variable \( V_N^*(2k-1) \) that is a measurable function of \( \{V_N(1), V_N(3), \ldots V_N(2k-1), U_k\} \) satisfying the three conditions below:

(a) \( V_N^*(2k-1) \) is independent of \( \{V_N(1), V_N(3), \ldots, V_N(2k-3)\} \);
(b) \( V_N^*(2k-1) \) has the same distribution as \( V_N(2k-1) \);
(c) for any \( \mu \) such that \( 0 < \mu \leq \|V_N(2k-1)\|_\infty < \infty \),

\[
\Pr(|V_N^*(2k-1) - V_N(2k-1)| > \mu) \leq 18(\|V_N(2k-1)\|_\infty/\mu)^{1/2} \sup |Pr(AB) - Pr(A)Pr(B)|,
\]

where the inequality follows by setting the \( \gamma \) in Lemma E.3 as infinity, and the supremum is over all possible sets \( A \) and \( B \), for \( A, B \) in the \( \sigma \)-field of events generated by \( \{V_N(1), V_N(3), \ldots, V_N(2k-3)\} \) and by \( V_N(2k-1) \), respectively. Most importantly, such construction of \( V_N^*(2k-1) \) guarantees that \( V_N^*(1), V_N^*(3), \ldots, V_N^*(2k-1) \) are mutually independent with each other based on condition (a) above. Up to here, we have established the approximation of the dependent random sequence \( \{V_N(2k-1)\} \) by the independent one \( \{V_N^*(2k-1)\} \).

**Step 2.2.** Without loss of generality, let \( q_N \) be an even number. Then,

\[
\Pr(|W_N'(w)| > \eta/2) = \Pr\left(\sum_{k=1}^{q_N/2} |V_N(2k-1) - V_N^*(2k-1)| + \sum_{k=1}^{q_N/2} V_N^*(2k-1) > \eta/2\right) \\
\leq \Pr\left(\sum_{k=1}^{q_N/2} V_N^*(2k-1) > \eta/4\right) + \Pr\left(\sum_{k=1}^{q_N/2} |V_N(2k-1) - V_N^*(2k-1)| > \eta/4\right) \\
:= R_{21}(w) + R_{22}(w).
\]

Firstly, we bound \( R_{21}(w) \) as follows. Denote \( r_i = |\Delta(i,N)| \), then \( \bar{r}_N = \sup_{1 \leq i \leq N} r_i \). Noting that \( \kappa(\cdot) \) is bounded, let \( \sup_{w^* \in \Omega_{w^*}} |\prod_{q=1}^{Q} \kappa(v_q)| = A_1 \) for some constant \( A_1 > 0 \). Then, by construction,

\[
|Q_{N,i}(w)| \leq 2A_1(NhQ)^{-1}, \quad |V_N(k)| \leq 2r_k A_1(NhQ)^{-1} \leq 2\bar{r}_N A_1(NhQ)^{-1}.
\]

Let \( \lambda_N = C[NhQ \ln(N)]^{1/2} \) and we have that for \( N \) large enough, by choosing \( C \) properly,

\[
\lambda_N |V_N(k)| = 2CA_1 \bar{r}_N \left(\frac{\ln(N)}{NhQ}\right)^{1/2} \leq 1/2,
\]
Given that the density function 

$$f(a) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2}\right),$$

From the Markov inequality, for any generic random variable \(X\), constants \(c\) and \(a > 0\), we have

$$\Pr(X > c) \leq \frac{E[\exp(aX)]}{\exp(ac)}.$$

Consequently, based on the independence of \(\{V_N^*(2k-1)\}_{k=1}^{\lfloor qN/2 \rfloor}\) and (B.50),

$$R_{21}(w) = \Pr\left(\sum_{k=1}^{\lfloor qN/2 \rfloor} V_N^*(2k-1) > \eta/4\right)$$

$$= \Pr\left(\sum_{k=1}^{\lfloor qN/2 \rfloor} V_N^*(2k-1) > \eta/4\right) + \Pr\left(-\sum_{k=1}^{\lfloor qN/2 \rfloor} V_N^*(2k-1) > \eta/4\right)$$

$$\leq \left\{ E\left[\exp\left(\lambda_N \sum_{k=1}^{\lfloor qN/2 \rfloor} V_N^*(2k-1)\right)\right] + E\left[\exp\left(-\lambda_N \sum_{k=1}^{\lfloor qN/2 \rfloor} V_N^*(2k-1)\right)\right]\right\} / \exp(\lambda_N\eta/4)$$

$$\leq \left\{ \prod_{k=1}^{\lfloor qN/2 \rfloor} E[\exp(\lambda_N V_N^*(2k-1))] + \prod_{k=1}^{\lfloor qN/2 \rfloor} E[\exp(-\lambda_N V_N^*(2k-1))]\right\} / \exp(\lambda_N\eta/4)$$

$$\leq 2 \prod_{k=1}^{\lfloor qN/2 \rfloor} E[\exp(\lambda_N^2 V_N^{*2}(2k-1))] / \exp(\lambda_N\eta/4)$$

$$\leq 2 \exp\left(-\lambda_N\eta/4 + \lambda_N^2 \sum_{k=1}^{\lfloor qN/2 \rfloor} E[V_N^{*2}(2k-1)]\right)$$

(B.51)

where the first inequality is obtained by letting \(a = \lambda_N\) and \(c = \eta/4\) in the Markov inequality. Due that \(\{V_N(2k-1)\}\) and \(\{V_N^*(2k-1)\}\) have identical probability and \(V_N(k) = \sum_{i \in S_k} Q_{N,i}\)

$$\sum_{k=1}^{\lfloor qN/2 \rfloor} E[\sum_{k=1}^{\lfloor qN/2 \rfloor} V_N^{*2}(2k-1)] = \sum_{k=1}^{\lfloor qN/2 \rfloor} E[V_N^2(2k-1)] = \sum_{k=1}^{\lfloor qN/2 \rfloor} \sum_{i,j \in S_{2k-1}} Cov(Q_{N,i}, Q_{N,j}).$$

Given that the density function \(f_{W_i}\) is uniformly bounded (Assumption 5.2 (b)), there exists a
constant $A_2$ such that $|f_{W_i^c}| < A_2$. Then, because

$$Q_{N,i} = \frac{1}{N} \left\{ K(W_i^c, w^c)1[W_i^d = w^d] - E \left[ K(W_i^c, w^c)1[W_i^d = w^d] \right] \right\},$$

we have

$$Var[Q_{N,i}] = E[Q_{N,i}^2] \leq \frac{1}{N^2} E \left[ K^2(W_i^c, w^c) \right] = \frac{1}{(N\eta Q)^2} \int \prod_{q=1}^Q \kappa^2 \left( \frac{w_q^c - w_q^c}{h} \right) f_{W_i^c}(w_q^c)dw_q^c$$

$$\leq \frac{A_2}{N^2 h Q} \prod_{q=1}^Q \kappa^2(v_q)dv_q = \frac{A_3}{N^2 h Q}, \quad (B.52)$$

with $A_3 = 2K_2^Q$ and $A_3 < \infty$ due that $\int \kappa^2(v)dv = K_2 < \infty$. Recall that $r_i = |\Delta(i, N)|$. By the Cauchy-Schwarz inequality and (B.52)

$$\sum_{k=1}^{qN/2} \sum_{i,j \in S_{2k-1}} Cov(Q_{N,i}, Q_{N,j}) \leq \sum_{k=1}^{qN/2} \sum_{i,j \in S_{2k-1}} |Cov(Q_{N,i}, Q_{N,j})| \leq \sum_{k=1}^{qN/2} \sum_{i,j \in S_{2k-1}} Var[Q_{N,i}]$$

$$\leq \frac{A_3}{2N^2 h Q} \sum_{k=1}^{qN/2} |S_{2k-1}|(|S_{2k-1}| - 1),$$

substituting $\sum_{k=1}^{qN/2} |S_{2k-1}| \leq N$ and $|S_{2k-1}| \leq r_{i2k-1}$ into the above inequality,

$$\sum_{k=1}^{qN/2} \sum_{i,j \in S_{2k-1}} Cov(Q_{N,i}, Q_{N,j}) \leq \frac{A_3}{2N^2 h Q} \left( \sum_{k=1}^{qN/2} r_{i2k-1}^2 + N \right) = \frac{A_4}{N h Q}, \quad (B.53)$$

for some constant $A_4 > 0$, because $\sum_{k=1}^{qN/2} r_{i2k-1}^2 \leq \sum_{k=1}^{qN/2} r_{i_k}^2 \leq \sum_{i=1}^{N} |\Delta(i, N)|^2 = O(N)$ (Assumption 5.2). Given (B.53), it is easy to see that (B.51) becomes to

$$R_{21}(w) \leq 2 \exp \left( -\frac{\lambda N \eta}{4} + \lambda^2 N h Q \frac{A_4}{N h Q} \right) = 2 \exp \left( -\frac{\lambda N \eta}{4} + A_4 \ln(N) \right). \quad (B.54)$$

Let $\eta = 4A_5[\ln(N)/(NhQ)]^{1/2}$ for some constant $A_5 > 0$. Then, we have $\lambda N \eta = A_5 \ln(N)$. We can bound $R_{21}(w)$ uniformly as

$$\sup_{w \in R_N} R_{21}(w) \leq 2 \exp((A_4 - A_5) \ln(N)) = 2N^{-\alpha}, \quad (B.55)$$

and we choose $A_5$ large enough such that $\alpha > 0$ with $\alpha = A_5 - A_4$.

At last, we deal with $R_{22}(w)$. Let $B_{2k-3} \in \sigma\{V_N(1), \ldots, V_N(2k-3)\}$, $B_{2k-1} \in \sigma\{V_N(2k-1)\}$ and

$$\alpha_{2k-1} = \sup_{B_{2k-3}, B'_{2k-1}} \left| Pr(B_{2k-3}, B'_{2k-1}) - Pr(B_{2k-3}) Pr(B'_{2k-1}) \right|. \quad \text{(5.2)}$$
Making use of (B.46), we can obtain that the reminder term

\[ R_{22}(w) = Pr \left( \sum_{k=1}^{qN/2} |V_N(2k-1) - V_N^*(2k-1)| > \eta/4 \right) \]

\[ \leq \sum_{k=1}^{qN/2} Pr \left( |V_N(2k-1) - V_N^*(2k-1)| > \frac{\eta}{2q_N} \right) \]

\[ \leq 18 \sum_{k=1}^{qN/2} \left( \frac{2qN\|V_N(2k-1)\|_\infty}{\eta} \right)^{1/2} \alpha_{2k-1}. \]

(B.56)

Furthermore, applying (B.48) and \( \eta = 4A_5[\ln(N)/(Nh^Q)]^{1/2} \) to the above inequality,

\[ R_{22}(w) \leq 18 \sum_{k=1}^{qN/2} \left( \frac{2qNA_1 r_{2k-1}}{\eta Nh^Q} \right)^{1/2} \alpha_{2k-1} \leq A_6 \left( \frac{qN\bar{r}_N}{[\ln(N)Nh^Q]^{1/2}} \right)^{1/2} \sum_{k=1}^{qN/2} \alpha_{2k-1} \]

\[ \leq A_6 \left( \frac{N}{\ln(N)} \right)^{1/2} \sum_{k=1}^{qN/2} \alpha_{2k-1}. \]

(B.57)

uniformly in \( w \) for some constant \( A_6 > 0 \), where the last line is due to \( \bar{r}_N = O([Nh^Q/\ln(N)]^{1/2}) \) and \( q_N \leq N \). Now, substitute (B.55) and (B.57) into (B.47),

\[ \sup_{w \in \Omega_W} Pr (|W_N'(w)| > \eta/2) \leq 2N^{-\alpha} + A_6 \left( \frac{N}{\ln(N)} \right)^{1/2} \sum_{k=1}^{qN/2} \alpha_{2k-1} \]

which, together with (B.45), further implies that

\[ Pr(R_2 > \eta) \leq 4L_N N^{-\alpha} + 2A_6 L_N \left( \frac{N}{\ln(N)} \right)^{1/2} \sum_{k=1}^{qN/2} \alpha_{2k-1}. \]

(B.58)

Let \( l_N = [\ln(N)h^{(Q+2)}/N]^{1/2} = \eta h^{Q+1} \rightarrow 0 \), then \( L_N = 1/l_N^Q = 1/[\eta h^{(Q+1)}] \rightarrow \infty \) as \( N \rightarrow \infty \). By properly choosing \( \alpha \), we can obtain the result that \( L_N N^{-\alpha} \) is summable, i.e. \( \sum_{N=1}^\infty L_N N^{-\alpha} < \infty \).

In addition, by Assumption 5.3, we know that \( L_N \left( \frac{N}{\ln(N)} \right)^{1/2} \sum_{k=1}^{qN/2} \alpha_{2k-1} \) is also summable. It then follows from the Borel-Cantelli lemma that

\[ R_2 = O(\eta) = O \left( \frac{[\ln(N)]^{1/2}}{Nh^Q} \right) \text{ almost surely.} \]

(B.59)

Together with (B.41) and (B.44), we arrive the conclusion that

\[ \sup_{w \in \Omega_W} \left| \hat{f}_{W_1}(w) - f_{W_1}(w) \right| = O_p \left( [\ln(N)/(Nh^Q)]^{1/2} + h^2 \right). \]

(B.60)
Proof of Corollary 5.3. We prove the desired result in two steps. Step 1 aims at the uniform convergence of \( \hat{f}_{|\mathcal{N}_i||Z_i|,\mathcal{N}_i^*}} \). Step 2 fulfills the proof by establishing the uniform convergence of \( \hat{f}_{S^*|\mathcal{N}_i^*||D_i,S_i,Z_i|,\mathcal{N}_i}} \).

Step 1. From (B.22) we know that \( F_{|\mathcal{N}_i||Y_i|,|Z_i||Z_i^*}} = F_{|\mathcal{N}_i||Z_i||Z_i^*}} \times T_{|Y_i||Z_i||Z_i^*}} \times F_{|\mathcal{N}_i||Z_i||Z_i^*}} \). Denote \( B(\gamma^0) := F_{|\mathcal{N}_i||Y_i|,|Z_i||Z_i^*}} \times F_{|\mathcal{N}_i||Z_i||Z_i^*}} \), and let \( \lambda(\gamma^0) \) and \( \psi(\gamma^0) \) represent the eigenvalues and eigenvectors of \( B(\gamma^0) \). Then, we have \( B(\gamma^0) - \lambda(\gamma^0) I_{K_T} \psi(\gamma^0) = 0 \).

Furthermore, recall that \( T_{|Y_i||Z_i||Z_i^*}} \) is a diagonal matrix with all entries on its diagonal strictly positive. It then yields from the eigendecomposition that for the eigenvalue \( \lambda(\gamma^0) = f_{Y_i|Z_i,\mathcal{N}_i^*}=n^*(y) \), its eigenvector is \( \psi(\gamma^0) = [f_{|\mathcal{N}_i||Z_i,\mathcal{N}_i^*}=n^*(n_1), ..., f_{|\mathcal{N}_i||Z_i,\mathcal{N}_i^*}=n^*(K_N)]^T \). Andrew et al. (1993) shows the existence of a neighborhood of \( \gamma^0 \) in the parameter space, denoted by \( \mathcal{M}_0 \), such that for any \( \gamma \in \mathcal{M}_0 \), there exist an eigenvalue function \( \lambda(\gamma) \) and an eigenvector function \( \psi(\gamma) \) that are both analytic functions of \( \gamma \). Given the uniform convergence of \( \hat{\gamma}_N \) to \( \gamma^0 \) proved in Theorem 5.2, we only need to consider the convergence of \( \psi(\gamma) \) over a small neighborhood of \( \gamma^0 \) such that \( \|\gamma - \gamma^0\|_\infty \leq \eta \) with \( \eta = o(1) \). The rest of the proof is exactly the same with the proof of Lemma 3 in Hu (2008), therefore ignored here due to space limitation. Let \( \hat{Y}_N := \psi(\hat{\gamma}_N) \) and \( \psi_0 := \psi(\gamma^0) \), then we can show the uniform convergence

\[
\sup_{\|\gamma_N - \gamma^0\|_\infty \leq \eta} \|\hat{Y}_N - \psi_0\|_\infty = O_p \left(\|\gamma_N - \gamma^0\|_\infty\right),
\]

\[
\sup_{\|\gamma_N - \gamma^0\|_\infty \leq \eta} \left\|\frac{\partial \psi(\gamma^0)}{\partial \gamma'} (\gamma_N - \gamma^0)\right\|_\infty = O_p \left(\|\gamma_N - \gamma^0\|_\infty^2\right).
\]

Step 2. Again, because of the uniform convergence of \( \hat{\gamma}_N \), in Step 2 we consider only a small neighborhood of \( \gamma^0 \). Denote \( \varphi = (\varphi_1, ..., \varphi_6)^T \) where each of its elements represents one probability distribution on the right hand side of the equation (3):

\[
\varphi_1 = f_{S_i|I_i^*,Z_i|,\mathcal{N}_i^*}|Z_i|, \quad \varphi_2 = f_{S_i^*|Z_i,\mathcal{N}_i^*}|Z_i|, \quad \varphi_3 = f_{|\mathcal{N}_i||Z_i,\mathcal{N}_i^*}|Z_i|, \quad \varphi_4 = f_{|\mathcal{N}_i^*||Z_i|},
\]

\[
\varphi_5 = f_{S_i|Z_i,\mathcal{N}_i|}, \quad \varphi_6 = f_{|\mathcal{N}_i||Z_i|}.
\]

where we actually have that \( \varphi_3 = \psi \). Given Proposition 4.2, \( \phi = \phi(\varphi) = \varphi_1 \varphi_2 \varphi_3 \varphi_4 / (\varphi_5 \varphi_6) \), which is a twice continuously differentiable function of \( \varphi \) by Assumption 5.2. Beside, its estimator is constructed by \( \hat{\varphi}_N = \phi(\hat{\varphi}_N) \) with true value \( \varphi^0 = \phi(\varphi^0) \). Let the true value of \( \varphi \) be \( \varphi^0 = \varphi(\gamma^0, \psi^0) = (\varphi_1^0, ..., \varphi_6^0)^T \) and let its plug-in estimator be \( \hat{\varphi}_N = \varphi(\hat{\gamma}_N, \hat{\psi}_N) = (\hat{\varphi}_1, ..., \hat{\varphi}_6, N)^T \). Then,

\[
\frac{d\phi(\varphi)}{d\varphi'} = \begin{pmatrix}
\varphi_2 \varphi_3 \varphi_4 & \varphi_1 \varphi_3 \varphi_4 & \varphi_1 \varphi_2 \varphi_4 & \varphi_1 \varphi_2 \varphi_3 & -\varphi_1 \varphi_2 \varphi_3 \varphi_4 & -\varphi_1 \varphi_2 \varphi_3 \varphi_4 \\
\varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 \\
\varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 \\
\varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 \\
\varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 \\
\varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 & \varphi_5 \varphi_6 \\
\end{pmatrix}.
\]

(B.61)

Recall that there exists a \( \epsilon > 0 \), such that \( \varphi_6^0 \) are uniformly bounded from below by \( \epsilon \) based on the condition stated in Corollary 5.3. We also know that \( \varphi_5^0 = C^*_n f_0^*(1) f_D(0)^{(n-s)} \) \( > \epsilon \) uniformly over \( \Omega_W \) for some constant \( \epsilon' \). Moreover, since \( \varphi_1 \) to \( \varphi_4 \) are all conditional probabilities of discrete random variables, their true values \( \varphi_1^0 \) to \( \varphi_4^0 \) all lie in \([0,1]\). When we consider a uniform \( o(1) \) neighborhood of \( \gamma^0 \), by the uniform convergence of \( \hat{\psi}_N \) in Step 1, we know that for large enough sample size, \( \varphi_{1,N} \) to \( \varphi_{4,N} \) are also uniformly bounded from above and \( \varphi_{5,N} \) and \( \varphi_{6,N} \) are uniformly bounded from below. Therefore, any intermediate value \( \hat{\varphi} \) between \( \varphi^0 \) and \( \varphi_N \) is uniformly
bounded. Thus, for the derivative in (B.61) evaluated at \( \tilde{\varphi} \), there exists some constant \( C > 0 \) such that \( \| d\phi(\tilde{\varphi})/d\varphi' \| \leq C \) uniformly over \( \Omega_W \). By the mean value theorem, we then have that

\[
\sup_{\| \tilde{\gamma}_N - \gamma^0 \|_\infty < \eta} \| \hat{\phi}_N - \phi^0 \|_\infty = \sup_{\| \tilde{\gamma}_N - \gamma^0 \|_\infty < \eta} \| \phi(\tilde{\gamma}_N) - \phi(\gamma^0) \|_\infty \\
\leq \sup_{\| \tilde{\gamma}_N - \gamma^0 \|_\infty < \eta} \left\| \frac{d\phi(\tilde{\gamma})}{d\varphi'} \right\| \| \tilde{\gamma}_N - \gamma^0 \|_\infty \\
\leq C \sup_{\| \tilde{\gamma}_N - \gamma^0 \|_\infty < \eta} \| \hat{\phi}_N - \phi^0 \|_\infty, \tag{B.62}
\]

where \( \tilde{\varphi} \) is an intermediate vector between \( \varphi^0 \) and \( \hat{\phi}_N \). Besides, because \( \hat{\phi}_N = \varphi(\hat{\gamma}_N, \hat{\psi}_N) \) and \( \varphi^0 = \varphi(\gamma^0, \psi^0) \), together with the fact that \( \varphi(\gamma, \psi) \) is continuously differentiable in \( (\gamma, \psi) \) with uniformly bounded first order derivative, we get that (B.62) can be further bounded by

\[
\sup_{\| \tilde{\gamma}_N - \gamma^0 \|_\infty < \eta} \| \hat{\phi}_N - \phi^0 \|_\infty \leq C' \left( \sup_{\| \tilde{\gamma}_N - \gamma^0 \|_\infty < \eta} \| \hat{\psi}_N - \psi \|_\infty + \| \tilde{\gamma}_N - \gamma^0 \|_\infty \right) = O_p \left( \| \tilde{\gamma}_N - \gamma^0 \|_\infty \right),
\]

for some constant \( C' > 0 \), and the last line is from in Step 1. Furthermore, recall that \( \phi = \phi(\psi) \), where \( \psi = \psi(\gamma, \varphi) \) and \( \varphi = \varphi(\gamma) \). Thus, \( \phi \) can be regarded as a function of \( \gamma \) only. Applying similar arguments, we can also obtain that

\[
\sup_{\omega \in \Omega_W} \left\| \hat{\phi}_N - \phi^0 - \frac{\partial \phi}{\partial \gamma}(\hat{\gamma}_N - \gamma^0) \right\|_\infty = O_p \left( \| \hat{\gamma}_N - \gamma^0 \|_\infty \right). \tag{B.63}
\]

**Proof of Theorem 5.4.** Now, from \( m(x; \theta, \phi) = \sum_{j=1}^{K_T} m^*(x_j; \theta) f_{T_i|x_i}(t_j) \) with \( x_j = (d, s_j, z, n_j) \) and \( t_j = (s_j, n_j) \), we can get

\[
\mathcal{L}_N(\theta, \hat{\phi}_N) - \mathcal{L}_N(\theta, \phi^0) \\
= \frac{1}{N} \sum_{i=1}^{N} \tau_i \left\{ \left[ Y_i - m \left( X_i; \theta, \hat{\phi}_N \right) \right]^2 - \left[ Y_i - m \left( X_i; \theta, \phi^0 \right) \right]^2 \right\} \\
= \frac{1}{N} \sum_{i=1}^{N} \tau_i \left[ m \left( X_i; \theta, \hat{\phi}_N \right) - m \left( X_i; \theta, \phi^0 \right) \right]^2 \\
- \frac{2}{N} \sum_{i=1}^{N} \tau_i \left[ Y_i - m \left( X_i; \theta, \phi^0 \right) \right] \left[ m \left( X_i; \theta, \hat{\phi}_N \right) - m \left( X_i; \theta, \phi^0 \right) \right] \\
= \frac{1}{N} \sum_{i=1}^{N} \tau_i \left\{ \sum_{j=1}^{K_T} m^*(x_{i,j}; \theta) \left[ \hat{f}_{T_i|x_i}(t_j) - f_{T_i|x_i}(t_j) \right] \right\}^2 \\
- \frac{2}{N} \sum_{i=1}^{N} \tau_i \left[ Y_i - m \left( X_i; \theta, \phi^0 \right) \right] \left\{ \sum_{j=1}^{K_T} m^*(x_{i,j}; \theta) \left[ \hat{f}_{T_i|x_i}(t_j) - f_{T_i|x_i}(t_j) \right] \right\}, \tag{B.64}
\]

where \( x_{i,j} = (D_i, s_j, Z_i, n_j) \). Because of the uniform convergence of \( \hat{\gamma}_N \), we only need to focus on
a small neighborhood of $\gamma^0$. Due to the boundedness of $\tau(x)$ and the Cauchy-Schwarz inequality,

$$\left| \mathcal{L}_N(\theta, \hat{\phi}_N) - \mathcal{L}_N(\theta, \phi^0) \right| \leq \frac{C}{N} \sum_{i=1}^{N} \sum_{j=1}^{K_T} m^*(x_{ij}; \theta)^2 \sum_{j=1}^{K_T} \left[ \hat{f}_{T^*_i|X_i}(t_j) - f_{T^*_i|X_i}(t_j) \right]^2$$

$$+ \frac{2}{N} \sum_{i=1}^{N} \sum_{j=1}^{K_T} \tau_i \left| Y_i - m(X_i; \theta, \phi^0) \right| \left| m^*(x_{ij}; \theta) \right| \left[ \hat{f}_{T^*_i|X_i}(t_j) - f_{T^*_i|X_i}(t_j) \right]$$

$$\leq C \left( \sup_{\|\hat{\gamma}_N - \gamma^0\|_\infty \leq \eta} \left\| \hat{\phi}_N - \phi^0 \right\|_\infty \right)^2 \frac{1}{N} \sum_{j=1}^{K_T} \sum_{i=1}^{N} m^*(x_{ij}; \theta)^2$$

$$+ 2 \sum_{\|\hat{\gamma}_N - \gamma^0\|_\infty \leq \eta} \left\| \hat{\phi}_N - \phi^0 \right\|_\infty \left[ \frac{1}{N} \sum_{j=1}^{K_T} \sum_{i=1}^{N} \tau_i \left| Y_i - m(X_i; \theta, \phi^0) \right|^2 \right]^{1/2}$$

$$\left[ \frac{1}{N} \sum_{i=1}^{N} \left| m^*(x_{ij}; \theta) \right|^2 \right]^{1/2} \right). \quad (B.65)$$

