Regression Adjustments for Disentangling Spillover Effects*

David M. Ritzwoller

Stanford GSB

Job Market Paper

Link to latest version

ABSTRACT. Empirical analyses that characterize the mechanisms that mediate spillover effects often do so by relating responses to a treatment, shock, or policy change with a measure of economic proximity, such as geographic distance, technological similarity, trade costs, or migration flows. Typically, such efforts are based on regressions that associate outcomes with proximity-weighted averages of the treatments received by other units. We show that regressions with this structure measure how the association between one unit's outcome and another unit's treatment correlates with the proximity between the two units. We then argue that, if the proximity measure of interest is associated with other channels that mediate spillover effects, causal interpretations of such relationships are susceptible to confounding. For instance, if technologically similar firms tend to be geographically proximate, then a positive association between technological similarity and the intensity of the productivity spillovers between firms might arise spuriously. We give conditions under which the effect of a proximity measure on spillover intensity can instead be recovered by regressing outcomes on averages of other units' treatments, reweighted by residualized versions of the proximity measure under consideration. We show that estimates obtained in this manner achieve the optimal rate of convergence and give a resampling procedure for constructing estimates of the associated uncertainty. We illustrate the proposed method with several empirical applications.

Keywords: Spillover Effects, Indirect Effects, Network Interference, Mediation

JEL: C21, C14, C15

Date: November 15, 2025

*Email: ritzwoll@stanford.edu.

I thank my advisors Guido Imbens and Joseph Romano for their support and guidance. I thank Nicholas Bloom, Jiafeng Chen, Stefano DellaVigna, Han Hong, Woojin Kim, Evan Munro, Kyle Myers, Stephen Redding, Brad Ross, Jaume Vives-i-Bastida, and Heidi Williams for helpful comments and conversations. I gratefully acknowledge funding from the National Science Foundation Graduate Research Fellowship, the Stanford GSB Academic Career Advancement Fellowship, and the Sloan Foundation Graduate Fellowship on the Fiscal and Economic Effects of Innovation and Productivity Policies, awarded through the National Bureau of Economic Research. Computational support was provided by the Data, Analytics, and Research Computing (DARC) group at the Stanford Graduate School of Business (RRI:SCR_022938).

1. INTRODUCTION

The indirect effects of economic actions are often multifaceted. For instance, one firm's investment in research and development may influence another firm's value through both product market competition and communication among researchers (Arrow, 1962; Dasgupta and Stiglitz, 1980; Griliches, 1992; Bloom et al., 2013). Likewise, productivity changes in one city may affect output in others through the exchange of goods, labor, and ideas (Marshall, 1890; Ellison et al., 2010; Allen and Arkolakis, 2014; Monte et al., 2018).

As a result, aggregate estimates of indirect effects can conflate distinct and potentially countervailing mechanisms. A shock to one location's natural resource endowment, for example, might benefit locations closely connected through labor markets, while harming locations linked through trade in intermediates, as factor reallocation toward the natural-resource sector crowds out manufacturing (i.e., "Dutch Disease"; Krugman, 1987; Allcott and Keniston, 2018). The indirect effects of changes to industrial policy (Greenstone et al., 2010; Giroud et al., 2024) and infrastructure investment (Kline and Moretti, 2014; Donaldson and Hornbeck, 2016), similarly, can depend on how—and how closely—locations are connected. As a consequence, efforts to disentangle the channels through which spillover effects propagate are frequently critical in the design and evaluation of economic policy.

Empirical analyses that characterize the mechanisms that mediate spillover effects often do so by relating responses to a shock or policy change with a chosen measure of economic proximity. For example, Moretti (2004) finds that the spillover effects of increases in human capital on productivity are larger for firms whose investments in research and development are concentrated in more similar technological areas.[1] Analyses with this structure confront an essential ambiguity. The association between spillover intensity and any one proximity measure may be confounded by correlation with other mediating channels. That is, if productivity spillovers are larger for technologically similar firms, and technologically similar firms are likely to be geographically proximate, should we infer that technological similarity, itself, affects the intensity of productivity spillovers?[2]

In this paper, we develop an econometric framework for identifying and estimating the effect of a measure of proximity on spillover intensity. To motivate our approach, we begin by giving a nonparametric decomposition, in the spirit of Yitzhaki (1996), Angrist (1998), and Angrist and Krueger (1999), of the implicit estimands of a class of regression specifications typically used in empirical applications that link spillover

[1] Results with an analogous structure are common. For example, Franklin et al. (2024) show that the spillover effects of localized public investment on wages are larger for locations that are more closely connected in labor markets. Myers and Lanahan (2022) find that the spillovers from venture grants to small firms are larger for firms that are technologically or geographically close. Bernstein et al. (2019) find that the spillover effects of bankruptcy are larger for firms that draw on more similar customer bases. Moscona and Seck (2024) show that the spillover effects of cash transfer programs on consumption are larger for more culturally similar individuals. Miguel and Kremer (2004), Feyrer et al. (2017), Diamond and McQuade (2019), and Egger et al. (2022) (among many other analogous examples) find that the spillover effects of de-worming programs, shocks to natural resource endowments, low income housing credits, and cash transfer programs, respectively, decrease with the geographic distance.

[2] This indeterminacy was highlighted saliently by Jaffe et al. (1993), who write "The most difficult problem confronted by the effort to test for spillover-localization is the difficulty of separating spillovers from correlations that may be due to a pre-existing pattern of geographic concentration of technologically related activities."

effects to measures of proximity. In particular, in such settings, researchers often report coefficients obtained from regressions that relate outcomes to proximity-weighted averages of other units' treatments—that is, through specifications similar to

$$Y_i = \alpha + \theta \cdot \Delta_i + \varepsilon_i \ , \quad \text{where} \quad \Delta_i = \sum_{j \neq i} D_{i,j} W_j \ . \tag{1.1}$$

Here, Y_i is an outcome, W_j is a treatment, and $D_{i,j}$ is a proximity measure of interest. See, for instance, Miguel and Kremer (2004), Conley and Udry (2010), Bloom et al. (2013), Acemoglu et al. (2016a), Cai and Szeidl (2024), and Franklin et al. (2024). In Section 2, we show that, under certain conditions, the coefficient θ on the covariate Δ_i identifies a convex average of the mixed-partial derivatives[3]

$$\frac{\partial^2}{\partial \delta \, \partial w} \, \mathbb{E}[Y_i \mid D_{i,j} = \delta, W_j = w] \ . \tag{1.2}$$

Written differently, such regressions measure how the association between one unit's outcome and another unit's treatment *correlates* with the proximity between the two units.[4]

Yet researchers who consider regressions of this kind are, in many instances, fundamentally interested in assigning a causal interpretation to such relationships. That is, often, substantive interest is in measuring the structural relationship between a measure of proximity and spillover intensity. For example, a central question animating the study of economic agglomeration concerns the role of geographic proximity in generating and magnifying spillover effects (Jaffe, 1986; Ellison and Glaeser, 1997; Ellison et al., 2010). To what extent is Silicon Valley's outsized success in the production of novel technologies a consequence of the close geographic proximity of innovative firms? Would industrial policies that encouraged greater clustering (e.g., tax credits, investments in transportation and housing infrastructure, etc.) increase the social returns to investment in research and development? This literature has long recognized, however, that productivity spillovers can also propagate in other ways. For example, ideas may move through scientific networks, and so productivity spillovers might thus be determined by the quantity of effort directed to related scientific problems, rather than through geography, per se (Jaffe, 1986; Jaffe et al., 1993).

As a consequence, if the proximity measure of interest is not the only channel that potentially mediates spillovers, then the parameter (1.2) does not necessarily measure changes in spillover effects induced by exogenous changes to proximity. That is, if the effects of changes in public sector investment in research and development W_j on the outcome Y_i are larger for pairs of firms i, j that are technologically similar, and technological similarity is positively correlated with geographic proximity $D_{i,j}$, then the parameter (1.2) does not recover the causal relationship between geographic proximity and productivity spillover intensity.

---

[3] In the result stated in Section 2, the conditional expectation in parameter (1.2) is replaced by a conditional expectation that effectively holds the "location" of the unit j fixed, and averages over other units i with $D_{i,j} = \delta$. Further notation will be introduced to make the distinction precise.

[4] We allow for the distributions of the treatments and proximity measures to be discrete, continuous, or to contain point masses. For instance, if the treatment variable is valued on $\{0, 1\}$, the parameter (1.2) is replaced by the derivative of the difference $\frac{\partial}{\partial \delta} \left(\mathbb{E}[Y_i \mid D_{i,j} = \delta, W_i = 1] - \mathbb{E}[Y_i \mid D_{i,j} = \delta, W_i = 0] \right)$.

We formalize this point in Section 3. In particular, we propose a class of parameters, analogous to the descriptive parameter (1.2), that reflect the causal relationship between a measure of proximity and spillover intensity. We refer to the elements of this class as “Spillover Proximity Gradients.” We show, with an example calibrated to data from Bloom et al. (2013), that when there are several, correlated measures of proximity that mediate spillover effects, unadjusted estimates of the causal relationship between spillover intensity and any one proximity measure can be badly biased.

Accordingly, in Section 4, we propose a regression-based approach for recovering convex averages of spillover proximity gradients. This proposal is premised on the observation that, in many applications, researchers observe additional, auxiliary measures of the features and economic proximity of the units in their sample. For instance, Myers and Lanahan (2022) measure how the spillover effects of small business loans depend on firms’ technological proximity, and also observe the geographic distance between each pair of firms. Formally, let X_i and $H_{i,j}$ denote vectors of covariates associated with the unit i and the pair of units i, j, respectively, where X_i and X_j can be included as sub-vectors of $H_{i,j}$.

The core methodological idea advanced in this paper is to adjust the regression (1.1) by residualizing the treatment and proximity measure of interest with appropriately chosen covariates and auxiliary proximity measures. The proposed method has two steps. First, the user constructs estimates of the conditional expectations

$$\pi_n(x) = \mathbb{E}[W_j \mid X_j = x] \quad \text{and} \quad \gamma_n(h) = \mathbb{E}[D_{i,j} \mid H_{i,j} = h] \, , \tag{1.3}$$

respectively. Denote these estimates by $\hat{\pi}_n(\cdot)$ and $\hat{\gamma}_n(\cdot)$. Second, the user solves the regression specification

$$Y_i = \alpha + \theta \cdot \hat{\Delta}_i^\star + \varepsilon_i \, , \quad \text{where} \quad \hat{\Delta}_i^\star = \sum_{j \neq i} \hat{D}_{i,j}^\star \hat{W}_j^\star \tag{1.4}$$

and the variables

$$\hat{W}_j^\star = W_j - \hat{\pi}_n(X_j) \quad \text{and} \quad \hat{D}_{i,j}^\star = D_{i,j} - \hat{\gamma}_n(H_{i,j}) \tag{1.5}$$

denote residualized versions of the treatment and proximity measure of interest.

This two-step estimator accounts for potential endogeneity in both the treatment W_j and the proximity measure $D_{i,j}$ by conditioning on observable covariates. Formally, we assume that the treatment W_i is unconfounded conditional on the covariates X_i, and that the covariates and auxiliary proximity measures $H_{i,j}$ are sufficient to stratify pairs of units in terms of any characteristics that jointly determine spillover intensity and the proximity measure of interest. For instance, in the setting considered in Myers and Lanahan (2022), the latter condition would be satisfied if, on average, firms in the same geographic location have the same response to changes in the productivity of firm j, up to their technological similarity to j.[5] In extensions deferred to the Appendix, we illustrate how to augment the specification (1.4) to accommodate settings in

[5] Roughly speaking, this is the identification strategy considered in Jaffe et al. (1993). It has been applied, implicitly or explicitly, by Bloom et al. (2013) and Myers and Lanahan (2022), among many others.

which treatment endogeneity is addressed using difference-in-differences or instrumental variable research designs, as well as settings in which instrumental variables are available for the proximity measure of interest.

In Section 5, we give conditions under which the estimator (1.4) is consistent. We enforce two main restrictions. First, we restrict attention to settings where the estimators $\hat{\pi}_n(\cdot)$ and $\hat{\gamma}_n(\cdot)$ are obtained from well-specified linear regressions. This is analogous to the "saturated specification" condition considered by Angrist (1998). Second, we assume that interference is determined by an underlying, unobserved measure of the proximity of each pair of units in a latent space. That is, we assume that interference can be substantial only among pairs of units that are close in this latent space, and that units are only ever proximate in observed measures if they are proximate in the latent measure. The converse, for the latter, need not hold. We focus our consideration on asymptotic regimes in which the number of units that are proximate in the latent space to each unit increases, so long as the fraction of such units converges to zero. The main novelty of this framework is that we allow for interference to be mediated by proximity in both observed and unobserved characteristics, rather than through the proximity measure of interest alone.

We show that, in such settings, coefficients obtained from regression specifications of the form (1.4) converge at a rate strictly slower than $O_p(n^{-1/2})$. In particular, the rate of convergence achieved by the class of estimators under consideration decreases as the extent of cross-unit interference increases. Nevertheless, we establish that this sub-parametric rate is optimal, up to constant factors, in a minimax sense.

In Section 6, we detail a procedure for quantifying the uncertainty associated with estimates obtained from regressions with the structure (1.4). Most approaches to inference in settings with interference take two routes. In the first, units are placed into disjoint groups, and clustered standard errors can be constructed under the assumption that interference across groups is sufficiently small (Leung, 2023). In the second, interference is assumed to be bounded above by a decreasing function of an observed proximity measure, and inferences can be constructed with kernel-based methods (Kojevnikov, 2021; Leung, 2022a).

Neither approach is applicable to our setting. In particular, as we allow for interference to be mediated by unobserved characteristics, units cannot necessarily be arranged into groups or otherwise organized around an observed proximity measure. Instead, we show that conservative estimates of uncertainty can be obtained via a procedure based on repeatedly computing the coefficient from a regression specification analogous to (1.4), where the outcomes have been replaced by the empirical residuals and the signs of the residualized treatments have been flipped at random. The implementation of the procedure does not require the choice of any bandwidths or tuning parameters, and places no a priori restrictions on the structure of the interference across units. This approach is closely related to the variance estimators proposed in Adao et al. (2019) and the randomization inference-based methods considered by Borusyak and Hull (2023). Relative to these procedures, we show that the proposed method is asymptotically conservative for tests concerning the value of the parameter targeted by the estimator (1.4) and is applicable to settings where spillovers are non-linear, heterogeneous, and potentially mediated by unobserved channels.

In Section 7, we illustrate the application of the proposed methodology to measure the relationship between productivity spillover intensity and technological and product market proximity, using data from Bloom et al. (2013). In line with the results reported in Bloom et al. (2013), we find that adjusting estimates of the relationship between spillover intensity and a given measure of proximity for correlated channels that also mediate spillovers meaningfully changes spillover effect estimates. We give two additional applications, using data from Hornbeck and Moretti (2024) and Franklin et al. (2024), in Appendix F.

This paper contributes to an extensive literature that considers the identification and estimation of spillover effects. Much of this research builds on the "exposure mapping" and "effective treatment" frameworks developed in Hudgens and Halloran (2008), Manski (2013), Toulis and Kao (2013), and Aronow and Samii (2017), which parameterize outcomes with correctly specified, low-dimensional summaries of the treatments received by other units. By contrast, and joining Leung (2022a), Sävje (2024), and Menzel (2025), we consider parameters that can be defined without placing such restrictions and do not assume that the complete structure of interference is known a priori (or that information sufficient to recover this structure is observed).[6]

The key distinction, in our asymptotic analysis, is that we assume that interference is determined by proximity in both the observed and unobserved characteristics of the population under consideration. The assumptions that we use to express this structure are closely related to the "latent surface models" considered by, for example, Hoff et al. (2002), McCormick and Zheng (2015), and Breza et al. (2020). We differ, in that our primary aim is not to recover the latent surface. Instead, we use the geometric structure associated with the latent surface as device in our theoretical analysis for organizing the influence of unobserved interference.

The main contribution of this paper is a method for characterizing the *causal* relationship between a measure of proximity and spillover intensity. This objective is related to, but distinct from, the descriptive problems considered in Pollmann (2023) and Wang et al. (2025), who aim to characterize the average spillover effect between pairs of units that are separated by a given distance (see also Munro et al. (2025) and Li and Wager (2022), who consider related problems). Put differently, this paper is distinguished by the fact that its explicit objective is to estimate the *effect* of interventions that affect the proximity between pairs of units.

We contribute, as well, to a growing literature that considers the estimation of spillover effects in observational data (see e.g., Leung and Loupos 2025; Auerbach and Tabord-Meehan 2025).[7] In this context, the main novelty of our analysis is that we consider models where measures of proximity are endogenous,

[6] See also Gao and Ding (2023), who apply the framework developed in Leung (2022a) to regression-based estimators.

[7] A distinct line of this literature develops methods for accommodating network interference in treatment selection, primarily through an assumption that a well-specified exposure mapping is unconfounded, conditional on a vector of covariates (see e.g., Veitch et al., 2019; Forastiere et al., 2021; Ogburn et al., 2024; Emmenegger et al., 2025; Auerbach and Tabord-Meehan, 2025). Leung and Loupos (2025) consider a more general model, where treatments are jointly determined, conditional on an observed network and covariates. In each case, the networks that mediate selection into treatment and spillovers between units are taken to be fixed and observed. Moreover, in contrast to the regression-based methods at the center of this paper, explicit specification, or estimation, of the network dependence in the treatments is required.

in the sense that pairs of units that are more proximate in a particular measure can be more likely to be proximate in other observed or unobserved factors that might also mediate spillover effects.[8]

Finally, we contribute to a broader literature concerning the interpretation and statistical analysis of regression specifications where the covariates of interest take the form of weighted averages (Adao et al., 2019; Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022b; Borusyak and Hull, 2023). In this literature, this paper is distinguished by the fact that we do not assume that the regressions under consideration are well-specified, and study inference for a class of parameters that can be defined nonparametrically.

Section 8 concludes. Proofs for results stated in Sections 2 to 4 are given in Appendices A and B. Proofs for results stated in Sections 5 and 6, and all supporting lemmas, are given in Appendices C and D, respectively. Appendices E and F give additional results and discussion that will be introduced at appropriate points throughout the paper.

2. A Nonparametric Decomposition of Spillover Regression

We consider settings with the following structure. There are n units. Each unit i in $[n] = \{1, \ldots, n\}$ is associated with a real-valued treatment W_i and a real-valued outcome Y_i. In turn, each pair of units i, j is associated with a variable $D_{i,j}$ that measures geographic or economic proximity. The main interest is in characterizing the relationship between the proximity measure $D_{i,j}$ and the spillover effect of the treatment W_j on the outcome Y_i. Settings with this structure occur throughout applied economics.[9]

Example 1 (Research and Development)**.** Firm level investments in R&D affect the market outcomes of other firms (Arrow, 1962; Griliches, 1992). Bloom et al. (2013) measure how "knowledge spillovers" vary with "technological proximity." The main hypothesis is that a firm's innovative activity has larger effects on the productivity of firms that operate in similar technological areas. In this application, the treatment W_j is the stock of R&D for firm j, the outcome Y_i is the market value of firm i, and the variable $D_{i,j}$ denotes technological proximity, which, following Jaffe (1986), is measured by the uncentered correlation (or "cosine similarity") of patent shares across USPTO technology classes.[10] This measure is widely adopted, see e.g., Ellison and Glaeser (1997), Moretti (2004), and Ellison et al. (2010). ■

Example 2 (Productivity)**.** Using city-level data from the U.S. Census, Hornbeck and Moretti (2024) argue that the indirect effect of a change in a cities' manufacturing productivity on the wages in another city depends on the propensity to migrate between the two cities. That is, productivity growth increases labor demand in city j, encouraging migration from city i, which reduces labor supply. Here, the treatment W_j denotes total

[8] We note that we use the term "endogenous networks" in a different sense than it is used in Gao (2024), who studies experimental settings where treatments can affect measures of proximity.

[9] To simplify exposition, we have ruled out applications where treatments and outcomes are measured for different units; see, e.g., Kovak (2013), Autor et al. (2013), and Adao et al. (2019), where treatments are assigned to industries and outcomes are measured at geographic units of aggregation. We treat this extension in Appendix E.1.

[10] Bloom et al. (2013) consider several alternative measures of technological proximity, largely based on reweighting or rescaling the inner product between patent shares. We focus our discussion on their baseline measure for the sake of simplicity.

factor productivity growth, the outcome Y_i denotes growth in average earnings, and the proximity measure $D_{i,j}$ denotes the pretreatment probability that an individual that migrates from city i relocates to city j. ■

Example 3 (Public Works)**.** Similarly, Franklin et al. (2024) study the spillover effects of a public works program implemented in Addis Ababa, Ethiopia. The program provided guaranteed public work to targeted households and was randomized across neighborhoods. In this setting, the treatment W_j indicates that neighborhood j has been assigned to the program and the outcome Y_i denotes the average earnings in neighborhood i. Franklin et al. (2024) argue that the indirect effect of the program on earnings depends on commuting flows. That is, individuals who participate in the program no longer commute to other neighborhoods, reducing private sector labor supply. In our notation, they posit that the effect of the program W_j on earnings Y_i is increasing in the pretreatment probability $D_{i,j}$ that a worker who works in neighborhood i lives in neighborhood j. ■

2.1 Spillover Regression

In the examples outlined above, the main interest is in characterizing the effect of economic proximity on spillover intensity. What, precisely, does this mean? To give empirical content to an answer to this question, we begin our analysis by considering the interpretation of regression specifications often used in applications with this structure. In particular, in such settings, researchers typically report coefficients obtained from regression specifications similar to

$$Y_i = \alpha + \theta \cdot \Delta_i + \varepsilon_i \ , \quad \text{where} \quad \Delta_i = \sum_{j \neq i} D_{i,j} W_j \tag{2.1}$$

measures the "exposure" of unit i to the treatments W_j through the proximity measure $D_{i,j}$.[11]

The interpretation of coefficients obtained in this way varies widely. In some cases, measures of proximity are observed directly in data and the specification (2.1) is posed without a strict model-based justification; see, for instance, Bloom et al. (2013), Cai et al. (2015), Fafchamps and Quinn (2018), Huber (2018), Zacchia (2020), Arque-Castells and Spulber (2022), Lerche (2025), and Bergeaud et al. (2025). In others, either the specification (2.1) or the measure $D_{i,j}$ are derived from (or approximate the solution to) models of spatial equilibrium; see e.g., Redding and Venables (2004), Hanson (2005), Conley and Udry (2010), Ahlfeldt et al. (2015), Acemoglu et al. (2016a), Acemoglu et al. (2016b), Rotemberg (2019), Helm (2020), Borusyak et al. (2022a), Franklin et al. (2024), Cai and Szeidl (2024), and Adao et al. (2025).

[11] Often, researchers also include the variable W_i as a covariate, and report the associated coefficient as a measure of the "direct effect." In other cases, the proximity measure of interest enters through some monotonic function $Q(D_{i,j})$. For example, when $D_{i,j}$ is a function of geographic distance, researchers often set $Q(\cdot)$ as a binary indicator, or vector of binary indicators, that $D_{i,j}$ is an element of some pre-defined interval; see, e.g, Miguel and Kremer (2004), Egger et al. (2022), Myers and Lanahan (2022), and Muralidharan et al. (2023). Similarly, in settings where the specification (2.1) has been derived from a model of spatial equilibrium, the function $Q(\cdot)$ often depends on various elasticities fixed at values chosen by referencing pre-existing literature; see e.g., Kovak (2013), Autor et al. (2013), Donaldson and Hornbeck (2016), and Adao et al. (2025). We treat several of these extensions in Appendix E.1, but focus our consideration on the specification (2.1) in the main text for the sake of parsimony.

The sentiment adopted in much of this literature is that such estimates recover structural quantities—parameters defined and justified in terms of the primitives and assumptions of an economic, or parametric, model, but that are not necessarily nonparametrically meaningful. In this paper, we advance an alternative perspective. We begin by giving a result, in the spirit of Yitzhaki (1996), Angrist (1998), and Angrist and Krueger (1999), that shows that, under certain conditions, the regression (2.1) identifies a weighted average of particular mixed-partial derivatives of the conditional expectation of the outcomes. This result will suggest that the coefficient $\hat{\theta}_n$ on the covariate Δ_i in the specification (2.1) can be interpreted as targeting a particular nonparametric, causal estimand, whose identification and estimation are the main subjects of this paper.

2.2 Baseline Model

Our objective is to give a nonparametric interpretation for coefficients obtained from regression specifications of the form (2.1). To do this, we impose several baseline conditions on the joint distribution of the treatments W_j and proximity measures $D_{i,j}$ across units. First, we posit that the proximity measure $D_{i,j}$ is generated by a function of some, potentially unobserved, features of the units i and j.

Assumption 2.1 (Latent Factor Structure)**.** *There exists a function $D(\cdot,\cdot)$, valued on the non-negative, real-valued set $\mathcal{D}$, and potentially unobserved coordinates $(S_i)_{i=1}^n$ on the set $\mathcal{S}$ such that the representation*

$$D_{i,j} = D(S_i, S_j) \; , \tag{2.2}$$

holds for each pair of units i,j.

If, for instance, the variable $D_{i,j}$ is a function of the geographic distance between units i and j, then S_i can be taken to denote the geographic coordinates of unit i. Likewise, in Example 1, technological proximity is given by the uncentered correlation of the shares of each firm's patents that belong to each USPTO technology class. In more abstract settings, like in Examples 2 and 3, the factors can be viewed more expansively. That is, if $D_{i,j}$ measures the probability of migrating between cities i and j, then the coordinate S_i could, for example, collect city i's geographic coordinates, average wage, and average amenity value.

The main restriction is that the proximity measure $D_{i,j}$ is not determined by any features inherent to the *pair* i, j. In Appendix E.1 we show that the result stated in this section can be extended to settings where $D_{i,j}$ is determined by the coordinates S_i, S_j as well as i.i.d., pair-specific errors. This includes, for instance, the latent position and graphon network models considered by, e.g., Hoff et al. (2002), Breza et al. (2020), and Li and Wager (2022). Nevertheless, ruling out pair-specific errors greatly simplifies the exposition of the identification and statistical results given in subsequent sections.