Because $(D_i, Z_i)$ is i.i.d., then $x_{ij} = (D_i, s_j, Z_i, n_j)$ is also i.i.d. for any given $j = 1, ..., K_T$. Then, by Assumption 5.4 and the uniform convergence of i.i.d. samples (Lemma 2.4 of Newey and MacFadden (1994))

$$\sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} m^*(x_{ij}; \theta)^2 \leq \sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} m^*(x_{ij}; \theta)^2 - E \left[ m^*(x_{ij}; \theta)^2 \right] + \sup_{\theta \in \Theta} E \left[ m^*(x_{ij}; \theta)^2 \right]$$

$$= O_p(1), \quad (B.66)$$

because $\sup_{\theta \in \Theta} E \left[ m^*(x_{ij}; \theta)^2 \right] \leq E[h_1(x_{ij})] < \infty$ by Assumption 5.4. Similarly, the uniform convergence of data with dependency neighborhood structure in Lemma E.2 leads to

$$\sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \tau_i \left| Y_i - m(X_i; \theta, \phi^0) \right|^2 = O_p(1), \quad (B.67)$$

because of Assumption 5.1 and Assumption 5.4 (v). Hence, we can conclude that

$$\sup_{\theta \in \Theta} \left| \mathcal{L}_N(\theta, \hat{\phi}_N) - \mathcal{L}_N(\theta, \phi^0) \right| = O_p \left( \sup_{\|\hat{\gamma}_N - \gamma^0\|_\infty \leq \eta} \left\| \hat{\phi}_N - \phi^0 \right\|_\infty \right) = O_p \left( \|\hat{\gamma}_N - \gamma^0\|_\infty \right). \quad (B.68)$$

Next, we show the uniform convergence of $\mathcal{L}_N(\theta, \phi^0)$ to $\mathcal{L}(\theta, \phi^0)$ by verifying the uniform law of large number for dependent data as in Lemma E.2. Firstly, condition (i), (ii), (iii) and (iv)-(c) of Lemma E.2 are trivially sanctified by Assumption 5.4 (i), (iii) and (v). Secondly, (iv) (a) of Lemma E.2 holds because of Assumption 5.1. In addition, we have that $1/N \sum_{i=1}^{N} |\Delta(i, N)| \leq 1/N \sum_{i=1}^{N} |\Delta(i, N)|^2 = O(1)$ as in Assumption 5.2. Hence, we have verified that all required conditions of Lemma E.2 are satisfied, implying

$$\sup_{\theta \in \Theta} \left| \mathcal{L}_N(\theta, \phi^0) - \mathcal{L}(\theta, \phi^0) \right| = \sup_{\theta \in \Theta} \left| \frac{1}{N} \sum_{i=1}^{N} \tau_i \left[ Y_i - m(X_i; \theta, \phi^0) \right]^2 - E \left[ \tau_i \left[ Y_i - m(X_i; \theta, \phi^0) \right]^2 \right] \right|$$
Then, making use of (B.68), (B.69) and Theorem 5.2, we can bound
\[
\sup_{\theta \in \Theta} \left| \mathcal{L}(\theta, \phi^0) - \mathcal{L}_N(\theta, \hat{\phi}_N) \right| = \sup_{\theta \in \Theta} \left| \mathcal{L}(\theta, \phi^0) - \mathcal{L}_N(\theta, \phi^0) + \mathcal{L}_N(\theta, \phi^0) - \mathcal{L}_N(\theta, \hat{\phi}_N) \right|
\leq \sup_{\theta \in \Theta} \left| \mathcal{L}(\theta, \phi^0) - \mathcal{L}_N(\theta, \phi^0) \right| + \sup_{\theta \in \Theta} \left| \mathcal{L}_N(\theta, \phi^0) - \mathcal{L}_N(\theta, \hat{\phi}_N) \right|
= \sup_{\theta \in \Theta} \left| \mathcal{L}(\theta, \phi^0) - \mathcal{L}_N(\theta, \phi^0) \right| + O_p \left( \|\tilde{\gamma}_N - \gamma^0\|_\infty \right)
= o_p(1).
\]

As assumed in Assumption 5.4, \(\theta^0\) uniquely minimizes the objective function \(\mathcal{L}(\theta, \phi^0)\) over \(\Theta\). Then, for any \(\delta > 0\), there exists a \(\epsilon > 0\) such that \(\|\hat{\theta}_N - \theta^0\| > \delta\) implies \(\mathcal{L}(\hat{\theta}_N, \phi^0) - \mathcal{L}(\theta^0, \phi^0) > \epsilon\). Thus, by the definition of \(\hat{\theta}_N\),
\[
Pr \left( \|\hat{\theta}_N - \theta^0\| > \delta \right) \leq Pr \left( \mathcal{L}(\hat{\theta}_N, \phi^0) - \mathcal{L}(\theta^0, \phi^0) > \epsilon \right)
\leq Pr \left( \mathcal{L}(\hat{\theta}_N, \phi^0) - \mathcal{L}(\hat{\theta}_N, \hat{\phi}_N) + \mathcal{L}(\hat{\theta}_N, \hat{\phi}_N) - \mathcal{L}(\theta^0, \phi^0) > \epsilon \right)
\leq Pr \left( \mathcal{L}(\hat{\theta}_N, \phi^0) - \mathcal{L}(\hat{\theta}_N, \hat{\phi}_N) + \mathcal{L}(\theta^0, \hat{\phi}_N) - \mathcal{L}(\theta^0, \phi^0) > \epsilon \right)
\leq Pr \left( \sup_{\theta \in \Theta} \left| \mathcal{L}(\theta, \phi^0) - \mathcal{L}_N(\theta, \hat{\phi}_N) \right| > \epsilon \right)
\to 0,
\]
where the last line is due to (B.70). It then follows from (B.71) that \(\|\hat{\theta}_N - \theta^0\| = o_p(1)\).

**Proof of Lemma 5.5.** (a) Based on Theorems 5.2 and 5.3, we know that \(\hat{\phi}_N = \phi(\tilde{\gamma}_N)\) and \(\tilde{\gamma}_N \overset{p}{\to} \gamma^0\). Hence, in what follows, we can establish the consistency of \(\frac{1}{N} \sum_{i=1}^N \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta'}\) in a small neighborhood of \(\gamma^0\). For a small constant \(\eta > 0\), by triangular inequality,
\[
\sup_{\|\hat{\gamma}_N - \gamma^0\|_\infty < \eta} \left| \frac{1}{N} \sum_{i=1}^N \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta'} - E \left[ \frac{\partial g(W_i; \theta^0, \phi^0)}{\partial \theta'} \right] \right|
\leq \sup_{\|\hat{\gamma}_N - \gamma^0\|_\infty < \eta} \left| \frac{1}{N} \sum_{i=1}^N \frac{\partial g(W_i; \tilde{\theta}_N, \hat{\phi}_N)}{\partial \theta'} - \frac{1}{N} \sum_{i=1}^N \frac{\partial g(W_i; \hat{\theta}_N, \phi^0)}{\partial \theta'} \right|
+ \frac{1}{N} \sum_{i=1}^N \frac{\partial g(W_i; \tilde{\theta}_N, \phi^0)}{\partial \theta'} - \frac{1}{N} \sum_{i=1}^N \frac{\partial g(W_i; \theta^0, \phi^0)}{\partial \theta'}
+ \frac{1}{N} \sum_{i=1}^N \frac{\partial g(W_i; \theta^0, \phi^0)}{\partial \theta'} - E \left[ \frac{\partial g(W_i; \theta^0, \phi^0)}{\partial \theta'} \right]
:= \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_3.
\]

Given (B.72), it suffices to show that \(\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3\) are all \(o_p(1)\). In what follows, we divide the rest of the proof into three steps.
Step 1. First, consider $H_1$. By definition of $g(W_i; \theta, \phi)$, we have

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta'} - \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \phi^0)}{\partial \theta'}
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \tau_i \left\{ \left[ Y_i - m(X_i; \hat{\theta}_N, \hat{\phi}_N) \right] \frac{d^2 m(X_i; \hat{\theta}_N, \hat{\phi}_N)}{d\theta d\theta'} - \left[ Y_i - m(X_i; \hat{\theta}_N, \phi^0) \right] \frac{d^2 m(X_i; \hat{\theta}_N, \phi^0)}{d\theta d\theta'} \right\}
\]

\[
- \frac{1}{N} \sum_{i=1}^{N} \tau_i \left\{ \frac{d m(X_i; \hat{\theta}_N, \hat{\phi}_N)}{d \theta} \frac{d m(X_i; \hat{\theta}_N, \phi^0)}{d \theta'} + \frac{d m(X_i; \hat{\theta}_N, \phi^0)}{d \theta} \frac{d m(X_i; \hat{\theta}_N, \phi^0)}{d \theta'} \right\}.
\]  \hspace{1cm} (B.73)

Making use of the identity $\hat{a} \hat{b} = (\hat{a} - a)b + a(\hat{b} - b) + (\hat{a} - a)(\hat{b} - b)$ and applying it to both terms on the right hand side of (B.73) give us

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta'} - \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \phi^0)}{\partial \theta'}
\]

\[
= - \frac{1}{N} \sum_{i=1}^{N} \tau_i \left[ m(X_i; \hat{\theta}_N, \hat{\phi}_N) - m(X_i; \hat{\theta}_N, \phi^0) \right] \frac{d^2 m(X_i; \hat{\theta}_N, \phi^0)}{d\theta d\theta'}
\]

\[
+ \frac{1}{N} \sum_{i=1}^{N} \tau_i \left[ Y_i - m(X_i; \hat{\theta}_N, \phi^0) \right] \left[ \frac{d^2 m(X_i; \hat{\theta}_N, \phi^0)}{d\theta d\theta'} - \frac{d^2 m(X_i; \hat{\theta}_N, \phi^0)}{d\theta d\theta'} \right]
\]

\[
- \frac{1}{N} \sum_{i=1}^{N} \tau_i \left[ m(X_i; \hat{\theta}_N, \hat{\phi}_N) - m(X_i; \hat{\theta}_N, \phi^0) \right] \left[ \frac{d^2 m(X_i; \hat{\theta}_N, \hat{\phi}_N)}{d\theta d\theta'} - \frac{d^2 m(X_i; \hat{\theta}_N, \phi^0)}{d\theta d\theta'} \right]
\]

\[
- \frac{1}{N} \sum_{i=1}^{N} \tau_i \left[ \frac{d m(X_i; \hat{\theta}_N, \hat{\phi}_N)}{d \theta} \frac{d m(X_i; \hat{\theta}_N, \phi^0)}{d \theta'} + \frac{d m(X_i; \hat{\theta}_N, \phi^0)}{d \theta} \frac{d m(X_i; \hat{\theta}_N, \phi^0)}{d \theta'} \right]
\]

\[
- \frac{1}{N} \sum_{i=1}^{N} \tau_i \left[ \frac{d m(X_i; \hat{\theta}_N, \hat{\phi}_N)}{d \theta} - \frac{d m(X_i; \hat{\theta}_N, \phi^0)}{d \theta} \right] \left[ \frac{d m(X_i; \hat{\theta}_N, \phi^0)}{d \theta'} - \frac{d m(X_i; \hat{\theta}_N, \phi^0)}{d \theta'} \right].
\]  \hspace{1cm} (B.74)

Recall that $m(X_i; \theta, \phi) = \sum_{j=1}^{K_t} m^*(x_{i,j}; \theta) f_{T_t^*|X_i}(t_j)$ and $x_{i,j} = (D_i, s_j, Z_i, n_j)$. We can further rewrite (B.74) as

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta'} - \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \phi^0)}{\partial \theta'}
\]

\[
= - \frac{1}{N} \sum_{i=1}^{N} \tau_i \left\{ \sum_{j=1}^{K_t} m^*(x_{i,j}; \hat{\theta}_N) \left[ \hat{f}_{T_t^*|X_i}(t_j) - f_{T_t^*|X_i}(t_j) \right] \right\} \sum_{j=1}^{K_t} \frac{d^2 m^*(x_{i,j}; \hat{\theta}_N)}{d\theta d\theta'} f_{T_t^*|X_i}(t_j)
\]

\[
+ \frac{1}{N} \sum_{i=1}^{N} \tau_i \left[ Y_i - m(X_i; \hat{\theta}_N, \phi^0) \right] \left\{ \sum_{j=1}^{K_t} \frac{d^2 m^*(x_{i,j}; \hat{\theta}_N)}{d\theta d\theta'} \left[ \hat{f}_{T_t^*|X_i}(t_j) - f_{T_t^*|X_i}(t_j) \right] \right\}.
\]
\[-\frac{1}{N} \sum_{i=1}^{N} \tau_i \left\{ \sum_{j=1}^{K_T} m^*(x_{i,j}; \tilde{\theta}_N) \left[ \hat{f}_{T_i|X_i}^* - f_{T_i|X_i}^* \right] \right\} \times \left\{ \sum_{j=1}^{K_T} \frac{d^2 m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta d\theta'} \left[ \hat{f}_{T_i|X_i}^* - f_{T_i|X_i}^* \right] \right\} \]
\[-\frac{1}{N} \sum_{i=1}^{N} \tau_i \left\{ \sum_{j=1}^{K_T} \frac{\partial m^*(x_{i,j}; \tilde{\theta}_N)}{\partial \theta} \left[ \hat{f}_{T_i|X_i}^* - f_{T_i|X_i}^* \right] \right\} \sum_{j=1}^{K_T} \frac{d m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta} \hat{f}_{T_i|X_i}^* \]
\[-\frac{1}{N} \sum_{i=1}^{N} \tau_i \left\{ \sum_{j=1}^{K_T} \frac{d m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta} \hat{f}_{T_i|X_i}^* \right\} \sum_{j=1}^{K_T} \frac{\partial m^*(x_{i,j}; \tilde{\theta}_N)}{\partial \theta'} \left[ \hat{f}_{T_i|X_i}^* - f_{T_i|X_i}^* \right] \]
\[-\frac{1}{N} \sum_{i=1}^{N} \tau_i \left\{ \sum_{j=1}^{K_T} \frac{\partial m^*(x_{i,j}; \tilde{\theta}_N)}{\partial \theta'} \left[ \hat{f}_{T_i|X_i}^* - f_{T_i|X_i}^* \right] \right\} \times \left\{ \sum_{j=1}^{K_T} \frac{\partial m^*(x_{i,j}; \tilde{\theta}_N)}{\partial \theta'} \left[ \hat{f}_{T_i|X_i}^* - f_{T_i|X_i}^* \right] \right\} \right\} . \quad (B.75)\]

Because that for a $k \times k$ matrix $A = ab'$ where $a, b \in \mathbb{R}^k$, then $\|A\| = \|a\|\|b\|$, the boundedness of $f_{T_i|X_i}$ and (B.75),

\[
\mathcal{H}_1 \leq C \sup_{\|\tilde{\gamma}_N - \gamma_0\|_\infty < \eta} \left\| \hat{\gamma}_N - \gamma_0 \right\|_\infty \frac{1}{N} \sum_{j,l=1}^{K_T} \sum_{i=1}^{N} \tau_i \left| m^*(x_{i,j}; \tilde{\theta}_N) \right| \left| \frac{d^2 m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta d\theta'} \right| 
\]
\[+ C \sup_{\|\tilde{\gamma}_N - \gamma_0\|_\infty < \eta} \left\| \hat{\gamma}_N - \gamma_0 \right\|_\infty \frac{1}{N} \sum_{j,l=1}^{K_T} \sum_{i=1}^{N} \tau_i \left| Y_i - m(X_i; \tilde{\theta}_N, \phi^0) \right| \left| \frac{d^2 m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta d\theta'} \right| 
\]
\[+ C \left( \sup_{\|\tilde{\gamma}_N - \gamma_0\|_\infty < \eta} \left\| \hat{\gamma}_N - \gamma_0 \right\|_\infty \right)^2 \frac{1}{N} \sum_{j,l=1}^{K_T} \sum_{i=1}^{N} \left| m^*(x_{i,j}; \tilde{\theta}_N) \right| \left| \frac{d^2 m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta d\theta'} \right| 
\]
\[+ 2C \sup_{\|\tilde{\gamma}_N - \gamma_0\|_\infty < \eta} \left\| \hat{\gamma}_N - \gamma_0 \right\|_\infty \frac{1}{N} \sum_{j,l=1}^{K_T} \sum_{i=1}^{N} \left| \frac{\partial m^*(x_{i,j}; \tilde{\theta}_N)}{\partial \theta} \right| \left| \frac{d m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta} \right| 
\]
\[C \left( \sup_{\|\tilde{\gamma}_N - \gamma_0\|_\infty < \eta} \left\| \hat{\gamma}_N - \gamma_0 \right\|_\infty \right)^2 \frac{1}{N} \sum_{j,l=1}^{K_T} \sum_{i=1}^{N} \left| \frac{\partial m^*(x_{i,j}; \tilde{\theta}_N)}{\partial \theta} \right| \left| \frac{d m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta} \right| 
\]
\[:= \mathcal{H}_{11} + \mathcal{H}_{12} + \mathcal{H}_{13} + \mathcal{H}_{14} + \mathcal{H}_{15}. \quad (B.76)\]

By the Cauchy-Schwarz inequality, we can further bound $\mathcal{H}_{11}$ as

\[
\mathcal{H}_{11} \leq C \sup_{\|\tilde{\gamma}_N - \gamma_0\|_\infty < \eta} \left\| \hat{\gamma}_N - \gamma_0 \right\|_\infty \sum_{j,l=1}^{K_T} \left[ \frac{1}{N} \sum_{i=1}^{N} \left| m^*(x_{i,j}; \tilde{\theta}_N) \right|^2 \right]^{1/2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left| \frac{d^2 m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta d\theta'} \right|^2 \right]^{1/2} 
\]
\[\leq O_p \left( \sup_{\|\tilde{\gamma}_N - \gamma_0\|_\infty < \eta} \left\| \hat{\gamma}_N - \gamma_0 \right\|_\infty \right). \]
where the second line is due to (B.66) and Lemma E.7, and the last line is because of Corollary 5.3. For $H_{12}$, it follows again from the Cauchy-Schwarz inequality and Corollary 5.3 that

$$H_{12} \leq o_p(1) \sum_{j=1}^{K_T} \left[ \frac{1}{N} \sum_{i=1}^{N} \tau_{ij} \left| Y_i - m(X_i; \hat{\theta}_N) \right|^2 \right]^{1/2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{d^2 m^*(x_{i,j}; \hat{\theta}_N)}{d\theta d\theta'} \right) \right]^{1/2} = o_p(1), \quad \text{(B.77)}$$

where the last line is due to the uniform convergence in (B.69) and that proved in Lemma E.7.

Given $H_{11} = o_p(1)$, it is apparent that $H_{13}$ is also an $o_p(1)$. Similarly, if we know that $H_{14} = o_p(1)$, then $H_{15} = o_p(1)$. Again, by the Cauchy-Schwarz inequality and Lemma E.7,

$$H_{14} \leq o_p(1) \sum_{j=1}^{K_T} \left[ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{dm^*(x_{i,j}; \hat{\theta}_N)}{d\theta} \right\|^2 \right]^{1/2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{dm^*(x_{i,j}; \hat{\theta}_N)}{d\theta'} \right\|^2 \right]^{1/2} = o_p(1), \quad \text{(B.79)}$$

Thus, based on (B.77), (B.78) and (B.79), we can conclude that $H_1 = o_p(1)$.

**Step 2.** Consider the term inside the absolute value in $H_2$

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \phi^0)}{\partial \theta'} - \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \theta^0, \phi^0)}{\partial \theta'}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \tau_{ij} \left\{ \left[ Y_i - m(X_i; \hat{\theta}_N, \phi^0) \right] \frac{\partial^2 m(X_i; \hat{\theta}_N, \phi^0)}{\partial \theta \partial \theta'} - \left[ Y_i - m(X_i; \theta^0, \phi^0) \right] \frac{\partial^2 m(X_i; \theta^0, \phi^0)}{\partial \theta \partial \theta'} \right\}$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \tau_{ij} \left( \frac{\partial m(X_i; \hat{\theta}_N, \phi^0)}{\partial \theta} \frac{\partial m(X_i; \hat{\theta}_N, \phi^0)}{\partial \theta'} - \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta} \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta'} \right). \quad \text{(B.80)}$$

Applying again the identity $\hat{a} \hat{b} - ab = (\hat{a} - a)b + a(\hat{b} - b) + (\hat{a} - a)(\hat{b} - b)$ to (B.80) and substituting

$$m(X_i; \theta, \phi) = \sum_{j=1}^{K_T} m^*(x_{i,j}; \theta)f_{T^*_i}(x_i(t_j))$$

give us

$$H_2 \leq \frac{C}{N} \sum_{i=1}^{N} \left[ \frac{m^*(x_{i,j}; \hat{\theta}_N) - m^*(x_{i,j}; \theta^0)}{\partial \theta} \left\| \frac{d^2 m^*(x_{i,j}; \theta^0)}{d\theta d\theta'} \right\| \right]$$

$$+ \frac{C}{N} \sum_{i=1}^{N} \sum_{j=1}^{K_T} \tau_{ij} \left| Y_i - m(X_i; \theta^0, \phi^0) \right| \left\| \frac{d^2 m^*(x_{i,j}; \hat{\theta}_N)}{d\theta d\theta'} - \frac{d^2 m^*(x_{i,j}; \theta^0)}{d\theta d\theta'} \right\|$$

$$+ \frac{C}{N} \sum_{i=1}^{N} \sum_{j=1}^{K_T} \left[ m^*(x_{i,j}; \hat{\theta}_N) - m^*(x_{i,j}; \theta^0) \right] \left\| \frac{d^2 m^*(x_{i,j}; \theta^0)}{d\theta d\theta'} - \frac{d^2 m^*(x_{i,j}; \theta^0)}{d\theta d\theta'} \right\|$$

$$+ \frac{2C}{N} \sum_{i=1}^{N} \sum_{j=1}^{K_T} \left\| \frac{dm^*(x_{i,j}; \hat{\theta}_N)}{d\theta} - \frac{dm^*(x_{i,j}; \theta^0)}{d\theta} \right\| \left\| \frac{dm^*(x_{i,j}; \theta^0)}{d\theta'} \right\|.$$
Then, we can rewrite $H$ leading to

Consequently, we know that the stated result holds.

Because the Cauchy-Schwarz inequality and Lemma E.7, it is easy to show $H_{21}$ to $H_{25}$ are all $o_p(1)$. Consequently, we know that $H_2 = o_p(1)$.

**Step 3.** Next, consider $H_3$. Let $g_r(W_i; \theta, \phi)$ be the $r$-th element in the column vector $g(W_i; \theta, \phi)$. Then, we can rewrite $H_3^2$ as

$$
H_3^2 = \left| \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g_r(W_i; \theta^0, \phi^0)}{\partial \theta} - E \left[ \frac{\partial g_r(W_i; \theta^0, \phi^0)}{\partial \theta} \right] \right|^2
$$

Because $E[\partial g_r(W_i; \theta^0, \phi^0)/\partial \theta] < \infty$ as in Assumption 5.5, the variance of $\partial g_r(W_i; \theta^0, \phi^0)/\partial \theta$ exists and is finite for all $r, q = 1, ..., d_\theta$. Then, the Chebyshev’s inequality implies

$$
Pr \left[ \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g_r(W_i; \theta^0, \phi^0)}{\partial \theta} - E \left[ \frac{\partial g_r(W_i; \theta^0, \phi^0)}{\partial \theta} \right] \right] > \epsilon
$$

$$
\leq Var \left[ \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g_r(W_i; \theta^0, \phi^0)}{\partial \theta} \right] / \epsilon^2
$$

$$
= \frac{1}{\epsilon^2 N^2} \sum_{i=1}^{N} \sum_{j \in \Delta(i, N)} Cov \left( \frac{\partial g_r(W_i; \theta^0, \phi^0)}{\partial \theta}, \frac{\partial g_r(W_j; \theta^0, \phi^0)}{\partial \theta} \right) + s.o.
$$

$$
\leq \frac{C}{\epsilon^2 N^2} \sum_{i=1}^{N} |\Delta(i, N)| + s.o.
$$

$$
= O \left( \frac{1}{\epsilon^2 N} \right).
$$

where the second equality comes from Assumption 5.1, and the last line is because that $1/N \sum_{i=1}^{N} |\Delta(i, N)| = O(1)$ (Assumption 5.2), and set $\epsilon$ such that $\epsilon \to 0$ and $\epsilon^2N \to \infty$ as $N \to \infty$. Thus,

$$
\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g_r(X_i; \theta^0, \phi^0)}{\partial \theta} - E \left[ \frac{\partial g_r(X_i; \theta^0, \phi^0)}{\partial \theta} \right] \overset{p}{\to} 0,
$$

for all $r, q = 1, ..., d_\theta,$

leading to $H_3 = o_p(1)$. Based on the results in the above three steps, we can make the conclusion that the stated result holds.

(b) This proof is analogous to the proof of Theorem 8.1 in Newey and MacFadden (1994). All the sufficient conditions are verified in the Lemmas E.8, E.9 and E.10. Recall that $\tilde{F}_W(w) = 1/N \sum_{i=1}^{N} 1[W_i \leq w]$ represents the empirical distribution and $\int \delta(w) d\tilde{F}_W(w) = 1/N \sum_{i=1}^{N} \delta(W_i)$. 

68
By Assumption 5.4 and the construction of $\delta(w)$, we have

\[
\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ g(W_i; \theta^0, \hat{\phi}_N) - g(W_i; \theta^0, \phi^0) - \delta(W_i) \right] \right\|
\leq \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ g(W_i; \theta^0, \hat{\phi}_N) - g(W_i; \theta^0, \phi^0) - G(W_i; \hat{\gamma}_N - \gamma^0) \right] \right\|
+ \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ G(W_i; \hat{\gamma}_N - \gamma^0) - \int G(w; \hat{\gamma}_N - \bar{\gamma}) dF_W(w) \right] \right\|
+ \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \int G(w; \hat{\gamma}_N - \bar{\gamma}) dF_W(w) - \int \delta(w) d\hat{F}_W(w) \right] \right\|
+ \sqrt{N} \left| \int \delta(w) d\hat{F}_W(w) - \int \delta(w) d\bar{F}_W(w) \right| = o_p(1),
\]

where the last line follows from Lemmas E.8, E.9 and E.10. ■

**Proof of Theorem 5.6.** By Assumption 5.4 and the construction of $\delta(w)$, we know that $E[\tilde{g}_i] = 0$. Since the dependency neighborhood $\Delta(i, N)$ is symmetric as in Assumption 5.7, we know that $\Sigma_N^\theta$ is symmetric: because for $\forall r, q = 1, 2, ..., d_{\theta}$, its $(r, q)$-th entry

\[
\sum_{i=1}^{N} \sum_{j \in \Delta(i, N)} E[\tilde{g}_{i,r} \tilde{g}_{j,q}] = \sum_{j=1}^{N} \sum_{i \in \Delta(j, N)} E[\tilde{g}_{i,r} \tilde{g}_{j,q}] = \sum_{i=1}^{N} \sum_{j \in \Delta(i, N)} E[\tilde{g}_{i,q} \tilde{g}_{j,r}],
\]

where the first equality follows from change of index and the second equality is due to the symmetry of $\Delta(i, N)$. Under Assumption 5.7, the sufficient conditions for the CLT under neighborhood dependent data required in Lemma E.6 are satisfied. Thus, we can show that $\left[ \Sigma_N^\theta \right]^{-1/2} S_N^\theta \xrightarrow{d} N(0, I_{d_{\theta}})$. Next, we show the asymptotic normality for $\sqrt{N}(\hat{\theta}_N - \theta^0)$.

From (12) and Lemma 5.5 (b), we have

\[
- \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{g}_i + o_p(1) \right] = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta^r} \sqrt{N}(\hat{\theta}_N - \theta^0).
\]

Since from Lemma 5.5 (a), we have that $\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta^r} \xrightarrow{P} E \left[ \frac{\partial g(W_i; \theta^0, \phi^0)}{\partial \theta^r} \right]$, where by Assumption 5.5 $E \left[ \frac{\partial g(W_i; \theta^0, \phi^0)}{\partial \theta^r} \right]$ is invertible. Thus, $\left[ \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta^r} \right]^{-1}$ exists for large enough $N$.

Moreover, recall that $\Omega_N$ is symmetric and $\Omega_N \xrightarrow{P} \Omega$ with $\Omega$ being positive definite and nonsingular. It indicates that $\Omega_N^{-1/2}$ also exists for large enough $N$. Then, because $\| \Omega_N^{-1/2} \| = O(1)$ and $\Omega_N^{-1/2} = \sqrt{N}[\Sigma_N^\theta]^{-1/2}$, we can obtain

\[
\sqrt{N}(\hat{\theta}_N - \theta^0) = - \left[ \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(W_i; \hat{\theta}_N, \hat{\phi}_N)}{\partial \theta^r} \right]^{-1} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{g}_i + o_p(1) \right].
\]
\begin{align*}
\Delta \Omega_2 \leq & \frac{1}{N} \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} \left( |\hat{g}_i - \tilde{g}_i| |\hat{g}_j' - \tilde{g}_j'| + |\tilde{g}_i| |\hat{g}_j - \tilde{g}_j'| + |\hat{g}_i - \tilde{g}_i| |\hat{g}_j'| \right) \\
\leq & \frac{1}{N} \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} \left( |\hat{g}_i - \tilde{g}_i| |\hat{g}_j' - \tilde{g}_j'| + |\tilde{g}_i| |\hat{g}_j - \tilde{g}_j'| + |\hat{g}_i - \tilde{g}_i| |\hat{g}_j'| \right) 
\end{align*}

Given (B.86), it suffices to show \( \Delta \Omega_1 = o_p(1) \) by verifying that (a) \( \tilde{g}_i \) and \( \hat{g}_i \) are bounded, and (b) \( \frac{1}{N} \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} |\hat{g}_i - \tilde{g}_i| = o_p(1) \).

Firstly, (a) is satisfied if \(|g(w; \theta, \phi) + \delta(w; \theta, \phi)| \) is uniformly bounded over \( \Omega_W \) and \( \Theta \times [0, 1] \). We know that \( m^*(x; \theta) \) is continuous differentiable in \( \theta \) to order three (Assumption 5.5) and \( \Theta \) is compact, implying for all \( x \in \Omega_X \)

\begin{align*}
|m^*(x; \theta)|, \quad \left| \frac{\partial m^*(x; \theta)}{\partial \theta} \right|, \quad \left| \frac{\partial^2 m^*(x; \theta)}{\partial \theta \partial \theta} \right| \quad \text{are bounded uniformly over \( \Theta \). (B.87)}
\end{align*}

Furthermore, since \( \nu(w; \theta, \gamma) \) is almost everywhere (a.e.) continuously differentiable in \( w^c \) (Assumption 5.6), it implies (by definition of \( \nu(w; \theta, \gamma) \)) that \( m^*(x; \theta) \) and \( \frac{\partial m^*(x; \theta)}{\partial \theta} \) are also continuous...
in \( w^c \) a.e. within the compact \( \Omega_{W^c} \). Therefore, for \( \forall \theta \in \Theta \),
\[
|m^*(x; \theta)|, \quad \left| \frac{\partial m^*(x; \theta)}{\partial \theta} \right|, \quad \left| \frac{\partial^2 m^*(x; \theta)}{\partial \theta \partial \theta'} \right|
\]
are bounded uniformly over \( \Omega_X \). [B.88]

Then, (B.87) and (B.88) together indicate the uniform boundedness of \( |m^*(x; \theta)| \) and its first and second derivatives over \( \Omega_X \) and \( \Theta \). Thus,
\[
\sup_{w \in \Omega_W, (\theta, \phi) \in \Theta \times [0, 1]} |g(w; \theta, \phi)| = \sup_{w \in \Omega_W, (\theta, \phi) \in \Theta \times [0, 1]} \left| \tau(x)(y - m(x; \theta, \phi)) \frac{\partial m^*(x; \theta)}{\partial \theta} \right| \\
\leq C \sup_{w \in \Omega_W, (\theta, \phi) \in \Theta \times [0, 1]} \left| \frac{\partial m^*(x; \theta)}{\partial \theta} \right| \leq C_1,
\]
where the first inequality is because the maximum of \( y \) and \( m(x; \theta, \phi) \) are finite since \( \Omega_{W^c} \) is compact, and \( \tau(\cdot) \) is bounded (Assumption 5.2).

For \( \delta(W_i; \theta, \phi) = \nu(W_i; \theta, \phi) - E[\nu(W_i; \theta, \phi)] \), with \( \nu(W_i; \theta, \phi) = \tau(X_i) \frac{\partial R(W_i, \theta, \phi)}{\partial \theta} \frac{\partial R(t, \gamma)}{\partial \gamma} \mathbf{1}_d \), and the \( d_\theta \times K_\tau \) vector
\[
\frac{\partial R(W_i; \theta, \phi)}{\partial \theta} = \begin{bmatrix}
- \frac{\partial m(X_i, \theta, \phi)}{\partial \theta} m^*(x_{i,1}; \theta) + (Y_i - m(X_i; \theta, \phi)) \frac{\partial m^*(x_{i,1}; \theta)}{\partial \theta} \\
\vdots \\
- \frac{\partial m(X_i, \theta, \phi)}{\partial \theta} m^*(x_{i,K_T}; \theta) + (Y_i - m(X_i; \theta, \phi)) \frac{\partial m^*(x_{i,K_T}; \theta)}{\partial \theta}
\end{bmatrix},
\]

it is easy to see that \( \delta(W_i; \theta, \phi) \) is a function of \( m^*(x; \theta) \), \( \frac{\partial m^*(x; \theta)}{\partial \theta} \), and \( \frac{\partial \phi(t, \gamma)}{\partial \gamma} \), and it is linear in \( \phi \). Moreover, \( \phi \) is the probability function of discrete random variables therefore strictly lies in \([0, 1]\). Hence, the above discussion together with the uniform boundedness of \( \frac{\partial \phi(t, \gamma)}{\partial \gamma} \) provided in the proof of Corollary 5.3 leads to \( \sup_{w \in \Omega_W, (\theta, \phi) \in \Theta \times [0, 1]} |\delta(w; \theta, \phi)| \leq C_2 \) for constant \( C_2 > 0 \). So far we have established that (a) holds.

Secondly, move on to (b). For \( \theta^*_N \) between \( \theta^0 \) and \( \hat{\theta}_N \), the triangular inequality and the mean value theorem lead to
\[
\left\| \hat{g}_i - \bar{g}_i \right\| \leq \left\| g(W; \hat{\theta}_N, \hat{\phi}_N) - g(W; \theta^0, \hat{\phi}_N) \right\| + \left\| g(W; \theta^0, \hat{\phi}_N) - g(W; \theta^0, \phi^0) \right\| \\
+ \left\| \delta(W; \hat{\theta}_N, \hat{\phi}_N) - \delta(W; \theta^0, \hat{\phi}_N) \right\| + \left\| \delta(W; \theta^0, \hat{\phi}_N) - \delta(W; \theta^0, \phi^0) \right\| \\
\leq \left\| \frac{\partial g(W; \theta^*_N, \hat{\phi}_N)}{\partial \theta'} \right\| \left\| \hat{\theta}_N - \theta^0 \right\| + \left\| \delta(W; \theta^0, \hat{\phi}_N) - g(W; \theta^0, \phi^0) \right\| \\
+ \left\| \frac{\partial \delta(W; \theta^*_N, \hat{\phi}_N)}{\partial \theta'} \right\| \left\| \hat{\theta}_N - \theta^0 \right\| + \left\| \delta(W; \theta^0, \hat{\phi}_N) - \delta(W; \theta^0, \phi^0) \right\|. \tag{B.89}
\]

Start from the first term of (B.89), when sample size is large enough (i.e. \( \hat{\phi}_N \) is close to \( \phi^0 \)),
\[
\left\| \frac{\partial g(W; \theta^*_N, \hat{\phi}_N)}{\partial \theta'} \right\| = \tau(X_i) \left[ - \frac{\partial m(X_i; \theta^*_N, \hat{\phi}_N)}{\partial \theta} \frac{\partial m(X_i; \theta^*_N, \hat{\phi}_N)}{\partial \theta'} \right]
\]
the third term of (B.89), by the dominated convergence theorem, we have

\[ C \left( \sum_{i,j=1}^{K_T} \left| \frac{\partial m^*(x_{i,j}; \theta_N^*)}{\partial \theta} \right| \right) \leq C_3, \tag{B.90} \]

where the last line is because of (B.87) and (B.88). For the second term of (B.89), it yields from the calculation in (E.17) that

\[
\left\| g(W_i; \theta^0, \hat{\phi}_N) - g(W_i; \theta^0, \phi^0) \right\| \leq \left\| \hat{\phi}_N - \phi^0 \right\|_{\infty} \left[ \sum_{i,j=1}^{K_T} \left| m^*(x_{i,j}; \theta^0) \right| \right] + \tau_i \left| Y_i - m(X_i; \theta^0, \phi^0) \right| + s.o.
\]

\[
\leq C \left\| \hat{\phi}_N - \phi^0 \right\|_{\infty} \left[ \sum_{i,j=1}^{K_T} \left| m^*(x_{i,j}; \theta^0) \right| \right] + s.o.
\]

\[
\leq C_4 \left\| \hat{\phi}_N - \phi^0 \right\|_{\infty}. \tag{B.91}
\]

where the second inequality is because of the compactness of \( \Omega \) which implies both \( m^*(x_{i,j}; \theta^0) \) and \( |Y_i - m(X_i; \theta^0, \phi^0)| \) are bounded, and the last inequality is due to (B.87) and (B.88). To bound the third term of (B.89), by the dominated convergence theorem, we have

\[
\frac{\partial \delta(W_i; \theta_N^*, \hat{\phi}_N)}{\partial \theta'} = \tau(X_i) \frac{\partial}{\partial \theta'} \left( \frac{\partial R(W_i; \theta_N^*, \hat{\phi}_N)}{\partial \theta} \frac{\partial \phi(t; \hat{\gamma}_N)}{\partial \gamma'} \right) - E \left[ \tau(X_i) \frac{\partial}{\partial \theta'} \left( \frac{\partial R(W_i; \theta_N^*, \hat{\phi}_N)}{\partial \theta} \frac{\partial \phi(t; \hat{\gamma}_N)}{\partial \gamma'} \right) \right].
\]

Based on similar arguments used to obtain (B.90) and the uniform boundedness of \( \frac{\partial \phi(\gamma)}{\partial \gamma'} \) over \( \gamma \in [0, 1] \) provided in the proof of Corollary 5.3, we can get \( \left\| \frac{\partial \delta(W_i; \theta_N^*, \hat{\phi}_N)}{\partial \theta'} \right\| \leq C_5 \) for some constant \( C_5 > 0 \). At last,

\[
\left\| \delta(W_i; \theta^0, \hat{\phi}_N) - \delta(W_i; \theta^0, \phi^0) \right\| \leq \left\| \tau(X_i) \right\| \left\| \frac{\partial R(W_i; \theta^0, \hat{\phi}_N)}{\partial \theta} \frac{\partial \phi(t; \hat{\gamma}_N)}{\partial \gamma'} \right\| + \left\| \frac{\partial \phi(t; \hat{\gamma}_N)}{\partial \gamma'} \right\| \leq C_6 \left\| \hat{\gamma}_N - \gamma^0 \right\|_{\infty}.
\]
Given the results above, by Corollary 5.3, (B.86) can be bounded as

\[ \Delta \Omega_1 \leq \frac{C}{N} \sum_{i=1}^{N} |\Delta(i, N)| \left( \|\hat{\theta}_N - \theta^0\| + \|\hat{\gamma}_N - \gamma^0\|_{\infty} \right) = o_p(1), \]

based on the consistency of \( \hat{\theta}_N \) and \( \hat{\gamma}_N \), and the fact that \( 1/N \sum_{i=1}^{N} |\Delta(i, N)| = O(1) \).

**Step 2.** Next, let us deal with \( \Delta \Omega_2 \). Based on (E.7) and Assumption 5.7 (v),

\[
E[\|\Delta \Omega_2\|^2] \leq \frac{d_\theta}{N^2} \left\| \sum_{i,k=1}^{N} \sum_{j \in \Delta(i, N)} \sum_{l \in \Delta(k, N)} E\left[ (\tilde{g}_i \tilde{g}_j' - E[\tilde{g}_i \tilde{g}_j'])'(\tilde{g}_k \tilde{g}_l' - E[\tilde{g}_k \tilde{g}_l']) \right] \right\|_{\infty} \\
\leq \frac{d_\theta}{N^2} o \left( \| [\Sigma_N^g]^2 \|_{\infty} \right) = o(1),
\]

where the last line comes from (E.10) that \( o(\| [\Sigma_N^g]^2 \|_{\infty}/N^2) = o(1) \). Hence, \( \| \hat{\Omega}_N - \Omega \| = o_p(1). \]

### C  Tables and Figures
Table 3: Statistics of Misclassified Links ($p^\omega = 0.6$, $p^V = \delta^V/N$)

(a) $r_{deg} = 5$

| $\delta^V$ | $1 - p^U$ (%) | $p^V$ (%) | $N$ | $|N_i|$ | $S_i$ | Misclassified links (%) |
|------------|--------------|----------|-----|--------|--------|--------------------------|
|            |              |          |     |        |        | avg. | max | avg. | max | 1 to 0 | 0 to 1 | total | ratio (%) |
| 0          | 20           | 0        | 1k  | 4.97  | 14.84  | 1.49 | 6.95 | 677.5 | 0    | 677.5  | 12.01 |
| 0          | 20           | 0        | 2k  | 5.04  | 15.63  | 1.51 | 7.51 | 1371  | 0    | 1371   | 12.00 |
| 0          | 20           | 0        | 5k  | 5.11  | 16.66  | 1.53 | 8.12 | 3482  | 0    | 3482   | 12.01 |
| 0.010      | 1k           | 0        | 4.97 | 14.84 | 1.49   | 6.95 | 677.5 | 0    | 677.5 | 12.01 |
| 0.005      | 2k           | 0.057   | 5.04 | 15.63 | 1.51   | 7.51 | 1371  | 0    | 1371  | 12.00 |
| 0.002      | 5k           | 0.005   | 5.11 | 16.66 | 1.53   | 8.12 | 3482  | 0    | 3482  | 12.01 |
| 0.050      | 1k           | 0.050   | 5.04 | 15.63 | 1.51   | 7.51 | 1371  | 0    | 1371  | 12.00 |
| 0.5        | 20           | 0.025   | 5.11 | 16.66 | 1.53   | 8.12 | 3482  | 0    | 3482  | 12.01 |
| 0.1        | 40           | 0.010   | 5.04 | 15.63 | 1.51   | 7.51 | 1371  | 0    | 1371  | 12.00 |
| 0.5        | 40           | 0.010   | 5.11 | 16.66 | 1.53   | 8.12 | 3482  | 0    | 3482  | 12.01 |

(b) $r_{deg} = 8$

| $\delta^V$ | $1 - p^U$ (%) | $p^V$ (%) | $N$ | $|N_i|$ | $S_i$ | Misclassified links (%) |
|------------|--------------|----------|-----|--------|--------|--------------------------|
|            |              |          |     |        |        | avg. | max | avg. | max | 1 to 0 | 0 to 1 | total | ratio (%) |
| 0          | 20           | 0        | 1k  | 7.85  | 20.50  | 2.36 | 9.08 | 1069  | 0    | 1069   | 12.02 |
| 0          | 20           | 0        | 2k  | 7.92  | 21.55  | 2.40 | 9.75 | 2177  | 0    | 2177   | 12.01 |
| 0          | 20           | 0        | 5k  | 8.12  | 22.83  | 2.44 | 10.58 | 5540  | 0    | 5540   | 12.01 |
| 0.010      | 1k           | 0        | 7.85 | 20.50 | 2.36   | 9.08 | 1069  | 0    | 1069  | 12.02 |
| 0.005      | 2k           | 0.050   | 7.92 | 21.55 | 2.40   | 9.75 | 2177  | 0    | 2177  | 12.00 |
| 0.002      | 5k           | 0.005   | 8.12 | 22.83 | 2.44   | 10.58 | 5540  | 0    | 5540  | 12.01 |
| 0.050      | 1k           | 0.050   | 7.85 | 20.50 | 2.36   | 9.08 | 1069  | 0    | 1069  | 12.02 |
| 0.5        | 20           | 0.025   | 8.12 | 22.83 | 2.44   | 10.58 | 5540  | 0    | 5540  | 12.01 |
| 0.1        | 40           | 0.010   | 8.12 | 22.83 | 2.44   | 10.58 | 5540  | 0    | 5540  | 12.01 |
| 0.5        | 40           | 0.010   | 8.12 | 22.83 | 2.44   | 10.58 | 5540  | 0    | 5540  | 12.01 |

Note: The results in this table can be applied to both $(|N_i|, S_i)$ and $(\tilde{|N_i|}, \tilde{S_i})$. 

74
Table 4: Estimation Results of Treatment Effect $\tau_d(0,0,3)$ ($p^N = p^{\tilde{N}} = 0.6,$ $p^V = \delta^V / N$)

(a) $r_{deg} = 5$

| $\delta^V$ | $(1 - p^U, p^V)$ | $(1 - p^U, p^V) \ N$ | SPE | Naive 1 | Naive 2 |
|------------|-----------------|------------------|-----|---------|---------|
|            | bias | sd  | mse | cr  | bias | sd  | mse | cr  | bias | sd  | mse | cr  |
| (20, 0.010)| 1k   | -0.060 | 0.349 | 0.125 | 0.931 | -0.073 | 0.292 | 0.091 | 0.943 | -0.063 | 0.294 | 0.090 | 0.935 |
| 0.1 (20, 0.005) | 2k   | -0.027 | 0.245 | 0.061 | 0.941 | -0.077 | 0.206 | 0.048 | 0.933 | -0.071 | 0.208 | 0.048 | 0.931 |
| (20, 0.002) | 5k   | -0.016 | 0.133 | 0.018 | 0.937 | -0.060 | 0.132 | 0.021 | 0.924 | -0.063 | 0.130 | 0.021 | 0.916 |
| (20, 0.050) | 1k   | -0.053 | 0.354 | 0.128 | 0.941 | -0.076 | 0.319 | 0.108 | 0.942 | -0.061 | 0.284 | 0.084 | 0.946 |
| 0.5 (20, 0.025) | 2k   | -0.032 | 0.243 | 0.060 | 0.941 | -0.097 | 0.219 | 0.057 | 0.925 | -0.061 | 0.205 | 0.046 | 0.939 |
| (20, 0.010) | 5k   | -0.028 | 0.133 | 0.019 | 0.942 | -0.083 | 0.139 | 0.026 | 0.909 | -0.062 | 0.133 | 0.022 | 0.922 |
| (40, 0.010) | 1k   | 0.075  | 0.538 | 0.296 | 0.950 | -0.035 | 0.405 | 0.165 | 0.952 | -0.016 | 0.390 | 0.153 | 0.948 |
| 0.1 (40, 0.005) | 2k   | 0.051  | 0.384 | 0.150 | 0.942 | -0.018 | 0.276 | 0.076 | 0.948 | -0.019 | 0.273 | 0.075 | 0.955 |
| (40, 0.002) | 5k   | 0.038  | 0.236 | 0.057 | 0.938 | -0.013 | 0.173 | 0.030 | 0.948 | 0.004  | 0.182 | 0.033 | 0.945 |
| (40, 0.050) | 1k   | 0.059  | 0.547 | 0.303 | 0.940 | -0.040 | 0.398 | 0.160 | 0.958 | -0.012 | 0.399 | 0.160 | 0.950 |
| 0.5 (40, 0.025) | 2k   | 0.040  | 0.368 | 0.137 | 0.941 | -0.047 | 0.280 | 0.081 | 0.954 | -0.015 | 0.283 | 0.080 | 0.953 |
| (40, 0.010) | 5k   | 0.022  | 0.219 | 0.048 | 0.940 | -0.052 | 0.189 | 0.038 | 0.944 | 0.012  | 0.175 | 0.031 | 0.952 |

(b) $r_{deg} = 8$

| $\delta^V$ | $(1 - p^U, p^V)$ | $(1 - p^U, p^V) \ N$ | SPE | Naive 1 | Naive 2 |
|------------|-----------------|------------------|-----|---------|---------|
|            | bias | sd  | mse | cr  | bias | sd  | mse | cr  | bias | sd  | mse | cr  |
| (20, 0.010)| 1k   | 0.060 | 0.574 | 0.334 | 0.953 | -0.161 | 0.527 | 0.304 | 0.941 | -0.142 | 0.507 | 0.277 | 0.949 |
| 0.1 (20, 0.005) | 2k   | -0.048 | 0.284 | 0.083 | 0.940 | -0.140 | 0.359 | 0.148 | 0.929 | -0.130 | 0.392 | 0.170 | 0.928 |
| (20, 0.002) | 5k   | -0.020 | 0.180 | 0.033 | 0.954 | -0.141 | 0.237 | 0.076 | 0.910 | -0.139 | 0.233 | 0.074 | 0.908 |
| (20, 0.050) | 1k   | -0.016 | 0.535 | 0.287 | 0.930 | -0.170 | 0.518 | 0.298 | 0.938 | -0.170 | 0.522 | 0.302 | 0.938 |
| 0.5 (20, 0.025) | 2k   | 0.019 | 0.399 | 0.160 | 0.954 | -0.141 | 0.394 | 0.175 | 0.935 | -0.155 | 0.361 | 0.154 | 0.934 |
| (20, 0.010) | 5k   | -0.019 | 0.169 | 0.029 | 0.963 | -0.162 | 0.241 | 0.084 | 0.899 | -0.144 | 0.243 | 0.080 | 0.902 |
| (40, 0.010) | 1k   | 0.383 | 0.792 | 0.774 | 0.934 | -0.119 | 0.776 | 0.617 | 0.946 | -0.120 | 0.756 | 0.585 | 0.942 |
| 0.1 (40, 0.005) | 2k   | 0.356 | 0.569 | 0.451 | 0.933 | -0.118 | 0.574 | 0.343 | 0.946 | -0.103 | 0.560 | 0.325 | 0.945 |
| (40, 0.002) | 5k   | 0.280 | 0.343 | 0.196 | 0.897 | -0.101 | 0.354 | 0.135 | 0.935 | -0.086 | 0.354 | 0.133 | 0.938 |
| (40, 0.010) | 1k   | 0.367 | 0.794 | 0.765 | 0.919 | -0.184 | 0.757 | 0.607 | 0.948 | -0.121 | 0.749 | 0.575 | 0.949 |
| 0.5 (40, 0.025) | 2k   | 0.323 | 0.556 | 0.413 | 0.934 | -0.148 | 0.552 | 0.326 | 0.937 | -0.115 | 0.550 | 0.316 | 0.950 |
| (40, 0.010) | 5k   | 0.211 | 0.342 | 0.162 | 0.910 | -0.154 | 0.348 | 0.145 | 0.928 | -0.103 | 0.362 | 0.141 | 0.945 |

Note: SPE lists the semiparametric estimation results proposed in Section 5.3. Estimates of Naive 1 are computed using OLS with $\{Y_i, D_i, S_i, Z_i, |N_i|\}_{i=1}^N$; and estimates of Naive 2 are computed using OLS with $\{Y_i, D_i, \tilde{S}_i, Z_i, |N_i|\}_{i=1}^N$. True value of the treatment effect $\tau_d(0,0,3) = 1$.  