Second, we assume that the treatments are i.i.d., centered, and orthogonal to the latent factors.

Assumption 2.2 (Exogeneity)**.** *The replicates $(S_i, W_i)_{i=1}^n$ are i.i.d. and satisfy the condition*

$$\mathbb{E}[W_i \mid S_i] = 0 \tag{2.3}$$

almost surely.

This assumption permits dependence in the outcomes across units, as would arise from spillover effects, but rules out cross-sectional dependence in the treatments. We impose this restriction because regression specifications of the form (2.6) are not suited to distinguish spillover effects "mediated by" the proximity measure $D_{i,j}$ from correlation between the treatments W_i and W_j.[12] This is a version of the identification problem studied in Manski (1993).

Assumption 2.2 also rules out settings where the mean of the treatment varies with S_i. Equivalent assumptions, in closely related models, are used in Adao et al. (2019), Borusyak et al. (2022b), and Borusyak and Hull (2023). In Section 4, we relax this aspect of Assumption 2.2 to hold after conditioning on available covariates. We have also stipulated that the treatments are centered. This is equivalent to the main methodological recommendation of Borusyak and Hull (2023) and is increasingly adopted in practice.

Third, we require that the association between the treatment and outcome of two distinct units is relatively insensitive to changes in their proximity to a third unit. To state this assumption, we introduce some useful notation. For a real-valued function f defined on the set $\mathcal{A}$ in $\mathbb{R}$, we use the short-hand $\partial_a f(a')$ to denote the derivative of $f(a)$ evaluated at a'. If the set $\mathcal{A}$ is connected, this is defined in the usual way. If the set $\mathcal{A}$ is not connected, then, following a convention set out in Kolesár and Plagborg-Møller (2024), we define $\partial_a f(a)$ on the convex hull of $\mathcal{A}$ by extending f via linear interpolation. That is, if $\mathcal{A} = \{0, 1\}$, then $\partial_a f(1/2) = f(1) - f(0)$. We use the notation $\partial^2_{a,b} f(a, b)$ for the analogous mixed-partial derivative.

Assumption 2.3 (Limited Second-Order Interference)**.** *For distinct units i, j, and k, it holds that*

$$\partial^2_{\delta,w}\, \mathbb{E}[Y_i \mid D_{k,j} = \delta, W_j = w, S_j = s] = o(n^{-1/2}) \tag{2.4}$$

uniformly over each δ, w, and s in their respective domains.

Assumption 2.3 is relatively mild, ruling out only situations where changes to the coordinate S_k can make a non-negligible difference on the relationship between the distinct units i and j. Finally, we require two mild regularity conditions concerning the smoothness of the conditional expectation of the outcome and the scaling of the proximity measure $D_{i,j}$. We defer the statement of these assumptions to Appendix A.

2.3 Decomposition

The following Theorem expresses the population solution to the regression (2.1) as a convex average of the mixed-partial derivatives of the conditional expectation

$$\mu(\delta, w \mid S_j) = \mathbb{E}[Y_i \mid D_{i,j} = \delta, W_j = w, S_j]\,, \tag{2.5}$$

[12] This choice of emphasis is common in models of network interference (see, e.g., Leung 2020; Auerbach 2022; Li and Wager 2022). The settings considered in Forastiere et al. (2021), Pollmann (2023), Ogburn et al. (2024), Emmenegger et al. (2025), and Leung and Loupos (2025) are exceptions. However, the approaches to estimation taken in these cases are substantially more complex, and require explicit specification (or, in the case of Leung and Loupos (2025), estimation) of the dependence across treatments.

defined for each δ and w in their respective domains. Let

$$L_n(\theta) = \min_{\alpha} \sum_{i=1}^{n} (Y_i - \alpha - \theta \cdot \Delta_i)^2 \tag{2.6}$$

denote the loss function associated with the regression specification (2.1).

Theorem 2.1. *Assume that the treatments and proximity measures have support contained in intervals $\overline{\mathcal{W}}$ and $\overline{\mathcal{D}}$ of constant length and that the outcomes are identically distributed. Under Assumptions 2.1 to 2.3, and two regularity conditions stated in Appendix A, the parameter $\overline{\theta}_n$ that minimizes the risk $\mathbb{E}[L_n(\theta)]$ admits the representation*

$$\overline{\theta}_n = \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}\Big[\lambda(\delta, w \mid S_j)\, \partial^2_{\delta,w} \mu(\delta, w \mid S_j)\Big]\, \mathrm{d}w\, \mathrm{d}\delta + o(n^{-1/2})\ , \tag{2.7}$$

where the weights

$$\lambda(\delta, w \mid S_j) = \frac{\operatorname{Cov}(\mathbb{I}\{D_{i,j} \geq \delta\}, D_{i,j} \mid S_j) \operatorname{Cov}(\mathbb{I}\{W_j \geq w\}, W_j \mid S_j)}{\mathbb{E}[\operatorname{Var}(D_{i,j} \mid S_j) \operatorname{Var}(W_j \mid S_j)]} \tag{2.8}$$

are convex, in the sense that

$$\lambda(\delta, w \mid S_j) \geq 0 \quad \text{and} \quad \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid S_j)]\, \mathrm{d}w\, \mathrm{d}\delta = 1\ , \tag{2.9}$$

respectively. That is, the weights (2.8) are positive almost surely and integrate to one in expectation.

Remark 2.1. Theorem 2.1 applies to settings where distributions of the treatments and proximity measures are discrete, have continuous support, or have point masses. For instance, if the treatments W_j are valued on $\{0, 1\}$ and $\operatorname{Var}(W_j \mid S_j)$ is constant almost surely, then the representation (2.7) can be re-expressed as the weighted average derivative

$$\overline{\theta}_n = \int_{\overline{\mathcal{D}}} \lambda(\delta \mid S_j)\, \partial_\delta \left(\mathbb{E}[Y_i \mid D_{i,j} = \delta, W_j = 1, S_j] - \mathbb{E}[Y_i \mid D_{i,j} = \delta, W_j = 0, S_j]\right) \mathrm{d}\delta + o(n^{-1/2})\ ,$$
$$\text{where} \quad \lambda(\delta \mid S_j) = \frac{\operatorname{Cov}(\mathbb{I}\{D_{i,j} \geq \delta\}, D_{i,j} \mid S_j)}{\mathbb{E}[\operatorname{Var}(D_{i,j} \mid S_j)]}\ , \tag{2.10}$$

by the convention set out prior to the statement of Assumption 2.3. Likewise, if, additionally, the proximity measure $D_{i,j}$ is supported on $\{0, 1\}$, and $\operatorname{Var}(D_{i,j} \mid S_j)$ is constant almost surely, then (2.7) takes the form of the double-difference

$$\begin{aligned}\overline{\theta}_n &= \left(\mathbb{E}[Y_i \mid D_{i,j} = 1, W_j = 1] - \mathbb{E}[Y_i \mid D_{i,j} = 1, W_j = 0]\right) \\ &\quad - \left(\mathbb{E}[Y_i \mid D_{i,j} = 0, W_j = 1] - \mathbb{E}[Y_i \mid D_{i,j} = 0, W_j = 0]\right) + o(n^{-1/2})\ ,\end{aligned} \tag{2.11}$$

by the same convention. ∎

Remark 2.2. Observe that the minimization over the intercept α is included as part of the definition of the loss function (2.6), and so occurs within the expectation used to define the parameter $\overline{\theta}_n$. This arrangement is nonstandard, but is essential to obtain the representation (2.7). In particular, by the Frisch-Waugh-Lovell

Theorem, we can write

$$\mathbb{E}[L_n(\theta)] = \mathbb{E}\left[\sum_{i=1}^{n}(Y_i - \theta\cdot\tilde{\Delta}_i)^2\right] ,$$

$$\text{where}\quad \tilde{\Delta}_i = \sum_{j\neq i}\left(D_{i,j} - \frac{1}{n}\sum_{k\neq j} D_{k,j}\right) W_j - \frac{1}{n}\sum_{k\neq i} D_{k,i}W_i \ , \tag{2.12}$$

and so, by the Projection Theorem, the parameter $\overline{\theta}_n$ admits the representation

$$\overline{\theta}_n = \frac{\sum_{i=1}^{n}\mathbb{E}[Y_i\tilde{\Delta}_i]}{\sum_{i=1}^{n}\mathbb{E}[(\tilde{\Delta}_i)^2]} \ . \tag{2.13}$$

Here, by including the construction of the intercept within the expectation, the proximity measures $D_{i,j}$ are effectively centered by the averages $\frac{1}{n}\sum_{k\neq j} D_{k,j}$ (up to a negligible error term introduced by the second term in (2.12)). That is, the proximity measures $D_{i,j}$ are *implicitly* centered conditional on the coordinate S_j. The error in the average $\frac{1}{n}\sum_{k\neq j} D_{k,j}$ about the expectation $\mathbb{E}[D_{k,j} \mid S_j]$ contributes the leading term of the error in the approximation (2.7), and is controlled by Assumption 2.3. This implicit centering causes each of the terms that appear in the representation (2.7) to condition on the coordinate S_j, which has important implications for the identification arguments pursued below. ■

This result can be best understood through comparison to the simpler setting where an outcome Y_i is regressed on a real-valued treatment W_i and a vector of covariates X_i—that is, through the specification

$$Y_i = \alpha + \theta\cdot W_i + \gamma^{\top} X_i + \varepsilon_i \ . \tag{2.14}$$

Building on results due to Yitzhaki (1996), Angrist (1998), Angrist and Krueger (1999), Kolesár and Plagborg-Møller (2024) show that, under conditions analogous to those used in Theorem 2.1, if the expectation of W_i conditional on the covariates X_i is linear, then the population coefficient on W_i admits the representation

$$\int_{\mathcal{W}} \mathbb{E}\Big[\lambda(w \mid X_i)\,\partial_w \mathbb{E}[Y_i \mid W_i = w, X_i]\Big]\,\mathrm{d}w \ , \tag{2.15}$$

$$\text{where}\quad \lambda(w \mid X_i) = \frac{\mathrm{Cov}(\mathbb{I}\{W_i \geq w\}, W_i \mid X_i)}{\mathbb{E}[\mathrm{Var}(W_i \mid X_i)]} \ .$$

Here, as before, the weight function $\lambda(w \mid X_i)$ is positive almost surely and integrates to one in expectation. Written differently, linear regression coefficients identify convex averages of the change in the outcome associated with incremental changes in the treatment.

Theorem 2.1 shows that, under certain conditions, an analogous, nonparametric interpretation can be given to coefficients obtained from spillover regressions.[13] That is, regression specifications of the form (2.1)

[13] Vazquez-Bare (2023) consider estimators similar to (2.1) in settings where there are a collection of groups, units within the same group receive the same treatment, and the proximity measure $D_{i,j}$ is binary and indicates that i and j are elements of the same group. He finds that the resultant estimate does not have a coherent causal interpretation. By contrast, we show that, if treatments are independent and identically distributed, the implicit estimand of the specification under consideration permits a clean representation.

identify convex averages of the change in the dependence of one unit's outcome on another unit's treatment associated with changes in the proximity of the two units.[14] This decomposition is purely descriptive. That is, we have placed no restrictions on the causal relationship between the treatments, proximity measures, and outcomes, and are only describing the joint distribution of their realized values.[15]

Linear regression coefficients invite causal interpretations, as measuring as the *effect* of W_i on Y_i. So too does the representation (2.7), as measuring the *effect* of the proximity measure $D_{i,j}$ on the intensity of the spillover effect of W_j on Y_i. In the remainder of the paper, we define causal parameters that encode this structure, state conditions under which they are identified, and develop procedures for augmenting the regression specification (2.6) to operationalize these conditions.

3. Spillover Proximity Gradients

In many settings in applied economics, the primary, substantive interest is in characterizing the causal relationship between a measure of proximity and the spillover effects of a shock, policy, or action. In particular, often, spillover effects are potentially mediated by several distinct economic phenomena, whose extent and intensity have different implications for welfare and policy. Recovering the structural relationship between responses to a treatment any one proximity measure, then, requires disentangling these channels.

Example 1 (Continued)**.** Consider, for instance, the problem of interest in Bloom et al. (2013). Aggregate estimates of the indirect effects of one firm's investment in R&D on other firms' values risk conflating two distinct mechanisms. The first mechanism gives rise to the core market failure in markets for innovation (Arrow, 1962). One firm's investment in R&D may reduce the cost of R&D for technologically similar firms. Such cost reductions are socially valuable, but are rarely internalized by the innovating firm. Thus, when this mechanism exists—and, in particular, when it is quantitatively meaningful—there is a risk of persistent underinvestment in R&D. The second mechanism concerns product-market interactions: investments in R&D by one firm may reduce sales for firms marketing substitutes and increase sales for firms marketing complements (Dasgupta and Stiglitz, 1980). The welfare and policy implications of the second channel have little overlap with the first, in the sense that the underlying economic phenomena are fundamentally different, and depend largely on market structure (Spence, 1984).

To estimate the effect of technological proximity on R&D spillover intensity, it is necessary to account for the potentially countervailing effect of product market rivalry. To operationalize this idea, Bloom et al. (2013) measure the "product market proximity" of firms i and j by computing the uncentered correlation $G_{i,j}$ of the share of each firms' sales across four-digit industry codes. Panel A of Figure 1 displays features of the joint distribution of technological proximity $D_{i,j}$ and product market proximity $G_{i,j}$ across the sample of firms

[14]As it can be shown that $\int\int \lambda(\delta, w \mid S_j)\,\mathrm{d}w\,\mathrm{d}\delta = \mathrm{Var}(D_{i,j} \mid S_j)\,\mathrm{Var}(W_j \mid S_j)$, units j that have larger conditional variances $\mathrm{Var}(W_j \mid S_j)$ and $\mathrm{Var}(D_{i,j} \mid S_j)$ tend to receive larger weights.

[15]The orthogonality condition stipulated by Assumption 2.2 is best motivated by the assumption that the treatments are randomly assigned. If this condition did not hold, the part of the weight $\lambda(\delta, w \mid S_j)$ associated with $D_{i,j}$ would vary with W_j, and would no longer necessarily integrate to one. We relax this assumption, by introducing covariates, in Section 4.

considered by Bloom et al. (2013).[16] Technological proximity increases substantially and systematically with product market proximity. ■

Example 2 (Continued)**.** In Hornbeck and Moretti (2024), interest centers on characterizing how the spillover effect of total factor productivity growth on wages varies with migration. Of course, other notions of economic proximity might also mediate indirect effects. For instance, suppose that we measure the proportion of output produced in city i that is purchased by firms in city j. Denote this measure by $G_{i,j}$. An increase in city i's productivity may increase demand for intermediate inputs produced by city j, increasing city j's labor demand in proportion to $G_{i,j}$. ■

Example 3 (Continued)**.** Franklin et al. (2024) measure how the indirect effects of a public works program depend on commuting flows. There are, again, other channels that might also mediate spillover effects. For example, as the program consists of publicly subsidized employment in small-scale activities that improve shared amenities (e.g., garbage disposal, greening of public spaces), businesses that are geographically close to the treated neighborhoods may experience increased demand, and so spillover effects might, just as well, be driven by geographic distance $G_{i,j}$. ■

In this section, we propose a class of parameters, which we refer to as spillover proximity gradients, that express the causal relationship between a measure of proximity and spillover intensity. We show that, when there are several, correlated measures of proximity that could potentially mediate spillover effects, unadjusted estimates of the causal relationship between spillover intensity and any one measure can be badly biased.

3.1 Potential Outcomes

Our interest is in estimating the effect of a measure of proximity on spillover intensity. To pose this objective formally, we must prescribe the potential value of each unit's outcome under counterfactual interventions to each other unit's treatment and proximity.

To this end, we impose a condition that enables us to intervene on the proximity measure of interest $D_{i,j}$, while holding constant any unobserved features of other units. We assume that each unit is associated with an additional, potentially unobserved, variable U_i. The variable U_i should be thought of as a unit-specific residual or latent factor. We refer to the variable

$$Z_i = (W_i, S_i, U_i) \tag{3.1}$$

as the *state* of unit i and let Z_{-i} collect the state of all units other than i.

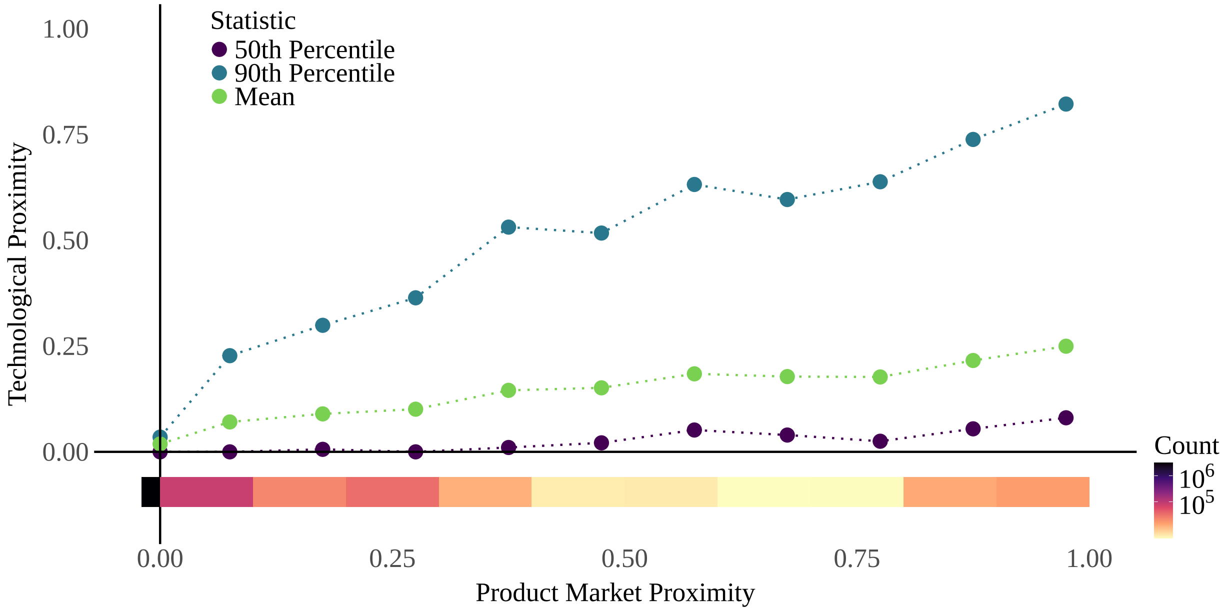
Assumption 3.1 (Anonymous Interference)**.** *Outcomes are generated through the structural equation*

$$Y_i = F(Z_i, Z_{-i}) \ , \tag{3.2}$$

where the function $F(\cdot, \cdot)$ is symmetric in the components of its second argument.

[16] We detail our treatment of these data, and display analogous figures for Examples 2 and 3, in Appendix F.

FIGURE 1. Correlation in Measures of Economic Proximity



Notes: Figure 1 displays features of the joint distribution of technological and product market proximity across a sample of publicly traded U.S. firms, using data from the replication packages associated with Bloom et al. (2013) and Lucking et al. (2019). The measures are computed using the shares of patents filed across USPTO technology classes and shares of sales across four digit industry codes from 1990 to 1995, respectively. Further details concerning the treatment of these data are given in Appendix F. The x-axis measures product market proximity and has been partitioned into bins. The means, 50th percentiles, and 90th percentiles of technological proximity within each bin are displayed using green, purple, and blue dots and are measured relative to the y-axis. A heatmap measuring the number of pairs of units in each bin is displayed below both the x-axis.

The state Z_i should be interpreted as collecting all of the features of the unit i that can impact any unit's outcome. Like Assumption 2.1, the main restriction is that the variable Y_i is not determined by any factors inherent to the pair i, j. That is, Assumption 3.1 ensures that the influence of unit j on unit i is "anonymous," in the sense that unit i's outcome depends only on the value of Z_j and not in any way on j's "identity."

Assumption 3.1 reduces the problem of defining potential outcomes associated with interventions to the proximity of another unit into the specification of choices for counterfactual values for the coordinate S_j. Any counterfactual coordinate consistent with the intervention $D_{i,j} = \delta$ is necessarily an element of the set

$$\mathcal{S}^{(i)}(\delta) = \{s : D_n(S_i, s) = \delta\} . \tag{3.3}$$

The convex combination

$$S_j^{(i)}(\delta) = S_j + \alpha_j^{(i)}(\delta)(S_i - S_j) , \quad \text{where} \quad \alpha_j^{(i)}(\delta) = \arg\min_{\alpha} \{|\alpha| : S_j + \alpha(S_i - S_j) \in \mathcal{S}^{(i)}(\delta)\} , \tag{3.4}$$

is a natural choice. If this choice is uniquely-defined, we can express the potential outcome

$$Y_{i,j}(\delta, w) = F(Z_i, Z_j(w, S_j^{(i)}(\delta)), Z_{-i,j}) , \quad \text{where} \quad Z_j(w, s) = (w, s, U_j) \tag{3.5}$$

denotes the counterfactual state of unit j that would be realized if the coordinate S_j were set to s and the treatment W_j were set to w. Here, the vector $Z_{-i,j}$ collects each unit's state, excluding the units i and j.

Unless stated otherwise, for the remainder of the paper, we use (3.5) to denote the potential outcome for unit i associated with the intervention that sets the treatment W_j to w and the proximity $D_{i,j}$ to δ. Thus, it remains to enforce structure that ensures that the choice (3.4) can be uniquely defined. To do this, we impose the following strengthening of Assumption 2.1.

Assumption 3.2 (Monotone Factor Structure)**.** *There exists a norm $\|\cdot\|$ on $\mathcal{S}$ such that*

$$D_{i,j} = D(S_i, S_j) = \overline{D}(\|S_i - S_j\|) \tag{3.6}$$

for some function $\overline{D}(\cdot)$ that is continuous and non-increasing.

It can be shown that, under Assumption 3.2, so long as the set (3.3) is non-empty, the choice (3.4) is uniquely defined.[17] Assumption 3.2 holds in settings where the proximity measure of interest is a decreasing function of geographic distance, such as pure iceberg trade costs. Panel A of Figure 2 illustrates the construction (3.4) in such a setting. Assumption 3.2 also holds in Example 1.

Example 1 (Continued)**.** In Bloom et al. (2013), technological proximity is given by

$$D_{i,j} = \frac{\langle S_i, S_j \rangle}{\|S_i\|_2 \|S_j\|_2} = 1 - \frac{1}{2} \left\| \frac{S_i}{\|S_i\|_2} - \frac{S_j}{\|S_j\|_2} \right\|_2^2 , \tag{3.7}$$

where S_i denotes the share of firm i's patents in each USPTO technology class and $\|\cdot\|_2$ and $\langle \cdot, \cdot \rangle$ denote the Euclidean norm and inner-product. Thus, Assumption 3.2 is satisfied. Panel B of Figure 2 illustrates a simplified version of the construction (3.4) for this case. ■

In Appendix E.2, we detail several approaches for extending Assumption 3.2 to settings where proximity measures are functions of many features of the units under consideration. For instance, commuting and migration probabilities, like those considered in Examples 2 and 3, are often modeled as being functions of geographic distance, as well as wages and amenity values (Monte et al., 2018). In essence, we address such settings by assuming that the measure $D_{i,j}$ is monotone in some feature of the unit j and conditioning on other information sufficient to isolate the dependence of $D_{i,j}$ on this feature.

3.2 Spillover Proximity Gradients

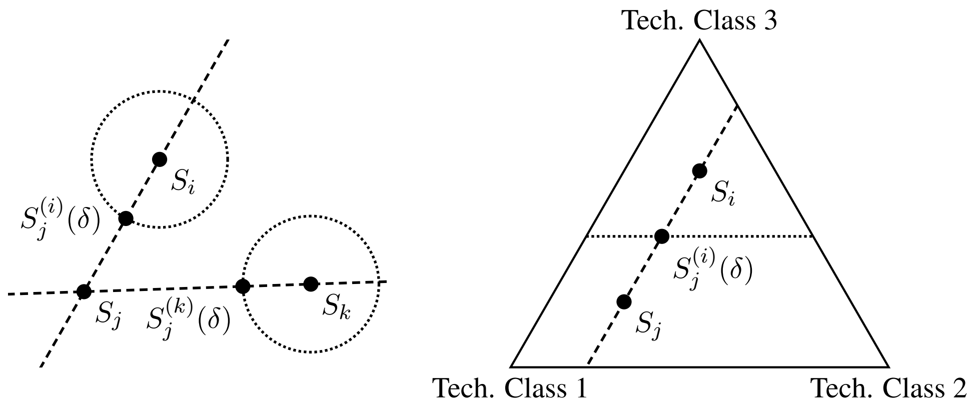
We give a framework for identifying and estimating averages of the variables

$$\partial^2_{\delta,w} Y_{i,j}(\delta, w) , \tag{3.8}$$

[17] In Appendix E.2, we extend Assumption 3.2 to accommodate setting where $\overline{D}(\cdot)$ has discontinuities. For example, in some settings, the proximity measure $D_{i,j}$ might take the form of an indicator that the distance $\|S_i - S_j\|$ is less than or equal to a specified radius. This can be addressed analogously to formalizations of regression discontinuity designs, by defining the potential outcome (3.5) as the limiting value as S_j approaches the boundary of $\mathcal{S}^{(i)}(0)$ along the line $S_j + \alpha(S_i - S_j)$. Assumption 3.2 can also be extended without loss to settings where $D_{i,j} = \overline{D}(S_i, \|S_i - S_j\|)$, so long as $\overline{D}(S_i, \cdot)$ remains monotone.

FIGURE 2. Counterfactual Coordinate Construction

Panel A: Geographic Proximity *Panel B: Technological Similarity*



Notes: Figure 2 illustrates the construction of the counterfactual coordinate $S_j^{(i)}(\delta)$ in two examples. In Panel A, the variables S_i measure geographic coordinates and the proximity measure of interest is a function of Euclidean distance. Panel B gives a simplified version of the setting consider by Bloom et al. (2013). Here, the coordinates S_i take values on a simplex and measure the proportion of each firm's patents filed in each of three technology classes. The proximity measure of interest is given by the inner product $D_{i,j} = \langle S_i, S_j \rangle$. In both cases, the level sets (3.3) are displayed with dotted lines. The values of the points $S_j + \alpha(S_i - S_j)$, as α varies, are displayed with dashed lines. The points (3.4), where these lines intersect, are marked with dots.