75
Table 5: Estimation Results of Treatment Effect $\tau_d(0, 1, 3)$ ($p^\omega = p^\phi = 0.6, \ p^V = \delta^V / N$)

(a) $r_{\text{deg}} = 5$

| $\delta^V$ | $(1 - p^U, p^V)$ | $(1 - p^U, p^V)$ | $N$ | SPE | Naive 1 | Naive 2 |
|------------|------------------|-----------------|-----|-----|---------|---------|
|            |                  |                 |     | bias| sd    | mse | cr |
| (20, 0.010)| 1k               | -0.068 0.313 0.102 0.948 | 0.131 0.279 0.095 0.921 | 0.120 0.268 0.086 0.936 |
| 0.1        | (20, 0.005)      | 2k               | -0.059 0.215 0.050 0.939 | 0.121 0.195 0.053 0.897 | 0.140 0.190 0.056 0.883 |
|            | (20, 0.002)      | 5k               | -0.052 0.126 0.019 0.930 | 0.132 0.122 0.032 0.815 | 0.133 0.126 0.034 0.818 |
| (20, 0.050)| 1k               | -0.066 0.323 0.108 0.944 | 0.078 0.283 0.086 0.941 | 0.133 0.270 0.090 0.920 |
| 0.5        | (20, 0.025)      | 2k               | -0.059 0.209 0.047 0.946 | 0.075 0.201 0.046 0.943 | 0.136 0.195 0.057 0.892 |
|            | (20, 0.010)      | 5k               | -0.057 0.114 0.016 0.931 | 0.081 0.124 0.022 0.907 | 0.135 0.115 0.031 0.778 |
| (40, 0.010)| 1k               | 0.040 0.528 0.281 0.953 | 0.287 0.405 0.247 0.885 | 0.318 0.408 0.268 0.882 |
| 0.1        | (40, 0.005)      | 2k               | 0.007 0.350 0.123 0.949 | 0.299 0.293 0.175 0.834 | 0.303 0.291 0.176 0.825 |
|            | (40, 0.002)      | 5k               | 0.001 0.209 0.044 0.957 | 0.305 0.181 0.126 0.600 | 0.316 0.183 0.133 0.581 |
| (40, 0.050)| 1k               | 0.054 0.522 0.276 0.946 | 0.255 0.393 0.219 0.898 | 0.303 0.411 0.261 0.892 |
| 0.5        | (40, 0.025)      | 2k               | 0.027 0.325 0.106 0.953 | 0.252 0.286 0.145 0.863 | 0.325 0.275 0.181 0.788 |
|            | (40, 0.010)      | 5k               | 0.003 0.196 0.039 0.952 | 0.248 0.181 0.094 0.730 | 0.322 0.185 0.138 0.590 |

(b) $r_{\text{deg}} = 8$

| $\delta^V$ | $(1 - p^U, p^V)$ | $(1 - p^U, p^V)$ | $N$ | SPE | Naive 1 | Naive 2 |
|------------|------------------|-----------------|-----|-----|---------|---------|
|            |                  |                 |     | bias| sd    | mse | cr |
| (20, 0.010)| 1k               | -0.024 0.516 0.267 0.953 | 0.071 0.445 0.203 0.941 | 0.083 0.446 0.205 0.944 |
| 0.1        | (20, 0.005)      | 2k               | -0.061 0.323 0.108 0.951 | 0.084 0.319 0.109 0.937 | 0.078 0.335 0.118 0.941 |
|            | (20, 0.002)      | 5k               | -0.089 0.190 0.044 0.939 | 0.077 0.200 0.046 0.936 | 0.087 0.202 0.048 0.938 |
| (20, 0.050)| 1k               | 0.062 0.629 0.400 0.966 | 0.062 0.448 0.204 0.955 | 0.075 0.454 0.212 0.943 |
| 0.5        | (20, 0.025)      | 2k               | -0.054 0.373 0.142 0.960 | 0.058 0.330 0.112 0.945 | 0.082 0.305 0.100 0.943 |
|            | (20, 0.010)      | 5k               | -0.096 0.177 0.041 0.937 | 0.044 0.208 0.045 0.942 | 0.078 0.210 0.050 0.927 |
| (40, 0.010)| 1k               | 0.329 0.813 0.768 0.932 | 0.267 0.703 0.565 0.938 | 0.279 0.709 0.581 0.933 |
| 0.1        | (40, 0.005)      | 2k               | 0.299 0.571 0.416 0.932 | 0.300 0.511 0.351 0.908 | 0.306 0.502 0.346 0.901 |
|            | (40, 0.002)      | 5k               | 0.173 0.336 0.143 0.916 | 0.272 0.318 0.175 0.877 | 0.285 0.322 0.185 0.851 |
| (40, 0.010)| 1k               | 0.300 0.814 0.752 0.934 | 0.244 0.655 0.488 0.939 | 0.298 0.700 0.579 0.933 |
| 0.5        | (40, 0.025)      | 2k               | 0.256 0.538 0.355 0.925 | 0.240 0.474 0.282 0.912 | 0.298 0.500 0.339 0.903 |
|            | (40, 0.010)      | 5k               | 0.119 0.327 0.121 0.937 | 0.218 0.316 0.147 0.888 | 0.284 0.330 0.190 0.863 |

Note: SPE lists the semiparametric estimation results proposed in Section 5.3. Estimates of Naive 1 are computed using OLS with $\{Y_i, D_i, S_i, Z_i, |N_i|\}_{1}^{N}$; and estimates of Naive 2 are computed using OLS with $\{Y_i, D_i, \tilde{S}_i, Z_i, |\tilde{N}_i|\}_{1}^{N}$. True value of the treatment effect $\tau_d(0, 1, 3) = 2$.  

76
Table 6: Estimation Results of Spillover Effect $\tau_s(1,0,3)$ ($p^\omega = p^{\hat{\omega}} = 0.6$, $p^V = \delta^V/N$)

(a) $r_{deg} = 5$

| $\delta^V$ | $(1-p^U,p^V)$ | $(1-p^U,p^V)$ | N | SPE | Naive 1 | Naive 2 |
|---------|----------------|----------------|----|-----|---------|---------|
|        | (%)            | (%)            |    | bias | sd      | mse     | cr     | bias | sd      | mse     | cr     |
| (20, 0.010) | 1k             | 0.035 0.488 0.240 0.957 | 0.270 0.214 0.119 0.749 | 0.366 0.215 0.180 0.582 |
| 0.1     | (20, 0.005)    | 2k             | 0.040 0.373 0.141 0.961 | 0.254 0.160 0.090 0.628 | 0.351 0.155 0.147 0.365 |
|         | (20, 0.002)    | 5k             | 0.073 0.209 0.049 0.945 | 0.252 0.102 0.074 0.306 | 0.342 0.100 0.127 0.074 |
| (20, 0.050) | 1k             | 0.050 0.543 0.297 0.947 | -0.096 0.220 0.058 0.924 | 0.355 0.211 0.170 0.596 |
| 0.5     | (20, 0.025)    | 2k             | 0.054 0.354 0.128 0.952 | -0.091 0.159 0.033 0.919 | 0.352 0.151 0.147 0.360 |
|         | (20, 0.010)    | 5k             | 0.082 0.209 0.051 0.930 | -0.104 0.105 0.022 0.841 | 0.346 0.102 0.130 0.089 |
| (40, 0.010) | 1k             | 0.051 0.750 0.565 0.945 | 0.436 0.283 0.270 0.650 | 0.533 0.301 0.375 0.562 |
| 0.1     | (40, 0.005)    | 2k             | 0.079 0.607 0.375 0.949 | 0.432 0.209 0.230 0.441 | 0.532 0.216 0.330 0.299 |
|         | (40, 0.002)    | 5k             | 0.165 0.351 0.150 0.923 | 0.415 0.138 0.192 0.144 | 0.507 0.138 0.276 0.046 |
| (40, 0.050) | 1k             | 0.037 0.753 0.568 0.952 | 0.082 0.289 0.090 0.948 | 0.519 0.309 0.364 0.611 |
| 0.5     | (40, 0.025)    | 2k             | 0.086 0.557 0.317 0.942 | 0.079 0.212 0.051 0.936 | 0.517 0.204 0.309 0.285 |
|         | (40, 0.010)    | 5k             | 0.175 0.369 0.167 0.927 | 0.057 0.138 0.022 0.932 | 0.505 0.138 0.274 0.044 |

(b) $r_{deg} = 8$

| $\delta^V$ | $(1-p^U,p^V)$ | $(1-p^U,p^V)$ | N | SPE | Naive 1 | Naive 2 |
|---------|----------------|----------------|----|-----|---------|---------|
|        | (%)            | (%)            |    | bias | sd      | mse     | cr     | bias | sd      | mse     | cr     |
| (20, 0.010) | 1k             | 0.076 0.861 0.748 0.935 | 0.653 0.367 0.561 0.544 | 0.780 0.354 0.733 0.392 |
| 0.1     | (20, 0.005)    | 2k             | -0.034 0.403 0.163 0.945 | 0.650 0.171 0.451 0.404 | 0.753 0.167 0.595 0.010 |
|         | (20, 0.002)    | 5k             | 0.119 0.891 0.809 0.943 | 0.265 0.368 0.205 0.885 | 0.774 0.372 0.737 0.459 |
| (20, 0.050) | 1k             | 0.029 0.746 0.557 0.941 | 0.263 0.263 0.138 0.833 | 0.764 0.267 0.655 0.181 |
| 0.5     | (20, 0.025)    | 2k             | -0.013 0.375 0.141 0.951 | 0.245 0.176 0.091 0.730 | 0.745 0.174 0.586 0.014 |
|         | (20, 0.010)    | 5k             | 1.270 1.023 2.659 0.789 | 1.260 0.535 1.872 0.335 | 1.348 0.517 2.085 0.254 |
| (40, 0.010) | 1k             | 1.040 0.831 1.772 0.796 | 1.244 0.367 1.683 0.104 | 1.314 0.379 1.869 0.063 |
| 0.1     | (40, 0.005)    | 2k             | 0.743 0.602 0.915 0.787 | 1.179 0.253 1.454 0.008 | 1.269 0.255 1.676 0.001 |
|         | (40, 0.002)    | 5k             | 1.171 1.045 2.462 0.823 | 0.873 0.506 1.019 0.581 | 1.356 0.516 2.106 0.231 |
| (40, 0.010) | 1k             | 0.993 0.854 1.715 0.843 | 0.814 0.368 0.798 0.401 | 1.298 0.369 1.822 0.053 |
| 0.5     | (40, 0.025)    | 2k             | 0.653 0.620 0.811 0.845 | 0.803 0.241 0.702 0.090 | 1.270 0.253 1.678 0.002 |
|         | (40, 0.010)    | 5k             | 0.653 0.620 0.811 0.845 | 0.803 0.241 0.702 0.090 | 1.270 0.253 1.678 0.002 |

Note: SPE lists the semiparametric estimation results proposed in Section 5.3. Estimates of Naive 1 are computed using OLS with $\{Y_i, D_i, S_i, Z_i, |N_i|\}_{i=1}^N$; and estimates of Naive 2 are computed using OLS with $\{Y_i, D_i, S_i, Z_i, |N_i|\}_{i=1}^N$. True value of the treatment effect $\tau_s(1,0,3) = 3$. 

77
Table 7: Estimation Results of Spillover Effect $\tau_s(1, 1, 3)$ ($p^\omega = \tilde{p}^\omega = 0.6$, $p^V = \delta^V/N$)  

(a) $r_{deg} = 5$

| $\delta^V$ | $(1 - p^U,p^V)$ | $(1 - p^U,p^V)$ | $N$ | SPE | Naive 1 | Naive 2 |
|------------|----------------|----------------|-----|-----|---------|---------|
|            | (%)            | (%)            |     |     |         |         |
|            |                 |                 | 1k  | 0.060 0.846 0.719 0.942 | 0.617 0.243 0.440 0.272 | 0.709 0.246 0.564 0.172 |
| 0.1        | (20, 0.010)    |                 | 2k  | 0.208 0.522 0.273 0.957 | 0.603 0.173 0.393 0.064 | 0.693 0.181 0.513 0.029 |
|            | (20, 0.005)    |                 | 5k  | 0.021 0.284 0.881 0.954 | 0.605 0.112 0.379 0.001 | 0.695 0.112 0.496 0.000 |
|            | (20, 0.050)    |                 | 1k  | 0.127 0.928 0.878 0.941 | 0.289 0.246 0.144 0.768 | 0.690 0.246 0.537 0.203 |
|            | (20, 0.025)    |                 | 2k  | 0.090 0.555 0.316 0.952 | 0.299 0.186 0.124 0.633 | 0.704 0.173 0.525 0.025 |
|            | (20, 0.010)    |                 | 5k  | 0.047 0.306 0.096 0.954 | 0.294 0.117 0.100 0.275 | 0.703 0.114 0.507 0.000 |
| (40, 0.010) |                 |                 | 1k  | 0.619 0.979 1.342 0.897 | 1.273 0.343 1.739 0.037 | 1.386 0.359 2.050 0.030 |
|            | (40, 0.005)    |                 | 2k  | 0.377 0.864 0.889 0.915 | 1.288 0.242 1.718 0.000 | 1.406 0.249 2.038 0.000 |
|            | (40, 0.002)    |                 | 5k  | 0.232 0.591 0.403 0.937 | 1.277 0.167 1.658 0.000 | 1.385 0.164 1.944 0.000 |
| (40, 0.050) |                 |                 | 1k  | 0.564 1.016 1.351 0.905 | 0.865 0.344 0.866 0.297 | 1.370 0.369 2.013 0.041 |
|            | (40, 0.025)    |                 | 2k  | 0.272 0.869 0.829 0.926 | 0.875 0.247 0.828 0.053 | 1.387 0.249 1.986 0.000 |
|            | (40, 0.010)    |                 | 5k  | 0.207 0.588 0.388 0.949 | 0.868 0.156 0.777 0.000 | 1.389 0.167 1.957 0.000 |

(b) $r_{deg} = 8$

| $\delta^V$ | $(1 - p^U,p^V)$ | $(1 - p^U,p^V)$ | $N$ | SPE | Naive 1 | Naive 2 |
|------------|----------------|----------------|-----|-----|---------|---------|
|            | (%)            | (%)            |     |     |         |         |
|            |                 |                 | 1k  | 0.382 1.416 2.150 0.931 | 1.320 0.419 1.919 0.119 | 1.456 0.403 2.281 0.062 |
| 0.1        | (20, 0.010)    |                 | 2k  | 0.139 1.055 1.133 0.950 | 1.345 0.321 1.912 0.013 | 1.480 0.304 2.282 0.007 |
|            | (20, 0.005)    |                 | 5k  | -0.053 0.621 0.389 0.938 | 1.369 0.191 1.912 0.000 | 1.489 0.196 2.256 0.000 |
|            | (20, 0.050)    |                 | 1k  | 0.498 1.411 2.339 0.940 | 0.921 0.412 1.019 0.396 | 1.474 0.424 2.353 0.066 |
|            | (20, 0.025)    |                 | 2k  | 0.029 1.003 1.007 0.944 | 0.928 0.293 0.947 0.111 | 1.470 0.304 2.255 0.004 |
|            | (20, 0.010)    |                 | 5k  | -0.019 0.627 0.393 0.950 | 0.931 0.201 0.908 0.002 | 1.476 0.201 2.220 0.000 |
| (40, 0.010) |                 |                 | 1k  | 2.673 1.634 9.812 0.647 | 3.000 0.618 9.379 0.001 | 3.120 0.631 10.136 0.001 |
|            | (40, 0.005)    |                 | 2k  | 2.185 1.432 6.826 0.704 | 2.974 0.443 9.041 0.000 | 3.113 0.449 9.892 0.000 |
|            | (40, 0.002)    |                 | 5k  | 1.437 1.152 3.391 0.779 | 2.997 0.301 9.072 0.000 | 3.125 0.294 9.851 0.000 |
| (40, 0.010) |                 |                 | 1k  | 2.540 1.698 9.335 0.724 | 2.482 0.607 6.526 0.012 | 3.125 0.613 10.140 0.001 |
|            | (40, 0.025)    |                 | 2k  | 2.135 1.460 6.692 0.727 | 2.459 0.423 6.225 0.000 | 3.109 0.441 9.860 0.000 |
|            | (40, 0.010)    |                 | 5k  | 1.262 1.175 2.974 0.837 | 2.489 0.279 6.274 0.000 | 3.120 0.295 9.821 0.000 |

Note: SPE lists the semiparametric estimation results proposed in Section 5.3. Estimates of Naive 1 are computed using OLS with \{Y_i, D_i, S_i, Z_i, |N|N\}_{i=1}^N; and estimates of Naive 2 are computed using OLS with \{Y_i, D_i, \tilde{S}_i, Z_i, |N|N\}_{i=1}^N. True value of the treatment effect $\tau_s(1, 1, 3) = 2.5$.  

78
| $q$ | $(1 - p^U, p^V)$ | $(1 - p^U, p^V)$ | SPE bias | SPE sd | SPE mse | SPE cr | Naive 1 bias | Naive 1 sd | Naive 1 mse | Naive 1 cr | Naive 2 bias | Naive 2 sd | Naive 2 mse | Naive 2 cr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.05 | (20, 0.002) | (20, 0) | $\tau^q_{0.0, 0.3}$ | 0.020 | 0.118 | 0.014 | 0.915 | -0.067 | 0.128 | 0.021 | 0.910 | -0.054 | 0.135 | 0.021 | 0.929 |
| | | | $\tau^q_{0.1, 0.3}$ | 0.145 | 0.184 | 0.055 | 0.865 | 0.249 | 0.097 | 0.072 | 0.273 | 0.347 | 0.099 | 0.130 | 0.063 |
| | | | $\tau^q_{1.0, 0.3}$ | 0.095 | 0.301 | 0.100 | 0.935 | 0.604 | 0.111 | 0.377 | 0.000 | 0.701 | 0.114 | 0.504 | 0.000 |
| | | | $\tau^q_{1.1, 0.3}$ | -0.001 | 0.110 | 0.013 | 0.934 | -0.065 | 0.129 | 0.021 | 0.917 | -0.062 | 0.131 | 0.021 | 0.923 |
| 0.1 | (20, 0.002) | (20, 0) | $\tau^q_{0.0, 0.3}$ | 0.021 | 0.110 | 0.013 | 0.934 | -0.065 | 0.129 | 0.021 | 0.917 | -0.062 | 0.131 | 0.021 | 0.923 |
| | | | $\tau^q_{0.1, 0.3}$ | 0.157 | 0.200 | 0.065 | 0.869 | 0.247 | 0.098 | 0.070 | 0.296 | 0.346 | 0.099 | 0.129 | 0.060 |
| | | | $\tau^q_{1.0, 0.3}$ | 0.101 | 0.295 | 0.097 | 0.920 | 0.596 | 0.109 | 0.368 | 0.000 | 0.696 | 0.115 | 0.497 | 0.000 |
| | | | $\tau^q_{1.1, 0.3}$ | 0.049 | 0.219 | 0.051 | 0.917 | -0.019 | 0.175 | 0.031 | 0.950 | -0.001 | 0.180 | 0.032 | 0.949 |
| 0.05 | (40, 0.002) | (40, 0) | $\tau^q_{0.0, 0.3}$ | 0.005 | 0.210 | 0.044 | 0.945 | 0.306 | 0.186 | 0.128 | 0.620 | 0.318 | 0.184 | 0.135 | 0.606 |
| | | | $\tau^q_{0.1, 0.3}$ | 0.245 | 0.329 | 0.168 | 0.877 | 0.424 | 0.130 | 0.196 | 0.101 | 0.513 | 0.136 | 0.281 | 0.032 |
| | | | $\tau^q_{1.0, 0.3}$ | 0.347 | 0.545 | 0.417 | 0.910 | 1.290 | 0.155 | 1.687 | 0.000 | 1.397 | 0.164 | 1.979 | 0.000 |
| | | | $\tau^q_{1.1, 0.3}$ | 0.070 | 0.225 | 0.055 | 0.930 | 0.005 | 0.182 | 0.033 | 0.956 | 0.014 | 0.174 | 0.030 | 0.951 |
| 0.1 | (40, 0.002) | (40, 0) | $\tau^q_{0.0, 0.3}$ | 0.010 | 0.208 | 0.043 | 0.954 | 0.307 | 0.187 | 0.129 | 0.613 | 0.313 | 0.182 | 0.131 | 0.586 |
| | | | $\tau^q_{0.1, 0.3}$ | 0.236 | 0.330 | 0.165 | 0.882 | 0.413 | 0.139 | 0.190 | 0.153 | 0.499 | 0.134 | 0.267 | 0.053 |
| | | | $\tau^q_{1.0, 0.3}$ | 0.344 | 0.509 | 0.377 | 0.886 | 1.282 | 0.161 | 1.670 | 0.000 | 1.388 | 0.157 | 1.951 | 0.000 |

Table 8: Robustness Check of Exclusion Restriction ($p^\omega = p^{\tilde{\omega}} = 0.6$, $p^V = \delta^V/N$, $p^{\tilde{V}} = 0$, $\delta^V = 0.1$, $N = 5k$)

(a) $r_{deg} = 5$

(b) $r_{deg} = 8$
Table 9: Robustness Check for One Type of Measurement Error ($p^U = p^\tilde{\gamma} = 0.6$, $p^V = \delta^V/N$, $p^\tilde{V} = \delta^V/N$, $N = 5k$)

(a) $r_{deg} = 5$

| $\delta^V$ | $\delta^V$ | $(1 - p^U, p^V)$ | $(1 - p^U, p^V)$ | SPE | Naive 1 | Naive 2 |
|------------|------------|-----------------|-----------------|-----|---------|---------|
|            |            |                 |                 | bias | sd      | mse     | cr   |
|            |            |                 |                 |      |         |         |      |
| 0.1 0.05   | (20, 0.002)| (20, 0.001)     |                 | $\tau_d(0, 0, 3)$ | -0.027 | 0.124 | 0.016 | 0.930 |
|            |            |                 |                 | $\tau_d(0, 1, 3)$ | -0.055 | 0.115 | 0.016 | 0.923 |
|            |            |                 |                 | $\tau_s(1, 0, 3)$ | 0.125  | 0.204 | 0.057 | 0.880 |
|            |            |                 |                 | $\tau_s(1, 1, 3)$ | 0.104  | 0.303 | 0.103 | 0.933 |
| 0.1 0.1    | (20, 0.002)| (20, 0.002)     |                 | $\tau_d(0, 0, 3)$ | -0.027 | 0.129 | 0.017 | 0.934 |
|            |            |                 |                 | $\tau_d(0, 1, 3)$ | -0.053 | 0.116 | 0.016 | 0.927 |
|            |            |                 |                 | $\tau_s(1, 0, 3)$ | 0.098  | 0.221 | 0.058 | 0.904 |
|            |            |                 |                 | $\tau_s(1, 1, 3)$ | 0.104  | 0.306 | 0.104 | 0.942 |
| 0.1 0.05   | (40, 0.002)| (40, 0.001)     |                 | $\tau_d(0, 0, 3)$ | 0.049  | 0.231 | 0.056 | 0.941 |
|            |            |                 |                 | $\tau_d(0, 1, 3)$ | -0.007 | 0.208 | 0.043 | 0.959 |
|            |            |                 |                 | $\tau_s(1, 0, 3)$ | 0.183  | 0.342 | 0.150 | 0.910 |
|            |            |                 |                 | $\tau_s(1, 1, 3)$ | 0.320  | 0.540 | 0.394 | 0.909 |
| 0.1 0.1    | (40, 0.002)| (40, 0.002)     |                 | $\tau_d(0, 0, 3)$ | 0.026  | 0.232 | 0.054 | 0.939 |
|            |            |                 |                 | $\tau_d(0, 1, 3)$ | -0.012 | 0.211 | 0.045 | 0.954 |
|            |            |                 |                 | $\tau_s(1, 0, 3)$ | 0.159  | 0.348 | 0.146 | 0.927 |
|            |            |                 |                 | $\tau_s(1, 1, 3)$ | 0.282  | 0.524 | 0.354 | 0.926 |

(b) $r_{deg} = 8$

| $\delta^V$ | $\delta^V$ | $(1 - p^U, p^V)$ | $(1 - p^U, p^V)$ | SPE | Naive 1 | Naive 2 |
|------------|------------|-----------------|-----------------|-----|---------|---------|
|            |            |                 |                 | bias | sd      | mse     | cr   |
|            |            |                 |                 |      |         |         |      |
| 0.1 0.05   | (20, 0.002)| (20, 0.001)     |                 | $\tau_d(0, 0, 3)$ | -0.052 | 0.202 | 0.044 | 0.956 |
|            |            |                 |                 | $\tau_d(0, 1, 3)$ | -0.097 | 0.201 | 0.050 | 0.943 |
|            |            |                 |                 | $\tau_s(1, 0, 3)$ | -0.040 | 0.419 | 0.177 | 0.955 |
|            |            |                 |                 | $\tau_s(1, 1, 3)$ | -0.027 | 0.669 | 0.448 | 0.949 |
| 0.1 0.1    | (20, 0.002)| (20, 0.002)     |                 | $\tau_d(0, 0, 3)$ | -0.061 | 0.180 | 0.036 | 0.959 |
|            |            |                 |                 | $\tau_d(0, 1, 3)$ | -0.105 | 0.186 | 0.046 | 0.934 |
|            |            |                 |                 | $\tau_s(1, 0, 3)$ | -0.049 | 0.379 | 0.146 | 0.955 |
|            |            |                 |                 | $\tau_s(1, 1, 3)$ | -0.026 | 0.644 | 0.415 | 0.949 |
| 0.1 0.05   | (40, 0.002)| (40, 0.001)     |                 | $\tau_d(0, 0, 3)$ | 0.264  | 0.338 | 0.184 | 0.903 |
|            |            |                 |                 | $\tau_d(0, 1, 3)$ | 0.169  | 0.318 | 0.130 | 0.917 |
|            |            |                 |                 | $\tau_s(1, 0, 3)$ | 0.728  | 0.646 | 0.947 | 0.937 |
|            |            |                 |                 | $\tau_s(1, 1, 3)$ | 1.359  | 1.150 | 3.168 | 0.902 |
| 0.1 0.1    | (40, 0.002)| (40, 0.002)     |                 | $\tau_d(0, 0, 3)$ | 0.254  | 0.349 | 0.187 | 0.909 |
|            |            |                 |                 | $\tau_d(0, 1, 3)$ | 0.165  | 0.330 | 0.136 | 0.916 |
|            |            |                 |                 | $\tau_s(1, 0, 3)$ | 0.737  | 0.636 | 0.948 | 0.931 |
|            |            |                 |                 | $\tau_s(1, 1, 3)$ | 1.446  | 1.134 | 3.377 | 0.787 |

80
Supplement to “Spillovers of Program Benefits with Mismeasured Networks”

Lina Zhang

Job Market Paper
(Click here for the latest version)

September 22, 2020

In this supplemental material, Section D provides sufficient conditions under which the existing nonparametric estimation of Leung (2020b) is still consistent, when no instrument variable for the true network is available. It is important for practitioners, because if one is not aware of the potential network mismeasurement in a setting with a limited misclassification rate, the findings in this section ensure that the standard nonparametric estimators are nevertheless likely to be consistent when the sample size is sufficiently large.\footnote{Department of Econometrics and Business Statistics, Monash University (lina.zhang@monash.edu).}

Section E introduces some useful lemmas which are used in the proof of the Appendix in the main text.

D Single Network Proxy

Denote the observed $N \times N$ adjacency matrix by $A$, with its $ij$-th entry $A_{ij} = 1$ if $j \in \mathcal{N}_i$. For $(d, s, z, n) \in \{0, 1\} \times \Omega_{S,Z,|\mathcal{N}|}$, define the conditional propensity score

$$e(d, s, z, n) = Pr\left(D_i = d, S_i = s \mid Z_i = z, |\mathcal{N}_i| = n\right).$$

The definition of $e(d, s, z, n)$ is akin to the propensity score of the multi-valued treatment (Imbens, 2000). Notably, $e(d, s, z, n)$ does not depend on the index $i$ even if the network is mismeasured, implying that it is identifiable from the observables.\footnote{Other studies using a single and imperfectly measured network include, e.g. Chandrasekhar and Lewis (2011), He and Song (2018), Lewbel, Qu, and Tang (2019), Sävje (2019) and Leung (2019a).}

**Theorem D.1** Under the assumptions in Proposition 3.1 and suppose the following conditions are satisfied.

\footnote{The conditional propensity score $e(d, s, z, n)$ does not depend on the index $i$, because under Assumptions 3.2 (a) and 3.4 (a), $e(d, s, z, n) = f_D^d(1)f_D(0)^{1-d}f_S|z=z,|\mathcal{N}_i|=n(s) = f_D^d(1)f_D(0)^{1-d} \times C_s^{n-1}f_D^d(1)f_D(0)^{n-s}$ where $f_D(d) := Pr(D_i = d)$. From the expression of $e(d, s, z, n)$, we can see that it is identical across all units in $\mathcal{P}$.}
(a) **(Strict Overlap)** There exists a constant $\xi \in (0, 1)$, such that $e(d, s, z, n) \in \{0, 1\} \times \Omega_{S, Z, |N|}$. 

(b) **(Lipschitz Condition)** For any $(d, z) \in \{0, 1\} \times \Omega_{Z}$ and all $(s, n), (s', n') \in \Omega_{S, |N|}$, there exist two Lipschitz constants $L_S$ and $L_{|N|}$ such that

$$
|m^*(d, s, z, n) - m^*(d, s', z, n)| < L_S |s - s'|,
$$

$$
|m^*(d, s, z, n) - m^*(d, s, z, n')| < L_{|N|} |n - n'|.
$$

(c) **(Misclassification Rate)** There exists some constant $\delta > 0$ such that

$$
\sup_{i \in \mathcal{P}, (z, n) \in \Omega_{Z, |N|}} E \left[ \|A^*_i - A_i\|^2 \mid Z_i = z, |N_i| = n \right] = O(N^{-\delta}).
$$

Then, we have $\|m_i - m^*\|_\infty = O(N^{-\delta}).$

Theorem D.1 states that the bias of the naive estimator $m_i$ relative to the CASF $m^*$ is negligible when the sample size is sufficiently large, if the observed network links suffer from only limited or mild level of mismeasurement. Such a condition holds, for example, if the misclassification of the network occurs only in a subset of an increasing number of individuals with rate $O(N^{\alpha})$ ($\kappa > \delta$), and the misclassification probability converges to zero uniformly $\sup_{i,j \in \mathcal{P}} E[|A^*_ij - A_{ij}|] = O(N^{-\kappa})$. It also holds in situations where links are misclassified only among a decreasing number of individuals, for example with decreasing rate $O(N^{-\delta})$, and the misclassification probability is fixed: $\sup_{i,j \in \mathcal{P}} E[|A^*_ij - A_{ij}|] = p$ for a constant $p \in [0, 1]$. Similar conditions have been considered in other papers studying mismeasured network; see e.g. Lewbel et al. (2019).³

This result is of practical interest, because it applies to many previous studies where the network misclassification may have been presented, while it was assumed not to be. The key assumption on the misclassification rate might be verified in some situations, for example, if an upper bound of the extent to which a certain form of the measurement error occurs is known and small.

Next, I show that the consistency of the nonparametric estimation for the CASF $m^*$ in Leung (2020b) is maintained, while its asymptotic convergence rate depends on the average network misclassification rate. For $\forall (d, s, z, n) \in \{0, 1\} \times \Omega_{S, Z, |N|}$, let

$$
\hat{m}(d, s, z, n) := \frac{\sum_{i=1}^{N} Y_i \hat{f}_{D_i, S_i, Z_i, |N_i|}(d, s, z, n)}{\sum_{i=1}^{N} \hat{f}_{D_i, S_i, Z_i, |N_i|}(d, s, z, n)}, \tag{D.1}
$$

where $\hat{f}_{D_i, S_i, Z_i, |N_i|}$ is the kernel density.

**Theorem D.2 (Consistency)** Let assumptions in Theorem D.1 and Assumption 5.1-5.3 hold. Suppose $m^*(d, s, z, n)$ is twice continuously differentiable in its continuous arguments. Then,

$$
\|\hat{m} - m^*\|_\infty = O_p \left( N^{-\delta} + h^2 + (NhQ)^{-1/2} \right).
$$

³The strict overlap assumption rules out the cluster randomized trial where groups are randomly assigned to be treated. It is likely to hold in situations where the network is sparse, and it might be violated in the presence of dense network. In the proof of Theorem D.1 (see Appendix), I also provide another sufficient assumption under which the result in Theorem D.1 still holds without relying on the strict overlap of the propensity score $e(d, s, z, n)$. I defer the large sample property of the estimator for $m_i$ in Section 5. The Lipschitz condition is satisfied by any bounded function $m^*$. 

82
It is worthy noticing that if all covariates are discrete, we can simply replace the kernel density \( \hat{f}_{D_i,S_i,Z_i|[N_i]}(d,s,z,n) \) in (D.1) by the indicator \( 1_i(d,s,z,n) := 1[D_i = d, S_i = s, Z_i = z, |N_i| = n], \) and the consistency still holds with the convergence rate \( N^{-\min(\delta, 1/2)} \). See more details in the proof of Theorem D.2.

D.1 Simulations for Nonparametric Estimation with Single Network Proxy

The data generating process is described in Section 6. Set the overall misclassification rate \( p^\omega = 0.6 \). We set the probabilities of false negative \( 1 - p^U = N^{-\delta} \) and false positive \( p^V = 50N^{-\delta - 1} \). Thus, the larger value of \( \delta \) is, the less misclassification of the network links. Notice that by setting \( p^V = 50N^{-\delta - 1} \), we let the misreported links maintain the similar sparsity to the actual network, which is common in many empirical applications. Similar design of the measurement errors is used in Lewbel et al. (2019) to study parametric estimation of the endogenous peer effects in a linear model. Table 10 reports, for different values of \( \delta \), the corresponding misclassification probabilities \( 1 - p^U \) and \( p^V \), average numbers of the observed degree and of the observed treated friends, and the average numbers of misclassified links. We can see that the total number of misclassified links decreases as \( \delta \) increasing. In addition, the ratio between the number of the misclassified links and the number of the actual links decreases with the sample size, varying from 295% to less than 1% when the network degree is relatively small \( (r_{deg} = 5) \), and varying from 197% to less than 1% when the network degree is relatively large \( (r_{deg} = 8) \).

Because covariate \( Z_i \) is binary, we proceed the nonparametric estimation of the CASF via

\[
\hat{m}(d, s, z, n) := \frac{\sum_{i=1}^{N} Y_i 1_i(d, s, z, n)}{\sum_{i=1}^{N} 1_i(d, s, z, n)},
\]

where recall that \( 1_i(d, s, z, n) := 1[D_i = d, S_i = s, Z_i = z, |N_i| = n] \). Table 11 presents the nonparametric estimation results, including bias, standard deviation (sd), mean squared error (mse), and the ratio between the mse of the feasible nonparametric estimation in (D.2) using the observed data \( \{Y_i, D_i, S_i, Z_i, |N_i|\}_{i=1}^{N} \) and the mse of the infeasible nonparametric estimation using the latent data \( \{Y_i, D_i, S_i^*, Z_i, |N_i^*|\}_{i=1}^{N} \) via replacing \( 1_i(d, s, z, n) \) in (D.2) by \( 1[D_i = d, S_i^* = s, Z_i = z, |N_i^*| = n] \).\(^4\) It is easy to observe several patterns from the results. First, for low misclassification rate \( (\delta \geq 0.5) \), the nonparametric estimates perform quite well as expected: as sample size increases, the bias, the sd and the mse decrease, and the ratio between two mean squared errors also decrease to one, in most cases. Second, when the misclassification rate is high \( (\delta < 0.5) \), the estimation bias become unstable across different sample sizes, and the ratio between mean squared errors is relatively high, implying biased and inaccurate estimation. Lastly, when \( r_{deg} \) increases from 5 to 8, i.e. the average degree increases, the sd becomes larger. It is intuitive because the larger average degree leads to less effective sample size for the nonparametric estimator in (D.2) at a given \( (d, s, z, n) \).

\(^4\)If the ratio between two mean squared errors is one, the feasible nonparametric estimation performs as good as the infeasible one. A larger (smaller) than one ratio means that the feasible nonparametric estimation produces larger (smaller) mean squared error than that of the infeasible estimation.
Table 10: Statistics of Misclassified Links \((p^* = 0.6)\)

**(a) \(r_{deg} = 5\)**

| \(N\) | \(\delta\) | \(1 - p^c\) (%) | \(p^c\) (%) | \(|N_1|\) avg. max | \(S_i\) avg. max | Misclassified links | 1 to 0 | 0 to 1 | total | ratio (%) |
|------|---------|-----------------|----------|----------------|----------------|-------------------|------|------|-------|----------|
| 1k   | 0.1     | 50.1            | 2.51     | 18.9           | 5.7            | 1698             | 14954| 16652| 295   |          |
|      | 0.3     | 12.6            | 0.63     | 9.0            | 2.7            | 427.1            | 3757 | 4184 | 74.2  |          |
|      | 0.5     | 3.16            | 0.16     | 6.5            | 1.9            | 106.5            | 943.6| 1050 | 18.6  |          |
|      | 0.7     | 0.79            | 0.04     | 5.9            | 1.8            | 26.8             | 236.1| 262.9| 4.66  |          |
|      | 0.9     | 0.20            | 0.01     | 5.7            | 1.7            | 6.812            | 59.56| 66.37| 1.18  |          |
| 2k   | 0.1     | 46.8            | 1.17     | 18.1           | 5.4            | 3218             | 27970| 31188| 272   |          |
|      | 0.3     | 10.2            | 0.26     | 8.4            | 2.5            | 702.6            | 6125 | 6828 | 59.6  |          |
|      | 0.5     | 2.24            | 0.06     | 6.3            | 1.9            | 153.4            | 1340 | 1494 | 13.0  |          |
|      | 0.7     | 0.49            | 0.01     | 5.9            | 1.8            | 33.7             | 292.5| 326.2| 2.85  |          |
|      | 0.9     | 0.11            | 0.00     | 5.8            | 1.7            | 8.222            | 70.61| 78.83| 0.27  |          |
| 5k   | 0.1     | 42.7            | 0.43     | 17.1           | 5.1            | 7433             | 63914| 71347| 245   |          |
|      | 0.3     | 7.77            | 0.08     | 7.9            | 2.4            | 1354             | 11646| 13000| 44.8  |          |
|      | 0.5     | 1.41            | 0.01     | 6.2            | 1.9            | 246.1            | 2119 | 2365 | 8.15  |          |
|      | 0.7     | 0.26            | 0.00     | 5.9            | 1.8            | 44.66            | 386.7| 431.3| 1.49  |          |
|      | 0.9     | 0.05            | 0.00     | 5.8            | 1.7            | 8.222            | 70.61| 78.83| 0.27  |          |

**(b) \(r_{deg} = 8\)**

| \(N\) | \(\delta\) | \(1 - p^c\) (%) | \(p^c\) (%) | \(|N_1|\) avg. max | \(S_i\) avg. max | Misclassified links | 1 to 0 | 0 to 1 | total | ratio (%) |
|------|---------|-----------------|----------|----------------|----------------|-------------------|------|------|-------|----------|
| 1k   | 0.1     | 50.1            | 2.51     | 18.9           | 5.7            | 2682             | 14904| 17587| 197   |          |
|      | 0.3     | 12.6            | 0.63     | 9.0            | 2.7            | 676.6            | 3744 | 4421 | 49.6  |          |
|      | 0.5     | 3.16            | 0.16     | 6.5            | 1.9            | 168.3            | 940.4| 1109 | 12.4  |          |
|      | 0.7     | 0.79            | 0.04     | 5.9            | 1.8            | 42.36            | 235.3| 277.7| 3.11  |          |
|      | 0.9     | 0.20            | 0.01     | 5.8            | 1.7            | 10.79            | 59.37| 70.17| 0.79  |          |
| 2k   | 0.1     | 46.8            | 1.17     | 20.5           | 6.1            | 5101             | 27923| 33025| 182   |          |
|      | 0.3     | 10.2            | 0.26     | 11.6           | 3.5            | 1114             | 6115 | 7229 | 39.8  |          |
|      | 0.5     | 2.24            | 0.06     | 9.6            | 2.9            | 243.2            | 1338 | 1581 | 8.70  |          |
|      | 0.7     | 0.49            | 0.01     | 9.2            | 2.8            | 53.34            | 292.0| 345.3| 1.90  |          |
|      | 0.9     | 0.11            | 0.00     | 9.1            | 2.7            | 11.81            | 63.83| 75.64| 0.42  |          |
| 5k   | 0.1     | 42.7            | 0.43     | 19.6           | 5.9            | 11827            | 63869| 75697| 164   |          |
|      | 0.3     | 7.77            | 0.08     | 11.1           | 3.3            | 2156             | 11638| 13794| 29.9  |          |
|      | 0.5     | 1.41            | 0.01     | 9.6            | 2.9            | 392.0            | 2117 | 2509 | 5.44  |          |
|      | 0.7     | 0.26            | 0.00     | 9.3            | 2.8            | 71.26            | 386.4| 457.6| 0.99  |          |
|      | 0.9     | 0.05            | 0.00     | 9.2            | 2.8            | 12.88            | 70.57| 83.44| 0.18  |          |

Note: All the statistics are obtained by averaging the 1000 replications; “1 to 0” is the total number of missing links (false negative); “0 to 1” lists the total number of misreported nonexistent links (false positive); “total” displays the total number of misclassified links including missing an existing link (1 to 0) and misreporting an nonexistent link (0 to 1); “ratio” is the ratio between the total number of misclassified links and the number of total links.
Table 11: Nonparametric Estimation of Treatment and Spillover Effects ($\rho^e = 0.6$)

(a) $r_{deg} = 5$

| $N$ | $\delta$ | $\tau_d(0,0,3)$ | $\tau_d(0,1,3)$ | $\tau_s(1,0,3)$ | $\tau_s(1,1,3)$ |
|-----|----------|-----------------|-----------------|-----------------|-----------------|
|     |          | bias | sd | mse | ratio | bias | sd | mse | ratio | bias | sd | mse | ratio |
| 0.1 |          | 0.04 | 1.04 | 1.08 | 2.91 | 0.09 | 1.39 | 1.94 | 2.06 | 0.03 | 0.70 | 0.50 | 2.72 |
| 0.3 | -0.04 | 0.99 | 0.99 | 2.67 | 0.09 | 1.39 | 1.95 | 2.07 | -0.15 | 0.75 | 0.59 | 3.23 |
| 1k  | 0.5     | -0.08 | 0.78 | 0.62 | 1.67 | -0.10 | 1.16 | 1.35 | 1.44 | -0.56 | 0.57 | 0.64 | 3.52 |
| 0.7 | -0.05  | 0.65 | 0.43 | 1.17 | -0.02 | 1.01 | 1.03 | 1.09 | -0.25 | 0.51 | 0.32 | 1.77 |
| 0.9 | -0.01 | 0.64 | 0.40 | 1.09 | 0.00 | 0.98 | 0.97 | 1.03 | -0.07 | 0.44 | 0.20 | 1.11 |
| 0.1 |          | 0.01 | 0.70 | 0.49 | 2.74 | 0.04 | 1.04 | 1.08 | 2.44 | 0.01 | 0.49 | 0.24 | 2.71 |
| 0.3 | -0.01  | 0.63 | 0.40 | 2.23 | 0.01 | 1.06 | 1.12 | 2.53 | -0.20 | 0.48 | 0.27 | 3.14 |
| 2k  | 0.5     | -0.02 | 0.49 | 0.24 | 1.37 | -0.05 | 0.80 | 0.65 | 1.46 | -0.44 | 0.38 | 0.34 | 3.84 |
| 0.7 | 0.02   | 0.43 | 0.19 | 1.05 | -0.01 | 0.69 | 0.48 | 1.08 | -0.12 | 0.31 | 0.11 | 1.31 |
| 0.9 | 0.01   | 0.41 | 0.17 | 0.96 | 0.01 | 0.66 | 0.44 | 0.99 | -0.03 | 0.32 | 0.10 | 1.15 |

(b) $r_{deg} = 8$

| $N$ | $\delta$ | $\tau_d(0,0,3)$ | $\tau_d(0,1,3)$ | $\tau_s(1,0,3)$ | $\tau_s(1,1,3)$ |
|-----|----------|-----------------|-----------------|-----------------|-----------------|
|     |          | bias | sd | mse | ratio | bias | sd | mse | ratio | bias | sd | mse | ratio |
| 0.1 |          | 0.07 | 1.60 | 2.55 | 1.78 | 0.03 | 1.53 | 2.33 | 0.96 | 0.07 | 1.42 | 2.02 | 1.25 |
| 0.3 | 0.05   | 1.57 | 2.46 | 1.71 | 0.32 | 1.78 | 3.27 | 1.35 | -0.01 | 1.39 | 1.93 | 2.25 |
| 1k  | 0.5     | -0.02 | 1.48 | 2.19 | 1.52 | -0.19 | 1.79 | 3.25 | 1.34 | -0.42 | 1.20 | 1.60 | 1.87 |
| 0.7 | 0.04   | 1.33 | 1.77 | 1.23 | 0.02 | 1.89 | 3.57 | 1.47 | -0.18 | 1.01 | 1.05 | 1.23 |
| 0.9 | -0.10  | 1.21 | 1.48 | 1.03 | 0.07 | 1.63 | 2.67 | 1.10 | -0.01 | 0.94 | 0.89 | 1.04 |
| 0.1 |          | 0.02 | 1.34 | 1.80 | 2.44 | 0.05 | 1.77 | 3.12 | 1.22 | -0.04 | 1.02 | 1.04 | 2.74 |
| 0.3 | -0.07  | 1.40 | 1.97 | 2.67 | 0.12 | 1.90 | 3.61 | 1.41 | -0.10 | 1.10 | 1.22 | 3.22 |
| 2k  | 0.5     | 0.01 | 1.10 | 1.20 | 1.63 | -0.06 | 1.76 | 3.10 | 1.21 | -0.32 | 0.80 | 0.74 | 1.96 |
| 0.7 | -0.01  | 0.91 | 0.83 | 1.13 | 0.04 | 1.55 | 2.41 | 0.94 | -0.08 | 0.66 | 0.45 | 1.18 |
| 0.9 | 0.04   | 0.87 | 0.76 | 1.03 | 0.03 | 1.54 | 2.39 | 0.93 | 0.01 | 0.62 | 0.39 | 1.03 |
| 0.1 |          | 0.01 | 0.95 | 0.90 | 2.88 | -0.19 | 1.63 | 2.71 | 1.71 | 0.02 | 0.66 | 0.44 | 2.58 |
| 0.3 | 0.00   | 0.93 | 0.87 | 2.78 | 0.01 | 1.78 | 3.16 | 1.99 | -0.23 | 0.67 | 0.50 | 2.94 |
| 5k  | 0.5     | -0.01 | 0.66 | 0.44 | 1.42 | -0.03 | 1.49 | 2.21 | 1.39 | -0.19 | 0.48 | 0.27 | 1.59 |
| 0.7 | 0.02   | 0.55 | 0.30 | 0.97 | -0.05 | 1.32 | 1.75 | 1.10 | -0.03 | 0.41 | 0.17 | 1.01 |
| 0.9 | 0.00   | 0.56 | 0.31 | 1.01 | 0.02 | 1.29 | 1.66 | 1.04 | -0.01 | 0.40 | 0.16 | 0.96 |

Note: The true values $\tau_d(0,0,3) = 1$, $\tau_d(0,1,3) = 2$, $\tau_s(1,0,3) = 3$ and $\tau_s(1,1,3) = 2.5$. Column “ratio” lists the ratio between the mse of the feasible estimation using the observed data $\{Y_i, D_i, S_i, Z_i, |N_i|^\}^N_{i=1}$ and the infeasible estimation using the latent data $\{Y_i, D_i, S_i^*, Z_i, |N_i^*|^\}^N_{i=1}$.
D.2 Proofs of Section D

**Proof of Theorem D.1.** We prove this theorem by three steps. Firstly, we show that the absolute difference between \( m_i(d, s, z, n) \) and \( m^*(d, s, z, n) \) is proportional to

\[
\Delta_A := E \left[ ||A_i^* - A_i||^2 \right] |D_i = d, S_i = s, Z_i = z, |N_i| = n].
\]

Secondly, we verify that condition (c) restricts the uniform convergence rate of \( \Delta_A \). At last, we present the alternative argument that relaxes the strict overlap condition of the propensity score.