---

measuring the incremental change in the spillover effect of W_j on Y_i as $D_{i,j}$ is increased. We refer to the variable (3.8) as a "spillover proximity gradient." Spillover proximity gradients are analogous to the descriptive parameters

$$\partial^2_{\delta,w}\, \mu(\delta, w \mid S_j) = \partial^2_{\delta,w} \mathbb{E}\left[Y_{i,j}(\delta, w) \mid W_j = w, D_{i,j} = \delta, S_j\right] \tag{3.9}$$

identified by the regression (2.6). However, unless the potential outcome inside the conditional expectation (3.9) is independent of the treatment W_j and proximity measure $D_{i,j}$, conditional on the factor S_j, the average

$$\mathbb{E}[\partial^2_{\delta,w} Y_{i,j}(\delta, w) \mid S_j] \tag{3.10}$$

and the parameter (3.9) will be, in general, different. In practice, this difference can be large.

Example 1 (Continued)**.** Suppose, for the sake of illustration, that the outcomes in Bloom et al. (2013) were generated by the specification

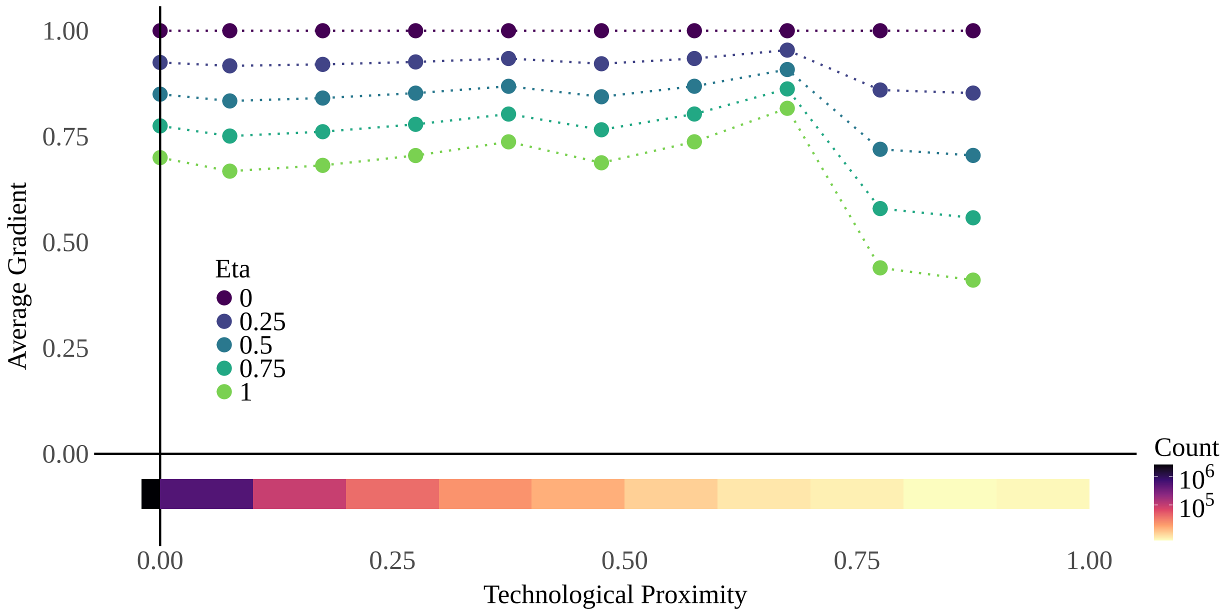
$$Y_i = \sum_{j \neq i} (\theta \cdot D_{i,j} - \eta \cdot G_{i,j}) W_j \ , \tag{3.11}$$

where we recall that $D_{i,j}$ and $G_{i,j}$ denote the technological and product market proximity between the firms i and j, respectively.[18] Assume also, for the sake of simplicity, that the treatments W_j are independent of

---

[18] In our notation, product market proximity is generated by $G_{i,j} = G(U_i, U_j)$. The product market proximity measure $G_{i,j}$ is given by the $G_{i,j} = \langle R_i, R_j \rangle / \|R_i\|_2 \|R_j\|_2$, where R_i denotes the share of firm i sales across four digit industry codes. Thus, the variable R_i is a component of the variable U_i.

FIGURE 3. Confounding in a Calibrated Example



Notes: Figure 3 displays estimates of the mixed-partial derivative (3.12), under the calibrated model (3.11), averaged over bins of the support of technological proximity. We estimate $\partial_\delta \mathbb{E}[G_{i,j} \mid D_{i,j} = \delta]$ at each bin center with a local-linear regression using a triangular kernel with bandwidth 0.1. We set the coefficient $\theta = 1$. Each series corresponds to a specified value of the coefficient η, which parametrizes the extent of the confounding with product market proximity. A heatmap measuring the number of pairs of units in each bin are displayed below the x-axis.

---

variables S_j and U_j. As the magnitude of the coefficient η increases in proportion to the coefficient θ, the confounding effect of product market proximity increases. To see this, Figure 3 displays estimates of the value of the quantity

$$\partial^2_{\delta,w}\mathbb{E}[Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, W_j = w] = \theta - \eta \cdot \partial_\delta \mathbb{E}[G_{i,j} \mid D_{i,j} = \delta] \tag{3.12}$$

averaged over bins of the support of technological proximity using the sample from Bloom et al. (2013). We set $\theta = 1$ and vary the value of η. By contrast, the average spillover proximity gradient satisfies

$$\mathbb{E}[\partial^2_{\delta,w} Y_{i,j}(\delta, w)] = \theta \tag{3.13}$$

irrespective of the coefficient η. Thus, in this sense and in this calibrated example, estimates of the effect of technological proximity on productivity spillover intensity that do not account for the potential that product market proximity affects spillover intensity can be severely negatively biased. ■

It is worth pausing to emphasize what spillover proximity gradients do and do not measure. First, averages of spillover proximity gradients are related to the "average indirect effect" and "marginal spatial effect,"

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\partial_w Y_{i,j}(D_{i,j}, w) \quad \text{and} \quad \mathbb{E}[\partial_w Y_{i,j}(D_{i,j}, w) \mid D_{i,j} = \delta]\,, \tag{3.14}$$

studied by Hu et al. (2022); Munro et al. (2025); Li and Wager (2022) and Pollmann (2023); Wang et al. (2025), respectively. The main difference, in both cases, is that we explicitly consider interventions to the proximity measure $D_{i,j}$. Second, spillover proximity gradients are related to the "average exposure contrasts," considered by, e.g., Aronow and Samii 2017, Sävje et al. 2021, Auerbach and Tabord-Meehan 2025, and Leung and Loupos 2025, where the main interest is in intervening on a low dimensional summary of the treatments and connections of "neighboring" units. Here, the low dimensional summary is often chosen to parameterize (or approximate) the global structure of interference across outcomes. By contrast, our aim is estimate average changes in one unit's outcome associated with changes to the treatment and proximity of another unit, while remaining agnostic, a priori, toward the global structure of interference.

Finally, in some cases, coefficients obtained from regressions of the form (2.1) have been interpreted as estimates of "total effects"; that is, parameters measuring the effect of changing the treatment of all units, in unison. Theorem 2.1 indicates that, without placing further structure on the data at hand, estimators based on specifications of the form (2.6) target a more modest quantity, again relating to the effect of changing the proximity and treatment of only one unit. Indeed, although, as we will see, average spillover proximity gradients can be estimated nonparametrically, recovering richer objects demands the imposition of further structure, either on the distribution of the treatments or on how how interference aggregates across units.[19]

4. Identification and Estimation

We propose a method for identifying and estimating averages of spillover proximity gradients. The core idea is to adjust the regressions considered in Section 2 by residualizing the treatment and proximity measure of interest with appropriately chosen covariates and auxiliary proximity measures. We give conditions that rationalize estimation strategies with this structure.

4.1 Regression Adjustment

Any approach for recovering averages of spillover proximity gradients must account for potential endogeneity in both the treatments and proximity measures of interest. We consider strategies that address these two sources of confounding by separately residualizing the two associated variables.

To this end, we assume that we additionally observe the vectors of covariates X_i and auxiliary proximity measures $G_{i,j}$. For instance, in Example 1, the covariates X_i might denote features of the firm i, like lagged values of R&D expenditure, and the proximity measure $G_{i,j}$ might denote the product market proximity and geographic distance between the firms i and j. Define the vector $H_{i,j} = (X_i, G_{i,j}, X_j)$. We impose the following convention to fit these measurements into the model proposed in Section 3. Let the vector

$$\overline{S}_i = (S_i, U_i) \tag{4.1}$$

[19] There are many papers in the network interference literature that aim to estimate total effects. In most cases, this is enabled by restricting attention to settings where the realized treatments are clustered (see e.g., Leung 2022b; Faridani and Niehaus 2022; Viviano et al. 2023) or by restricting the ways in which interference aggregates across units (Munro, 2025).

collect the two latent factors associated with unit i.

Assumption 4.1. *There exists a function $H(\cdot,\cdot)$ such that $H_{i,j} = H(\overline{S}_i, \overline{S}_j)$.*

Assumption 4.1 stipulates that the covariates $H_{i,j}$ are functions of the unit-specific factors $\overline{S}_i$ and $\overline{S}_j$. This condition rules out covariates that summarize features of units other than i and j.

The proposed method has two steps. First, the user constructs estimates of the conditional expectations

$$\pi_n(x) = \mathbb{E}[W_j \mid X_j = x] \quad \text{and} \quad \gamma_n(h) = \mathbb{E}[D_{i,j} \mid H_{i,j} = h] \,, \tag{4.2}$$

respectively. Denote these estimates by $\hat{\pi}_n(\cdot)$ and $\hat{\gamma}_n(\cdot)$. Second, the user computes the coefficients from the regression specification

$$Y_i = \alpha + \theta \cdot \hat{\Delta}_i^\star + \varepsilon_i \,, \quad \text{where} \quad \hat{\Delta}_i^\star = \sum_{j \neq i} \hat{D}_{i,j}^\star \hat{W}_j^\star \,. \tag{4.3}$$

Here, the variables

$$\hat{W}_i^\star = W_i - \hat{\pi}_n(X_j) \quad \text{and} \quad \hat{D}_{i,j}^\star = D_{i,j} - \hat{\gamma}_n(H_{i,j}) \tag{4.4}$$

denote residualized versions of the treatment and proximity measure of interest.[20] In this section, we give conditions under which endogeneity in the treatments and proximity measures is addressed by the use of the residualized variables (4.4).

Remark 4.1. The estimator obtained by solving the residualized regression (4.3) differs from the types of regressions usually encountered in empirical practice. In particular, specifications of the form

$$Y_i = \alpha + \theta \cdot \Delta_i + \xi \cdot \Gamma_i + \varepsilon_i \,, \quad \text{where} \quad \Delta_i = \sum_{j \neq i} D_{i,j} W_j \quad \text{and} \quad \Gamma_i = \sum_{j \neq i} G_{i,j} W_j \,, \tag{4.5}$$

are considered by Conley and Udry (2010), Bloom et al. (2013), Acemoglu et al. (2016a), and Lerche (2025), among many others. (Here, for the sake of simplicity, we have omitted consideration of the covariates X_i.) In contrast to cross-sectional regressions, the residualized spillover regression (4.3) and the "long" spillover regression (4.5) are not numerically equivalent and do not identify the same parameter, even when the expectation of $D_{i,j}$ conditional on $G_{i,j}$ is linear. In Appendix E.3, we show that, under this linearity condition, the two regressions identify the same parameter only if treatments are homoskedastic. ■

We focus the majority of our discussion on addressing endogeneity in measures of proximity, as our approaches for handling endogeneity in treatments are standard. In particular, in the main text, we assume that treatments are as-good-as randomly assigned conditional on the covariates X_j. In Appendix E.4, we show that this condition can be replaced with analogous assumptions based on parallel trends or instrumental variable identification strategies by making appropriate changes to the structure of the estimator (4.3).

[20] In practice, different sets of covariates X_i can be used to residualize the treatments and proximity measures. In particular, in settings where the coordinates S_i are observed, they can be used to residualize the treatments, but should not be used (without being coarsened in some way) to residualize the proximity measures, as otherwise there would be no variability in the residuals $\hat{D}_{i,j}^\star$. We assume that the same set of covariates are used in both cases, for the sake of simplicity.

Assumption 4.2 (Conditional Ignorability in Treatments). *The replicates $(W_i, \overline{S}_i)_{i=1}^n$ are i.i.d. and satisfy the condition*

$$\mathbb{E}[W_j \mid X_j, \overline{S}_j] = \mathbb{E}[W_j \mid X_j] \tag{4.6}$$

almost surely.

Assumption 4.2 is a direct generalization of Assumption 2.2. Like Assumption 2.2, cross-sectional dependence in the treatments has been ruled out.[21] In this case, however, the means of the treatments are allowed to vary with the latent factor, so long as this heterogeneity can be explained by the observed covariates X_j.

4.2 Decomposition, Redux

To discipline our discussion, we first state an extension to Theorem 2.1. In particular, we consider the infeasible objective function

$$L_n^\star(\theta) = \min_\alpha \sum_{i=1}^n (Y_i - \alpha - \theta \cdot \Delta_i^\star)^2 \ , \quad \text{where} \quad \Delta_i^\star = \sum_{j \neq i} D_{i,j}^\star W_j^\star \tag{4.7}$$

is constructed using the oracle residuals

$$W_i^\star = W_i - \pi_n(X_i) \quad \text{and} \quad D_{i,j}^\star = D_{i,j} - \gamma_n(H_{i,j}) \ . \tag{4.8}$$

We give conditions under which regressions of the form (4.7) identify convex averages of mixed-partial derivatives of the conditional expectation

$$\mu(\delta, w \mid H_{i,j}, \overline{S}_j) = \mathbb{E}[Y_i \mid D_{i,j} = \delta, W_j = w, H_{i,j}, \overline{S}_j] \ . \tag{4.9}$$

In Section 5, in turn, we give conditions under which the coefficient obtained from the regression (4.3) is consistent for the parameter that minimizes the expectation of the objective function (4.7).

We require a final condition that justifies our approach to residualizing the proximity measure. In particular, we assume that any unobserved heterogeneity in the relationship between the proximity measure and the covariates is additively separable.

Assumption 4.3 (Additively Separable Heterogeneity). *There exists a function $\psi_n(\cdot)$ such that*

$$\mathbb{E}[D_{i,j} \mid H_{i,j}, S_j] = \gamma_n(H_{i,j}) + \psi_n(S_j) \tag{4.10}$$

almost surely.

[21] Peer effects in treatment selection are considered by Forastiere et al. (2021), Ogburn et al. (2024), and Emmenegger et al. (2025), who study models where treatment is as-good-as randomly assigned conditional on a low-dimensional vector collecting features of neighboring units. Leung and Loupos (2025) consider a more general model where treatments are jointly determined, conditional on an observed network and covariates. In each case, in contrast to the regression-based approaches at the center of this paper, explicit specification, or estimation, of the cross-sectional dependence in treatments and network dependence across outcomes is required.

This condition allows us to residualize the proximity measures $D_{i,j}$ with an estimate of the expectation $\gamma_n(H_{i,j}) = \mathbb{E}[D_{i,j} \mid H_{i,j}]$, rather than with an estimate of the expectation $\mathbb{E}[D_{i,j} \mid H_{i,j}, S_j]$. The former can be obtained using the full sample, whereas the latter would need to be re-estimated for each unit j.[22] In particular, recall from the discussion in Remark 2.2 that regressions with the structure (2.6) effectively center the proximity measures $D_{i,j}$ conditional on the latent factors S_j. Regressions with the structure (4.7) center the residualized proximity measures $D^\star_{i,j}$ analogously. Thus, Assumption 4.3 ensures that, by centering the residual $D^\star_{i,j}$ conditional on the latent factor S_j, we effectively center the proximity measure $D_{i,j}$, conditional on both the covariates $H_{i,j}$ and the latent factor S_j.

The following Theorem gives a direct generalization of Theorem 2.1. As before, we require several moment conditions. We additionally impose a mild generalization of Assumption 2.3. In each case, we now condition on the covariates $H_{i,j}$. Again, we defer the statement of these conditions to Appendix A.

Theorem 4.1. *Assume that the treatments and proximity measures have support contained in intervals $\overline{\mathcal{W}}$ and $\overline{\mathcal{D}}$ of constant length and that the outcomes are identically distributed. Under Assumptions 2.1 and 4.1 to 4.3, and two regularity conditions stated in Appendix A, the parameter*

$$\overline{\theta}^\star_n = \arg\min_\theta \mathbb{E}[L^\star_n(\theta)] \tag{4.11}$$

admits the representation

$$\begin{aligned} \overline{\theta}^\star_n &= \theta^\star_n + o(n^{-1/2})\,, \quad \text{where} \\ \theta^\star_n &= \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}\Big[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j)\, \partial^2_{\delta,w}\, \mu(\delta, w \mid H_{i,j}, \overline{S}_j)\Big]\, \mathrm{d}w\, \mathrm{d}\delta\,, \end{aligned} \tag{4.12}$$

and the weights

$$\lambda(\delta, w \mid H_{i,j}, \overline{S}_j) = \frac{\mathrm{Cov}(\mathbb{I}\{D_{i,j} \geq \delta\}, D_{i,j} \mid H_{i,j}, S_j)\, \mathrm{Cov}(\mathbb{I}\{W_j \geq w\}, W_j \mid \overline{S}_j)}{\mathbb{E}[\mathrm{Var}(D_{i,j} \mid H_{i,j}, S_j)\, \mathrm{Var}(W_j \mid \overline{S}_j)]} \tag{4.13}$$

are convex, in the sense that they are positive almost surely and integrate to one in expectation.

Corollary 4.1. *Assume that the conditions of Theorem 4.1 hold. If, additionally, Assumption 3.1 holds and the potential outcomes $(\delta, w) \mapsto Y_{i,j}(\delta, w)$ have continuous first and second partial derivatives, then the parameter $\theta^\star_n$, defined in (4.12), can be re-expressed as*

$$\theta^\star_n = \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}\Big[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j)\, \partial_\delta\, \mathbb{E}[\partial_w\, Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, H_{i,j}, \overline{S}_j]\Big]\, \mathrm{d}w\, \mathrm{d}\delta\,, \tag{4.14}$$

where the weight function $\lambda(\delta, w \mid H_{i,j}, \overline{S}_j)$ is defined in (4.13).

[22] If Assumption 4.3 is unlikely to be accurate, then the results stated in this section will hold, unchanged, if an estimate of $\mathbb{E}[D_{i,j} \mid H_{i,j}, S_j]$ is used to residualize the proximity measure $D_{i,j}$, in place of the conditional expectation $\gamma_n(H_{i,j})$. Such an estimate can be obtained by regressing $D_{i,j}$ on $H_{i,j}$ separately for each unit j. In our empirical applications, we prefer to impose Assumption 4.3 because such unit-by-unit estimates are less precise and are more computationally expensive.

Theorem 4.1 is entirely analogous to Theorem 2.1, except that every quantity now additionally conditions on the covariates $H_{i,j}$ and the latent factor $\overline{S}_j$.[23] Corollary 4.1, says that, if additionally Assumption 3.1 holds, then regressions of the form (4.7) identify convex averages of the parameter

$$\partial_\delta \, \mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, H_{i,j}, \overline{S}_j] \; . \tag{4.15}$$

As a consequence, if the covariates $H_{i,j}$ and latent coordinate $\overline{S}_j$ are sufficient to hold constant features of the unit i that might covary with both its proximity $D_{i,j}$ to unit j and its response to j's treatment, then the regression specification (4.7) recovers convex averages of spillover proximity gradients.

4.3 Conditional Ignorability

Our baseline approach for addressing endogeneity in measures of proximity centers on the following conditional ignorability assumption.

Assumption 4.4 (Conditional Ignorability in Proximity)**.** *The potential spillover effect $\partial_w Y_{i,j}(\delta, w)$ associated with an intervention to the treatment of unit j on the outcome of unit i is mean independent of the realized proximity measure $D_{i,j}$, conditional on the covariates $H_{i,j}$ and the latent factor $\overline{S}_j$; that is*

$$\mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, H_{i,j}, \overline{S}_j] = \mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid H_{i,j}, \overline{S}_j] \tag{4.16}$$

almost surely for each δ, w in the support of $D_{i,j}$ and W_j.

Under Assumption 4.4, the parameter $\theta_n^\star$, targeted by regressions with the structure (4.7), identifies a convex average of spillover proximity gradients.

Corollary 4.2. *Assume that the conditions of Corollary 4.1 hold. If, additionally, Assumptions 3.2 and 4.4 hold, then the parameter $\theta_n^\star$, defined in (4.12), admits the representation*

$$\theta_n^\star = \int_{\mathcal{D}} \int_{\mathcal{W}} \mathbb{E}\Big[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j) \partial^2_{\delta,w} Y_{i,j}(\delta, w)\Big] \, \mathrm{d}w \, \mathrm{d}\delta \; , \tag{4.17}$$

where the weight function $\lambda(\delta, w \mid H_{i,j}, \overline{S}_j)$ is defined in (4.13).

Example 1 (Continued)**.** In the context of Bloom et al. (2013), the condition (4.16) stipulates that, if we fix a firm j, once we account for any observable features X_i and, e.g., the product market proximity $G_{i,j}$, then technological proximity $D_{i,j}$ doesn't vary systematically with any other quantities that might affect how firm i's value responds to R&D by firm j. This condition is satisfied, for instance, in the model (3.11) used to generate Figure 3. Bloom et al. (2013) use language suggestive of this assumption, arguing that their

[23] As it can be shown that $\int\int \lambda(\delta, w \mid H_{i,j}, \overline{S}_j) \, \mathrm{d}w \, \mathrm{d}\delta = \mathrm{Var}(D_{i,j} \mid H_{i,j}, S_j) \, \mathrm{Var}(W_j \mid \overline{S}_j)$, the weighted average (4.17) tends to place larger weight on the pairs of units i, j associated with larger values of $\mathrm{Var}(D_{i,j} \mid H_{i,j}, S_j)$. This echos classical results from Angrist (1998) and Angrist and Krueger (1999) associated with contexts without interference.

identification rests on “substantial independent variation in the two measures,” namely product market and technological proximity.[24] ■

Assumption 4.4 asks that the covariates $H_{i,j}$ stratify units i in terms of the characteristics that influence the intensity of the spillover effects they receive from j's treatment, up to the realized proximity measure $D_{i,j}$. To visualize this restriction, return to Panel A of Figure 2. The figure displays three coordinates S_i, S_j, and S_k. Assume that i and k have the same covariates and the same relation to j in terms of the auxiliary proximity measures, that is $G_{i,j} = G_{k,j}$. Assumption 4.4 implies that the difference in the proximity measures $D_{i,j}$ and $D_{k,j}$ cannot be correlated with any differences in the characteristics of the units i and k that influence the intensity of the spillover effects that they receive from the unit j. So, for instance, if all else equal, units that are further north tend to receive larger spillovers from unit j (perhaps, because they are closer in terms of some unobserved characteristic), and also tend to be closer to S_j, then, unless the covariates $H_{i,j}$ and $\overline{S}_j$ fix unit i's north-south coordinate, Assumption 4.4 will not hold.

As in other contexts in causal inference, conditional ignorability is a strong assumption. That is, in any particular setting, it may not be too difficult to hypothesize further channels that might also mediate spillover effects. For instance, Lychagin et al. (2016), incorporate geography into the empirical framework developed in Bloom et al. (2013), arguing that technological and geographic proximity are highly correlated and that R&D spillovers are potentially mediated by geographic proximity (see also Arque-Castells and Spulber (2022) and Zacchia (2020) who study how R&D spillovers vary with voluntary technology transfers and patent co-authorship networks, respectively). Nevertheless, we center our consideration on conditional ignorability because it represents the simplest departure from the assumption that spillovers are entirely mediated by the proximity measure of interest that enables identification, and because, in our view, it best reflects the approaches to disentangling spillover effects that have been taken in empirical practice. For instance, Lerche (2025) measures how the spillover effects of investment tax credits vary with input-output linkages, but controls for proximity in labor market flows. See also Conley and Udry (2010), Acemoglu et al. (2016a), Rotemberg (2019), Helm (2020), and Cai and Szeidl (2024) who consider similar problems in other settings.

Moreover, conditioning on observables is an essential building block for more nuanced identification strategies. Concretely, in Appendix E.5, we demonstrate how to adapt our framework to treat settings where suitable instrumental variables are available. For example, Fafchamps and Quinn (2018) conduct an experiment where social connections between managers of African manufacturing firms are exogenously formed. They then measure how business practices propagate through the exogenous linkages. Roughly speaking, under restrictions resembling standard conditions for the validity of linear two-stage least-squares estimators (Imbens and Angrist, 1994; Angrist et al., 2000), the residualized proximity measures can be replaced in the estimator (4.3) with their values predicted on the basis of the instrument.

[24] Bloom et al. (2013) support this assertion with a series of illustrative examples, writing e.g., “In our sample period, IBM was technologically close to Intel, Motorola, and Apple . . . However, while IBM is close to Apple in product market space, . . . [it] is not close to Intel and Motorola.”

5. CONSISTENCY

In this section, we give conditions under which the class of estimators proposed in Section 4 are consistent for convex averages of spillover proximity gradients. We focus our consideration on settings where each unit's outcome can be potentially affected by the treatments, locations, and unobserved characteristics of many other units. We show that, in such settings, coefficients obtained from regression specifications of the form (4.3) converge at a rate strictly slower than $O_p(n^{-1/2})$. In particular, the rate of convergence achieved by the estimator under consideration decreases as the extent of cross-unit interference increases. Nevertheless, we establish that, under our regularity conditions, this sub-parametric rate is optimal, in a minimax sense.

Throughout, the quantities c and C denote universal, positive constants, whose values are allowed to change in each appearance. For two sequences $f(n)$ and $g(n)$, the relations $f(n) \lesssim g(n)$ and $f(n) \asymp g(n)$ indicate that $f(n) \leq Cg(n)$ and $cg(n) \leq f(n) \leq Cg(n)$ for all sufficiently large n, respectively.