**Step 1.** Recall that Proposition 3.1 demonstrates

\[
m_i(d, s, z, n) = E \left[ m^*(d, S_i^*, z, |N_i^*|) \right] |D_i = d, S_i = s, Z_i = z, |N_i| = n].
\]

In addition, we can get the following conditional expectation

\[
E \left[ m^*(d, S_i, z, |N_i|) \right] |D_i = d, S_i = s, Z_i = z, |N_i| = n]] = E \left[ m^*(d, s, z, n) \right] |D_i = d, S_i = s, Z_i = z, |N_i| = n]
\]

\[
= m^*(d, s, z, n).
\]

Based on (D.3) and (D.4),

\[
m_i(d, s, z, n) - m^*(d, s, z, n) = E \left[ m^*(d, S_i^*, z, |N_i^*|) - m^*(d, S_i, z, |N_i|) \right] |D_i = d, S_i = s, Z_i = z, |N_i| = n]] = E \left[ m^*(d, S_i, z, |N_i|) - m^*(d, s, z, n) \right] |D_i = d, S_i = s, Z_i = z, |N_i| = n]
\]

By the Lipschitz condition of \( m^* \), we know that

\[
|m^*(d, S_i^*, z, |N_i^*|) - m^*(d, S_i, z, |N_i|)| \leq |m^*(d, S_i^*, z, |N_i^*|) - m^*(d, S_i, z, |N_i|)| + |m^*(d, S_i, z, |N_i|) - m^*(d, S_i, z, |N_i|)|
\]

\[
= |S_i^* - S_i|^c + L_{\mathcal{N}} |N_i^* - |N_i||.
\]

For any generic set \( A \), let \( A^c \) be its complement. Then,

\[
|S_i^* - S_i| = \left| \sum_{D_j \in \mathcal{N}_i} D_j - \sum_{D_j \in \mathcal{N}_i} D_j \right| = \left| \sum_{D_j \in \mathcal{N}_i \cap (\mathcal{N}_i^*)^c} D_j - \sum_{D_j \in \mathcal{N}_i \cap (\mathcal{N}_i^*)^c} D_j \right|
\]

\[
\leq \left| \sum_{D_j \in \mathcal{N}_i \cap (\mathcal{N}_i^*)^c} D_j \right| + \left| \sum_{D_j \in \mathcal{N}_i \cap (\mathcal{N}_i^*)^c} D_j \right|
\]

\[
\leq |\mathcal{N}_i^* \cap \mathcal{N}_i^c| + |\mathcal{N}_i \cap (\mathcal{N}_i^*)^c|
\]

\[
= |(\mathcal{N}_i^* \cap \mathcal{N}_i^c) \cup (\mathcal{N}_i \cap (\mathcal{N}_i^*)^c)|,
\]

where the second inequality is because \( D_j \) can only take values in \{0, 1\}, and the last equality is due that sets \( \mathcal{N}_i^* \cap \mathcal{N}_i^c \) and \( \mathcal{N}_i \cap (\mathcal{N}_i^*)^c \) are mutually exclusive. Moreover, because the cardinality
of the set \( (N_i^* \cap N_i^c) \cup (N_i \cap (N_i^*)^c) \) is \( \sum_{j \in P} |A^*_{ij} - A_{ij}| \), we have that
\[
|S_i^* - S_i| \leq \sum_{j \in P} |A^*_{ij} - A_{ij}| = \sum_{j \in P} |A^*_{ij} - A_{ij}|^2 = \|A^*_i - A_i\|^2. \tag{D.7}
\]

In addition, by definition of \( |N_i^*| \) and \( |N_i| \) we know that
\[
|\|N_i^*\| - |N_i|\| = \left| \sum_{j \in P} A^*_{ij} - \sum_{j \in P} A_{ij} \right| \leq \sum_{j \in P} |A^*_{ij} - A_{ij}| = \|A^*_i - A_i\|^2. \tag{D.8}
\]

Substitute (D.6), (D.7) and (D.8) into (D.5), for some constant \( C > 0 \)
\[
|m_i(d, s, z, n) - m^*(d, s, z, n)|
\leq E \left[ L_S |S_i^* - S_i| + L_N |N_i^* - |N_i|| \left| D_i = d, S_i = s, Z_i = z, |N_i| = n \right. \right]
\leq CE \left[ \|A^*_i - A_i\|^2 \left| D_i = d, S_i = s, Z_i = z, |N_i| = n \right. \right]. \tag{D.9}
\]

**Step 2.** By the law of iterated expectation,
\[
E \left[ \|A^*_i - A_i\|^2 \left| Z_i = z, |N_i| = n \right. \right] = \sum_{(d, s) \in \{0, 1\} \times \Omega_S} E \left[ \|A^*_i - A_i\|^2 \left| D_i = d, S_i = s, Z_i = z, |N_i| = n \right. \right] f_{D_i, S_i|Z_i = z, |N_i| = n}(d, s). \tag{D.10}
\]

Besides, under the strict overlap, we have \( e(d, s, z, n) = f_{D_i, S_i|Z_i = z, |N_i| = n}(d, s) \) \( > \xi > 0 \) uniformly for all \( (d, s, z, n) \in \{0, 1\} \times \Omega_S, |N_i| \). Consequently, from (D.10),
\[
0 \leq \sum_{(d, s) \in \{0, 1\} \times \Omega_S} E \left[ \|A^*_i - A_i\|^2 \left| D_i = d, S_i = s, Z_i = z, |N_i| = n \right. \right]
\leq E \left[ \|A^*_i - A_i\|^2 \left| Z_i = z, |N_i| = n \right. \right],
\]
which, together with the condition (c), implies that
\[
\sup_{i \in P, (d, s, z, n) \in \{0, 1\} \times \Omega_S, |N_i|} E \left[ \|A^*_i - A_i\|^2 \left| D_i = d, S_i = s, Z_i = z, |N_i| = n \right. \right] = O(N^{-\delta}). \tag{D.11}
\]

Therefore, we can conclude from (D.9) that
\[
\sup_{i \in P, (d, s, z, n) \in \{0, 1\} \times \Omega_S, |N_i|} \left| m_i(d, s, z, n) - m^*(d, s, z, n) \right| = O(N^{-\delta}),
\]
which, by the definition of \( \|\cdot\|_\infty \), implies \( \|m_i - m^*\|_\infty = O(N^{-\delta}) \).

**Step 3.** Now, consider the following assumption.

**Assumption D.1 (Misclassification Rate)** There exists some constant \( \delta > 0 \) such that
\[
\sup_{i \in P, (d, s, z, n) \in \{0, 1\} \times \Omega_S, |N_i|} E \left[ \|A^*_i - A_i\|^2 \left| D_i = d, S_i = s, Z_i = z, |N_i| = n \right. \right] = O(N^{-\delta}).
\]

87
Assumption D.1 directly asserts the uniform convergence rate of the conditional probability of 
\[ \|A_i^* - A_i\|^2 \] given \( (D_i = d, S_i = s, Z_i = z, |N_i| = n) \). Then, under Assumption D.1 and (D.9), we can get the desired result without relying on the strict overlap of the propensity score.

**Proof of Theorem D.2.** Denote \( X_i = (D_i, S_i, Z_i, |N_i|) \) and \( x = (d, s, z, n) \). Define \( u_i = Y_i - E[Y_i|X_i] = Y_i - m_i(X_i) \). Without loss of generality, suppose \( Z_i \) is continuous. Let \( \hat{f}^{ker}(x) := 1[D_i = d, S_i = s, |N_i| = n]K(\frac{z - z_i}{h}) \) and \( \hat{f}_X(x) := 1/N \sum_{i=1}^N \hat{f}^{ker}(x) \). Then, by Theorem 5.2, we know that \( |\hat{f}_X(x) - f_X(x)| = o_p(1) \). To establish the consistency, we rewrite

\[
\hat{m}(x) - m^*(x) = \frac{1}{N} \sum_{i=1}^N [m_i(X_i) - m^*(x)] \hat{f}^{ker}(x) = \frac{\hat{M}_1(x) + \hat{M}_2(x)}{\hat{f}_X(x)},
\]

where

\[
\hat{M}_1(x) = \frac{1}{N} \sum_{i=1}^N [m_i(X_i) - m^*(x)] f^{ker}_i(x), \quad \hat{M}_2(x) = \frac{1}{N} \sum_{i=1}^N u_i f^{ker}_i(x).
\]

By notation abuse, let \( Q \) denote the dimension of \( Z_i \) and let \( \kappa(z/h) = \prod_{q=1}^Q \kappa(z_q/h) \), where \( z = (z_1, ..., z_Q) \in \mathbb{R}^Q \). In addition, since \( m^*(x) \) and \( f_X(x) \) are twice continuously differentiable in the argument \( z \), by Taylor expansion, for \( v \in \mathbb{R}^Q \)

\[
f_X(d, s, z + hv, n) = f_X(x) + h \frac{\partial f_X(x)}{\partial z} v + h^2 v' \frac{\partial^2 f_X(x)}{\partial z \partial z'} v, \tag{D.12}
\]

\[
m^*(d, s, z + hv, n) = m^*(x) + h \frac{\partial m^*(x)}{\partial z} v + h^2 v' \frac{\partial^2 m^*(x)}{\partial z \partial z'} v,
\]

where \( \tilde{z} \) is between \( z \) and \( z + hv \), and \( \tilde{x} = (d, s, \tilde{z}, n) \). Let \( x_1 = (d_1, s_1, z_1, n_1) \). Due that \( X_i \) is identically distributed,

\[
E \left[ \hat{M}_1(x) \right] = E \left\{ [m_i(X_i) - m^*(x)] \hat{f}^{ker}_i(x) \right\}
\]

\[
= \frac{1}{h^Q} \sum_{(d_1, s_1, n_1) \in \{0,1\} \times \Omega_{S_i|X_i}} \int [m_i(x_1) - m^*(x)] 1[d_1 = d, s_1 = s, n_1 = n] \kappa \left( \frac{z_1 - z}{h} \right) f_X(x_1) dz_1
\]

\[
= \frac{1}{h^Q} \int [m_i(d, s, z_1, n) - m^*(x)] \kappa \left( \frac{z_1 - z}{h} \right) f_X(d, s, z_1, n) dz_1
\]

\[
= \int [m_i(d, s, z + hv, n) - m^*(x)] \kappa(v) f_X(d, s, z + hv, n) dv
\]

\[
+ \int [m^*(d, s, z + hv, n) - m^*(x)] \kappa(v) f_X(d, s, z + hv, n) dv, \tag{D.13}
\]

where the first term on the right hand side of (D.13) can be bounded as below:

\[
\left| \int [m_i(d, s, z + hv, n) - m^*(d, s, z + hv, n)] \kappa(v) f_X(d, s, z + hv, n) dv \right|
\]
\[ \leq \int |m_i(d,s,z+h,v,n) - m^*(d,s,z+h,v,n)| \kappa(v) f_X(x,d,s,z+h,v,n)dv \]
\[ \leq \|m_i - m^*\|_\infty \left( f_X(x) \int \kappa(v) dv + O(h^2) \right) \]
\[ = O(N^{-\delta}), \quad (D.14) \]

with the second inequality follows from the expansion of \( f_X \) in (D.12), \( \int v\kappa(v)dv = 0 \) and the boundedness of the second derivative of \( f_X \) to \( z \), and the last equality is due to \( \int \kappa(v)dv = 1 \) and Theorem D.1. In addition, for the second term of (D.13), by the expansions in (D.12)

\[ \int \left[ m^*(d,s,z+h,v,n) - m^*(x) \right] \kappa(v) f_X(x,d,s,z+h,v,n)dv \]
\[ = h^2 \int \left[ v' \frac{\partial f_X(x)}{\partial z} \frac{\partial m^*(x)}{\partial z'} v + f_X(x)v' \frac{\partial^2 m^*(x)}{\partial z \partial z'} v \right] \kappa(v) dv + O(h^3) \]
\[ = O(h^2), \quad (D.15) \]

where the last line \( O(h^2) \) comes from the boundedness of \( f_X \) and its first order derivative (Assumption 5.2), the compactness of \( \Omega_X \) and the continuity of the first and second order derivatives of \( m^* \) in \( z \). Given (D.13), (D.14) and (D.15),

\[ E \left[ \hat{M}_1(x) \right] = O(N^{-\delta} + h^2). \quad (D.16) \]

Next, we tackle the variance of \( \hat{M}_1(x) \). For notation simplicity, let \( \hat{M}_{1,i}(x) := [m_i(X_i) - m^*(x)] \hat{f}_i^{ker}(x) \). Then, \( \hat{M}_1(x) = 1/N \sum_{i=1}^N \hat{M}_{1,i}(x) \) and by Assumption 5.1

\[ Var[\hat{M}_1(x)] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j\in \Delta(i,N)} \text{Cov} (\hat{M}_{1,i}(x), \hat{M}_{1,j}(x)) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j\notin \Delta(i,N)} \text{Cov} (\hat{M}_{1,i}(x), \hat{M}_{1,j}(x)) \]
\[ = \frac{1}{N^2} \sum_{i=1}^N \sum_{j\in \Delta(i,N)} \text{Cov} (\hat{M}_{1,i}(x), \hat{M}_{1,j}(x)) + \text{s.o.,} \]

where we use s.o. to denote the terms of smaller order. By Assumption 5.2, since \( 1/N \sum_{i=1}^N |\Delta(i,N)| \leq 1/N \sum_{i=1}^N |\Delta(i,N)|^2 = O(1) \), we can get

\[ \frac{1}{N^2} \sum_{i=1}^N \sum_{j\in \Delta(i,N)} \text{Cov} (\hat{M}_{1,i}(x), \hat{M}_{1,j}(x)) \leq \frac{1}{N^2} \sum_{i=1}^N \sum_{j\in \Delta(i,N)} \text{Var} [\hat{M}_{1,i}(x)] \]
\[ \leq \frac{1}{N} \text{Var} [\hat{M}_{1,i}(x)] \frac{1}{N} \sum_{i=1}^N |\Delta(i,N)| \]
\[ = \frac{1}{N} \text{Var} [\hat{M}_{1,i}(x)] \cdot O(1). \quad (D.17) \]
Moreover, by change of variables and simple algebra

\[
\frac{1}{N} \text{Var} \left[ \hat{M}_{1i}(x) \right] \leq \frac{1}{N} E \left[ [m_i(X_i) - m^*(x)]^2 (\hat{f}_i^{ker}(x))^2 \right] \\
= \frac{1}{NhQ} \int \left[ m_i(d, s, z_1, n) - m^*(x) \right]^2 \kappa^2 \left( \frac{z_1 - z}{h} \right) f_{X_i}(d, s, z_1, n)dz_1 \\
= \frac{1}{NhQ} \int \left[ m_i(d, s, z + hv, n) - m^*(x) \right]^2 \kappa^2 (v) f_{X_i}(d, s, z + hv, n)dv \\
= \frac{1}{NhQ} \int \left[ m_i(d, s, z + hv, n) - m^*(d, s, z + hv, n) \right]^2 \kappa^2 (v) f_{X_i}(d, s, z + hv, n)dv \\
+ \frac{1}{NhQ} \int \left[ m^*(d, s, z + hv, n) - m^*(x) \right]^2 \kappa^2 (v) f_{X_i}(d, s, z + hv, n)dv \\
+ \frac{2}{NhQ} \int \left[ m_i(d, s, z + hv, n) - m^*(d, s, z + hv, n) \right] \times \left[ m^*(d, s, z + hv, n) - m^*(x) \right] \kappa^2 (v) f_{X_i}(d, s, z + hv, n)dv \\
:= VM_1 + VM_2 + VM_3.
\]

Let us start from $VM_1$. By Assumption 5.2 that $\int \kappa^2(v)dv = K_2$ and Theorem D.1,

\[
VM_1 \leq \frac{1}{Nh^Q} \| m_i - m^* \|_\infty^2 \int \kappa^2(v)dv = O \left( (Nh^Q)^{-1}N^{-2\delta} \right). \tag{D.18}
\]

Next, for $VM_2$, based on (D.12) and the boundedness of the derivatives,

\[
VM_2 = \frac{1}{Nh^Q} \int \left[ h \frac{\partial m^*(x)}{\partial z'} v + h^2 v' \frac{\partial^2 m^*(x)}{\partial z' \partial z} v \right]^2 \left[ f_{X_i}(x) + h \frac{\partial f_{X_i}(x)}{\partial z'} v + h^2 v' \frac{\partial^2 f_{X_i}(x)}{\partial z' \partial z} v \right] \kappa^2 (v)dv \\
= \frac{h^2}{Nh^Q} f_{X_i}(x) \int v' \frac{\partial m^*(x)}{\partial z} \frac{\partial m^*(x)}{\partial z'} v \kappa^2 (v)dv + s.o. \\
= O \left( (Nh^Q)^{-1}h^2 \right). \tag{D.19}
\]

According to the Cauchy-Schwarz inequality, we can then get $VM_3 = O((Nh^Q)^{-1}N^{-\delta}h)$, which together with (D.18) and (D.19) implies that

\[
\text{Var} \left[ \hat{M}_1(x) \right] = O \left( \frac{(N^{-\delta} + h)^2}{Nh^Q} \right). \tag{D.20}
\]

Thus, based on (D.16), (D.20) and the fact that $Nh^Q \to \infty$ as $N \to \infty$ (Assumption 5.2),

\[
\hat{M}_1(x) = O_p \left( N^{-\delta} + h^2 + (N^{-\delta} + h)/(Nh^Q)^{1/2} \right) = O_p \left( N^{-\delta} + h^2 + h(Nh^Q)^{-1/2} \right). \tag{D.21}
\]

Next, let us deal with $\hat{M}_2(x)$. Observe that $E[\hat{M}_2(x)] = 0$ and by Assumption 5.1,

\[
E \left[ \hat{M}_2^2(x) \right] = \frac{1}{N^2} E \left[ \left( \sum_{i=1}^N u_i \hat{f}_{i}^{ker}(x) \right)^2 \right] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j \in \Delta(i,N)} \text{Cov} \left( u_i \hat{f}_{i}^{ker}(x), u_j \hat{f}_{j}^{ker}(x) \right) + s.o.
\]
where, similar to (D.17) and because \( u_i \) may not be identically distributed,

\[
\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} Cov \left( u_i f_{i}^{ker}(x), u_j f_{j}^{ker}(x) \right) \leq \frac{1}{N} \max_{i \in \mathcal{P}} Var \left[ u_i f_{i}^{ker}(x) \right] O(1).
\]

For all \( i \in \mathcal{P} \), let \( \sigma_i^2(X_i) := Var[u_i|X_i] \), where the subscript \( i \) is used to capture the possibly non-identical distribution of \( Y_i \) given \( X_i \). By the law of iterated expectation and (D.12),

\[
\frac{1}{N} Var \left[ u_i f_{i}^{ker}(x) \right] = \frac{1}{N h^2} E \left[ u_i^2 1[D_i = d, S_i = s, N_i = n] \kappa^2 \left( \frac{Z_i - z}{h} \right) \right]
\]

\[
= \frac{1}{N h^2} E \left[ \sigma_i^2(X_i) 1[D_i = d, S_i = s, N_i = n] \kappa^2 \left( \frac{Z_i - z}{h} \right) \right]
\]

\[
= \frac{1}{N h^2} \int \sigma_i^2(d, s, z, n) \kappa^2 \left( \frac{z - z_1}{h} \right) f_{X_i}(d, s, z, n) dz_1
\]

\[
= \frac{1}{N h^2} \int \sigma_i^2(d, s, z + h, n) \kappa^2(v) f_{X_i}(d, s, z + h, n) dv
\]

\[
\leq \frac{C}{Nh^2},
\]

(D.22)

where the last line is by Assumption 5.2 that \( E[Y_i^2|X_i = x] < \infty \) as \( \Omega_W \) is compact, and \( f_{X_i}(x) \) is bounded for \( \forall x \in \Omega_X \). Thus, according to (D.22), we know that \( \hat{M}_2(x) = O_p((Nh^Q)^{-1/2}) \), which together with (D.21) indicates that

\[
\hat{m}(x) - m^*(x) = \frac{\hat{M}_1(x) + \hat{M}_2(x)}{f_{X_i}(x) + o_p(1)} = O_p \left( N^{-\delta} + h^2 + (Nh^Q)^{-1/2} \right).
\]

Thus, the above discussion fulfills the proof of Theorem D.2.

In what follows, we consider the case when covariates in \( Z_i \) are all discrete. We want to show that replacing the kernel density in \( \hat{m}(x) \) by the indicator \( 1_i(x) = 1[X_i = x] \) gives a consistent estimator of \( m^*(x) \). Firstly, for the numerator, because of Assumption 5.1 we have

\[
Var \left[ \frac{1}{N} \sum_{i=1}^{N} Y_i 1_i(x) \right] = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} Cov \left( Y_i 1_i(x), Y_j 1_j(x) \right) + s.o.
\]

By the compactness of \( \Omega_W \), we know that \( Var[Y_i 1_i(x)] < \infty \) for any \( x \) and all \( i \in \mathcal{P} \),

\[
Var \left[ \frac{1}{N} \sum_{i=1}^{N} Y_i 1_i(x) \right] \leq \frac{C}{N^2} \sum_{i=1}^{N} |\Delta(i, N)| = O(N^{-1}),
\]

91
for $1/N \sum_{i=1}^N |\Delta(i, N)| = O(1)$. Then, by Chebyshev's inequality, for any $\epsilon > 0$

$$
Pr \left( \left| \frac{1}{N} \sum_{i=1}^N Y_i 1_i(x) - \frac{1}{N} \sum_{i=1}^N E[Y_i 1_i(x)] \right| > \epsilon \right) \leq \frac{\text{Var} \left[ \frac{1}{N} \sum_{i=1}^N Y_i 1_i(x) \right]}{\epsilon^2} = O(N^{-1}),
$$

which implies that

$$
\left| \frac{1}{N} \sum_{i=1}^N Y_i 1_i(x) - \frac{1}{N} \sum_{i=1}^N E[Y_i 1_i(x)] \right| = O_p(N^{-1/2}). \quad (D.23)
$$

Recall that $X_i$ is identically distributed. Then, Theorem D.1 leads to

$$
\left| \frac{1}{N} \sum_{i=1}^N E[Y_i 1_i(x)] - m^*(x) f_{X_i}(x) \right| = \left| \frac{1}{N} \sum_{i=1}^N E[Y_i | X_i = x] f_{X_i}(x) - m^*(x) f_{X_i}(x) \right|
\leq \frac{1}{N} \sum_{i=1}^N |m_i(x) - m^*(x)| f_{X_i}(x)
\leq O_p(N^{-\delta}). \quad (D.24)
$$

Given (D.23) and (D.24), we have

$$
\left| \frac{1}{N} \sum_{i=1}^N Y_i 1_i(x) - m^*(x) f_{X_i}(x) \right|
\leq \left| \frac{1}{N} \sum_{i=1}^N Y_i 1_i(x) - \frac{1}{N} \sum_{i=1}^N E[Y_i 1_i(x)] \right| + \left| \frac{1}{N} \sum_{i=1}^N E[Y_i 1_i(x)] - m^*(x) f_{X_i}(x) \right|
= O_p(N^{-1/2} + N^{-\delta}). \quad (D.25)
$$

Because $E[1_i(x)] = f_{X_i}(x)$, replacing $Y_i$ in (D.23) by a constant one,

$$
\left| \frac{1}{N} \sum_{i=1}^N 1_i(x) - f_{X_i}(x) \right| = O_p(N^{-1/2}). \quad (D.26)
$$

Let $\hat{\beta}(x) = \frac{1}{N} \sum_{i=1}^N Y_i 1_i(x)$, $\beta(x) = m^*(x) f_{X_i}(x)$, and $\hat{f}_{X_i}(x) = \frac{1}{N} \sum_{i=1}^N 1_i(x)$. Based on (D.25) and (D.26)

$$
\frac{\sum_{i=1}^N Y_i 1_i(x)}{\sum_{i=1}^N 1_i(x)} - m^*(x) = \frac{\hat{\beta}(x)}{\hat{f}_{X_i}(x)} - \frac{\beta(x)}{f_{X_i}(x)} = \frac{(\hat{\beta}(x) - \beta(x)) f_{X_i}(x) + \beta(x)(f_{X_i}(x) - \hat{f}_{X_i}(x))}{\hat{f}_{X_i}(x)f_{X_i}(x)}
\leq \frac{(\hat{\beta}(x) - \beta(x)) f_{X_i}(x) + \beta(x)(f_{X_i}(x) - \hat{f}_{X_i}(x))}{f_{X_i}^2(x) + o_p(1)}
= O_p(N^{-1/2} + N^{-\delta}).
$$
E Useful Lemmas

This section introduces some useful lemmas which are used in the proofs of Appendix Section B.

**Lemma E.1** Denote $\mathcal{H}$ as a set of measurable functions such that $|h| \leq 1$ for $\forall h \in \mathcal{H}$, and denote $\text{sign}(x) = 1[x \geq 0] – 1[x < 0]$ for any real value $x$. For any random variables $X$ and $Z$, a solution to $\max_{h \in \mathcal{H}} |E[Xh(Z)]|$ is $h(Z) = \text{sign}(E[X|Z])$, and $\max_{h \in \mathcal{H}} |E[Xh(Z)]| = E[X \text{sign}(X|Z)]$.

**Proof of Lemma E.1.** By the law of iterated expectation

$$|E[Xh(Z)]| = \left| \int E[X|Z]h(Z) dPr(Z) \right| \leq \int |E[X|Z]h(Z)| dPr(Z) \leq \int |E[X|Z]| dPr(Z).$$

Then, by $|E[X|Z]| = E[X|Z] \text{sign}(E[X|Z])$, it is clear that $h(Z) = \text{sign}(E[X|Z])$. ■

**Lemma E.2 (Uniform Law of Large Number under Dependency Neighborhood)** For any function $b : \Omega_W \times \Theta \mapsto \mathbb{R}^p$, if the following conditions hold

(i) $\Theta$ is compact;

(ii) $b(w; \theta)$ is continuous in $\theta$ over $\Theta$;

(iii) there exists a function $h(w)$ with $\|b(w; \theta)\| \leq h(w)$ for all $\theta \in \Theta$ and $E[h(W_i)] < \infty$;

(iv) for some constant $\eta > 0$, define

$$u(w; \theta, \eta) = \sup_{\theta' \in \Theta, \|\theta' - \theta\| < \eta} \|b(w; \theta') - b(w; \theta)\|,$$

$$\Sigma_N^b(\theta) = \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} \text{Cov}(b(W_i; \theta), b(W_j; \theta)),$$

$$\Sigma_N^u(\theta, \eta) = \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} \text{Cov}(u(W_i; \theta, \eta), u(W_j; \theta, \eta)).$$

(a) for all $\theta \in \Theta$ and any fixed $\eta$,

$$\left\| \sum_{i=1}^{N} \sum_{j \not\in \Delta(i,N)} \text{Cov}(b(W_i; \theta), b(W_j; \theta)) \right\| = o\left(\|\Sigma_N^b(\theta)\|\right),$$

$$\sum_{i=1}^{N} \sum_{j \not\in \Delta(i,N)} \text{Cov}(u(W_i; \theta, \eta), u(W_j; \theta, \eta)) = o\left(\Sigma_N^u(\theta, \eta)\right).$$

(b) $1/N \sum_{i=1}^{N} |\Delta(i, N)| = O(1)$; (c) $\sup_{\theta \in \Theta} E[\|b(W_i; \theta)\|^{2+\delta}] < C$ for some constants $\delta > 0$ and $C > 0$, and all $i$;

then $E[b(W_i; \theta)]$ is continuous in $\theta$ and $\sup_{\theta \in \Theta} \left\| \frac{1}{N} \sum_{i=1}^{N} \{b(W_i; \theta) - E[b(W_i; \theta)]\} \right\|_{L^p} \to 0$.

**Proof of Lemma E.2.** This proof is based on the proof of Lemma 1 in Tauchen (1985). Let $b_r(W_i; \theta)$ be the $r$-th element in vector $b(W_i; \theta)$, $r = 1, 2, ..., p$. Define a matrix $\Lambda_{ij}(\theta)$ such that its $rq$-th entry is $\text{corr}(b_r(W_i; \theta), b_q(W_j; \theta))$, $r, q = 1, 2, ..., p$. Denote a diagonal matrix $V_i(\theta) = \text{diag}(\text{Var}[b_1(W_i; \theta)], ..., \text{Var}[b_p(W_i; \theta)])$. 93
By condition (iv) (c), for all \(i\) and given \(\eta\), there exist constants \(C_1, C_2 > 0\) such that 
\[\sup_{\theta \in \Theta} \text{Var}[b_r(W_i; \theta)] < C_1\] for all \(r = 1, \ldots, p\), and \(\sup_{\theta \in \Theta} \text{Var}[u(W_i; \theta, \eta)] < C_2\). Then,

\[
\|\Sigma_N^b(\theta)\| \leq \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} \|\text{Cov}(b(W_i; \theta), b(W_j; \theta))\| \\
\leq \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} \|V_i(\theta)^{1/2} A_{ij}(\theta) V_j(\theta)^{1/2}\| \\
\leq C_1 p \sum_{i=1}^{N} |\Delta(i, N)| = O(N),
\]

where the last line follows from \(1/N \sum_{i=1}^{N} |\Delta(i, N)| = O(1)\) in condition (iv) (b). Similarly, \(\Sigma_N^u(\theta, \eta) = O(N)\). Applying Chebyshev’s inequality, we have that for any \(\epsilon > 0\)

\[
\text{Pr} \left( \left\| \frac{1}{N} \sum_{i=1}^{N} \{b(W_i; \theta) - E[b(W_i; \theta)]\} \right\| > \epsilon \right) \leq \frac{1}{\epsilon^2 N^2} E \left[ \left\| \sum_{i=1}^{N} \{b(W_i; \theta) - E[b(W_i; \theta)]\} \right\|^2 \right] \\
= \frac{1}{\epsilon^2 N^2} tr \left( \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} \text{Cov}(b(W_i; \theta), b(W_j; \theta)) + \sum_{i=1}^{N} \sum_{j \notin \Delta(i,N)} \text{Cov}(b(W_i; \theta), b(W_j; \theta)) \right) \\
= \frac{p}{\epsilon^2 N^2} \left( \|\Sigma_N^b(\theta)\| + \text{s.o.} \right) \\
= O \left( \frac{1}{\epsilon^2 N} \right),
\]

where the second equality comes from that \(tr(A) \leq p\|A\|_\infty \leq p\|A\|\) for any \(p \times p\) square matrix \(A\), and the third equality is due to condition (iv) (a). By choosing \(\epsilon\) such that \(\epsilon \to 0\) and \(\epsilon^2 N \to \infty\) as \(N \to \infty\), we can get

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \{b(W_i; \theta) - E[b(W_i; \theta)]\} \right\| = o_p(1).
\]

Similar arguments can be used to show that \(\frac{1}{N} \sum_{i=1}^{N} \{u(W_i; \theta, \eta) - E[u(W_i; \theta, \eta)]\} = o_p(1)\). By condition (ii) the continuity of \(b(w; \theta)\) in \(\theta\), we have that with fixed \(\theta\), \(\lim_{\eta \to 0} u(w; \theta, \eta) = 0\) as \(\eta \to 0\). Thus, by dominated convergence theorem, for any \(\epsilon > 0\), there exists a \(\bar{\eta}(\theta)\) such that

\[E[u(W_i; \theta, \eta)] \leq \epsilon, \text{ whenever } \eta \leq \bar{\eta}(\theta). \quad (E.1)\]

Let \(B(\theta)\) be an open ball of radius \(\bar{\eta}(\theta)\) about \(\theta\). Due to the compactness of \(\Theta\), there exist a finite sequence of open balls \(B_k := B(\theta_k)\) with \(k = 1, 2, \ldots, K\) such that \(\Theta \subset \bigcup_{k=1}^{K} B_k\). Let \(\eta_k = \bar{\eta}(\theta_k)\) and \(u_k = E[u(W_i; \theta_k, \eta_k)]\). By (E.1) and dominated convergence theorem, if \(\theta \in B_k\) then \(u_k \leq \epsilon\) and \(\|E[b(W_i; \theta)] - E[b(W_i; \theta')]\| \leq \epsilon\). Next, for \(\forall \theta \in \Theta\), there exists a \(k\) such that \(\theta \in B_k\), then

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} b(W_i; \theta) - E[b(W_i; \theta)] \right\| \leq \epsilon.
\]
\[
\leq \frac{1}{N} \sum_{i=1}^{N} \|b(W_i; \theta) - b(W_i; \theta_k)\| + \left\| \frac{1}{N} \sum_{i=1}^{N} b(W_i; \theta_k) - E[b(W_i; \theta)] \right\| \\
+ \left\| E[b(W_i; \theta)] - E[b(W_i; \theta)] \right\|
\leq \frac{1}{N} \sum_{i=1}^{N} u(W_i; \theta_k, \eta_k) - u_k + u_k + \left\| E[b(W_i; \theta_k)] - E[b(W_i; \theta)] \right\| + \epsilon
\leq 4\epsilon
\]

whenever \( N \geq \tilde{N}_k(\epsilon) \), by \( u_k \leq \epsilon \). Thus, whenever \( N \geq \max_k \tilde{N}_k(\epsilon) \), we have that

\[
\sup_{\theta \in \Theta} \left\| \frac{1}{N} \sum_{i=1}^{N} \{b(W_i; \theta) - E[b(W_i; \theta)]\} \right\| \leq 4\epsilon.
\]

\[\blacksquare\]

**Lemma E.3 (Theorem 3 of Bradley et al. (1983))** Suppose \( X \) and \( Y \) are random variables taking their values on a Borel space \( \Gamma \) and \( \mathbb{R} \), respectively. Suppose \( U \) is a uniform \([0, 1]\) random variable independent of \((X, Y)\). Suppose \( \mu \) and \( \gamma \) are positive numbers such that \( \mu \leq \|Y\|_{\gamma} < \infty \). Let \( \|Y\|_{\gamma} = (E[|Y|^\gamma])^{1/\gamma} \). Then there exists a real-valued random variable \( Y^* = g(X, Y, U) \) where \( g \) is a measurable function from \( \Gamma \times \mathbb{R} \times [0, 1] \) into \( \mathbb{R} \), such that

(i) \( Y^* \) is independent of \( X \);

(ii) the probability distributions of \( Y^* \) and \( Y \) are identical;

(iii) \( Pr(|Y^* - Y| \geq \mu) \leq 18(\|Y\|_{\gamma}/\mu)^{\gamma/(2\gamma+1)}[\alpha(\mathcal{B}(X), \mathcal{B}(Y))]^{2\gamma/(2\gamma+1)} \),

where for any two \( \sigma \)-fields \( \mathcal{B}_1, \mathcal{B}_2 \), \( \alpha(\mathcal{B}_1, \mathcal{B}_2) = \sup |Pr(\mathcal{B}_1 \cap \mathcal{B}_2) - Pr(\mathcal{B}_1)Pr(\mathcal{B}_2)| \).

The following lemmas are pioneered by Stein (1986) and utilized in e.g. Chen et al. (2010), Ross (2011) and Goldstein and Rinott (1996) among others, to derive central limit theorems for dependency graphs. We re-state them here such that the proofs are self-contained.

**Lemma E.4 (Meckes et al. (2009) Lemma 1)** Let \( Z \in \mathbb{R}^p \) be a standard normal random vector with mean zero and covariance matrix \( I_d \).

(i) If a function \( f : \mathbb{R}^p \mapsto \mathbb{R} \) is twice continuously differentiable with compact support, then

\[
E \left[ \text{tr} \left( \frac{d^2 f(Z)}{dzdz'} \right) - Z' \frac{df(Z)}{dz} \right] = 0.
\]

(ii) If a random vector \( X \in \mathbb{R}^p \) is such that

\[
E \left[ \text{tr} \left( \frac{d^2 f(X)}{dxdx'} \right) - X' \frac{df(X)}{dx} \right] = 0
\]

for every \( f \in C^2(\mathbb{R}^p) \) that is twice continuously differentiable with finite absolute mean value \( E[|\text{tr} (d^2 f(X)/dxdx') - X' df(X)/dx|] < \infty \), then \( X \) is a standard normal random vector.

95
Lemma E.5 (Goldstein and Rinott (1996) Lemma 3.1) Let $Z \in \mathbb{R}^p$ be a standard normal random vector and let $h : \mathbb{R}^p \mapsto \mathbb{R}$ have three bounded derivatives. Define $(T_u h)(x) = E[h(x e^{-u + \sqrt{1 - e^{-2u}} Z})]$ for $x \in \mathbb{R}^p$. Then $f(x) = - \int_0^\infty [T_u h(x) - E[h(Z)]] du$ solves

$$tr \left( \frac{d^2 f(x)}{dx dx'} \right) - x' \frac{df(x)}{dx} = h(x) - E[h(Z)].$$

In addition, for any $k$-th partial derivative we have that

$$\left| \frac{\partial^k f(x)}{\prod_{j=1}^k \partial x_j} \right| \leq \frac{1}{k} \sup_{x \in \Omega} \left\| \frac{d^2 h(x)}{dx dx'} \right\|_\infty.$$

Further, for any $\lambda \in \mathbb{R}^p$ and positive definite $p \times p$ matrix $\Sigma$, then $f^*$, denoted by the change of variable $f^*(x) := f(\Sigma^{-1/2}(x - \lambda))$ solves

$$tr \left( \Sigma \nabla^2 f^*(x) \right) - (x - \lambda)' \nabla f^*(x) = h(\Sigma^{-1/2}(x - \lambda)) - E[h(Z)],$$

and

$$\left| \frac{\partial^k f^*(x)}{\prod_{j=1}^k \partial x_j} \right| \leq \frac{p^k}{k} \left\| \Sigma^{-1/2} \right\|_\infty \left\| \nabla^k h \right\|_\infty.$$

The lemma below is based on Theorem 1.4 of Goldstein and Rinott (1996) which aims at providing a bound on the distance to normality for any sum of dependent random vectors whose dependence structure is formed via dependency neighborhoods.

Lemma E.6 (Multivariate CLT under Dependency Neighborhood) Let $\{W_i\}_{i=1}^N$ be random vectors in $\mathbb{R}^p$ with $E[W_i] = 0$ and $Z \in \mathbb{R}^p$ be a standard normal random vector. Denote

$$S_N = \sum_{i=1}^N W_i \quad \text{and} \quad \Sigma_N = \sum_{i=1}^N \sum_{j \in \Delta(i,N)} E[W_i W_j].$$

In addition, denote $S_i^c = \sum_{j \notin \Delta(i,N)} W_j$. Assume $\Sigma_N$ is symmetric positive definite. If the following conditions hold,

(i) there exists a finite, strictly positive-definite and symmetric $p \times p$ matrix $\Omega$ such that $\| \frac{1}{N} \Sigma_N - \Omega \| \rightarrow 0$ as $N \rightarrow \infty$;

(ii) (a) $\left\| \sum_{i=1}^N \sum_{j,k \in \Delta(i,N)} E \left[ \| \text{vec}(W_i W_j') W_k' \| \right] \right\|_\infty = o \left( \left\| \Sigma_N^{3/2} \right\|_\infty \right)$;

(b) $\left\| \sum_{i,k=1}^N \sum_{l \in \Delta(k,N)} \left[ (W_i W_j' - E[W_i W_j])' (W_k W_l' - E[W_k W_l']) \right] \right\|_\infty = o \left( \left\| \Sigma_N^2 \right\|_\infty \right)$;

(c) $\left\| \sum_{i=1}^N \sum_{j \notin \Delta(i,N)} \text{Cov} (W_i, W_j) \right\|_\infty = o \left( \left\| \Sigma_N \right\|_\infty \right)$;

(d) $E \left[ W_i S_i | S_i^c \right] \geq 0$ for all $i \in \mathcal{P}$;
then $\Sigma_N^{-1/2} S_N \xrightarrow{d} \mathcal{N}(0, I_p)$.

**Proof of Lemma E.6.** Denote $S_{i,q}$ as the $q$-th element of $S_i$. Let $h : \mathbb{R}^p \to \mathbb{R}$ be a function with bounded mixed partial derivatives up to order three. Denote $\nabla^k h$ the $k$-th derivative of $h$. Let $\nabla_r f^*(x) = \partial f^*(x)/\partial x_r$ and $\nabla_{rq}^2 f^*(x) = \partial^2 f^*(x)/\partial x_r \partial x_q$. It follows directly from the proof of Theorem 1.4 in Goldstein and Rinott (1996) that

$$
\left| E \left[ h \left( \Sigma_N^{-1/2} S_N \right) - E[h(Z)] \right] \right| 
\leq \frac{p^2}{2} \left\| \Sigma_N^{-1/2} \right\|_2^2 \left\| \nabla^2 h \right\|_\infty \sum_{r,q=1}^p \left( W_{i,r} W_{j,q} - E[W_{i,r} W_{j,q}] \right)^2 
+ \sum_{r=1}^p \sum_{i=1}^N E[W_{i,r} \nabla_r f^*(S_i^c)] 
+ \frac{p^3}{6} \left\| \Sigma_N^{-1/2} \right\|_3^3 \left\| \nabla^3 h \right\|_\infty \sum_{r,q,u=1}^p \sum_{i=1}^N E \left[ W_{i,r} \sum_{j \in \Delta(i,N)} W_{j,q} \sum_{k \in \Delta(i,N)} W_{k,u} \right],
$$

(E.3)

where $f^*$ is defined as in Lemma E.5. Consider the second term on the right hand side of (E.3)

$$
\left| \sum_{r=1}^p \sum_{i=1}^N E[W_{i,r} \nabla_r f^*(S_i^c)] \right| 
\leq \sum_{r=1}^p \sum_{i=1}^N E \left| W_{i,r} [\nabla_r f^*(S_i^c) - \nabla_r f^*(0)] \right| 
+ \sum_{r=1}^p \sum_{i=1}^N E \left| \nabla_r f^*(0) \right| 
= \sum_{r,q=1}^p \sum_{i=1}^N E \left[ W_{i,r} S_{i,q} r \nabla_{rq}^2 f^*(\tilde{S}_i^c) \right],
$$

(E.4)

where $\tilde{S}_i^c$ is between $S_i^c$ and 0 and the last equality comes from the mean value theorem and the fact that $E[W_{i,r}] = 0$. Without loss of generality, suppose there exists a function $\tilde{f}$ such that $\tilde{S}_i^c = \tilde{f}(S_i^c)$. Then, we can further bound (E.4) as below:

$$
\left| \sum_{r,q=1}^p \sum_{i=1}^N E \left[ W_{i,r} S_{i,q} r \nabla_{rq}^2 f^*(\tilde{S}_i^c) \right] \right| 
\leq \frac{p^2}{2} \left\| \Sigma_N^{-1/2} \right\|_2^2 \left\| \nabla^2 h \right\|_\infty \sum_{r,q=1}^p \sum_{i=1}^N \left\{ E[W_{i,r} S_{i,q} r \nabla_{rq}^2 f^*(\tilde{f}(S_i^c))] \right\},
$$

(E.5)

where the inequality is because of Lemma E.1 and $|\nabla_{rq}^2 f^* \circ \tilde{f}| \leq \frac{p^2}{2} \left\| \Sigma_N^{-1/2} \right\|_2^2 \left\| \nabla^2 h \right\|_\infty$ by Lemma E.5. Therefore, we have that

$$
\left| E \left[ h \left( \Sigma_N^{-1/2} S_N \right) - E[h(Z)] \right] \right| 
\leq \frac{p^2}{2} \left\| \Sigma_N^{-1/2} \right\|_2^2 \left\| \nabla^2 h \right\|_\infty \sum_{r,q=1}^p \sum_{i=1}^N \left\{ E[W_{i,r} S_{i,q} r \nabla_{rq}^2 f^*(\tilde{f}(S_i^c))] \right\},
$$

97
\[
\begin{align*}
\leq & \frac{p^2}{2} \left\| \sum_{i=1}^{N} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right\|_2 \\
+ & \frac{p^2}{2} \left\| \sum_{i=1}^{N} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right\|_2 \\
+ & \frac{p^3}{6} \left\| \sum_{i=1}^{N} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right\|_3 \\
& \left[ E \left[ \left( \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right)^2 \right] \right]^{1/2} \left( \sum_{r,q=1}^{p} \right)^{1/2} \\
= & pE \left[ \left\| \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right\|_2 \right]^{1/2},
\end{align*}
\]
for some constant \( p > 0 \). Let us start from the first term. By the Cauchy-Schwarz inequality

\[
\begin{align*}
\sum_{r,q=1}^{p} E \left[ \left( \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right)^2 \right] \leq & \left( \sum_{r,q=1}^{p} E \left[ \left( \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right)^2 \right] \right)^{1/2} \left( \sum_{r,q=1}^{p} \right)^{1/2} \\
= & pE \left[ \left\| \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right\|_2 \right]^{1/2},
\end{align*}
\]
where

\[
\begin{align*}
E \left[ \left\| \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right\|_2 \right] \\
= & E \left[ \left( \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right)^\prime \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right] \\
= & \left( \sum_{r,q=1}^{p} \right)^{1/2} \left( \sum_{r,q=1}^{p} \right)^{1/2} \\
\leq & p \left\| \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} (W_i \cdot W_{j,q} - E[W_i \cdot W_{j,q}]) \right\|_\infty. (E.7)
\end{align*}
\]

Besides, since \( E [W_i, W_j] S_i^c \geq 0 \) for all \( i = 1, ..., N \) and \( r = 1, ..., p \), the second term becomes to

\[
\sum_{r,q=1}^{p} \sum_{i=1}^{N} E \left[ W_i, S_i^c \right] = \sum_{r,q=1}^{p} \sum_{i=1}^{N} \sum_{j \notin \Delta(i,N)} Cov(W_i, W_{j,q}) \leq p^2 \left\| \sum_{i=1}^{N} \sum_{j \notin \Delta(i,N)} Cov(W_i, W_j) \right\|_\infty. (E.8)
\]

For the last term, we can obtain

\[
\sum_{r,q,u=1}^{p} \sum_{i=1}^{N} E \left[ W_i, \sum_{j \in \Delta(i,N)} W_{j,q} \sum_{k \in \Delta(i,N)} W_{k,u} \right] \leq \sum_{r,q,u=1}^{p} \sum_{i=1}^{N} \sum_{j,k \in \Delta(i,N)} E \left[ W_i, W_{j,q} W_{k,u} \right]
\]
Moreover, since $\|N^{-1}\Sigma_N - \Omega\| \to 0$, implying that there exist $\epsilon, \tau$ such that

$$0 < \epsilon \leq \frac{1}{N}\lambda_{\min}(\Sigma_N) \leq \frac{1}{N}\lambda_{\max}(\Sigma_N) < \tau < \infty.$$  

In addition, by the property of norm and the symmetry of $\Sigma_N$, we have that

$$\|\Sigma_{N}^{-1/2}\|_{\infty}^{2} \leq \|\Sigma_{N}^{-1/2}\|_{\infty}^{2} = tr(\Sigma_{N}^{-1}) = \sum_{r=1}^{p} \lambda_{r}^{-1}(\Sigma_N) \leq p\lambda_{\min}(\Sigma_N) = O(N^{-1}),$$

where $\lambda_{r}(\Sigma_N)$ means the $r$-th largest eigenvalue of matrix $\Sigma_N$. Similarly,

$$\|\Sigma_{N}\|_{\infty}^{2} = O(N^{2}), \quad \|\Sigma_{N}^{3/2}\|_{\infty}^{2} = O(N^{3}), \quad \|\Sigma_{N}^{2}\|_{\infty}^{2} = O(N^{4}).$$  

(E.10)

Now, plugging (E.7), (E.8) and (E.9) into (E.6) gives us

$$\left|E \left[ h \left( \Sigma_{N}^{-1/2}S_{N} \right) - E[h(Z)] \right] \right| \leq C \left\| \Sigma_{N}^{-1/2} \right\|_{\infty}^{2} \left\| \sum_{i,k=1}^{N} \sum_{j \in \Delta(i,N)} \sum_{l \in \Delta(k,N)} E \left[ (W_{i}W_{j}' - E[W_{i}W_{j}']) (W_{k}W_{l}' - E[W_{k}W_{l}']) \right] \right\|_{\infty}^{1/2}$$

$$+ C \left\| \Sigma_{N}^{-1/2} \right\|_{\infty}^{2} \left\| \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} \text{Cov}(W_{i},W_{j}) \right\|_{\infty} + C \left\| \Sigma_{N}^{-1/2} \right\|_{\infty}^{3} \left\| \sum_{i=1}^{N} \sum_{j,k \in \Delta(i,N)} E \left[ \|\text{vec}(W_{i}W_{j}')W_{k}\| \right] \right\|_{\infty}$$

$$= \left\| \Sigma_{N}^{-1/2} \right\|_{\infty}^{4} o \left( \left\| \Sigma_{N}^{2} \right\|_{\infty}^{1/2} + \|\Sigma_{N}\|_{\infty} \right) + \left\| \Sigma_{N}^{-1/2} \right\|_{\infty}^{3} o \left( \left\| \Sigma_{N}^{3/2} \right\|_{\infty} \right)$$

$$= o(1),$$

implying that $\Sigma_{N}^{-1/2}S_{N} \overset{d}{\to} N(0, I_{p})$. ■

In what follows, we first present several lemmas that will be used to show the asymptotic properties of the jacobian and hessian matrix of the objective function.

**Lemma E.7** Under Assumptions 5.4, 5.5 and the i.i.d. of $x_{i,j}$ across $i$ for any given $j = 1, 2, \ldots, K_{T}$, we have that

$$\sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{d^{2}m^{*}(x_{i,j}; \theta)}{d\theta d\theta'} \right\|^{2} = O_{p}(1); \quad \sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{dm^{*}(x_{i,j}; \theta)}{d\theta} \right\|^{2} = O_{p}(1);$$

and for $\tilde{\theta}_{N} \overset{p}{\to} \theta^{0}$,

$$\frac{1}{N} \sum_{i=1}^{N} \left\| m^{*}(x_{i,j}; \tilde{\theta}_{N}) - m^{*}(x_{i,j}; \theta^{0}) \right\|^{2} = o_{p}(1);$$

99
Proof of Lemma E.7. By Assumption 5.5 and the uniform convergence of i.i.d. samples (Lemma 2.4 of Newey and MacFadden (1994))

\[
\begin{align*}
&\sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{d^2 m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta d\theta'} \right\|^2 \\
&\leq \sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{d^2 m^*(x_{i,j}; \theta)}{d\theta d\theta'} \right\|^2 - E \left[ \left\| \frac{d^2 m^*(x_{i,j}; \theta)}{d\theta d\theta'} \right\|^2 \right] + \sup_{\theta \in \Theta} E \left[ \left\| \frac{d^2 m^*(x_{i,j}; \theta)}{d\theta d\theta'} \right\|^2 \right] \\
&= o_p(1) + \sup_{\theta \in \Theta} E \left[ \left\| \frac{d^2 m^*(x_{i,j}; \theta)}{d\theta d\theta'} \right\|^2 \right]. \quad (E.11)
\end{align*}
\]

Because \( \sup_{\theta \in \Theta} E \left[ \left\| \frac{d^2 m^*(x_{i,j}; \theta)}{d\theta d\theta'} \right\|^2 \right] \leq E[H_1(x_{i,j})] < \infty \) by Assumption 5.5, (E.11) becomes to

\[
\sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{d^2 m^*(x_{i,j}; \theta)}{d\theta d\theta'} \right\|^2 = O_p(1). \quad (E.12)
\]

Similar arguments can be used to show that \( \sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{dm^*(x_{i,j}; \theta)}{d\theta} \right\|^2 = O_p(1) \). Besides, the mean value theorem gives

\[
\begin{align*}
\frac{1}{N} \sum_{i=1}^{N} \left\| m^*(x_{i,j}; \tilde{\theta}_N) - m^*(x_{i,j}; \theta^0) \right\|^2 &= \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\partial m^*(x_{i,j}; \tilde{\theta}_N)}{\partial \theta} (\tilde{\theta}_N - \theta^0) \right\|^2 \\
&\leq \sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\partial m^*(x_{i,j}; \theta)}{\partial \theta} \right\|^2 \left\| \tilde{\theta}_N - \theta^0 \right\|^2 \\
&= o_p(1), \quad (E.13)
\end{align*}
\]

for \( \tilde{\theta}_N \) between \( \tilde{\theta}_N \) and \( \theta^0 \). Similarly, we can also obtain that

\[
\begin{align*}
\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\partial m^*(x_{i,j}; \tilde{\theta}_N)}{\partial \theta} - \frac{\partial m^*(x_{i,j}; \theta^0)}{\partial \theta} \right\|^2 &\leq \sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\partial m^*(x_{i,j}; \theta)}{\partial \theta} \right\|^2 \left\| \tilde{\theta}_N - \theta^0 \right\|^2 = o_p(1). \quad (E.14)
\end{align*}
\]

Moreover, since

\[
\frac{d^2 m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta_i d\theta_q} - \frac{d^2 m^*(x_{i,j}; \theta^0)}{d\theta_i d\theta_q} = \frac{\partial}{\partial \theta'} \left( \frac{d^2 m^*(x_{i,j}; \tilde{\theta}_N)}{d\theta_i d\theta_q} \right) (\tilde{\theta}_N - \theta^0),
\]
by the uniformly bounded third derivative of $m^*(x; \theta)$,

$$
\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{d^2 m^* (x_{i,j}; \hat{\theta}_N)}{d\theta d\theta'} - \frac{d^2 m^* (x_{i,j}; \theta^0)}{d\theta d\theta'} \right\|^2 \leq \frac{1}{N} \sum_{r,q=1}^{d_q} \sum_{i=1}^{N} \left\| \frac{d^2 m^* (x_{i,j}; \hat{\theta}_N)}{d\theta_r d\theta_q} - \frac{d^2 m^* (x_{i,j}; \theta^0)}{d\theta_r d\theta_q} \right\|^2 \\
\leq \frac{1}{N} \sum_{r,q=1}^{d_q} \sum_{i=1}^{N} \left\| \frac{\partial}{\partial \theta'} \left( \frac{d^2 m^* (x_{i,j}; \hat{\theta}_N)}{d\theta_r d\theta_q} \right) \right\|^2 \left\| \hat{\theta}_N - \theta^0 \right\|^2 \\
= O_p \left( \left\| \hat{\theta}_N - \theta^0 \right\|^2 \right) = o_p(1). \tag{E.15}
$$

Lemmas E.8 to E.10 show the key steps for establishing the asymptotics for the jacobian of the objective function. The proofs are based on Section 8 of Newey and MacFadden (1994) and extended to adopt data under dependency-neighborhoods structure.