5.1 Linear Residualization

To facilitate these results, and in keeping with our emphasis on regression-based estimation procedures, we restrict attention to settings where the estimators $\hat{\pi}_n$ and $\hat{\gamma}_n$ are obtained from least-squares regressions. In particular, with an abuse of notation, we consider the estimator

$$\hat{\theta}_n^\star = \arg\min_\theta \min_\alpha \sum_{i=1}^n (Y_i - \alpha - \theta \cdot \hat{\Delta}_i^\star)^2 \, , \quad \text{where} \quad \hat{\Delta}_i^\star = \sum_{j \neq i} \hat{D}_{i,j}^\star \hat{W}_j^\star \tag{5.1}$$

has been constructed using the empirical residuals

$$\hat{W}_i^\star = W_i - \hat{\pi}_n^\top X_i \quad \text{and} \quad \hat{D}_{i,j}^\star = D_{i,j} - \hat{\gamma}_n^\top H_{i,j} \tag{5.2}$$

centered with the linear regression coefficients

$$\hat{\pi}_n = \arg\min_\pi \sum_{i=1}^n (W_i - \pi^\top X_i)^2 \quad \text{and} \quad \hat{\gamma}_n = \arg\min_\gamma \sum_{i=1}^n \sum_{j \neq i}^n (D_{i,j} - \gamma^\top H_{i,j})^2 \, , \tag{5.3}$$

respectively. Here, we adopt the convention that both X_i and $H_{i,j}$ contain intercepts.

Moreover, we assume that the estimators (5.3) are well-specified.

Assumption 5.1 (Fixed Dimension, Linear Nuisance Parameters)**.** *The nuisance parameters $\pi_n(x)$ and $\gamma_n(h)$ are linear in their arguments and the covariate vector $H_{i,j}$ has a fixed dimension.*

The linearity enforced by Assumption 5.1 plays an essential role in our arguments. In settings without interference, the use of more generic nuisance parameter estimators is increasingly common. Often, such approaches are enabled by the use of sample-splitting (Chernozhukov et al., 2018). Sample-splitting is not generally appropriate for our setting, because the structure of the cross-sectional dependence across units is not known a priori.[25] As a practical matter, the covariate vector $H_{i,j}$ can be constructed by taking a basis expansion or binning the observable covariates, so long as the number of components remains small.

5.2 Regularity Conditions

Our asymptotic analysis centers on the idea that cross-unit interference is non-negligible only for pairs of units that are "proximate," either through the measure of interest or through other observed or unobserved characteristics. To give structure to this idea, we assume that there exists a symmetric, binary-valued function

$$A_{i,j} = A_n(\overline{S}_i, \overline{S}_j) = A_n(\overline{S}_j, \overline{S}_i) \in \{0, 1\} \tag{5.4}$$

of the latent factors that indicates whether the units i and j are proximate. The latent proximity indicator $A_{i,j}$ is used to encode two relationships: interference arises only among proximate pairs of units, and the measure of interest $D_{i,j}$ is nonzero only for such pairs. For the latter, the converse is not necessarily true. We express these assumptions formally as follows.

Assumption 5.2 (Latent Spatial Structure)**.** *Define the sets*

$$L_n(S_j) = \{s : A_n(s, \overline{S}_j) = 1\} \quad \textit{and} \quad L_n^{(1)}(\overline{S}_i, \overline{S}_j) = L_n(\overline{S}_i) \cap L_n(\overline{S}_j) \tag{5.5}$$

for each pair of latent factors $\overline{S}_i$, $\overline{S}_j$. *There exists a sequence* ρ_n, *with* $\rho_n \to 0$ *and* $n\rho_n \to \infty$, *that satisfies*

$$P\{\overline{S}_j \in L_n(\overline{S}_i) \mid \overline{S}_i\} \asymp \rho_n \quad \textit{and} \quad P\{P\{\overline{S}_k \in L_n^{(1)}(\overline{S}_i, \overline{S}_j) \mid \overline{S}_i, \overline{S}_j\} > 0\} = O(\rho_n)\ , \tag{5.6}$$

uniformly almost surely, such that:

(i) *Outcomes are nearly mean independent of the states of distant units, in the sense that*

$$\textit{if} \quad A_{i,j} = 0, \quad \textit{then} \quad \mathbb{E}[Y_i \mid Z_j, Z_i] - \mathbb{E}[Y_i \mid Z_i] = o(\rho_n^{1/2}) \tag{5.7}$$

uniformly almost surely.

(ii) *The proximity measure* $D_{i,j}$ *is equal to zero for distant units and is non-negligible, with a non-vanishing probability, for proximate units, in the sense that*

$$P\{D_{i,j} = 0 \mid A_{i,j} = 0\} = 1 \quad \textit{and} \quad \mathrm{Var}(D_{i,j} \mid A_{i,j} = 1, S_j) \asymp 1 \tag{5.8}$$

hold uniformly almost surely.

(iii) *Covariates* $H_{i,j}$ *associated with distant units are not predictive of proximity, in the sense that*

$$\textit{if} \quad A_{i,j} = 0\ , \quad \textit{then} \quad \mathbb{E}[D_{i,j} \mid H_{i,j}, S_j] = O(\rho_n)\ , \tag{5.9}$$

uniformly almost surely.

---

[25] Leung and Loupos (2025) consider a more complicated nuisance parameter estimator in a setting with interference. Their argument is not directly applicable to our setting, as our data, nuisance parameters, and estimator have different structures.

The sequence ρ_n measures the expected fraction of units that are proximate to any given unit j. In the first part of the relation (5.6), we have imposed the condition that this fraction can vary across units only up to a constant factor. The conditions $\rho_n \to 0$ and $n\rho_n \to \infty$ restrict our consideration to asymptotic sequences in which the number of units that are proximate to each given unit grows without bound, but remains small relative to the sample size. The second part of the condition (5.6) stipulates that the fraction of pairs of units i, j that could feasibly be proximate to a third unit k is also of order $O(\rho_n)$. This restriction should be interpreted as encoding the idea that i and j can only *ever* share a common neighbor (in the sense that $P\{\overline{S}_k \in L_n^{(1)}(\overline{S}_i, \overline{S}_j) \mid \overline{S}_i, \overline{S}_j\}$ has a positive probability) when they themselves are sufficiently "close" in the latent space. This event, that i and j are sufficiently close to feasibly share a neighbor, cannot, in turn, be *much* more likely to occur than the event that i and j are, themselves, neighbors.

For the sake of concreteness, consider the setting where the latent proximity indicators are generated by

$$A_n(\overline{S}_i, \overline{S}_j) = \mathbb{I}\{\|\overline{S}_i - \overline{S}_j\| \leq \phi_n\} \ , \tag{5.10}$$

where $\|\cdot\|$ is some norm (or semi-norm) on the latent factors and ϕ_n is a sequence of bandwidths. Networks with the structure (5.10) are referred to as "random geometric graphs," and are widely studied in the network interference literature (Penrose, 2003; Leung, 2022a). In this case, the second part of the condition (5.6) will be satisfied when the measure of the latent factors is doubling with respect to the norm $\|\cdot\|$, in the sense that $P\{\|\overline{S}_i - \overline{S}_j\| \leq 2\phi_n\} \lesssim P\{\|\overline{S}_i - \overline{S}_j\| \leq \phi_n\}$. Doubling is a common feature of well-behaved measures, including distributions with densities bounded above and below on finite-dimensional Euclidean spaces (Heinonen, 2001). In this sense, the condition (5.6) asks that the latent proximity indicator $A_{i,j}$ behaves like its name, that is, it should reflect an underlying measure of proximity.[26]

Part (i) of Assumption 5.2 stipulates that only units that are proximate to a given unit can have non-negligible effects on the expected value of its outcome. This restriction is closely related to the "Approximate Neighborhood Interference" framework developed in Leung (2022a), and adopted in, e.g., Leung (2022b) and Leung and Loupos (2025).[27] Crucially, in our setting, interference is parameterized by the latent proximity measure $A_{i,j}$, rather than through the observed proximity measure $D_{i,j}$, per se. That is, we explicitly allow for spillover effects, and other forms of interference, to be mediated by other observed or unobserved characteristics. Like in Leung (2022a), because we have not imposed the restriction that outcomes that are sufficiently far away are independent, our framework accommodates general forms of "endogenous" spillover effects, in the sense of Manski (1993).

[26] In the region $n^{-1/2} \lesssim \rho_n$, the second part of the relation (5.6) will fail in sparse graphon models without a geometric structure (like those considered in Li and Wager (2022)). That is, in that context, in general, the size of the intersection of the neighborhoods of two randomly drawn nodes is of order $n\rho_n^2$.

[27] The Approximate Neighborhood Interference framework is more expressive than Assumption 5.2, in that, there, consistency is established by measuring how the variability in a unit's outcome changes as one progressively resamples the treatments of all units whose distances exceed an increasing threshold. By contrast, our approach (effectively) fixes a single threshold and measures dependence by resampling the latent factor of just one unit whose distance exceeds that threshold.

In each of our running examples, the proximity measure $D_{i,j}$ is equal to zero for most pairs of units i, j. That is, the measure $D_{i,j}$ is sparse. For instance, in our treatment of the data associated with Example 1, 92% of pairs firms never file patents in the same technology class, and so their technological proximity is equal to zero.[28] Parts (ii) and (iii) of Assumption 5.2 relate the indicator $A_{i,j}$ to the sparsity of the proximity measure of interest and its association with the covariates. In particular, Part (ii) requires that the measure $D_{i,j}$ is only ever non-zero for units whose latent factors are proximate. Part (iii), in turn, requires that values of the covariates $H_{i,j}$ that predict large values of $D_{i,j}$ only occur among proximate units. In other words, the binary, latent proximity measure $A_{i,j}$ respects the structure of the observed proximity measures $D_{i,j}$ and $H_{i,j}$, while allowing for interference between units that are only proximate through unobserved characteristics.

Likewise, many of the covariates $H_{i,j}$ that are potentially relevant to our running examples are equal to zero for the majority of pairs of units. To accommodate such situations, we require the following condition concerning the sparsity and correlation of the components of the covariates.

Assumption 5.3 (Well-Behaved Covariate Sparsity)**.**
(i) *Each component of the covariate vector $H_{i,j} = (H_{i,j}^{(k)})_{k=1}^{p}$ is bounded by a constant almost surely. Moreover, each of the components of the covariate vector X_i has variance bounded below by a constant.*
(ii) *For each component $H_{i,j}^{(k)}$ of the covariate vector $H_{i,j}$ there exists a sequence $\kappa_{n,k} \gtrsim \rho_n$ such that*

$$P\{H_{i,j}^{(k)} \neq 0 \mid \overline{S}_j\} \gtrsim \kappa_{n,k} \ , \quad P\{H_{i,j}^{(k)} \neq 0 \mid \overline{S}_i\} \gtrsim \kappa_{n,k} \ , \quad \textit{and} \quad \mathrm{Var}(H_{i,j}^{(k)}) \gtrsim \kappa_{n,k} \ . \tag{5.11}$$

The heterogeneity in the sparsity, in turn, is well-behaved, in the sense that

$$\sum_{l=1}^{p} |K_{k,l}^{-1}| = O(1) \quad \textit{and} \quad \sum_{l=1}^{p} \sqrt{\frac{\kappa_{n,k}}{\kappa_{n,l}}} |K_{k,l}^{-1}| = O(1) \tag{5.12}$$

both hold for each k, where $K_{k,l} = \mathrm{Corr}(H_{i,j}^{(k)}, H_{i,j}^{(l)})$ denotes the k,l component of $K = \mathrm{Corr}(H_{i,j})$.

The restriction (5.11) stipulates that none of the covariates in $H_{i,j}$ are strictly more sparse than the proximity measure of interest. This condition is necessary to ensure that the variability in the estimator $\hat{\gamma}_n$ does not, on its own, determine the rate of convergence of the estimator $\hat{\theta}_n^{\star}$. The first part of the restriction (5.12) requires that different covariates are not so strongly collinear that the inverse covariance structure becomes unstable. The second part, in effect, ensures that each component of $\hat{\gamma}_n$ can be estimated at a rate that is determined by its own sparsity, rather than being driven by association with other covariates with different levels of sparsity.

In turn, we require a condition that limits the curvature of the function $F(\cdot)$.

Assumption 5.4 (Curvature)**.** *Let the variable Z_j' denote an independent copy of Z_j. Construct $Z^{(j)}$ by replacing Z_j with Z_j' in the collection $Z = (Z_i)_{i=1}^{n}$. Let $Z^{(j,k)}$ be constructed analogously, by replacing Z_j*

[28] In Examples 2 and 3, the measure $D_{i,j}$, which measures the probability of migrating or commuting between the pair of location, is smaller than 0.01 for 93% and 92% pairs of locations, respectively

and Z_k with Z'_j and Z'_k. Define the difference operators

$$\nabla_j f(Z) = f(Z) - f(Z^{(j)}) \quad \text{and} \quad \nabla^2_{j,k} f(Z) = \nabla_j \nabla_k f(Z) \,, \tag{5.13}$$

respectively, for any real-valued function $f(\cdot)$. The bounds

$$\begin{aligned} &\mathbb{E}[\mathbb{E}[(\nabla^2_{j,k} F(Z_i, Z_{-i}))^2 \mid A_{i,j} = 1, A_{i,k} = 1, Z^{(j,k)}]] = O(n^{-2}\rho_n^{-3}) \,, \\ &\mathbb{E}[\mathbb{E}[(\nabla^2_{j,k} F(Z_i, Z_{-i}))^2 \mid A_{i,j} = 1, A_{i,k} = 0, Z^{(j,k)}]] = O(n^{-2}\rho_n^{-1}) \,, \quad \text{and} \\ &\mathbb{E}[\mathbb{E}[(\nabla^2_{j,k} F(Z_i, Z_{-i}))^2 \mid A_{i,j} = 0, A_{i,k} = 0, Z^{(j,k)}]] = O(n^{-2}\rho_n) \end{aligned} \tag{5.14}$$

hold.

The quantity $\nabla^2_{j,k} F(Z_i, Z_{-i})$ characterizes how the association between i's outcome and j's state varies as k's state changes, and can be interpreted like a randomized version of a mixed-partial derivative. The first bound can be thought of as describing the proportion of pairs of units j, k proximate to i for which this curvature can be on the order of a constant. That is, the curvature can be non-zero and large for all pairs of units proximate to i, so long as $\rho_n \lesssim n^{-2/3}$. As the quantity ρ_n becomes larger, this proportion decreases. For instance, when $\rho_n = n^{-1/2}$, each unit will have, on average, n pairs of neighbors. The bound (5.14) implies $n^{1/2}$ of these pairs can be associated curvature on the order of a constant. The second two bounds pertain to the case that only j is proximate to i and that neither j nor k are proximate to i, respectively, and, consequently, get progressively stronger. We discuss this condition in further detail in Appendix E.6.[29]

Finally, we require a set of additional, mild moment bounds that restrict the variability of the outcomes on certain events. These bounds are satisfied if, for instance, the random variables $\mathbb{E}[Y_i | Z_i, Z_j]$ and $\mathbb{E}[Y_i | Z_i]$ are bounded almost surely. To facilitate exposition, we defer the statement of this assumption to Appendix C.

5.3 Consistency

The following Theorem demonstrates that, under the conditions developed in this section, the estimator $\hat{\theta}^\star_n$, obtained by minimizing the objective function (4.3), is consistent for the parameter $\theta^\star_n$, characterized in Theorem 4.1. The rate of convergence is determined by the sequence ρ_n, introduced in Assumption 5.2.

Theorem 5.1. *Suppose that the conditions of Theorem 4.1 hold. If, additionally, Assumption 3.1, Assumptions 5.1 to 5.4, and an additional regularity condition stated in Appendix C hold, then*

$$\hat{\theta}^\star_n = \theta^\star_n + O_p(\rho_n^{1/2}) \,, \tag{5.15}$$

where $\theta^\star_n$ is defined in (4.12).

[29] We note that Assumption 5.4 is satisfied by models where outcomes are given by structural equations of the form $Y_i = F_i(W_i, \frac{1}{n\rho_n}\Delta_i)$, where $F_i(\cdot,\cdot)$ is some unit-specific function, so long as the second derivative of $F_i(W_i, \cdot)$ does not grow too quickly. Models with this structure are considered in, e.g., Li and Wager (2022).

As a consequence, if the conditions of Corollary 4.2 are satisfied, the estimator $\hat{\theta}_n^\star$ is consistent for a convex averages of spillover proximity gradients. Observe that none of the conditions needed to establish identification are used to establish consistency. That is, under the assumptions developed in this section, even if the identifying conditions developed in Section 4 fail, the estimator $\hat{\theta}_n^\star$ will still consistently recover the descriptive parameter $\theta_n^\star$.

Remark 5.1. Theorem 5.1 follows by first showing that, under the maintained assumptions, the estimator $\hat{\theta}_n^\star$ can be approximated by the statistic

$$\frac{\sum_{i=1}^n Y_i \tilde{\Delta}_i^\star}{\sum_{i=1}^n (\tilde{\Delta}_i^\star)^2} , \quad \text{where} \quad \tilde{\Delta}_i = \sum_{j \neq i} \tilde{D}_{i,j}^\star W_j^\star \quad \text{and} \quad \tilde{D}_{i,j}^\star = D_{i,j} - \mathbb{E}[D_{i,j} \mid H_{i,j}, S_j] , \tag{5.16}$$

up to an error of order $O_p(n^{-1/2})$. Assumptions 5.1 and 5.3 are essential for this step. The numerator and denominator of (5.16), in turn, satisfy the approximations

$$\frac{1}{n^2} \sum_{i=1}^n Y_i \tilde{\Delta}_i^\star = \mathbb{E}[Y_i \tilde{D}_{i,j}^\star W_j^\star] + O_p(\rho_n^{3/2}) \quad \text{and} \tag{5.17}$$

$$\frac{1}{n^2} \sum_{i=1}^n (\tilde{\Delta}_i^\star)^2 = \mathbb{E}[\mathrm{Var}(D_{i,j} \mid H_{i,j}, S_j) \, \mathrm{Var}(W_j \mid \overline{S}_j)] + O_p(\rho_n^{3/2}) , \tag{5.18}$$

respectively. Roughly speaking, Theorem 5.1 then follows from the facts that the parameter $\theta_n^\star$ can be expressed as the quotient of the expectations in (5.17) and (5.18) and that the expectation in the approximation to the denominator can be bounded from below by ρ_n.

The most nonstandard aspect of this argument pertains to the approximation (5.17). This bound relies on the application of the Hoeffding type decomposition

$$Y_i = \mathbb{E}[Y_i \mid Z_i] + \sum_{j \neq i} (\mathbb{E}[Y_i \mid Z_j, Z_i] - \mathbb{E}[Y_i \mid Z_i]) + R_i \tag{5.19}$$

to approximate each outcome. We control the remainder terms R_i by pairing Assumption 5.4 with a suitable, higher-order Efron-Stein type inequality, due to Bobkov et al. (2019). After plugging the decomposition (5.19) into the numerator of (5.16), the leading term can be represented as a degenerate U-statistic of order three (and likewise for the leading term in the denominator). The variability of these quantities can then be controlled through Assumption 5.2. ■

5.4 Optimality

Theorem 5.1 indicates that the rate of convergence of the estimator $\hat{\theta}_n^\star$ is at most $\rho_n^{1/2}$. Recall that, by Assumption 5.2, the sequence ρ_n is asymptotically, strictly larger than $1/n$. That is, on the sequences under consideration, each unit has an increasing number of neighbors. Thus, the rate of convergence implied by Theorem 5.1 is strictly slower than $n^{-1/2}$, the rate usually achieved by estimates of scalar parameters in settings without interference.

In this section, we show that, despite this, the rate $\rho_n^{1/2}$ is the best that any estimator can achieve, in a minimax sense, in data satisfying our regularity conditions. For the purposes of this subsection, we restrict attention to the setting without covariates X_i or $G_{i,j}$. That is, we are not asking the more complicated question whether our estimator makes efficient use of the information made available through the covariates. Instead, we ask whether the estimator $\hat{\theta}_n^\star$ depends optimally on the extent of cross-unit interference.

To state the result, we require three additional pieces of notation. First, let the set $\mathcal{P}_n(\rho_n)$ denote the class of distributions for the collection of random variables $(Y_i, Z_i)_{i=1}^n$, such that the conditions of Theorem 5.1 are satisfied with the sequence ρ_n. Second, we let $\theta_n^\star(P)$ denote the parameter $\theta_n^\star$, evaluated at the distribution P in $\mathcal{P}_n(\rho_n)$. Third, we let the set $\hat{\Theta}$ collect all, potentially randomized, real-valued functions of the observable data $(Y_i, W_i, D_i)_{i=1}^n$, where the vector $D_i = (D_{i,j})_{j\neq i}$ collects the proximity between i and each other unit.[30]

With this in place, we obtain the following lower bound on the minimax risk

$$\mathfrak{M}_n(\rho_n) = \inf_{\hat{\theta}\in\hat{\Theta}} \sup_{P\in\mathcal{P}_n(\rho_n)} \mathbb{E}_P \left[(\hat{\theta} - \theta_n^\star(P))^2\right]^{1/2} , \tag{5.20}$$

where the dependence of the estimators $\hat{\theta}$ on the observable data is left implicit.

Theorem 5.2. *It holds that*

$$\mathfrak{M}_n(\rho_n) \gtrsim \rho_n^{1/2} , \tag{5.21}$$

for all sufficiently large n.

By Theorem 5.2, there are no estimators that converge to the parameter $\theta_n^\star$, in root-mean-squared error, uniformly more quickly, up to constant factors, than the regression-based estimator $\hat{\theta}_n^\star$ proposed in Section 4.

Remark 5.2. The lower bound (5.21) is obtained from an application of the Le Cam (1973) two-point method (see e.g., Wainwright 2019 for a textbook treatment). The argument follows by explicitly constructing two distributions in $\mathcal{P}_n(\rho_n)$ that are close in the total-variation norm, but are associated with diverging values of the parameter $\theta_n^\star(P)$. Our construction is based on partitioning the units into ρ_n^{-1} groups, such that units are only proximate to other units in their group, and specifying the outcomes and latent factors in such a way that units within the same group have exactly the same outcome. In effect, in this construction, there are ρ_n^{-1} i.i.d. observations, giving rise to the minimax rate $\rho_n^{1/2}$ in a natural way.[31] ■

6. INFERENCE

Existing approaches to inference in settings with interference take two routes. In some cases, units are partitioned into disjoint groups. If interference across groups is negligible, clustered standard errors can be

[30] The same result will hold, by an identical argument, if we allow the estimators in the set $\hat{\Theta}$ to use the coordinates S_i.

[31] A similar construction is noted, heuristically, in Li and Wager (2022), in order to suggest that the upper bounds achieved in that paper are unimprovable. In effect, Theorem 5.2 formalizes that observation. We note that the upper bounds in Li and Wager (2022), who consider a different but closely related setting, are restricted to the case that $n^{-1/2} \lesssim \rho_n$. By contrast, in our setting, we give matching upper and lower bounds over the full regime $n\rho_n \to \infty$.

constructed in the usual way (Leung, 2023). In others, interference is assumed to be bounded above by a decreasing function of an observed proximity measure, and so kernel-based methods are appropriate (see e.g., Kojevnikov 2021; Leung 2022a; Wang et al. 2025).

Neither approach is directly applicable to our setting. In particular, as we allow interference to be mediated by unobserved characteristics, units cannot necessarily be arranged into groups or otherwise organized around an observed proximity measure. Thus, in this section, we propose an alternative, resampling-based procedure for giving conservative estimates of the uncertainty associated with the estimators proposed in Section 4.

6.1 Construction

Our proposal is based on resampling the coefficient from a regression specification analogous to (5.1), where the outcomes have been replaced by the empirical residuals, in data where the signs of the (residualized) treatments have been flipped at random. Formally, define the empirical residual

$$\hat{\varepsilon}_i = Y_i - \hat{\alpha}_n - \hat{\theta}_n^\star \hat{\Delta}_i^\star \tag{6.1}$$

for each unit i in $1, ..., n$, where $\hat{\alpha}_n$ denotes the intercept from the regression (5.1) used to construct the estimator $\hat{\theta}_n^\star$. Let the vector $V = (V_i)_{i=1}^n$ collect a sequence of i.i.d. Rademacher random variables. That is, each coordinate V_i is independently, uniformly distributed on $\{-1, 1\}$. Let $\hat{\phi}_n^\star(V)$ denote the coefficient obtained from the regression specification

$$\hat{\phi}_n^\star(V) = \arg\min_{\phi} \min_{\alpha} \sum_{i=1}^n \left(\hat{\varepsilon}_i - \alpha - \phi \cdot \hat{\Delta}_i^\star(V)\right)^2 , \quad \text{where} \quad \hat{\Delta}_i^\star(V) = \sum_{j \neq i} \hat{D}_{i,j}^\star W_j^\star V_j \tag{6.2}$$

denotes a version of the treatment exposure $\hat{\Delta}_i^\star$ where the signs of the residualized treatments $\hat{W}_j^\star$ have been flipped at random. We give conditions under which the distribution of the coefficients $\hat{\phi}_n(V)$, over the randomness in the signs V, can be used to construct conservative estimates of the uncertainty associated with the estimator $\hat{\theta}_n^\star$.

In particular, let $V^{(1)}, \dots, V^{(B)}$ denote B independent copies of the vector V. We show that a conservative, asymptotically level α, confidence interval for the parameter $\theta_n^\star$ can be obtained by reporting the interval

$$\hat{\theta}_n^\star \pm z_{1-\alpha/2} \hat{\sigma}_n , \quad \text{where} \quad \hat{\sigma}_n^2 = \frac{2}{B} \sum_{b=1}^B \left(\hat{\phi}_n^\star(V^{(b)})\right)^2 \tag{6.3}$$

and $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of a standard Gaussian distribution. Likewise, the statistic $\hat{\sigma}_n$ can be used to report a conservative standard error. We show that scaling the variance estimator by a factor of 2, in this way, is essential, in the sense that, in some (non-pathological) configurations, the resultant confidence interval is asymptotically exact. This proposal is summarized in Algorithm 1.