**Lemma E.8 (Linearization)** Under assumptions in Lemma 5.5 (b), there exists a function $G(\cdot; \gamma) : \Omega_W \mapsto \mathbb{R}^{d_q}$ which is linear in $\gamma$ and satisfies

$$
\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ g(W_i; \theta^0, \phi_N) - g(W_i; \theta^0, \phi^0) - G(W_i; \hat{\gamma}_N - \gamma^0) \right] \right\| = o_p(1).
$$

**Proof of Lemma E.8.** Recall that $g(W_i; \theta, \phi) = \tau_i [Y_i - m(X_i; \theta, \phi)] \frac{\partial m(X_i; \theta, \phi)}{\partial \theta}$. Then,

$$
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(W_i; \theta^0, \phi_N) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(W_i; \theta^0, \phi^0) \\
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \left[ [Y_i - m(X_i; \theta^0, \phi_N)] \frac{\partial m(X_i; \theta^0, \phi_N)}{\partial \theta} - [Y_i - m(X_i; \theta^0, \phi^0)] \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta} \right], \tag{E.16}
$$

where making use of the identity $\hat{a}b - ab = (\hat{a} - a)b + a(\hat{b} - b) + (\hat{a} - a)(\hat{b} - b)$ leads to

$$
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(W_i; \theta^0, \phi_N) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(W_i; \theta^0, \phi^0) \\
= - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \left[ m(X_i; \theta^0, \hat{\phi}_N) - m(X_i; \theta^0, \phi^0) \right] \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta} \\
+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \left[ Y_i - m(X_i; \theta^0, \phi^0) \right] \left[ \frac{\partial m(X_i; \theta^0, \hat{\phi}_N)}{\partial \theta} - \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta} \right] \\
- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \left[ m(X_i; \theta^0, \hat{\phi}_N) - m(X_i; \theta^0, \phi^0) \right] \left[ \frac{\partial m(X_i; \theta^0, \hat{\phi}_N)}{\partial \theta} - \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta} \right] \\
= - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \sum_{j=1}^{K_T} m^*(x_{i,j}; \theta^0) \left[ \hat{f}_{T_i^j} | X_i(t_j) - f_{T_i^j} | X_i(t_j) \right] \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta}.
$$
+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \left[ Y_i - m(X_i; \theta^0, \phi^0) \right] \sum_{j=1}^{K_T} \frac{\partial m^*(x_{i,j}; \theta^0)}{\partial \theta} \left[ \hat{f}_{T^*_t|X}(t_j) - f_{T^*_t|X}(t_j) \right] \\
- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \sum_{j=1}^{K_T} m^*(x_{i,j}; \theta^0) \left[ \hat{f}_{T^*_t|X}(t_j) - f_{T^*_t|X}(t_j) \right] \sum_{j=1}^{K_T} \frac{\partial m^*(x_{i,j}; \theta^0)}{\partial \theta} \left[ \hat{f}_{T^*_t|X}(t_j) - f_{T^*_t|X}(t_j) \right] \\
:= G_1 + G_2 + G_3. \tag{E.17}

Firstly, consider $G_3$. By the Cauchy-Schwarz inequality, (B.66) and Lemma E.7,

\begin{align*}
\|N^{-1/2} G_3\| &\leq C \frac{N}{N} \sum_{i=1}^{N} \sum_{j=1}^{K_T} \left| m^*(x_{i,j}; \theta^0) \left[ \hat{f}_{T^*_t|X}(t_j) - f_{T^*_t|X}(t_j) \right] \right| \frac{\partial m^*(x_{i,i}; \theta^0)}{\partial \theta} \left[ \hat{f}_{T^*_t|X}(t_i) - f_{T^*_t|X}(t_i) \right] \\
&\leq \left( \sup_{\|\gamma_N - \gamma^0\|_\infty < \eta} \left| \phi_N - \phi^0 \right| \right) \frac{2}{N} \left[ \sum_{i=1}^{N} \sum_{j=1}^{K_T} \left| m^*(x_{i,j}; \theta^0) \right| \frac{\partial m^*(x_{i,i}; \theta^0)}{\partial \theta} \right] \\
&\leq \left( \sup_{\|\gamma_N - \gamma^0\|_\infty < \eta} \left| \phi_N - \phi^0 \right| \right) \frac{2}{N} \left[ \sum_{i=1}^{N} \left| m^*(x_{i,j}; \theta^0) \right|^2 \left[ \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\partial m^*(x_{i,i}; \theta^0)}{\partial \theta} \right| \right]^2 \right]^{1/2} \\
&= O_p \left( \|\gamma_N - \gamma^0\|_\infty^2 \right). \tag{E.18}
\end{align*}

Thus, given (E.18) we can get that $\|G_3\| = O_p(N^{1/2}\|\gamma_N - \gamma^0\|_\infty^2) = o_p(1)$ by Assumption 5.6.

Next, let us consider $G_1 + G_2$. Recall that the $1 \times K_T$ row vector $R(W_i; \theta, \phi)$ is defined as

\[
R(W_i; \theta, \phi) = \left[ Y_i - m(X_i; \theta, \phi) \right] m^*(x_{i,1}; \theta) \\
\vdots \\
\left[ Y_i - m(X_i; \theta, \phi) \right] m^*(x_{i,K_T}; \theta).
\]

Denote $\phi(t; \gamma_N) = [\hat{f}_{T^*_t|X}(t_1), ..., \hat{f}_{T^*_t|X}(t_{K_T})]'$ and $\phi(t; \gamma^0) = [f_{T^*_t|X}(t_1), ..., f_{T^*_t|X}(t_{K_T})]'$. Then, simple calculations yield that

\[
G_1 + G_2 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \left[ \sum_{j=1}^{K_T} \left[ Y_i - m(X_i; \theta^0, \phi^0) \right] \frac{\partial m^*(x_{i,j}; \theta^0)}{\partial \theta} \\
- m^*(x_{i,j}; \theta^0) \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta} \left[ \hat{f}_{T^*_t|X}(t_j) - f_{T^*_t|X}(t_j) \right] \right] \\
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \left[ \frac{\partial}{\partial \theta} R(W_i; \theta^0, \phi^0) \left( \phi(t; \gamma_N) - \phi(t; \gamma^0) \right) \right] \\
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \left[ \frac{\partial}{\partial \theta} R(W_i; \theta^0, \phi^0) \frac{\partial \phi(t; \gamma^0)}{\partial \gamma} (\gamma_N - \gamma^0) \right] + G_R, \tag{E.19}
\]
where the reminder term

$$G_R := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \frac{\partial}{\partial \theta} R(W_i; \theta^0, \phi^0) \left[ \phi(t; \gamma_N) - \phi(t; \gamma^0) - \frac{\partial \phi(t; \gamma^0)}{\partial \gamma'} (\gamma_N - \gamma^0) \right]$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \left[ \nabla R_1 + \nabla R_2 \right] \left[ \phi(t; \gamma_N) - \phi(t; \gamma^0) - \frac{\partial \phi(t; \gamma^0)}{\partial \gamma'} (\gamma_N - \gamma^0) \right],$$

with $\frac{\partial}{\partial \theta} R(W_i; \theta^0, \phi^0) := \nabla R_1 + \nabla R_2$ and

$$\nabla R_1 = \left[ Y_i - m(X_i; \theta^0, \phi^0) \right] \left[ \frac{\partial m^*(x_i; \theta^0)}{\partial \theta} \ldots \frac{\partial m^*(x_i; \theta^0)}{\partial \theta} \right],$$

$$\nabla R_2 = - \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta} \left[ m^*(x_{i,1}; \theta^0) \ldots m^*(x_{i,K_T}; \theta^0) \right].$$

Next, we show that $G_R = o_p(1)$. Due to Theorem 5.2, we can focus on a small neighborhood of $\gamma^0$ and bound the reminder term as follows:

$$\|N^{-1/2}G_R\| \leq \sup_{\|\gamma_N - \gamma^0\|_\infty < \eta} \left\| (\hat{\phi}_N - \phi^0) - \frac{\partial \phi(\gamma^0)}{\partial \gamma'} (\gamma_N - \gamma^0) \right\| \frac{1}{N} \sum_{i=1}^{N} \tau_i \|\nabla R_1 + \nabla R_2\|$$

$$\leq O_p \left( \|\hat{\gamma}_N - \gamma^0\|_\infty^2 \right) \left[ \frac{1}{N} \sum_{i=1}^{N} \tau_i \|\nabla R_1\| + \frac{1}{N} \sum_{i=1}^{N} \tau_i \|\nabla R_2\| \right],$$

where the $O_p(\|\hat{\gamma}_N - \gamma^0\|_\infty^2)$ is due to (B.63), and applying the Cauchy-Schwarz inequality to each of the term inside the bracket leads to

$$\frac{1}{N} \sum_{i=1}^{N} \tau_i \|\nabla R_1\| \leq \frac{1}{N} \sum_{i=1}^{N} \tau_i \left| Y_i - m(X_i; \theta^0, \phi^0) \right| \left\| \left[ \frac{\partial m^*(x_{i,1}; \theta^0)}{\partial \theta} \ldots \frac{\partial m^*(x_{i,1}; \theta^0)}{\partial \theta} \right] \right\|$$

$$\leq C \left[ \frac{1}{N} \sum_{i=1}^{N} \tau_i \left| Y_i - m(X_i; \theta^0, \phi^0) \right|^2 \right]^{1/2} \left[ \frac{1}{N} \sum_{j=1}^{K_T} \sum_{i=1}^{N} \left\| \frac{\partial m^*(x_{i,j}; \theta^0)}{\partial \theta} \right\|^2 \right]^{1/2}$$

$$= O_p(1),$$

where the last line follows from (B.67) and Lemma E.7. Similarly, from the Cauchy-Schwarz inequality and Lemma E.7, we can also get that

$$\frac{1}{N} \sum_{i=1}^{N} \tau_i \|\nabla R_2\| \leq C \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta} \right\| \left\| \left[ m^*(x_{i,1}; \theta^0) \ldots m^*(x_{i,K_T}; \theta^0) \right] \right\|$$

$$\leq C \left[ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\partial m(X_i; \theta^0, \phi^0)}{\partial \theta} \right\|^2 \right]^{1/2} \left[ \frac{1}{N} \sum_{j=1}^{K_T} \sum_{i=1}^{N} m^*(x_{i,j}; \theta^0)^2 \right]^{1/2}$$

$$= O_p(1),$$

103
Therefore, it yields from the above results and Assumption 5.6 that
\[ \|G_R\| = O_p \left( N^{1/2} \|\hat{\gamma}_N - \gamma^0\|_\infty^2 \right) = o_p(1). \] (E.20)

To fulfill this proof and find the function \( G \), let \( \tilde{\nu}(W_i) := \tau_i \left[ \frac{\partial}{\partial \theta}(W_i; \theta^0, \phi^0) \frac{\partial \phi(t; \gamma^0)}{\partial \gamma} \right] \) and \( G(W_i; \gamma) = \tilde{\nu}(W_i)\gamma = \tau_i \left[ \frac{\partial}{\partial \theta}(W_i; \theta^0, \phi^0) \frac{\partial \phi(t; \gamma^0)}{\partial \gamma} \right] \gamma \), then by construction \( G(W_i; \gamma) \) is linear in \( \gamma \). Moreover, based on (E.18) and (E.20),
\[ \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ g(W_i; \theta^0, \hat{\phi}_N) - g(W_i; \theta^0, \phi^0) - G(W_i; \hat{\gamma}_N - \gamma^0) \right] \right\| \leq \|G_3\| + \|G_R\| = o_p(1). \] (E.21)

Lemma E.9 (Stochastic Equicontinuity) Let \( F_W(w) \) be the true probability distribution function of \( W_i \). Suppose assumptions in Lemma 5.5 (b) hold, then
\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ G(W_i; \hat{\gamma}_N - \gamma^0) - \int G(w; \hat{\gamma}_N - \gamma^0) dF_W(w) \right] = o_p(1). \]

Proof of Lemma E.9. By the linearity of \( G(w; \gamma) = \tilde{\nu}(w)\gamma \) in \( \gamma \), we can get
\[ \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ G(W_i; \hat{\gamma}_N - \gamma^0) - \int G(w; \hat{\gamma}_N - \gamma^0) dF_W(w) \right] \right\| = \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \tilde{\nu}(W_i) - E[\tilde{\nu}(W_i)] \right] (\hat{\gamma}_N - \gamma^0) \right\| \leq C \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \tilde{\nu}(W_i) - E[\tilde{\nu}(W_i)] \right] \right\| \|\hat{\gamma}_N - \gamma^0\|_\infty. \] (E.22)

Denote \( \tilde{\nu}_r(W_i) \) as the \( r \)-th entry of the vector \( \tilde{\nu}(W_i) \). Then,
\[ E \left[ \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \tilde{\nu}(W_i) - E[\tilde{\nu}(W_i)] \right] \right\|^2 \right] = \frac{1}{N} E \left[ \sum_{i=1}^{N} \left( \tilde{\nu}(W_i) - E[\tilde{\nu}(W_i)] \right) \right] \left[ \sum_{i=1}^{N} \left( \tilde{\nu}(W_i) - E[\tilde{\nu}(W_i)] \right) \right]\]
\[ = \frac{1}{N} \sum_{r=1}^{d_0} \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} Cov \left( \tilde{\nu}_r(W_i), \tilde{\nu}_r(W_j) \right) + \sum_{i=1}^{N} \sum_{j \notin \Delta(i,N)} Cov \left( \tilde{\nu}_r(W_i), \tilde{\nu}_r(W_j) \right) \]
\[ = \frac{1}{N} \sum_{r=1}^{d_0} \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} Cov \left( \tilde{\nu}_r(W_i), \tilde{\nu}_r(W_j) \right) + s.o., \] (E.23)
where the last line comes from Assumption 5.1. Note that due to \( \text{Var}[\tilde{\nu}_r(W_i)] < \infty \) as in Assumption 5.6 and \( 1/N \sum_{i=1}^{N} |\Delta(i, N)| = O(1), \)
\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{j \in \Delta(i, N)} \text{Cov} \left( \tilde{\nu}_r(W_i), \tilde{\nu}_r(W_j) \right) \leq \frac{C}{N} \sum_{i=1}^{N} |\Delta(i, N)| = O(1). \tag{E.24}
\]
Given (E.24), together with the consistency \( \|\hat{\gamma}_N - \gamma^0\|_{\infty} = o_p(1) \), we know that
\[
\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ G(W_i; \hat{\gamma}_N - \gamma^0) - \int G(w; \hat{\gamma}_N - \gamma^0) dF_W(w) \right] \right\| = o_p(1). \tag{E.25}
\]

\[\text{Lemma E.10 (Mean-square Differentiability)}\] Under assumptions in Lemma 5.5 (b), there exists a function \( \delta : \Omega_W \mapsto \mathbb{R}^{d_{\phi}} \) such that
\[
\int G(w; \hat{\gamma}_N - \gamma) dF_W(w) = \int \delta(w) d\hat{F}_W(w),
\]
\[
\sqrt{N} E \left[ \left\| \int \delta(w) d\hat{F}_W(w) - \int \delta(w) d\hat{F}_W(w) \right\| \right] = o(1),
\]
where \( \hat{F}_W(w) \) is the kernel estimator of \( F_W(w) \) and \( \hat{F}_W(w) := 1/N \sum_{i=1}^{N} 1[W_i \leq w] \) is the empirical distribution of \( W_i \).

\[\text{Proof of Lemma E.10}.\] Following the derivations of Theorem 8.1 or (Theorem 8.11) in Newey and MacFadden (1994), it is apparent from the linearity of \( G(w; \gamma) \) in \( \gamma \) that and the law of iterated expectation,
\[
\int G(w; \gamma) dF_W(w) = \int \nu(w) \gamma dw,
\]
where recall that the \( d_{\phi} \times d_{\gamma} \) matrix \( \nu(w) \) is defined as \( \nu(w) = E \left[ \tau(X_i) \frac{\partial}{\partial \theta} \mathcal{R}(W_i; \theta^0, \phi^0) \frac{\partial \phi(t, \gamma)}{\partial \gamma} \big|_{\gamma = \gamma(w)} 1_{W_i \leq w} \right] \).
In addition, let \( \delta(w) := \nu(w) - E[\nu(w)] \), we have
\[
\int G(w; \hat{\gamma}_N - \gamma) dF_W(w) = \int \delta(w) d\hat{F}_W(w),
\]
with \( \hat{F}_W(w) \) being the kernel estimator of the distribution of \( W_i \).

At last, recall the empirical distribution \( \hat{F}_W(w) = 1/N \sum_{i=1}^{N} 1[W_i \leq w] \). By an abuse of notation, we denote \( \kappa(\tilde{w}_r - \tilde{w}_r^c) := \prod_{q=1}^{Q} \kappa(\tilde{w}_r - \tilde{w}_r^c) \). Consider the difference between the two integrals \( \delta(F) \) defined as below, which can be interpreted as a smoothing bias term,
\[
\delta(F) := \int \delta(w) d\hat{F}_W(w) - \int \delta(w) d\hat{F}_W(w)
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} \left[ \sum_{w^d \in \Omega_W} \nu(w) \tilde{\nu}_r^c(w) dw^c - \nu(W_i) \right]
\]
\[ \frac{1}{N} \sum_{i=1}^{N} \left[ \sum_{w_i \in \Omega_w} \int \nu(w) \frac{1}{h^Q} 1[W_i = w_i^d] \prod_{q=1}^{Q} \kappa \left( \frac{w_q^c - W_q^c}{h} \right) dw^c - \nu(W_i) \right] \]
\[ = \frac{1}{N} \sum_{i=1}^{N} \int \nu(w^c, W_i^d) \frac{1}{h^Q} \prod_{q=1}^{Q} \kappa \left( \frac{w_q^c - W_q^c}{h} \right) dw^c - \nu(W_i) \]
\[ = \frac{1}{N} \sum_{i=1}^{N} \int \left[ \nu(W_i^c + hv, W_i^d) - \nu(W_i) \right] \prod_{q=1}^{Q} \kappa(v_q) dv \]
\[ := \frac{1}{N} \sum_{i=1}^{N} \delta_i(F). \quad (E.26) \]

Because the identical distribution of \( W_i \) across \( i \), it follows from (E.26) that
\[ \sqrt{N} E[\delta(F)] = \sqrt{N} E[\delta_i(F)] \]
\[ = \sqrt{N} E \left[ \int \nu(W_i^c + hv, W_i^d) \prod_{q=1}^{Q} \kappa(v_q) dv - \nu(W_i) \right] \]
\[ = \sqrt{N} \int \int \nu(\tilde{w}^c + hv, \tilde{w}^d) \prod_{q=1}^{Q} \kappa(v_q) dv dF_W(\tilde{w}) - \sqrt{N} \int \nu(w) dF_W(w) \]
\[ = \sqrt{N} \int \int \nu(\tilde{w}^c, \tilde{w}^d) \prod_{q=1}^{Q} \kappa(v_q) dv dF_W(\tilde{w} - hv, \tilde{w}^d) - \sqrt{N} \int \nu(w) dF_W(w) \]
\[ = \sqrt{N} \left\{ \int \int \nu(\tilde{w}^c, \tilde{w}^d) \prod_{q=1}^{Q} \kappa(v_q) dv dF_W(\tilde{w} - hv, \tilde{w}^d) - \int \int \nu(w) \prod_{q=1}^{Q} \kappa(v_q) dv dF_W(w) \right\} \]
\[ = \sqrt{N} \left\{ \sum_{w_i \in \Omega_w} \int \nu(w) \left[ f_{W_i^c, W_i^d}(w^c - hv, w^d) - f_{W_i^c, W_i^d}(w^c, w^d) \right] \prod_{q=1}^{Q} \kappa(v_q) dv \right\}, \quad (E.27) \]

which together with (B.39) and Assumption 5.6, implies that
\[ \sqrt{N} \| E[\delta(F)] \| \]
\[ \leq \sqrt{N} \sum_{w_i \in \Omega_w} \int \| \nu(w) \| \left\| \int \left[ f_{W_i^c, W_i^d}(w^c - hv, w^d) - f_{W_i^c, W_i^d}(w^c, w^d) \right] \prod_{q=1}^{Q} \kappa(v_q) dv \right\| dw^c \]
\[ \leq C \sqrt{N} h^2 \sum_{w_i \in \Omega_w} \int \| \nu(w) \| dw^c \]
\[ = o(1). \quad (E.28) \]
Next, let \( \delta(F) = (\delta_1(F), \ldots, \delta_d(F))' \) with \( \delta_r(F) = 1/N \sum_{i=1}^{N} \delta_{r,i}(F) \) and consider

\[
E \left[ \left\| \sqrt{N} \delta(F) - \sqrt{N} E[\delta(F)] \right\|^2 \right] = \sum_{r=1}^{d} E \left[ \left\| \sqrt{N} \delta_r(F) - \sqrt{N} E[\delta_r(F)] \right\|^2 \right] 
= N \sum_{r=1}^{d} E \left[ \left\| \frac{1}{N} \sum_{i=1}^{N} (\delta_{r,i}(F) - E[\delta_{r,i}(F)]) \right\|^2 \right] 
= \frac{1}{N} \sum_{r=1}^{d} \sum_{i=1}^{N} \sum_{j \in \Delta(i,N)} Cov(\delta_{r,i}(F), \delta_{r,j}(F)) + s.o., \tag{E.29}
\]

where the last line follows from Assumption 5.1. Due to the identical distribution of \( W_i \) and (E.26), we can bound the covariance in (E.29) by

\[
|Cov(\delta_{r,i}(F), \delta_{r,j}(F))| \leq Var[\delta_{r,i}(F)] \leq E[|\delta_{r,i}(F)|^2] 
= E \left[ \left( \int \nu_r(W_i^c + hv, W_i^d) - \nu_r(W_i) \prod_{q=1}^{Q} \kappa(v_q) dv \right)^2 \right]. \tag{E.30}
\]

From Assumption 5.2 we know that \( \int x \kappa(x) dx = 0 \) and \( \int x^2 \kappa(x) dx = K_2 \), and Assumption 5.6 that \( \nu(w) \) is twice continuously differentiable in \( w^c \). Expanding \( \nu_r(W_i^c + hv, W_i^d) \) around \( W_i^c \), then there exists a constant \( C > 0 \) such that

\[
|Cov(\delta_{r,i}(F), \delta_{r,j}(F))| \leq h^4 E \left[ \left( \int \nu' \frac{\partial \nu_r(W_i^c + w^*, W_i^d)}{\partial w^c} v \prod_{q=1}^{Q} \kappa(v_q) dv \right)^2 \right] \leq Ch^4. \tag{E.31}
\]

Substituting (E.31) into (E.29), since \( 1/N \sum_{i=1}^{N} |\Delta(i,N)| = O(1) \) as in Assumption 5.2,

\[
E \left[ \left\| \sqrt{N} \delta(F) - \sqrt{N} E[\delta(F)] \right\|^2 \right] = O(h^4) = o(1). \tag{E.32}
\]

Based on (E.28) and (E.32), since both the mean and variance of \( \sqrt{N} \delta(F) \) are \( o(1) \), by Chebyshev’s inequality, it follows directly that \( E \left[ \left\| \sqrt{N} \delta(F) \right\|^p \right] \to 0. \]