Algorithm 1: Estimation and Inference

Input: Outcome Y_i, treatment W_i, and proximity measure of interest $D_{i,j}$

Covariates X_i and auxiliary proximity measures $G_{i,j}$

Compute the linear regression coefficients

$$(\hat{\alpha}_{w,n}, \hat{\pi}_n) = \arg\min_{\alpha,\pi} \sum_{i=1}^{n} (W_i - \alpha - \pi^\top X_i)^2 \quad \text{and}$$

$$(\hat{\alpha}_{d,n}, \hat{\gamma}_n) = \arg\min_{\alpha,\gamma} \sum_{i=1}^{n} \sum_{j\neq i}^{n} (D_{i,j} - \alpha - \gamma^\top H_{i,j})^2$$

and residuals

$$\hat{W}_i^\star = W_i - \hat{\alpha}_{w,n} - \hat{\pi}_n^\top X_i \quad \text{and} \quad \hat{D}_{i,j}^\star = D_{i,j} - \hat{\alpha}_{d,n} - \hat{\gamma}_n^\top H_{i,j} \ ,$$

where $H_{i,j} = (X_i, G_{i,j}, X_j)$

Compute the linear regression coefficients

$$(\hat{\alpha}_n, \hat{\theta}_n^\star) = \arg\min_{\alpha,\theta} \sum_{i=1}^{n} (Y_i - \alpha - \theta \cdot \hat{\Delta}_i^\star)^2 \ , \quad \text{where} \quad \hat{\Delta}_i^\star = \sum_{j\neq i} \hat{D}_{i,j}^\star \hat{W}_j^\star \ ,$$

and the residuals

$$\hat{\varepsilon}_i = Y_i - \hat{\alpha}_n - \hat{\theta}_n^\star \hat{\Delta}_i^\star$$

for b *in* $1, \ldots, B$ **do**

  Draw a vector $V^{(b)} = (V_i^{(b)})_{i=1}^n$ of i.i.d. Rademacher random variables

  Compute the linear regression coefficient

$$\hat{\phi}_n^\star(V^{(b)}) = \arg\min_{\phi} \min_{\alpha} \sum_{i=1}^{n} \left(\hat{\varepsilon}_i - \alpha - \phi \cdot \hat{\Delta}_i(V^{(b)})\right)^2 \ , \quad \text{where}$$

$$\hat{\Delta}_i^\star(V^{(b)}) = \sum_{j\neq i} \hat{D}_{i,j}^\star \hat{W}_j^\star V_j^{(b)}$$

end

Return the spillover proximity gradient estimate $\hat{\theta}_n^\star$ and conservative variance estimate

$$\hat{\sigma}_n^2 = \frac{2}{B} \sum_{b=1}^{B} \left(\hat{\phi}_n^\star(V^{(b)})\right)^2$$

Notes: Algorithm 1 details the construction of the residualized estimator of convex averages of spillover proximity gradients, and associated standard error and confidence interval. We recommend taking B on the order of 2000.

The confidence interval (6.3) is closely related to both the plug-in methods developed in Adao et al. (2019) and the randomization inference-based procedures considered by Borusyak and Hull (2023). Relative to the former procedure, which has a related structure but it not based on resampling, we show that our proposal is conservative even when the regression (5.1) is not well-specified. The latter procedure, in our context, would entail comparing the estimator $\hat{\theta}_n^\star$ to its randomization distribution constructed by repeatedly permuting the residualized treatments.[32] Methods of this form have the appeal of ensuring exact error control for tests of strong null hypotheses in finite samples (Ritzwoller et al., 2024), but, unlike our proposal, will not necessarily control the level of tests of the value of the parameter $\theta_n^\star$.

6.2 Gaussian Approximation

We begin by characterizing the limiting distribution of the estimator $\hat{\theta}_n^\star$. To facilitate exposition, We require three additional conditions. We defer the statements of these conditions to Appendix C. The first condition is a mild strengthening of the second relation in (5.6), stated in Assumption 5.2. The second ensures that, when suitably scaled, the variance of the leading term of the estimator $\hat{\theta}_n^\star$ is bounded away from zero. The final restriction is a mild strengthening of the moment bounds needed for Theorem 5.1.

Theorem 6.1. *Define the population residual and conditional expectation*

$$\varepsilon_i = Y_i - \alpha_n - \theta_n^\star \cdot \Delta_i^\star \quad \text{and} \quad \tilde{\varepsilon}_{i,j} = \mathbb{E}[\varepsilon_i \mid Z_j, Z_i] - \mathbb{E}[\varepsilon_i \mid Z_i] \; , \tag{6.4}$$

where α_n is the mean of Y_i and $\theta_n^\star$ is defined in Theorem 4.1. Define the sequence

$$\sigma_n^2 = \frac{1}{2\mathbb{E}[(\tilde{D}_{i,j}^\star W_j^\star)^2]^2} \operatorname{Var}(\mathbb{E}[\tilde{\varepsilon}_{k,i}\tilde{D}_{k,j}^\star \mid Z_i, Z_j] W_j^\star + \mathbb{E}[\tilde{\varepsilon}_{k,j}\tilde{D}_{k,i}^\star \mid Z_j, Z_i] W_i^\star) \; . \tag{6.5}$$

Suppose that the conditions of Theorem 5.1 hold. If three additional regularity conditions, stated in Appendix C, hold, then

$$\sigma_n^{-1}(\hat{\theta}_n^\star - \theta_n^\star) \xrightarrow{d} \mathsf{N}(0,1) \; , \quad \text{where} \quad \sigma_n \asymp \rho_n^{1/2} \; . \tag{6.6}$$

Remark 6.1. As discussed in Remark 5.1, the leading order terms of both the numerator and denominator of the estimator $\hat{\theta}_n^\star$ can be expressed as degenerate U-statistics of order three. In general, the limiting distributions of degenerate U-statistics are non-Gaussian, and instead can be represented as mixtures of chi-squared random variables, where the mixing weights are determined by the eigenvalues of each U-statistic's kernel (see e.g., Serfling 1980 for a textbook treatment). In effect, Theorem 6.1 follows by showing that, under our regularity conditions (particularly, Assumptions 5.2 and 5.4) these eigenvalues become sufficiently diffuse that the mixture of chi-squared variables converges to a Gaussian distribution (see Hall 1984 for discussion of arguments with this structure). Functionally, this step is enabled by a general Berry-Esseen type central limit theorem for degenerate U-statistics, given in Liu et al. (2025). ■

[32] Results similar to those given in this section for a related procedure where the residualized treatments in the regression (6.2) are permuted, rather than randomly sign-flipped, will follow from analogous methods, although the details of the supporting arguments would be substantially more involved.

It is worth emphasizing that the leading order term in the variance of the estimator $\hat{\theta}_n^\star$ is determined by the distribution of the residual-like quantity

$$\tilde{\varepsilon}_{i,j} = (\mathbb{E}[Y_i \mid Z_j, Z_i] - \mathbb{E}[Y_i \mid Z_i]) - \theta_n^\star \tilde{D}_{i,j}^\star W_j^\star \,, \tag{6.7}$$

defined in (6.4). This variable can be interpreted as the contribution of the unit j to unit i's outcome, net any linear dependence on the quantity $\tilde{D}_{i,j}^\star W_j^\star$.

6.3 Consistent Inference

The following theorem establishes that the limiting distribution of the regression coefficient $\hat{\phi}_n^\star(V)$, defined in (6.2), over the random signs V, is closely related to the limiting, unconditional distribution of the estimator $\hat{\theta}_n^\star$, characterized in Theorem 6.1. In particular, the conditional distribution of the sign-flipped coefficient can be used to bound the variance of the unconditional distribution of the baseline estimator.

Theorem 6.2. *Define the sequence*

$$\varphi_n^2 = \frac{1}{\mathbb{E}[(\tilde{D}_{i,j}^\star W_j^\star)^2]^2} \operatorname{Var}(\mathbb{E}[\tilde{\varepsilon}_{k,i} \tilde{D}_{k,j}^\star \mid Z_i, Z_j] W_j^\star) \tag{6.8}$$

and the cumulative distribution function

$$\hat{R}_n(x) = \frac{1}{2^n} \sum_{v \in \{-1,1\}^n} \mathbb{I}\{\varphi_n^{-1} \hat{\phi}_n^\star(v) \leq x\} \,. \tag{6.9}$$

If the conditions of Theorem 6.1 hold, then

$$\rho_n \lesssim \sigma_n^2 \leq 2\varphi_n^2 \lesssim \rho_n \quad \textit{and} \quad \hat{R}_n(x) \xrightarrow{p} \Phi(x) \tag{6.10}$$

for all $x \in \mathbb{R}$, where $\Phi(x)$ denotes the standard normal cumulative distribution function.

As the vector V is uniformly distributed on the set $\{-1,1\}^n$, the function $\hat{R}_n(\cdot)$ gives the cumulative distribution function, conditional on the observed data, for a suitably scaled version of the coefficient $\hat{\phi}_n^\star(V)$. The statement (6.10) implies that both the sampling law of the baseline estimator and the conditional law of the sign-flipped estimator are approximately normal, with scales ordered so that the former is no larger than twice the latter. As a consequence, the variability of the resampling distribution (6.9) can be used to provide an upper bound for the sampling variability of the baseline estimator. That is, using twice the resampling variance yields a conservative standard error and the confidence interval (6.3) controls the asymptotic level.

Observe that the difference between the variance σ_n^2 of the baseline estimator $\hat{\theta}_n^\star$ and the conditional variance φ_n^2 of the sign-flipped estimator $\hat{\phi}_n^\star(V)$ is given by the covariance

$$\frac{1}{\mathbb{E}[(\tilde{D}_{i,j}^\star W_j^\star)^2]^2} \operatorname{Cov}(\mathbb{E}[\tilde{\varepsilon}_{k,i} \tilde{D}_{k,j}^\star \mid Z_i, Z_j] W_j^\star, \mathbb{E}[\tilde{\varepsilon}_{k,j} \tilde{D}_{k,i}^\star \mid Z_j, Z_i] W_i^\star) \,. \tag{6.11}$$

As the two terms appearing in this covariance are identically distributed, the Cauchy-Schwarz inequality implies that the quantity (6.11) is bounded above by the variance φ_n^2. This bound yields the factor of two appearing in the relation (6.10) and the conservative variance estimate $\hat{\sigma}_n^2$.

The covariance (6.11) can be close (or equal) to the variance φ_n^2 in reasonable data generating distributions. The simulation considered in the following subsection gives a concrete illustration of this point. Heuristically, the covariance (6.11) is larger when the channels that mediate spillovers from j to i and from i to j are more correlated. That is, if the outcome Y_i depends on the treatment W_j though the (potentially, unobserved) proximity measure $G_{i,j}$, and likewise for Y_j, W_i, and $G_{j,i}$, then the covariance (6.11) increases with the correlation between the proximity measures $G_{i,j}$ and $G_{j,i}$.

6.4 Simulation

We conclude this section by reporting the results of a simple simulation, designed to illustrate the settings under which the variance estimator $\hat{\sigma}_n^2$ is more, or less, conservative. For the sake of simplicity, we assume that there are no covariates X_i or proximity measures $G_{i,j}$ and that the treatments W_i are uniformly distributed on $\{-1/2, 1/2\}$. In turn, we assume that the coordinates S_i are uniformly distributed on the integers $\mathcal{M}_n = \{0, \dots, m_n - 1\}$ and the proximity measure of interest is given by $D_{i,j} = \mathbb{I}\{S_i = S_j\}$. In this setting, the sequence ρ_n is given by m_n^{-1}. We take $m_n = \sqrt{n}$ for the sake of concreteness. The factors S_i and U_i are perfectly correlated, in the sense that all units with the same value of S_i will have the same value of U_i. The marginal distribution of the variables U_i is multinomial on $\mathcal{M}_n$, with parameter proportional to

$$p^{(\eta)} = (1-\eta)p^{(0)} + \eta p^{(1)} \ , \tag{6.12}$$

where $p^{(0)}$ places mass m_n^{-1} on each coordinate and $p^{(1)}$ places mass m_n^{-1} and $1 - m_n^{-1}$ on the coordinates zero and $m/2$, respectively.[33] Here, the free parameter η will be used to vary the correlation (6.11).

Spillovers are mediated by an additional, unobserved proximity measure $G'_{i,j}$, which is determined by the i.i.d. factors U_i through the specification

$$G'_{i,j} = \mathbb{I}\{(S_i - S_j) \bmod m_n = U_i\} \ . \tag{6.13}$$

In particular, outcomes are generated by the specification

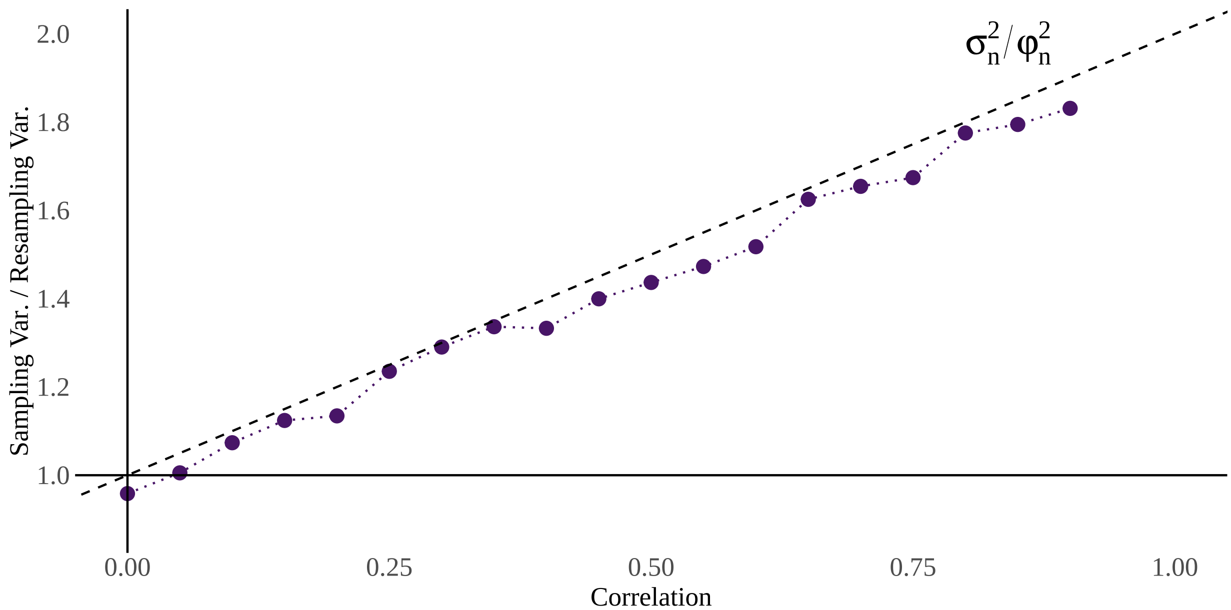
$$Y_i = \sum_{j \neq i} (\theta_n^\star (D_{i,j} - \rho_n) + G'_{i,j}) W_j \ . \tag{6.14}$$

Observe that, when $\eta = 0$, the measure $G'_{i,j}$ is uncorrelated with the measure $G'_{j,i}$, and that when $\eta = 1$ the two quantities are almost surely equal.

Theorems 6.1 and 6.2 imply that the ratio of the sampling variance of the estimator $\hat{\theta}_n^\star$ to the conditional variance of the sign-flipped estimator $\hat{\phi}_n^\star(V)$ should be well approximated by sequence σ_n^2/φ_n^2. In this model,

[33] That is, the variables U_i have a multinomial distribution with parameter given by a version of $p^{(\eta)}$ normalized to sum to one.

FIGURE 4. Observed and Predicted Variance Ratios



Notes: Figure 4 compares observed and predicted values of the ratio of the sampling variance of the estimator $\hat{\theta}_n^\star$ to the conditional variance of the sign-flipped estimator $\hat{\phi}_n^\star(V)$, in the simulation environment detailed in Section 6.4. The observed values of this ratio are displayed with purple dots. The value of this ratio predicted by Theorems 6.1 and 6.2 is displayed with a black dashed line. The x-axis displays values of the correlation (6.15). At each value of this correlation on a grid between 0 and 0.95, we use $5,000$ simulation replications with $B = 500$ random sign vectors $V^{(b)}$ used per replication.

---

it can be shown that the sequence σ_n^2/φ_n^2 is equal to one plus the correlation

$$\mathrm{Corr}(\mathbb{E}[\tilde{\varepsilon}_{k,i}\tilde{D}_{k,j}^\star \mid Z_i, Z_j]W_j^\star, \mathbb{E}[\tilde{\varepsilon}_{k,j}\tilde{D}_{k,i}^\star \mid Z_j, Z_i]W_i^\star) = \eta^2 + O(m_n^{-1})\ . \tag{6.15}$$

This prediction is bourne out in finite-samples. In particular, Figure 4 displays estimates of the ratio of the sampling variance of the estimator $\hat{\theta}_n^\star$ to the conditional variance of the sign-flipped estimator $\hat{\phi}_n^\star(V)$, for the case that $n = 1600$ and $m_n = \sqrt{n} = 40$, over a range of values for the correlation (6.15). The ratio σ_n^2/φ_n^2 is juxtaposed with a dashed line.

When the parameter η is equal to zero, the conditional variance of the sign-flipped estimator is very close to the sampling variance of estimator $\hat{\theta}_n^\star$. As η increases, the ratio of the two quantiles increases in proportion to the correlation (6.15), as predicted by Theorems 6.1 and 6.2. When η is close to one, two times the conditional variance of the sign-flipped estimator, that is, the variance estimate $\hat{\sigma}_n^2$, provides a good approximation to the sampling variance of the estimator $\hat{\theta}_n^\star$.

7. EMPIRICAL APPLICATION

We now illustrate the application of the methodology developed in this paper to Example 1. In particular, building on the data and empirical framework considered in Bloom et al. (2013), we characterize the relationship between technological similarity, product market similarity, and productivity spillover intensity.

For the sake of parsimony, we detail our treatment of these data in Appendix F, where we also illustrate applications to Examples 2 and 3.[34]

We consider a sample of publicly traded U.S. firms. Here, the treatment variable W_i is a binary indicator that firm i spent more than 100 million dollars on R&D between between 1996 and 2000. Following Bloom et al. (2013), we consider two proximity measures of interest. First, we set the proximity measure $D_{i,j}$ as the uncentered correlation (or "cosine similarity") between i and j's vectors of patent shares across USPTO technology classes, computed from all patents filed between 1996 and 2000. Second, we set $D_{i,j}$ as the uncentered correlation between i and j's vectors of sales shares across four digit industry codes, again from 1996 to 2000. We consider four choices for the outcome variable Y_i: market value (i.e., Tobin's Q), patent cites, sales, and R&D expenditure, each averaged between 2001 and 2005. We use the logarithm of each outcome variable. Results are displayed in Table 7.1. Panels A and B consider Technological and Product Market Proximity, respectively.

The first row in both panels displays the realized value of the estimator $\hat{\theta}_n^\star$, associated with each outcome, when neither the treatments nor the proximity measures of interest have been residualized (but the treatments are centered). Causal interpretations of these estimates are susceptible to two forms of confounding.

First, R&D expenditure might be endogenous. To address this, we residualize the treatments with lagged values of the treatment and outcome variables as well as fixed effects for the technology subcategory where each firm filed the majority of its patents. This is a simplification of the empirical strategy taken in Bloom et al. (2013).[35] The updated coefficients are displayed in the second row of Table 7.1. Second, the proximity measures of interest might be correlated with other factors that mediate spillover effects. To address this, in the third row, we additionally residualize the proximity measure of interest with the other proximity measure. We also consider a specification where we additionally control for geographic proximity. The standard error $\hat{\sigma}_n$, defined in (6.3), is reported below each estimate.

These estimates indicate that, across nearly all outcomes, firms that are more technologically similar capture larger R&D spillovers. Together, these findings are consistent with the idea that investments in R&D reduce the cost of technologically related investment. The exceptional outcome is sales: across each specification, the coefficient is not statistically different from zero. As these specifications control for product market proximity, this is intuitive.

The sign of the relationship between product market proximity and spillover intensity depends on the outcome. Increases in product market proximity are associated with decreased R&D spillovers on market value and patent citations and increased R&D spillovers on sales. The decrease in market value is consistent with the idea that investors interpret the investments of a rival firm as likely to depress long-term firm growth

[34] We rely on data from the replication package associated with Lucking et al. (2019), who improve the coverage of the data considered in Bloom et al. (2013), both in terms of the number of firms and years that are available.

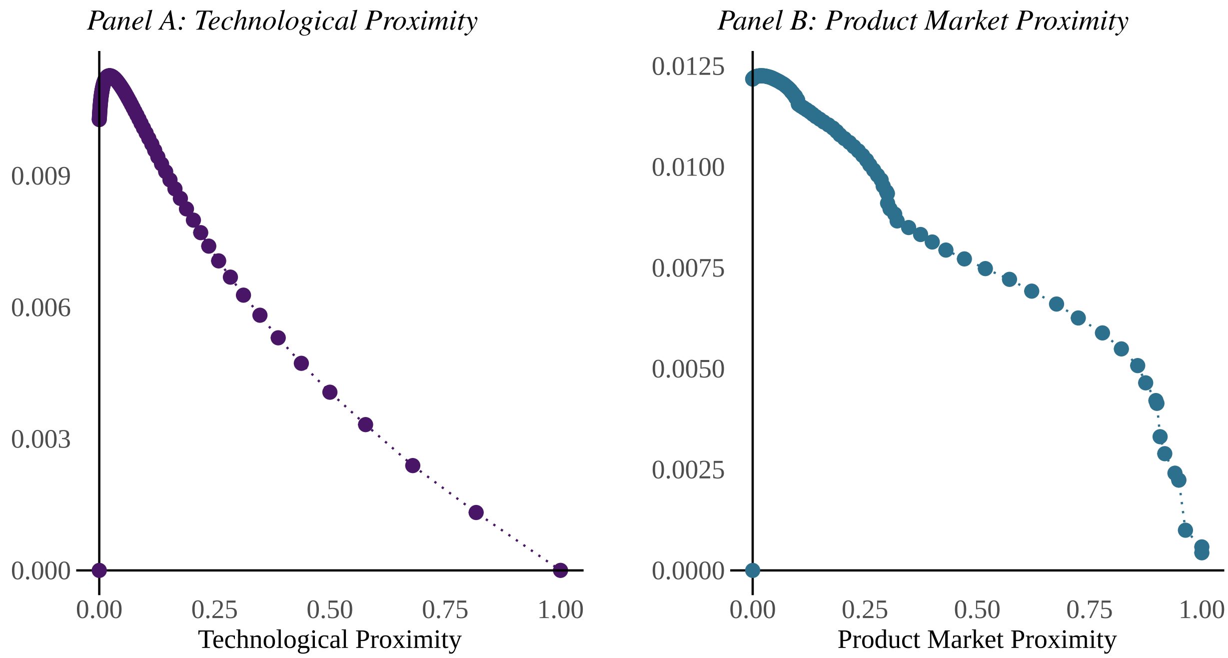
[35] Bloom et al. (2013) consider a similar panel specification, and so include firm fixed effects. We use a cross-sectional specification for the sake of simplicity. Bloom et al. (2013) also consider an instrument based on tax-induced changes in the user cost of R&D capital, which varies across states. We do not consider this approach, because these changes are highly correlated for pairs of firms whose R&D is concentrated in similar places (i.e., an analogue of Assumption 4.2 is violated).

TABLE 7.1. Application to R&D Spillover Estimation

		Value	Patent Cites	Sales	R&D Exp.
Panel A: Technological Proximity					
Unadjusted		0.052**	0.119**	0.106**	0.252**
		(0.012)	(0.024)	(0.022)	(0.012)
Resid. Treatment		0.201**	0.395**	0.272	0.647**
		(0.087)	(0.156)	(0.188)	(0.100)
Resid. Proximity	Prod. Market Prox.	0.227**	0.384**	0.171	0.636**
		(0.079)	(0.129)	(0.181)	(0.100)
	+ Geographic Prox.	0.231**	0.381**	0.142	0.626**
		(0.074)	(0.131)	(0.183)	(0.104)
Number of firms		1214	852	1294	1276
Panel B: Product Market Proximity					
Unadjusted		0.051	0.068	0.123**	0.248**
		(0.034)	(0.050)	(0.060)	(0.018)
Resid. Treatment		-0.146	0.013	0.626**	0.412
		(0.249)	(0.203)	(0.186)	(0.407)
Resid. Proximity	Technological Prox.	-0.294**	-0.153*	0.435**	-0.135
		(0.142)	(0.088)	(0.157)	(0.237)
	+ Geographic Prox.	-0.283*	-0.163*	0.414**	-0.141
		(0.146)	(0.087)	(0.168)	(0.245)
Number of firms		1370	839	1466	1364

Notes: Table 7.1 displays values of the estimator $\hat{\theta}_n^\star$ obtained by minimizing the objective function (4.3), using two choices for the proximity measure of interest, a variety of choices for covariates, and four choices for the outcome variable. The standard error $\hat{\sigma}_n$, defined in Algorithm 1, is displayed in parentheses below each estimate. One or two asterisks are placed beside each estimate to denote significance at the 10% and 5% levels, respectively. The estimates are obtained using data from the replication package associated with Lucking et al. (2019). Further details are given in Appendix F. In each case, the treatment variable W_i denotes a binary indicator that firm i spent more than one hundred million dollars on R&D between 1996 and 2000. The proximity measures $D_{i,j}$ are the uncentered correlation between patent shares filed between 1991 and 1995 across USPTO technology classes and the uncentered correlation between sales shares between 1991 and 1995 across four digit industry codes, respectively. The outcomes are averaged between 2001 and 2005, and enter in logs. In the first row, neither the treatments nor proximity measures have been residualized. In the second row, the treatments are residualized using lagged values of the treatment and outcome variables as well as fixed effects for the technology subcategory where each firm filed the majority of its patents. In the third and fourth rows, the proximity measure has been residualized using various measures of proximity, in addition to the covariates used to residualize the treatments. The estimators $\hat{\pi}_n(\cdot)$ and $\hat{\gamma}_n(\cdot)$ are both obtained from linear regressions.

FIGURE 5. Proximity Weight Estimates



Notes: Figure 5 displays estimates of the parameter $\text{Cov}(\mathbb{I}\{D_{i,j} \geq \delta\}, D_{i,j})$ across the support proximity measures of interest for the estimates displayed in Table 7.1. Each series has been normalized so that the points add to one. Panels A and B, display the weight estimates for technological proximity and product market proximity, respectively. In both cases, we measure $\text{Cov}(\mathbb{I}\{D_{i,j} \geq \delta\}, D_{i,j})$ at each of the percentiles of the distribution of the proximity measure $D_{i,j}$, conditional on $D_{i,j}$ being non-zero.

---

and earnings. The decrease in patent citations is consistent with models of cumulative innovation, in which older technologies are displaced—here, quite literally in citation patterns—by newer variants (Scotchmer, 2004). Increases in sales suggest some short-term strategic complementarity. It is worth cautioning that these estimates aggregate across many industries with heterogeneous market structures.

To unpack these results, Figure 5 displays estimates of the covariance $\text{Cov}(\mathbb{I}\{D_{i,j} \geq \delta\}, D_{i,j})$ for values of δ across the support of the two proximity measures of interest. Up to heterogeneity across covariates, this quantity determines the weight function (4.13) characterized in Theorem 4.1. These estimates indicate that much of the weight in the coefficients displayed in Table 7.1 is placed on pairs of firms whose technological and product market proximities are relatively small.

8. CONCLUSION

The spillover effects of economic actions often propagate through distinct, and potentially countervailing channels. This paper proposes a framework for identifying and estimating the causal relationship between a given measure of economic proximity—such as geographic distance, technological similarity, trade, or migration flows—and the intensity of the spillover effects of a treatment, shock, or policy change. Our consideration centers on widely-applied regressions that relate outcomes to proximity-weighted averages of the treatments assigned to other units. We show that coefficients obtained from such regressions admit a nonparametric interpretation—as targeting convex averages of parameters that measure how the association between one unit's outcome and another unit's treatment correlates with the proximity measure of interest.

The core premise of this paper is that when the proximity measure of interest is associated with other channels that also mediate spillover effects, causal interpretations of such relationships are susceptible to confounding. For example, Bloom et al. (2013) find that the spillover effects of investments in research and development are larger for pairs of firms whose research is concentrated in more technologically similar areas. If technologically similar firms tend to be geographically close, such a relationship might arise, spuriously, from spillover effects that are determined by geography alone.

The main contribution of this paper is a regression-based method for adjusting such estimates to account for correlated channels that might also transmit spillover effects. In particular, we give conditions under which the effect of a given proximity measure on spillover intensity can be recovered by regressing outcomes on averages of other units' treatments, reweighted by versions of the proximity measure that have been residualized with available auxiliary proximity measures. That is, in the application to Bloom et al. (2013), the propagation of spillovers through geography can be accounted for by residualizing technological similarity with geographic distance. We show that estimates obtained in this way are consistent, and optimal, in a minimax sense, and propose a new resampling based approach for constructing conservative estimates of the associated uncertainty. We illustrate, in data from Bloom et al. (2013), that the proposed adjustments can make a meaningful difference in practice.

Several extensions to the results presented in this paper have the potential to be particularly impactful. First, throughout, we have assumed that treatments are assigned independently across units. Further research should consider the extent to which peer effects in treatment selection can be accommodated. Leung and Loupos (2025) give a general treatment of a related problem. Second, our main approach for addressing endogeneity in measures of proximity is based on a conditional ignorability assumption. We have given a brief treatment of an extension to settings where suitable instrumental variables are available.[36] Further consideration of such settings, both theoretically and empirically, would be useful. Finally, the estimands of interest in this paper are local, in the sense that they relate to the spillover effects of changing the treatment and proximity of only one unit. It would be valuable to consider how to use such estimates as inputs into the recovery of more complicated counterfactuals that require the prescription of greater economic structure.

[36]For instance, Ellison et al. (2010) instrument geographic co-location in the United States with geographic co-location in the United Kingdom. Similarly, Fafchamps and Quinn (2018) implement an experiment that introduces exogenous variation in the social ties across business leaders, and measure the propagation of management practices through these ties. See Appendix E.5.

References

Acemoglu, D., Akcigit, U., and Kerr, W. (2016a). Networks and the macroeconomy: An empirical exploration. *Nber macroeconomics annual*, 30(1):273–335.

Acemoglu, D., Autor, D., Dorn, D., Hanson, G. H., and Price, B. (2016b). Import competition and the great us employment sag of the 2000s. *Journal of Labor Economics*, 34(S1):S141–S198.

Adao, R., Arkolakis, C., and Esposito, F. (2025). General equilibrium effects in space: Theory and measurement. Technical report, National Bureau of Economic Research.

Adao, R., Kolesár, M., and Morales, E. (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics*, 134(4):1949–2010.

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The economics of density: Evidence from the berlin wall. *Econometrica*, 83(6):2127–2189.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598.

Allcott, H. and Keniston, D. (2018). Dutch disease or agglomeration? the local economic effects of natural resource booms in modern america. *The Review of Economic Studies*, 85(2):695–731.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Angrist, J. D. (1998). Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica*, 66(2):249.

Angrist, J. D., Graddy, K., and Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies*, 67(3):499–527.

Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of labor economics*, volume 3, pages 1277–1366. Elsevier.

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

Arque-Castells, P. and Spulber, D. F. (2022). Measuring the private and social returns to r&d: Unintended spillovers versus technology markets. *Journal of Political Economy*, 130(7):1860–1918.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In *Readings in industrial economics: Volume two: Private enterprise and state intervention*, pages 219–236. Springer.

Auerbach, E. (2022). Identification and estimation of a partially linear regression model using network data. *Econometrica*, 90(1):347–365.

Auerbach, E. and Tabord-Meehan, M. (2025). The local approach to causal inference under network interference. *arXiv preprint arXiv:2105.03810*.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American economic review*, 103(6):2121–2168.

Bergeaud, A., Guillouzouic, A., Henry, E., and Malgouyres, C. (2025). From public labs to private firms: Magnitude and channels of local r&d spillovers. *The Quarterly Journal of Economics*.

Bernstein, S., Colonnelli, E., Giroud, X., and Iverson, B. (2019). Bankruptcy spillovers. *Journal of Financial Economics*, 133(3):608–633.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393.

Bobkov, S. G., Götze, F., and Sambale, H. (2019). Higher order concentration of measure. *Communications in Contemporary Mathematics*, 21(03):1850043.

Borusyak, K., Dix-Carneiro, R., and Kovak, B. (2022a). Understanding migration responses to local shocks. *Available at SSRN 4086847*.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Borusyak, K., Hull, P., and Jaravel, X. (2022b). Quasi-experimental shift-share research designs. *The Review of economic studies*, 89(1):181–213.

Breza, E., Chandrasekhar, A. G., McCormick, T. H., and Pan, M. (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 110(8):2454–2484.

Cai, J., Janvry, A. D., and Sadoulet, E. (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108.

Cai, J. and Szeidl, A. (2024). Indirect effects of access to finance. *American Economic Review*, 114(8):2308–2351.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, pages C1–C68.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American economic review*, 100(1):35–69.

Cox, D. R. (1958). Planning of experiments.

Dasgupta, P. and Stiglitz, J. (1980). Industrial structure and the nature of innovative activity. *The Economic Journal*, 90(358):266–293.

Diamond, R. and McQuade, T. (2019). Who wants affordable housing in their backyard? an equilibrium analysis of low-income property development. *Journal of Political Economy*, 127(3):1063–1117.

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A “market access” approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Duchi, J. (2024). Statistics and information theory. *URL: https://web.stanford.edu/class/stats311/lecture-notes.pdf.*, 2:23.

Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *The Annals of Statistics*, pages 586–596.

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., and Walker, M. (2022). General equilibrium effects of cash transfers: experimental evidence from kenya. *Econometrica*, 90(6):2603–2643.

Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of political economy*, 105(5):889–927.

Ellison, G., Glaeser, E. L., and Kerr, W. R. (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *American Economic Review*, 100(3):1195–1213.

Emmenegger, C., Spohn, M.-L., Elmer, T., and Bühlmann, P. (2025). Treatment effect estimation with observational network data using machine learning. *Journal of Causal Inference*, 13(1):20230082.

Fafchamps, M. and Quinn, S. (2018). Networks and manufacturing firms in africa: Results from a randomized field experiment. *The World Bank Economic Review*, 32(3):656–675.

Faridani, S. and Niehaus, P. (2022). Linear estimation of average global effects. *arXiv preprint arXiv:2209.14181*.

Feyrer, J., Mansur, E. T., and Sacerdote, B. (2017). Geographic dispersion of economic shocks: Evidence from the fracking revolution. *American Economic Review*, 107(4):1313–1334.

Forastiere, L., Airoldi, E. M., and Mealli, F. (2021). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534):901–918.

Franklin, S., Imbert, C., Abebe, G., and Mejia-Mantilla, C. (2024). Urban public works in spatial equilibrium: Experimental evidence from ethiopia. *American Economic Review*, 114(5):1382–1414.

Gao, M. (2024). Endogenous interference in randomized experiments. *arXiv preprint arXiv:2412.02183*.

Gao, M. and Ding, P. (2023). Causal inference in network experiments: regression-based analysis and design-based properties.

Giroud, X., Lenzu, S., Maingi, Q., and Mueller, H. (2024). Propagation and amplification of local productivity spillovers. *Econometrica*, 92(5):1589–1619.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.

Graham, B. S. (2016). Homophily and transitivity in dynamic network formation. Technical report, National Bureau of Economic Research.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of political economy*, 118(3):536–598.

Griliches, Z. (1992). The search for r&d spillovers. *The Scandinavian Journal of Economics*, pages S29–S47.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The nber patent citation data file: Lessons, insights and methodological tools.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of multivariate analysis*, 14(1):1–16.

Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of international economics*, 67(1):1–24.

Head, K. and Mayer, T. (2011). Gravity, market potential and economic development. *Journal of Economic Geography*, 11(2):281–294.

Heinonen, J. (2001). *Lectures on analysis on metric spaces*. Springer Science & Business Media.

Helm, I. (2020). National industry trade shocks, local labour markets, and agglomeration spillovers. *The Review of Economic Studies*, 87(3):1399–1431.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, pages 293–325.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *Ann. Math. Statist.*, 23(4):169–192.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098.

Hoover, D. N. (1979). Relations on probability spaces and arrays of. *Preprint, Institute for Advanced Study*.

Hornbeck, R. and Moretti, E. (2024). Estimating who benefits from productivity growth: local and distant effects of city productivity growth on wages, rents, and inequality. *Review of Economics and Statistics*, 106(3):587–607.

Hu, Y., Li, S., and Wager, S. (2022). Average direct and indirect causal effects under interference. *Biometrika*, 109(4):1165–1172.

Huber, K. (2018). Disentangling the effects of a banking crisis: Evidence from german firms and counties. *American Economic Review*, 108(3):868–898.

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the american statistical association*, 103(482):832–842.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Jackson, M. O. and Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915.

Jaffe, A. B. (1986). Technological opportunity and spillovers of r & d: Evidence from firms' patents, profits, and market value. *The American Economic Review*, pages 984–1001.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3):577–598.

Kline, P. and Moretti, E. (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *The Quarterly journal of economics*, 129(1):275–331.

Kojevnikov, D. (2021). The bootstrap for network dependent processes. *arXiv preprint arXiv:2101.12312*.

Kolesár, M. and Plagborg-Møller, M. (2024). Dynamic causal effects in a nonlinear world: the good, the bad, and the ugly. *arXiv preprint arXiv:2411.10415*.

Kovak, B. K. (2013). Regional effects of trade reform: What is the correct measure of liberalization? *American Economic Review*, 103(5):1960–1976.

Krugman, P. (1987). The narrow moving band, the dutch disease, and the competitive consequences of mrs. thatcher: Notes on trade in the presence of dynamic scale economies. *Journal of development Economics*, 27(1-2):41–55.

Lachièze-Rey, R. and Peccati, G. (2017). New berry–esseen bounds for functionals of binomial point processes. *The Annals of Applied Probability*, 27(4):1992–2031.

Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, pages 38–53.

Lehmann, E. L. and Romano, J. P. (2022). *Testing statistical hypotheses*, volume 4. Springer.

Lerche, A. (2025). Direct and indirect effects of investment tax incentives. *American Economic Review*.

Leung, M. P. (2020). Treatment and spillover effects under network interference. *Review of Economics and Statistics*, 102(2):368–380.

Leung, M. P. (2022a). Causal inference under approximate neighborhood interference. *Econometrica*, 90(1):267–293.

Leung, M. P. (2022b). Rate-optimal cluster-randomized designs for spatial interference. *The Annals of Statistics*, 50(5):3064–3087.

Leung, M. P. (2023). Network cluster-robust inference. *Econometrica*, 91(2):641–667.

Leung, M. P. and Loupos, P. (2025). Graph neural networks for causal inference under network confounding. *arXiv preprint arXiv:2211.07823*.

Li, S. and Wager, S. (2022). Random graph asymptotics for treatment effect estimation under network interference. *The Annals of Statistics*, 50(4):2334–2358.

Liu, S.-H., Shao, Q.-M., and Shi, H. (2025). Berry-esseen bounds for degenerate u-statistics with application to distance correlation. *Science China Mathematics*, pages 1–36.

Lucking, B., Bloom, N., and Van Reenen, J. (2019). Have r&d spillovers declined in the 21st century? *Fiscal Studies*, 40(4):561–590.

Lychagin, S., Pinkse, J., Slade, M. E., and Reenen, J. V. (2016). Spillovers in space: does geography matter? *The Journal of Industrial Economics*, 64(2):295–335.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23.

Marshall, A. (1890). *Principles of Economics*. Macmillan, London.

McCormick, T. H. and Zheng, T. (2015). Latent surface models for networks using aggregated relational data. *Journal of the American Statistical Association*, 110(512):1684–1695.

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica*, 89(5):2143–2188.

Menzel, K. (2025). Fixed-population causal inference for models of equilibrium. *arXiv preprint arXiv:2501.19394*.

Miguel, E. and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.

Monte, F., Redding, S. J., and Rossi-Hansberg, E. (2018). Commuting, migration, and local employment elasticities. *American Economic Review*, 108(12):3855–3890.

Moretti, E. (2004). Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review*, 94(3):656–690.

Moscona, J. and Seck, A. A. (2024). Age set versus kin: Culture and financial ties in east africa. *American Economic Review*, 114(9):2748–2791.

Munro, E. (2025). Causal inference under interference through designed markets. *arXiv preprint arXiv:2504.07217*.

Munro, E., Wager, S., and Xu, K. (2025). Treatment effects in market equilibrium. *arXiv preprint arXiv:2109.11647*, 5.

Muralidharan, K., Niehaus, P., and Sukhtankar, S. (2023). General equilibrium effects of (improving) public employment programs: Experimental evidence from india. *Econometrica*, 91(4):1261–1295.

Myers, K. R. and Lanahan, L. (2022). Estimating spillovers from publicly funded r&d: Evidence from the us department of energy. *American Economic Review*, 112(7):2393–2423.

Ogburn, E. L., Sofrygin, O., Diaz, I., and Van der Laan, M. J. (2024). Causal inference for social network data. *Journal of the American Statistical Association*, 119(545):597–611.

Penrose, M. (2003). *Random geometric graphs*, volume 5. OUP Oxford.

Pollmann, M. (2023). Causal inference for spatial treatments. *arXiv preprint arXiv:2011.00373*.

Redding, S. and Venables, A. J. (2004). Economic geography and international inequality. *Journal of international Economics*, 62(1):53–82.

Ritzwoller, D. M., Romano, J. P., and Shaikh, A. M. (2024). Randomization inference: Theory and applications. *arXiv preprint arXiv:2406.09521*.

Rotemberg, M. (2019). Equilibrium effects of firm subsidies. *American Economic Review*, 109(10):3475–3513.

Rubin, D. B. (1980). Discussion of a paper by D. Basu. *Journal of the American Statistical Association*, 75(371):591–593.

Sävje, F. (2024). Causal inference with misspecified exposure mappings: separating definitions and assumptions. *Biometrika*, 111(1):1–15.

Sävje, F., Aronow, P., and Hudgens, M. (2021). Average treatment effects in the presence of unknown interference. *Annals of statistics*, 49(2):673.

Scotchmer, S. (2004). *Innovation and incentives*. MIT press.

Serfling, R. (1980). Approximation theorems of mathematical statistics. *Wiley series in probability and mathematical statistics*.

Spence, M. (1984). Cost reduction, competition, and industry performance. *Econometrica*, 52(1):101–122.

Toulis, P. and Kao, E. (2013). Estimation of causal peer influence effects. In *International conference on machine learning*, pages 1489–1497. PMLR.

Vazquez-Bare, G. (2023). Identification and estimation of spillover effects in randomized experiments. *Journal of Econometrics*, 237(1):105237.

Veitch, V., Wang, Y., and Blei, D. (2019). Using embeddings to correct for unobserved confounding in networks. *Advances in Neural Information Processing Systems*, 32.

Viviano, D., Lei, L., Imbens, G., Karrer, B., Schrijvers, O., and Shi, L. (2023). Causal clustering: design of cluster experiments under network interference. *arXiv preprint arXiv:2310.14983*.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press.

Wang, Y., Samii, C., Chang, H., and Aronow, P. (2025). Design-based inference for spatial experiments under unknown interference. *The Annals of Applied Statistics*, 19(1):744 – 768.

Yitzhaki, S. (1996). On using linear regressions in welfare economics. *Journal of Business & Economic Statistics*, 14(4):478–486.

Zacchia, P. (2020). Knowledge spillovers through networks of scientists. *The Review of economic studies*, 87(4):1989–2018.

Supplemental Appendix to:

Regression Adjustments for Disentangling Spillover Effects*

David M. Ritzwoller

Stanford GSB

Contents

Notation

Throughout the proofs, the indices i, l, and q are taken to be distinct, unless otherwise noted. The quantities c and C denote universal, positive constants, whose values are allowed to change in each appearance. For two sequences $f(n)$ and $g(n)$, the relations $f(n) \lesssim g(n)$ and $f(n) \asymp g(n)$ indicate that $f(n) \leq Cg(n)$ and $cg(n) \leq f(n) \leq Cg(n)$ for all sufficiently large n. The set Π_m denotes the set of permutations of the set $[m] = \{1, \ldots, m\}$, treated as bijections.

*Date: November 15, 2025

APPENDIX A. PROOFS FOR DECOMPOSITION THEOREMS

In this Appendix, we give the proofs for Theorems 2.1 and 4.1. Both results hold under two additional regularity conditions, stated in Appendix A.1. Proofs for these results, and each of the supporting Lemmas, are given in Appendices A.2 and A.3, respectively.

A.1 Additional Regularity Conditions

Theorem 2.1 holds under two, additional regularity conditions. The first condition stipulates that various conditional expectations of the outcome have first and second derivatives that exist and are on the order of a constant.

Assumption A.1. *For any unit i and distinct units j and k, the map*

$$(\delta, w) \mapsto \mathbb{E}[Y_i \mid D_{k,j} = \delta, W_j = w, S_j] \tag{A.1}$$

is twice continuously differentiable, almost surely. Moreover:
(i) *For any units i and j, it holds that*

$$\partial_w \, \mathbb{E}[Y_i \mid W_j = w, S_j = s] = O(1) \tag{A.2}$$

uniformly over each w and s in their respective domains.
(ii) *For distinct units i, j, and k, it holds that*

$$\partial^2_{\delta,w} \, \mathbb{E}[Y_i \mid D_{i,j} = \delta, W_j = w, S_j = s] = O(1) \tag{A.3}$$

uniformly over each δ, w, and s in their respective domains.

The second condition asks that the conditional expectation of the proximity measure of interest is not much larger than its variance. This restriction is used to ensure that several of the terms in the error in the representation (2.7) are negligible. We additionally ask that the conditional variance of treatments is on the order of a constant.

Assumption A.2. *It holds that*

$$\mathbb{E}[D_{i,l} \mid S_l] \lesssim \mathrm{Var}(D_{i,l} \mid S_l) \quad \textit{and} \quad \mathrm{Var}(W_l \mid S_l) \asymp 1 \tag{A.4}$$

uniformly almost surely.

The regularity conditions needed for Theorem 4.1 are direct generalizations of Assumptions 2.3, A.1, and A.2. In particular, each restriction additionally condition on the covariates $H_{i,j}$.

Assumption A.3. *For any unit i and distinct units j and k, the map*

$$(\delta, w) \mapsto \mathbb{E}[Y_i \mid D_{k,j} = \delta, W_j = w, H_{i,j}, \overline{S}_j] \tag{A.5}$$

is twice continuously differentiable, almost surely. Moreover:
(i) *For any units i and j, it holds that*

$$\partial_w \, \mathbb{E}[Y_i \mid W_j = w, X_j^{\mathsf{out}}, \overline{S}_j] = O(1) \tag{A.6}$$

uniformly over each w in its domain.

(ii) *For distinct units i, j, and k, it holds that*

$$\partial^2_{\delta,w}\,\mathbb{E}[Y_i \mid D_{i,j}=\delta, W_j=w, H_{i,j}, \overline{S}_j] = O(1) \tag{A.7}$$

uniformly over each δ and w in their respective domains.

(iii) *For distinct units i, j, and k, it holds that*

$$\partial^2_{\delta,w}\,\mathbb{E}[Y_i \mid D_{k,j}=\delta, W_j=w, H_{i,j}, \overline{S}_j] = o(n^{-1/2}) \tag{A.8}$$

uniformly over each δ and w in their respective domains.

Assumption A.4. *It holds that*

$$\mathbb{E}[D_{i,l} \mid H_{i,l}, S_l] \lesssim \mathrm{Var}(D_{i,l} \mid H_{i,l}, S_l) \quad \textit{and} \quad \mathrm{Var}(W_l \mid \overline{S}_j) \asymp 1 \tag{A.9}$$

uniformly almost surely.

A.2 Proofs for Theorems 2.1 and 4.1

Theorems 2.1 and 4.1 are founded on three general Lemmas. The first Lemma facilitates the evaluation of several expectations that recur throughout the argument.

Lemma A.1. *The random variables $(Y_i)_{i=1}^n$ are identically distributed. The array $(M_{i,j})_{i,j\in[n]}$ satisfies*

$$M_{i,j} = M_n(S_i, S_j) \tag{A.10}$$

and is supported on an interval whose radius is bounded by a constant. The random variables $(S_i, W_i)_{i=1}^n$ are i.i.d. and satisfy the condition

$$\mathbb{E}[W_i \mid S_i] = 0\,. \tag{A.11}$$

Moreover, the conditional variance $\mathrm{Var}(W_i \mid S_i)$ is bounded by a constant almost surely. Define the terms

$$\tilde{\Xi}_i = \sum_{j\neq i}\left(M_{i,j} - \frac{1}{n}\sum_{k\neq j} M_{k,j}\right) W_j - \frac{1}{n}\sum_{k\neq i} M_{k,i}W_i \quad \textit{and} \quad \tilde{M}_{i,l} = M_{i,l} - \mathbb{E}[M_{i,l} \mid S_l]\,. \tag{A.12}$$

It holds that

$$\frac{1}{n^2}\sum_{i=1}^n \mathbb{E}[Y_i\hat{\Xi}_i] = \frac{n-2}{n}\left(\mathbb{E}[Y_i\tilde{M}_{i,l}W_l] - \mathbb{E}[Y_i\tilde{M}_{q,l}W_l]\right) \tag{A.13}$$
$$+\frac{n-2}{n^2}\mathbb{E}[Y_iM_{q,l}W_l] - \frac{n-1}{n^2}\mathbb{E}[Y_iM_{l,i}W_i] + \frac{1}{n^2}\mathbb{E}[Y_iM_{i,l}W_l]$$

and

$$\frac{1}{n^2}\sum_{i=1}^n \mathbb{E}[\hat{\Xi}_i^2] - \frac{n-2}{n}\mathbb{E}\left[\mathrm{Var}(M_{i,l} \mid S_l)\,\mathrm{Var}(W_l \mid S_l)\right] = O(n^{-2})\,, \tag{A.14}$$

respectively.

The second Lemma gives a decomposition of the expectations of products of three random variables that satisfy particular restrictions on their conditional dependencies. The result relies on an application of a method of argument due to Kolesár and Plagborg-Møller (2024), who build on ideas due to Yitzhaki (1996) and Angrist and Krueger (1999).

Lemma A.2. *Fix the random variables Y, W, Z, and U. Assume that the distributions of W and Z, conditioned U, have support contained in the bounded sets $\mathcal{W}$ and $\mathcal{Z}$ almost surely and that the restriction*

$$W \perp\!\!\!\perp Z \mid U \tag{A.15}$$

holds. If the random variable Y is real-valued and the first derivatives and second mixed-partial derivative of the conditional expectation $(z, w) \mapsto \mathbb{E}[Y \mid Z = z, W = w, U]$ exist and are uniformly bounded on $\mathcal{Z} \times \mathcal{W}$, then the equalities

$$\mathbb{E}[Y(W - \mathbb{E}[W \mid U])] = \int_{\mathcal{W}} \mathbb{E}[\mathbb{E}[\mathbb{I}\{W \geq w\}(W - \mathbb{E}[W \mid U]) \mid U]\partial_w \mathbb{E}[Y \mid W = w, U]]\,\mathrm{d}w \tag{A.16}$$

and

$$\begin{aligned}
&\mathbb{E}[Y(Z - \mathbb{E}[Z \mid U])(W - \mathbb{E}[W \mid U])] \\
&= \int_{\mathcal{Z}} \int_{\mathcal{W}} \mathbb{E}[\mathbb{E}[\mathbb{I}\{W \geq w\}(W - \mathbb{E}[W \mid U]) \mid U]\mathbb{E}[\mathbb{I}\{Z \geq z\}(Z - \mathbb{E}[Z \mid U]) \mid U] \\
&\qquad \partial^2_{z,w}\mathbb{E}[Y \mid Z = z, W = w, U]]\,\mathrm{d}w\,\mathrm{d}z
\end{aligned} \tag{A.17}$$

hold.

The third Lemma gives a helpful, well-known decomposition of particular conditional variances.

Lemma A.3. *Fix the random variables Z and U. If Z is a real-valued and supported on $\mathcal{Z}$, then*

$$\mathrm{Cov}(\mathbb{I}\{Z \geq z\}, Z \mid U) \geq 0 \quad \text{and} \quad \int_{\mathcal{Z}} \mathrm{Cov}(\mathbb{I}\{Z \geq z\}, Z \mid U)\,\mathrm{d}z = \mathrm{Var}(Z \mid U)\,, \tag{A.18}$$

almost surely, for each z in $\mathcal{Z}$.

A.2.1 Proof of Theorem 2.1. Define the risk

$$R_n(\theta) = \mathbb{E}\left[\min_{\alpha \in \mathbb{R}} \left\{\sum_{i=1}^{n} (Y_i - \alpha - \theta \cdot \Delta_i)^2\right\}\right]\,. \tag{A.19}$$

By the Frisch-Waugh-Lovell Theorem, we can write

$$R_n(\theta) = \mathbb{E}\left[\sum_{i=1}^{n} \left((Y_i - \tilde{Y}_n) - \theta \cdot \tilde{\Delta}_i\right)^2\right]\,, \tag{A.20}$$

where $\tilde{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and

$$\begin{aligned}
\tilde{\Delta}_i &= \sum_{j \neq i} D_{i,j} W_j - \frac{1}{n}\sum_{k=1}^{n}\sum_{m \neq k} D_{k,m} W_m \\
&= \sum_{j \neq i} D_{i,j} W_j - \sum_{m=1}^{n} \frac{1}{n} \sum_{k \neq m} D_{k,m} W_m = \sum_{j \neq i}\left(D_{i,j} - \frac{1}{n}\sum_{k \neq j} D_{k,j}\right) W_j - \frac{1}{n}\sum_{k \neq i} D_{k,i} W_i\,.
\end{aligned} \tag{A.21}$$

Thus, the parameter $\overline{\theta}_n$ admits the representation

$$\overline{\theta}_n = \frac{\sum_{i=1}^n \mathbb{E}[(Y_i - \tilde{Y}_n)\tilde{\Delta}_i]}{\sum_{i=1}^n \mathbb{E}[\tilde{\Delta}_i^2]} = \frac{\sum_{i=1}^n \mathbb{E}[Y_i \tilde{\Delta}_i]}{\sum_{i=1}^n \mathbb{E}[\tilde{\Delta}_i^2]} \tag{A.22}$$

by the Projection Theorem and the fact that $\sum_{i=1}^n \tilde{\Delta}_i = 0$ by construction.

We begin by simplifying the representation (A.22) through an application of Lemma A.1. Observe that the conditions of Lemma A.1 are implied by Assumptions 2.1 and 2.2. Hence, we obtain the expansion

$$\begin{aligned}
\overline{\theta}_n &= \frac{\mathbb{E}[Y_i \tilde{D}_{i,q} W_q] - \mathbb{E}[Y_i \tilde{D}_{l,q} W_q]}{\mathbb{E}[\operatorname{Var}(D_{i,q} \mid S_i)\operatorname{Var}(W_i \mid S_i)]} \\
&+ \frac{1}{n}\frac{\mathbb{E}[Y_i \tilde{D}_{l,q} W_q] - \mathbb{E}[Y_i \tilde{D}_{l,i} W_i]}{\mathbb{E}[\operatorname{Var}(D_{i,q} \mid S_q)\operatorname{Var}(W_q \mid S_q)]} \\
&+ \frac{1}{n}\frac{\mathbb{E}[\mathbb{E}[Y_i W_q \mid S_q]\mathbb{E}[D_{l,q} \mid S_q]] - \mathbb{E}[\mathbb{E}[Y_q W_q \mid S_q]\mathbb{E}[D_{l,q} \mid S_q]}{\mathbb{E}[\operatorname{Var}(D_{i,q} \mid S_q)\operatorname{Var}(W_q \mid S_q)]} \\
&+ \frac{1}{n^2}\frac{\mathbb{E}[Y_i \tilde{D}_{i,q} W_q]}{\mathbb{E}[\operatorname{Var}(D_{i,q} \mid S_q)\operatorname{Var}(W_q \mid S_q)]} + \frac{1}{n^2}\frac{\mathbb{E}[\mathbb{E}[Y_i W_q \mid S_q]\mathbb{E}[D_{i,q} \mid S_q]]}{\mathbb{E}[\operatorname{Var}(D_{i,q} \mid S_q)\operatorname{Var}(W_q \mid S_q)]} + O(n^{-2}) ,
\end{aligned} \tag{A.23}$$

where

$$\tilde{D}_{i,l} = D_{i,l} - \mathbb{E}[D_{i,l} \mid S_l] . \tag{A.24}$$

We further re-express the parameter $\overline{\theta}_n$ through several applications of Lemma A.2. Observe that Assumptions 2.1 and 2.2 imply that

$$W_l \perp\!\!\!\perp D_{i,l} \mid S_l \quad \text{and} \quad \mathbb{E}[W_l \mid S_l] = 0 . \tag{A.25}$$

Thus, Lemma A.2 implies that

$$\begin{aligned}
\mathbb{E}[Y_i \tilde{D}_{i,q} W_q] &= \int_{\mathcal{D}}\int_{\mathcal{W}} \mathbb{E}[\mathbb{E}[\mathbb{I}\{W_q \geq w\} W_q \mid S_q]\mathbb{E}[\mathbb{I}\{D_{i,q} \geq \delta\}\tilde{D}_{i,q} \mid S_q] \\
&\qquad \partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{i,q} = \delta, W_q = w, S_q]]\,\mathrm{d}w\,\mathrm{d}\delta , \\
\mathbb{E}[Y_i \tilde{D}_{l,q} W_q] &= \int_{\mathcal{D}}\int_{\mathcal{W}} \mathbb{E}[\mathbb{E}[\mathbb{I}\{W_q \geq w\} W_q \mid S_q]\mathbb{E}[\mathbb{I}\{D_{l,q} \geq \delta\}\tilde{D}_{l,q} \mid S_q] \\
&\qquad \partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{l,q} = \delta, W_q = w, S_q]]\,\mathrm{d}w\,\mathrm{d}\delta , \\
\mathbb{E}[Y_i W_q \mid S_q] &= \int_{\mathcal{W}} \mathbb{E}[\mathbb{I}\{W_q \geq w\} W_q \mid S_q]\partial_w \mathbb{E}[Y_i \mid W_q = w, S_q]\,\mathrm{d}w, \quad \text{and} \\
\mathbb{E}[Y_q W_q \mid S_q] &= \int_{\mathcal{W}} \mathbb{E}[\mathbb{I}\{W_q \geq w\} W_q \mid S_q]\partial_w \mathbb{E}[Y_q \mid W_q = w, S_q]\,\mathrm{d}w ,
\end{aligned} \tag{A.26}$$

respectively. Consequently, by plugging each of the representations (A.26) into the expression (A.23), we find that

$$\begin{aligned}
\overline{\theta}_n = &\int_{\mathcal{D}}\int_{\mathcal{W}} \mathbb{E}[\lambda(\delta, w \mid S_q)(\partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{i,q} = \delta, W_q = w, S_q] \\
&\qquad - \partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{l,q} = \delta, W_q = w, S_q])]\,\mathrm{d}w\,\mathrm{d}\delta \\
&+ \frac{1}{n}\int_{\mathcal{D}}\int_{\mathcal{W}} \mathbb{E}[\lambda(\delta, w \mid S_q)(\partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{l,q} = \delta, W_q = w, S_q]
\end{aligned} \tag{A.27}$$

$$- \partial^2_{\delta,w}\mathbb{E}[Y_q \mid D_{l,q} = \delta, W_q = w, S_q])]] \, \mathrm{d}w \, \mathrm{d}\delta$$

$$+ \frac{1}{n^2} \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid S_q) \partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{i,q} = \delta, W_q = w, S_q]] \, \mathrm{d}w \, \mathrm{d}\delta$$

$$+ \frac{1}{n} \int_{\overline{\mathcal{W}}} \mathbb{E}\Bigg[\lambda'(w \mid S_q)\Bigg(\left(\frac{n+1}{n}\right) \partial_w \mathbb{E}[Y_i \mid W_q = w, S_q]$$

$$- \partial_w \mathbb{E}[Y_q \mid W_q = w, S_q]\Bigg)\Bigg] \, \mathrm{d}w + O(n^{-2}) \, ,$$

where

$$\lambda(\delta, w \mid S_l) = \frac{\mathrm{Cov}(\mathbb{I}\{D_{i,l} \geq \delta\}, D_{i,l} \mid S_l) \, \mathrm{Cov}(\mathbb{I}\{W_l \geq w\}, W_l \mid S_l)}{\mathbb{E}[\mathrm{Var}(D_{i,l} \mid S_l) \, \mathrm{Var}(W_l \mid S_l)]} \quad \text{and} \tag{A.28}$$

$$\lambda'(w) = \frac{\mathbb{E}[D_{i,l} \mid S_l] \, \mathrm{Cov}(\mathbb{I}\{W_l \geq w\}, W_l \mid S_q)}{\mathbb{E}[\mathrm{Var}(D_{i,l} \mid S_l) \, \mathrm{Var}(W_l \mid S_l)]} \, ,$$

respectively. Now, observe that Lemma A.3 implies that

$$\lambda(\delta, w \mid S_l) \geq 0 \quad \text{and} \quad \mathbb{E}\left[\int_{\overline{\mathcal{D}}} \int_{\mathcal{W}} \lambda(\delta, w \mid S_l) \, \mathrm{d}w \, \mathrm{d}\delta\right] = 1 \, , \tag{A.29}$$

respectively. Thus, Assumptions 2.3 and A.1 imply that

$$\int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid S_j) \partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{l,q} = \delta, W_q = w, S_q]] \, \mathrm{d}w \, \mathrm{d}\delta = o(n^{-1/2}) \, , \tag{A.30}$$

$$\frac{1}{n} \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid S_q)(\partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{l,q} = \delta, W_q = w, S_q]$$

$$- \partial^2_{\delta,w}\mathbb{E}[Y_q \mid D_{l,q} = \delta, W_q = w, S_q])]] \, \mathrm{d}w \, \mathrm{d}\delta = o(n^{-1/2}) \, , \quad \text{and}$$

$$\frac{1}{n^2} \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid S_q) \partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{i,q} = \delta, W_q = w, S_q]] \, \mathrm{d}w \, \mathrm{d}\delta = O(n^{-2}) \, .$$

In turn, Assumption A.2, Lemma A.3, and Cauchy-Schwarz imply that $\lambda'(w \mid S_l) \geq 0$ and

$$\mathbb{E}[\lambda'(w \mid S_l)] \lesssim \frac{\mathbb{E}[\mathbb{E}[D_{i,l} \mid S_l] \sqrt{\mathrm{Var}(W_l \mid S_l)}]}{\mathbb{E}[\mathrm{Var}(D_{i,l} \mid S_l) \, \mathrm{Var}(W_l \mid S_l)]}$$

$$\lesssim \frac{\mathbb{E}[\mathrm{Var}(D_{i,l} \mid S_l) \sqrt{\mathrm{Var}(W_l \mid S_l)}]}{\mathbb{E}[\mathrm{Var}(D_{i,l} \mid S_l) \, \mathrm{Var}(W_l \mid S_l)]} = O(1) \tag{A.31}$$

uniformly over w. Thus, Assumption A.1, and the fact that the treatments and distances are defined on sets whose supports are bounded by a constant, imply that

$$\frac{1}{n} \int_{\overline{\mathcal{W}}} \mathbb{E}\Bigg[\lambda'(w \mid S_q)\Bigg(\left(\frac{n+1}{n}\right) \partial_w \mathbb{E}[Y_i \mid W_q = w, S_q] \tag{A.32}$$

$$- \partial_w \mathbb{E}[Y_q \mid W_q = w, S_q]\Bigg)\Bigg] \, \mathrm{d}w = O(n^{-1}) \, .$$

Hence, we obtain the representation

$$\overline{\theta}_n = \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid S_q) \partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{i,q} = \delta, W_q = w, S_q]] \, \mathrm{d}w \, \mathrm{d}\delta + o(n^{-1/2}) \, , \tag{A.33}$$

by plugging the bounds (A.30) and (A.32) into (A.27). The convexity of the weights is verified by the statement (A.29), completing the proof. ■

A.2.2 Proof of Theorem 4.1. By the same steps used to derive the representation (A.22), the parameter $\overline{\theta}_n^\star$ admits the representation

$$\overline{\theta}_n^\star = \frac{\sum_{i=1}^n \mathbb{E}[Y_i \tilde{\Delta}_i^\star]}{\sum_{i=1}^n \mathbb{E}[(\tilde{\Delta}_i^\star)^2]} \tag{A.34}$$

where

$$\tilde{\Delta}_i^\star = \sum_{j \neq i} \left(D_{i,j}^\star - \frac{1}{n} \sum_{k \neq j} D_{k,j}^\star \right) W_j^\star - \frac{1}{n} \sum_{k \neq i} D_{k,i}^\star W_i^\star \ . \tag{A.35}$$

Observe that Assumptions 2.1 and 4.1 imply that $(D_{i,j}^\star)_{i,j \in [n]}$ satisfies a representation of the form (A.10), with latent coordinates $\overline{S}_i$, and that the random variables $(\overline{S}_i, W_i^\star)_{i=1}^n$ are i.i.d., and satisfy the condition

$$\begin{aligned} \mathbb{E}[W_i^\star \mid \overline{S}_i] &= \mathbb{E}[\mathbb{E}[W_i^\star \mid X_i, \overline{S}_i] \mid \overline{S}_i] \\ &= \mathbb{E}[\mathbb{E}[W_i \mid X_i, \overline{S}_i] \mid S_i] - \mathbb{E}[\mathbb{E}[W_i \mid X_i] \mid \overline{S}_i] \\ &= \mathbb{E}[\mathbb{E}[W_i \mid X_i] \mid \overline{S}_i] - \mathbb{E}[\mathbb{E}[W_i \mid X_i] \mid \overline{S}_i] = 0 \ , \end{aligned} \tag{A.36}$$

almost surely. Now, define

$$\tilde{D}_{i,q}^\star = D_{i,q}^\star - \mathbb{E}[D_{i,q}^\star \mid S_q] \tag{A.37}$$

and observe that Assumption 4.3 implies that

$$\begin{aligned} \mathbb{E}[D_{i,q} - \mathbb{E}[D_{i,q} \mid H_{i,q}] \mid H_{i,q}, S_q] &= \mathbb{E}[D_{i,q} \mid H_{i,q}, S_q] - \mathbb{E}[D_{i,q} \mid H_{i,q}] \\ &= \psi_n(S_q) \\ &= \mathbb{E}[\psi_n(S_q) \mid S_q] \\ &= \mathbb{E}[\mathbb{E}[D_{i,q} \mid H_{i,j}, S_q] - \mathbb{E}[D_{i,q} \mid H_{i,q}] \mid S_q] \\ &= \mathbb{E}[D_{i,q} - \mathbb{E}[D_{i,q} \mid H_{i,q}] \mid S_q] \ . \end{aligned} \tag{A.38}$$

Consequently, we have that

$$\begin{aligned} \tilde{D}_{i,q}^\star &= D_{i,q} - \mathbb{E}[D_{i,q} \mid H_{i,q}] - \mathbb{E}[D_{i,q} - \mathbb{E}[D_{i,q} \mid H_{i,q}] \mid S_q] \\ &= D_{i,q} - \mathbb{E}[D_{i,q} \mid H_{i,q}] - \mathbb{E}[D_{i,q} - \mathbb{E}[D_{i,q} \mid H_{i,q}] \mid H_{i,q}, S_q] \\ &= D_{i,q} - \mathbb{E}[D_{i,q} \mid H_{i,q}, S_q] = D_{i,q} - \mathbb{E}[D_{i,q} \mid H_{i,q}, \overline{S}_q] \ , \end{aligned} \tag{A.39}$$

almost surely. Thus, the conditions of Lemma A.1 are satisfied with $D_{i,j}^\star$ taking the role of $M_{i,j}$ and $W_i^\star$ taking the role of W_i. Hence, we obtain the expansion

$$\begin{aligned} \overline{\theta}_n^\star &= \frac{\mathbb{E}[Y_i \tilde{D}_{i,q}^\star W_q^\star] - \mathbb{E}[Y_i \tilde{D}_{l,q}^\star W_q^\star]}{\mathbb{E}[\mathrm{Var}(D_{l,q}^\star \mid \overline{S}_q) \,\mathrm{Var}(W_q^\star \mid \overline{S}_q)]} \\ &\quad + \frac{1}{n} \frac{\mathbb{E}[Y_i \tilde{D}_{l,q}^\star W_q^\star] - \mathbb{E}[Y_i \tilde{D}_{l,i}^\star W_i^\star]}{\mathbb{E}[\mathrm{Var}(D_{l,q}^\star \mid \overline{S}_q) \,\mathrm{Var}(W_q^\star \mid \overline{S}_q)]} + \frac{1}{n^2} \frac{\mathbb{E}[Y_i \tilde{D}_{i,q}^\star W_q^\star]}{\mathbb{E}[\mathrm{Var}(D_{l,q}^\star \mid \overline{S}_q) \,\mathrm{Var}(W_q^\star \mid \overline{S}_q)]} \end{aligned} \tag{A.40}$$

$$+\frac{n+1}{n^2}\frac{\mathbb{E}[\mathbb{E}[Y_i W_q^\star \mid G_{l,q},\overline{S}_q]\mathbb{E}[D_{l,q}^\star \mid H_{l,q},\overline{S}_q]]}{\mathbb{E}[\mathrm{Var}(D_{l,q}^\star \mid \overline{S}_q)\,\mathrm{Var}(W_q^\star \mid \overline{S}_q)]}$$
$$-\frac{1}{n}\frac{\mathbb{E}[\mathbb{E}[Y_q W_q^\star \mid H_{l,q},\overline{S}_q]\mathbb{E}[D_{l,q}^\star \mid H_{l,q},\overline{S}_q]]}{\mathbb{E}[\mathrm{Var}(D_{l,q}^\star \mid \overline{S}_q)\,\mathrm{Var}(W_q^\star \mid \overline{S}_q)]}+O(n^{-2})\ .$$

Now, observe that Assumption 4.2 implies that

$$W_q \perp\!\!\!\perp D_{l,q} \mid S_q, H_{l,q}, X_q \quad \text{and} \quad \mathbb{E}[W_q \mid X_q] = \mathbb{E}[W_q \mid S_q, H_{l,q}] \tag{A.41}$$

respectively. Thus, Lemma A.2 implies that

$$\mathbb{E}[Y_i \tilde{D}_{i,q}^\star W_q^\star] = \int_{\mathcal{D}}\int_{\mathcal{W}} \mathbb{E}[\mathbb{E}[\mathbb{I}\{W_q \geq w\}W_q^\star \mid \overline{S}_q]\mathbb{E}[\mathbb{I}\{D_{i,q} \geq \delta\}\tilde{D}_{i,q}^\star \mid H_{i,q},\overline{S}_q]\}$$
$$\partial_{\delta,w}^2 \mathbb{E}[Y_i \mid D_{i,q} = \delta, W_q = w, H_{i,q},\overline{S}_q]]\,\mathrm{d}w\,\mathrm{d}\delta\ , \tag{A.42}$$
$$\mathbb{E}[Y_i \tilde{D}_{l,q} W_q^\star] = \int_{\mathcal{D}}\int_{\mathcal{W}} \mathbb{E}[\mathbb{E}[\mathbb{I}\{W_q \geq w\}W_q^\star \mid \overline{S}_q]\mathbb{E}[\mathbb{I}\{D_{l,q} \geq \delta\}\tilde{D}_{l,q}^\star \mid H_{i,q},\overline{S}_q]$$
$$\partial_{\delta,w}^2 \mathbb{E}[Y_i \mid D_{l,q} = \delta, W_q = w, H_{i,q},\overline{S}_q]]\,\mathrm{d}w\,\mathrm{d}\delta\ ,$$
$$\mathbb{E}[Y_i W_q^\star \mid H_{i,q}, X_q, \overline{S}_q] = \int_{\mathcal{W}} \mathbb{E}[\mathbb{I}\{W_q \geq w\}W^\star q \mid \overline{S}_q]\partial_w \mathbb{E}[Y_i \mid W_q^\star = w, H_{i,q},\overline{S}_q]\,\mathrm{d}w, \quad \text{and}$$
$$\mathbb{E}[Y_q W_q^\star \mid \overline{S}_q] = \int_{\mathcal{W}} \mathbb{E}[\mathbb{I}\{W_q^\star \geq w\}W_q^\star \mid \overline{S}_q]\partial_w \mathbb{E}[Y_q \mid W_q^\star = w, H_{i,q},\overline{S}_q]\,\mathrm{d}w\ ,$$

respectively. Consequently, by plugging each of the representations (A.42) into the expression (A.40), we find that

$$\overline{\theta}_n^\star = \int_{\overline{\mathcal{D}}}\int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{l,q},\overline{S}_q)(\partial_{\delta,w}^2 \mathbb{E}[Y_l \mid D_{i,q} = \delta, W_q = w, H_{l,q},\overline{S}_q] \tag{A.43}$$
$$- \partial_{\delta,w}^2 \mathbb{E}[Y_i \mid D_{l,q} = \delta, W_q = w, H_{l,q},\overline{S}_q])]\,\mathrm{d}w\,\mathrm{d}\delta$$
$$+\frac{1}{n}\int_{\overline{\mathcal{D}}}\int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{l,q},\overline{S}_q)(\partial_{\delta,w}^2 \mathbb{E}[Y_i \mid D_{l,q} = \delta, H_{l,q},\overline{S}_q]$$
$$- \partial_{\delta,w}^2 \mathbb{E}[Y_q \mid D_{l,q} = \delta, W_q = w, H_{l,q},\overline{S}_q])]]\,\mathrm{d}w\,\mathrm{d}\delta$$
$$+\frac{1}{n^2}\int_{\overline{\mathcal{D}}}\int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{l,q},\overline{S}_q)\partial_{\delta,w}^2 \mathbb{E}[Y_i \mid D_{i,q} = \delta, W_q = w, H_{l,q},\overline{S}_q]]\,\mathrm{d}w\,\mathrm{d}\delta$$
$$+\frac{1}{n}\int_{\overline{\mathcal{W}}} \mathbb{E}\Bigg[\lambda'(w \mid H_{l,q},\overline{S}_q)\Bigg(\left(\frac{n+1}{n}\right)\partial_w \mathbb{E}[Y_i \mid W_q = w, H_{l,q},\overline{S}_q]$$
$$- \partial_w \mathbb{E}[Y_q \mid W_q = w, H_{l,q},\overline{S}_q]\Bigg)\Bigg]\,\mathrm{d}w + O(n^{-2})\ ,$$

where

$$\lambda(\delta, w \mid H_{l,q},\overline{S}_q) = \frac{\mathrm{Cov}(\mathbb{I}\{D_{l,q} \geq \delta\}, D_{l,q} \mid H_{l,q},\overline{S}_q)\,\mathrm{Cov}(\mathbb{I}\{W_l \geq w\}, W_l \mid \overline{S}_q)}{\mathbb{E}[\mathrm{Var}(D_{l,q}^\star \mid \overline{S}_q)\,\mathrm{Var}(W_q^\star \mid \overline{S}_q)]} \quad \text{and} \tag{A.44}$$
$$\lambda'(w \mid X_q, H_{l,q},\overline{S}_q) = \frac{\mathbb{E}[\mathbb{E}[D_{l,q}^\star \mid H_{l,q},\overline{S}_q]\,\mathrm{Cov}(\mathbb{I}\{W_q \geq w\}, W_q \mid \overline{S}_q)}{\mathbb{E}[\mathrm{Var}(D_{l,q}^\star \mid \overline{S}_q)\,\mathrm{Var}(W_q^\star \mid \overline{S}_q)]}\ ,$$

respectively. Thus, it suffices to show that all but the first quantity in (A.43) are sufficiently small.

To this end, observe that Lemma A.3 implies that

$$\int_{\overline{\mathcal{W}}} \operatorname{Cov}(\mathbb{I}\{W_q \geq w\}, W_q \mid \overline{S}_q)\, \mathrm{d}w = \operatorname{Var}(W_q \mid X_q, \overline{S}_q) \quad \text{and} \tag{A.45}$$

$$\int_{\overline{\mathcal{D}}} \operatorname{Cov}(\mathbb{I}\{D_{l,q} \geq \delta\}, D_{l,q} \mid H_{l,q}, \overline{S}_q)\, \mathrm{d}\delta = \operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q)\ .$$

Thus, it holds that

$$\begin{aligned}
&\mathbb{E}\left[\int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \lambda(\delta, w \mid H_{l,q}, \overline{S}_q)\, \mathrm{d}w\, \mathrm{d}\delta\right] \qquad\qquad (A.46)\\
&= \frac{\mathbb{E}\left[\int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\operatorname{Cov}(\mathbb{I}\{D_{l,q} \geq \delta\}, D_{l,q} \mid H_{l,q}, \overline{S}_q) \operatorname{Cov}(\mathbb{I}\{W_q \geq w\}, W_q \mid \overline{S}_q)]\, \mathrm{d}w\, \mathrm{d}\delta\right]}{\mathbb{E}[\operatorname{Var}(D^\star_{l,q} \mid \overline{S}_q) \operatorname{Var}(W^\star_q \mid \overline{S}_q)]}\\
&= \frac{\mathbb{E}\left[\mathbb{E}[\mathbb{E}[\int_{\overline{\mathcal{D}}} \operatorname{Cov}(\mathbb{I}\{D_{l,q} \geq \delta\}, D_{l,q} \mid H_{l,q}, \overline{S}_q)\, \mathrm{d}\delta \mid \overline{S}_q] \mathbb{E}[\int_{\overline{\mathcal{W}}} \operatorname{Cov}(\mathbb{I}\{W_q \geq w\}, W_q \mid \overline{S}_q)\, \mathrm{d}w \mid \overline{S}_q]]\right]}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q] \operatorname{Var}(W_q \mid \overline{S}_q)]}\\
&= \frac{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q] \operatorname{Var}(W_q \mid \overline{S}_q)]}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q] \operatorname{Var}(W_q \mid \overline{S}_q)]} = 1\ ,
\end{aligned}$$

where the second equality follows from several applications of Fubini. Likewise, Lemma A.3 implies that

$$\lambda(\delta, w \mid H_{l,q}, \overline{S}_q) \geq 0\ , \tag{A.47}$$

almost surely. Hence, Assumption A.3 implies that

$$\begin{aligned}
&\int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{l,q}, \overline{S}_q) \partial^2_{\delta,w} \mathbb{E}[Y_i \mid D_{l,q} = \delta, W_q = w, H_{l,q}, \overline{S}_q]]\, \mathrm{d}w\, \mathrm{d}\delta = o(n^{-1/2})\ , \qquad (A.48)\\
&\frac{1}{n} \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid X_q, H_{l,q}, \overline{S}_q)(\partial^2_{\delta,w} \mathbb{E}[Y_i \mid D_{l,q} = \delta, W_q = w, H_{l,q}, \overline{S}_q]\\
&\qquad - \partial^2_{\delta,w} \mathbb{E}[Y_q \mid D_{l,q} = \delta, W_q = w, H_{l,q}, \overline{S}_q])]]\, \mathrm{d}w\, \mathrm{d}\delta = O(n^{-1})\ ,
\end{aligned}$$

and

$$\frac{1}{n^2} \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{l,q}, \overline{S}_q) \partial^2_{\delta,w} \mathbb{E}[Y_i \mid D_{i,q} = \delta, W_q = w, H_{l,q}, \overline{S}_q]]\, \mathrm{d}w\, \mathrm{d}\delta = O(n^{-2})\ , \tag{A.49}$$

respectively, and it so remains to bound the fourth term in (A.43).

Observe that we can decompose

$$\begin{aligned}
\lambda'(w \mid H_{l,q}, \overline{S}_q) &= \frac{\mathbb{E}[D_{l,q} \mid H_{l,q}, \overline{S}_q] \operatorname{Cov}(\mathbb{I}\{W_q \geq w\}, W_q \mid \overline{S}_q)}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q] \operatorname{Var}(W_q \mid \overline{S}_q)]} \qquad (A.50)\\
&\quad - \frac{\mathbb{E}[D_{l,q} \mid H_{l,q}] \operatorname{Cov}(\mathbb{I}\{W_q \geq w\}, W_q \mid \overline{S}_q)}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q] \operatorname{Var}(W_q \mid \overline{S}_q)]}\ .
\end{aligned}$$

Lemma A.3 and the fact that the distances $D_{l,q}$ are positive imply that both terms in (A.50) are positive almost surely. Moreover, it holds that

$$\frac{\mathbb{E}[\mathbb{E}[D_{l,q} \mid H_{l,q}, \overline{S}_q] \operatorname{Cov}(\mathbb{I}\{W_q \geq w\}, W_q \mid \overline{S}_q)]}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q] \operatorname{Var}(W_q \mid \overline{S}_q)]}$$

$$\lesssim \frac{\mathbb{E}[\mathbb{E}[\mathbb{E}[D_{l,q} \mid H_{l,q}, \overline{S}_q] \mid \overline{S}_q]\sqrt{\operatorname{Var}(W_q \mid \overline{S}_q)}]}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q]\operatorname{Var}(W_q \mid \overline{S}_q)]}$$

$$\lesssim \frac{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q]\sqrt{\operatorname{Var}(W_q \mid \overline{S}_q)}]}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q]\operatorname{Var}(W_q \mid \overline{S}_q)]} = O(1) \tag{A.51}$$

uniformly in w, where the first inequality follows from Lemma A.3 and Cauchy-Schwarz and the second inequality follows from Assumption A.4. Analogously, it holds that

$$\frac{\mathbb{E}[\mathbb{E}[D_{l,q} \mid H_{l,q}]\operatorname{Cov}(\mathbb{I}\{W_q \geq w\}, W_q \mid \overline{S}_q)]}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q]\operatorname{Var}(W_q \mid \overline{S}_q)]}$$

$$\leq \frac{\mathbb{E}[\mathbb{E}[\mathbb{E}[\mathbb{E}[D_{l,q} \mid H_{l,q}, \overline{S}_q] \mid H_{l,q}] \mid \overline{S}_q]\sqrt{\operatorname{Var}(W_q \mid \overline{S}_q)}]}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q]\operatorname{Var}(W_q \mid \overline{S}_q)]}$$

$$\lesssim \frac{\mathbb{E}[\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid H_{l,q}] \mid \overline{S}_q]\sqrt{\operatorname{Var}(W_q \mid \overline{S}_q)}]}{\mathbb{E}[\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q) \mid \overline{S}_q]\operatorname{Var}(W_q \mid \overline{S}_q)]}$$

$$\lesssim \frac{\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q)]}{\mathbb{E}[\operatorname{Var}(D_{l,q} \mid H_{l,q}, \overline{S}_q)]} = O(1) \tag{A.52}$$

uniformly in w, where the first inequality follows from Lemma A.3 and Cauchy-Schwarz and the second and third inequalities follow from Assumption A.4. Thus, by Assumption A.3 and the fact that the treatments and distances are defined on sets whose supports are bounded by a constant, we find that

$$\frac{1}{n}\int_{\overline{\mathcal{W}}} \mathbb{E}\Bigg[\lambda'(w \mid H_{l,q}, \overline{S}_q)\Bigg(\left(\frac{n+1}{n}\right)\partial_w \mathbb{E}[Y_i \mid W_q = w, H_{l,q}, \overline{S}_q] - \partial_w \mathbb{E}[Y_q \mid W_q = w, H_{l,q}, \overline{S}_q]\Bigg)\Bigg]\,\mathrm{d}w = O(n^{-1})\,. \tag{A.53}$$

Hence, we obtain the representation

$$\overline{\theta}^{\star}_n = \int_{\overline{\mathcal{D}}}\int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_j, S_j)\partial^2_{\delta,w}\,\mu(\delta, w \mid H_j, S_j)]\,\mathrm{d}w\,\mathrm{d}\delta + o(n^{-1/2})\,, \tag{A.54}$$

by plugging the bounds (A.48) and (A.53) into (A.43). The convexity of the weights is verified by the statements (A.46) and (A.47), completing the proof. ■

A.3 Proofs for Supporting Lemmas

A.3.1 Proof of Lemma A.1. We begin with the equality (A.13). Consider the decomposition

$$\hat{\Xi}_i = \sum_{j\neq i}\left(M_{i,j} - \frac{1}{n}\sum_{k\neq j} M_{k,j}\right)W_j - \frac{1}{n}\sum_{k\neq i} M_{k,i}W_i = Q_i^{(0)} + Q^{(1)} - Q_i^{(2)} + Q^{(3)}\,, \tag{A.55}$$

where

$$Q_i^{(0)} = \sum_{j\neq i}\left(M_{i,j} - \frac{1}{n-1}\sum_{k\neq j} M_{k,j}\right)W_j\,, \qquad Q^{(1)} = \frac{1}{n(n-1)}\sum_{j\neq i}\sum_{k\notin\{i,j\}} M_{k,j}W_j \tag{A.56}$$

$$Q_i^{(2)} = \frac{1}{n}\sum_{j\neq i} M_{j,i}W_i\ , \quad \text{and} \qquad\qquad Q^{(3)} = \frac{1}{n(n-1)}\sum_{j\neq i} M_{i,j}W_j\ . \tag{A.57}$$

We can evaluate

$$\begin{aligned}\mathbb{E}[Y_iQ_i^{(0)}] &= \sum_{j\neq i}\mathbb{E}[Y_i\tilde{M}_{i,j}W_j] - \frac{1}{n-1}\sum_{j\neq i}\sum_{k\neq j}\mathbb{E}[Y_i\tilde{M}_{k,j}W_j]\\ &= \frac{(n-2)}{n-1}\sum_{j\neq i}\mathbb{E}[Y_i\tilde{M}_{i,j}W_j] - \frac{1}{n-1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\mathbb{E}[Y_i\tilde{M}_{k,j}W_j]\\ &= (n-2)\left(\mathbb{E}[Y_i\tilde{M}_{i,q}W_q] - \mathbb{E}[Y_i\tilde{M}_{l,q}W_q]\right) \end{aligned}\tag{A.58}$$

and

$$\mathbb{E}[Y_i(Q^{(1)} - Q_i^{(2)} + Q^{(3)})] = \frac{n-2}{n}\mathbb{E}[Y_iM_{q,l}W_l] - \frac{n-1}{n}\mathbb{E}[Y_iM_{l,i}W_i] + \frac{1}{n}\mathbb{E}[Y_iM_{i,l}W_l]\ , \tag{A.59}$$

respectively. Hence, we have that

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n\mathbb{E}[Y_i\hat{\Xi}_i] &= \frac{n-2}{n}\left(\mathbb{E}\left[Y_i\tilde{M}_{i,l}W_l\right] - \mathbb{E}\left[Y_i\tilde{M}_{q,l}W_l\right]\right)\\ &\quad + \frac{n-2}{n^2}\mathbb{E}[Y_iM_{q,l}W_l] - \frac{n-1}{n^2}\mathbb{E}[Y_iM_{l,i}W_i] + \frac{1}{n^2}\mathbb{E}[Y_iM_{i,l}W_l]\end{aligned}\tag{A.60}$$

as required.

Next, we verify the inequality (A.14). Observe that

$$\begin{aligned}\mathbb{E}[Q_i^{(0)}Q_i^{(1)}] &= \frac{1}{n(n-1)}\sum_{j\neq i}\sum_{j'\neq i}\sum_{k'\notin\{i,j'\}}\mathbb{E}\left[\left(M_{i,j} - \frac{1}{n-1}\sum_{k\neq j}M_{k,j}\right)M_{k',j'}^\top W_jW_j'\right]\\ &= \frac{1}{n(n-1)}\sum_{j\neq i}\sum_{k'\notin\{i,j\}}\mathbb{E}\left[\left(\tilde{M}_{i,j} - \frac{1}{n-1}\sum_{k\neq j}\tilde{M}_{k,j}\right)M_{k',j}\operatorname{Var}(W_j\mid S_j)\right]\\ &= -\frac{1}{n(n-1)^2}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\mathbb{E}[\tilde{M}_{k,j}\tilde{M}_{k,j}\operatorname{Var}(W_j\mid S_j)]\\ &= \frac{n-2}{n(n-1)}\mathbb{E}[\tilde{M}_{i,l}^2\operatorname{Var}(W_l\mid S_l)]\end{aligned}\tag{A.61}$$

as the variables W_j are mutually independent and satisfy the restrictions (A.11) and the variables $M_{i,j}$ and $M_{l,j}$ are independent conditional on S_j. We can evaluate

$$\mathbb{E}[Q_i^{(0)}Q_i^{(2)}] = 0 \quad \text{and} \quad \mathbb{E}[Q_i^{(0)}Q^{(3)}] = \frac{n-2}{n(n-1)}\mathbb{E}[\tilde{M}_{i,l}^2\operatorname{Var}(W_l\mid S_l)] \tag{A.62}$$

analogously. As a consequence, it holds that

$$\begin{aligned}&\mathbb{E}[\hat{\Xi}_i^2] - \operatorname{Var}(Q_i^{(0)})\\ &\lesssim \frac{1}{n^2(n-1)^2}\sum_{j\neq i}^n\mathbb{E}\left[\sum_{k\notin\{i,j\}}M_{k,j}^2W_j^2\right] + \frac{1}{n^2}\mathbb{E}\left[\sum_{j\neq i}M_{j,i}^2W_i^2\right]\end{aligned}$$

$$+\frac{1}{n^2(n-1)^2}\sum_{j\neq i}\mathbb{E}\left[M_{i,j}^2W_j^2\right]+\frac{n-2}{n(n-1)}\mathbb{E}[\tilde{M}_{i,l}^2\operatorname{Var}(W_l\mid S_l)]=O(n^{-1})\,, \tag{A.63}$$

by the Cauchy-Schwarz inequality and the fact that the supports of the variables W_j and $M_{i,j}$ are bounded by a constant. To conclude the proof, observe that we can evaluate

$$\operatorname{Var}(Q_i^{(0)})=\mathbb{E}\left[\left(\sum_{j\neq i}\left(M_{i,j}-\frac{1}{n-1}\sum_{k\neq j}M_{k,j}\right)W_j\right)^2\right]=(n-2)\mathbb{E}\left[\tilde{M}_{i,l}^2\operatorname{Var}(W_l\mid S_l)\right]\;, \tag{A.64}$$

by the same steps as before. Hence, the inequality (A.14) follows from (A.63) and (A.64). ■

A.3.2 Proof of Lemma A.2. The result is based on two applications of the following Lemma, obtained from a method of argument developed in Kolesár and Plagborg-Møller (2024).

Lemma A.4. *Fix the random variables X, Y, and U. If the distribution of X, conditioned on U, has support contained in the bounded set $\mathcal{X}$ in $\mathbb{R}$ almost surely, Y is real-valued, and the conditional expectation $\mathbb{E}[Y\mid X=x,Z]$ is locally absolutely continuous and uniformly bounded on $\mathcal{X}$, then the equality*

$$\operatorname{Cov}(Y,X\mid U)=\int_{\mathcal{X}}\mathbb{E}[\mathbb{I}\{X\geq x\}(X-\mathbb{E}[X\mid U])\mid U]\,\partial_x\mathbb{E}[Y\mid X=x,U]\,\mathrm{d}x \tag{A.65}$$

holds almost surely.

Observe that the equality (A.16) follows immediately from Lemma A.4. To verify the second equality, observe that

$$\mathbb{E}[Y(Z-\mathbb{E}[Z\mid U])(W-\mathbb{E}[W\mid U])]=\mathbb{E}[Y(Z-\mathbb{E}[Z\mid U,W])(W-\mathbb{E}[W\mid U])] \tag{A.66}$$

$$=\mathbb{E}[\operatorname{Cov}(Y,Z\mid U,W)(W-\mathbb{E}[W\mid U])], \tag{A.67}$$

as W is independent of Z conditional on U. Lemma A.4 implies that

$$\operatorname{Cov}(Y,Z\mid U,W)=\int_{\mathcal{Z}}\mathbb{E}[\mathbb{I}\{Z\geq z\}(Z-\mathbb{E}[Z\mid U,W])\mid U]\,\partial_z\mathbb{E}[Y\mid Z=z,U,W]\,\mathrm{d}z\,. \tag{A.68}$$

Thus, we can evaluate

$$\begin{aligned}
&\mathbb{E}[Y(Z-\mathbb{E}[Z\mid U])(W-\mathbb{E}[W\mid U])]\\
&=\mathbb{E}\left[\left(\int_{\mathcal{Z}}\mathbb{E}[\mathbb{I}\{Z\geq z\}(Z-\mathbb{E}[Z\mid U,W])\mid U,W]\,\partial_z\mathbb{E}[Y\mid Z=z,U,W]\,\mathrm{d}z\right)(W-\mathbb{E}[W\mid U])\right]\\
&=\int_{\mathcal{Z}}\mathbb{E}\left[\mathbb{E}[\mathbb{I}\{Z\geq z\}(Z-\mathbb{E}[Z\mid U,W])\mid U,W](W-\mathbb{E}[W\mid U])\,\partial_z\mathbb{E}[Y\mid Z=z,U,W]\right]\mathrm{d}z\\
&=\int_{\mathcal{Z}}\mathbb{E}\left[\mathbb{E}[\mathbb{I}\{Z\geq z\}(Z-\mathbb{E}[Z\mid U])\mid U]\mathbb{E}\left[(W-\mathbb{E}[W\mid U])\,\partial_z\mathbb{E}[Y\mid Z=z,U,W]\mid U\right]\right]\mathrm{d}z\,,
\end{aligned} \tag{A.69}$$

where the second equality follows from Fubini's Theorem and the third equality follows from the fact that W is independent of Z conditional on U. Lemma A.4 again implies that

$$\begin{aligned}
&\mathbb{E}\left[\partial_z\mathbb{E}[Y\mid Z=z,U,W](W-\mathbb{E}[W\mid U])\mid U\right]\\
&=\operatorname{Cov}(\partial_z\mathbb{E}[Y\mid Z=z,U,W],W\mid U)
\end{aligned}$$

$$= \int_{\mathcal{W}} \mathbb{E}[\mathbb{I}\{W \geq w\}(W - \mathbb{E}[W \mid U]) \mid U] \partial^2_{z,w} \mathbb{E}[Y \mid Z = z, W = w, U] \mid U] \, \mathrm{d}w \, . \tag{A.70}$$

By plugging (A.70) into (A.69), and applying Fubini's Theorem, we obtain

$$\begin{aligned}
&\mathbb{E}[Y(Z - \mathbb{E}[Z \mid U])(W - \mathbb{E}[W \mid U])] \\
&= \int_{\mathcal{Z}} \int_{\mathcal{W}} \mathbb{E}[\mathbb{E}[\mathbb{I}\{W \geq w\}(W - \mathbb{E}[W \mid U]) \mid U] \mathbb{E}[\mathbb{I}\{Z \geq z\}(Z - \mathbb{E}[Z \mid U]) \mid U] \\
&\qquad \partial^2_{z,w} \mathbb{E}[Y \mid Z = z, W = w, U]] \, \mathrm{d}w \, \mathrm{d}z \, ,
\end{aligned} \tag{A.71}$$

as required. ■

A.3.3 Proof of Lemma A.3. First, observe that

$$\begin{aligned}
&\mathrm{Cov}(\mathbb{I}\{Z \geq z\}, Z \mid U) \\
&= \mathbb{E}[\mathbb{I}\{Z \geq z\}(Z - \mathbb{E}[Z \mid U]) \mid U] \\
&= \mathrm{Var}(\mathbb{I}\{Z \geq z\} \mid U)(\mathbb{E}[(Z - \mathbb{E}[Z \mid U]) \mid Z \geq z, U] - \mathbb{E}[(Z - \mathbb{E}[Z \mid U]) \mid Z < z, U]) \geq 0 \, ,
\end{aligned} \tag{A.72}$$

as required. In turn, by Fubini's theorem, it holds that

$$\begin{aligned}
&\int_{\mathcal{Z}} \mathrm{Cov}(\mathbb{I}\{Z \geq z\}, Z \mid U) \, \mathrm{d}z \\
&= \mathbb{E}\left[\int_{\mathcal{Z}} \mathbb{I}\{Z \geq z\}(Z - \mathbb{E}[Z \mid U]) \, \mathrm{d}z \mid U\right] = \mathrm{Var}(Z \mid U) \, ,
\end{aligned} \tag{A.73}$$

as required. ■

A.3.4 Proof of Lemma A.4. Define the conditionally centered random variable

$$\tilde{X} = X - \mathbb{E}[X \mid U] \, . \tag{A.74}$$

The fact that $\mathbb{E}[\tilde{X} \mid U] = 0$ implies that

$$\mathbb{E}[\mathbb{I}\{X \geq x\}\tilde{X} \mid U] = -\mathbb{E}[\mathbb{I}\{X < x\}\tilde{X} \mid U] \tag{A.75}$$

almost surely for each x in $\mathcal{X}$. Thus, choosing an arbitrary point x_0 in $\mathcal{X}$, we obtain

$$\begin{aligned}
&\int_{\mathcal{X}} \mathbb{E}[\mathbb{I}\{X \geq x\}\tilde{X} \mid U] \partial_x \mathbb{E}[Y \mid X = x, U] \, \mathrm{d}x \\
&\qquad = \int_{\mathcal{X}} \mathbb{E}[\mathbb{I}\{X \geq x \geq x_0\}\tilde{X} \mid U] \partial_x \mathbb{E}[Y \mid X = x, U] \, \mathrm{d}x \\
&\qquad - \int_{\mathcal{X}} \mathbb{E}[\mathbb{I}\{X < x < x_0\}\tilde{X} \mid U] \partial_x \mathbb{E}[Y \mid X = x, U] \, \mathrm{d}x \, .
\end{aligned} \tag{A.76}$$

Now, we can evaluate

$$\begin{aligned}
&\int_{\mathcal{X}} \mathbb{E}[\mathbb{I}\{X \geq x \geq x_0\}\tilde{X} \mid U] \partial_x \mathbb{E}[Y \mid X = x, U] \, \mathrm{d}x \\
&= \mathbb{E}\left[\int_{\mathcal{X}} \mathbb{I}\{X \geq x \geq x_0\}\tilde{X} \partial_x \mathbb{E}[Y \mid X = x, U] \, \mathrm{d}x \mid Z\right]
\end{aligned}$$

$$= \mathbb{E}\left[\mathbb{I}\{X \geq x_0\}\tilde{X}(\mathbb{E}[Y \mid X, U] - \mathbb{E}[Y \mid X = x_0, U]) \mid Z\right] \tag{A.77}$$

almost surely, where the first equality follows from Fubini's Theorem and the second equality follows from the Fundamental Theorem of Calculus, both of which are justified by the fact that the domain of X and the derivative $\partial_x \mathbb{E}[Y \mid X = x, U]$ are uniformly bounded, almost surely. By an identical argument we have that

$$\int_{\mathcal{X}} \mathbb{E}[\mathbb{I}\{X < x < x_0\}\tilde{X} \mid U]\partial_x \mathbb{E}[Y \mid X = x, U] \, \mathrm{d}x$$
$$= \mathbb{E}\left[\mathbb{I}\{X < x_0\}\tilde{X}(\mathbb{E}[Y \mid X = x_0, U] - \mathbb{E}[Y \mid X, U]) \mid U\right] \tag{A.78}$$

almost surely. Plugging (A.77) and (A.78) into (A.76) , we find that

$$\int_{\mathcal{X}} \mathbb{E}[\mathbb{I}\{X \geq x\}\tilde{X} \mid U]\partial_x \mathbb{E}[Y \mid X = x, U] \, \mathrm{d}x$$
$$= \mathbb{E}\left[\tilde{X}\mathbb{E}[Y \mid X, U] \mid U\right] = \mathrm{Cov}(Y, X \mid U) \,, \tag{A.79}$$

as required. ■

Appendix B. Proofs for Identification Results

In this Appendix, we give proofs for the corollaries stated in Section 4. Throughout, let the sets

$$\mathcal{D}(H_{i,j}, S_j) \quad \text{and} \quad \mathcal{W}(\overline{S}_j) \tag{B.1}$$

denote the support sets of the proximity measure $D_{i,j}$, conditional on $H_{i,j}$ and S_j, and the treatment W_j, conditional on $\overline{S}_j$, respectively. Define the set

$$\mathcal{F} = \mathcal{F}(H_{i,j}, \overline{S}_j) = \mathcal{D}(H_{i,j}, S_j) \times \mathcal{W}(\overline{S}_j) \,. \tag{B.2}$$

with an abuse of notation.

B.1 Proof of Corollary 4.1

Observe that the choice (3.4) is uniquely-defined on the event that $D_{i,j} = \delta$. Thus, Assumption 3.1 implies that, in this case, the potential outcome

$$Y_{i,j}(\delta, w) = F(Z_i, Z_j(S_j^{(i)}(\delta), w), Z_{-i,j}) \,, \quad \text{where} \quad Z_j(w, s) = (w, s, U_j) \,, \tag{B.3}$$

is uniquely-defined. Observe that Assumption 4.2 implies that the collections Z_i and $Z_{-i,j}$ are independent of treatment W_j. Thus, for each (δ, w) in $\mathcal{F}$, it holds that

$$\begin{aligned}\mathbb{E}[Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, W_j = w, H_{i,j}, \overline{S}_j] &= \mathbb{E}[F(Z_i, Z_j(\xi_j^{(i)}(\delta), w), Z_{-i,j}) \mid D_{i,j} = \delta, W_j = w, H_{i,j}, \overline{S}_j] \\ &= \mathbb{E}[F(Z_i, Z_j(\xi_j^{(i)}(\delta), w), Z_{-i,j}) \mid D_{i,j} = \delta, H_{i,j}, \overline{S}_j] \\ &= \mathbb{E}[Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, H_{i,j}, \overline{S}_j] \,. \end{aligned} \tag{B.4}$$

Hence, we can evaluate

$$\theta_n^\star = \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j)\partial^2_{\delta,w} \mathbb{E}[Y_{i,j} \mid D_{i,j} = \delta, W_j = w, H_{i,j}, \overline{S}_j]] \, \mathrm{d}w \, \mathrm{d}\delta$$

$$= \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j) \partial^2_{\delta,w} \mathbb{E}[Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, W_j = w, H_{i,j}, \overline{S}_j]] \, \mathrm{d}w \, \mathrm{d}\delta$$
$$= \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j) \partial_\delta \mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, H_{i,j}, \overline{S}_j]] \, \mathrm{d}w \, \mathrm{d}\delta \; , \tag{B.5}$$

where the first two equalities follows by definition and the third equality follows by (B.4) and the Leibniz integral rule. Observe that, by the definition of the weight function $\lambda(\delta, w \mid H_{i,j}, S_j)$ and the convention used to define the operators ∂_δ, ∂_w, and $\partial^2_{\delta,w}$, the final equality in (B.5) only requires the application of the equality (B.4) at points (δ, w) in $\mathcal{F}$. ■

B.2 Proof of Corollary 4.2

Fix a scalar δ in $\mathcal{D}(H_{i,j}, S_j)$. Observe that the level set

$$\mathcal{S}^{(i)}(\delta) = \{s : D_n(S_i, s) = \delta\} \; . \tag{B.6}$$

is non-empty by definition. As the function $\overline{D}(\cdot)$ is continuous and monotone, by Assumption 3.2, the set

$$\{r \geq 0 : \overline{D}(\cdot) = \delta\} \tag{B.7}$$

is a closed interval $[r_-, r_+]$. Now, observe that

$$S_j + \alpha(S_i - S_j) \in \mathcal{S}^{(i)}(\delta) \quad \text{if and only if} \quad |1 - \alpha| \|S_i - S_j\| \in [r_-, r_+] \; . \tag{B.8}$$

If the distance $\|S_i - S_j\| = 0$, then every α is feasible, and $\alpha^{(i)}_j(\delta)$ is uniquely defined as 0. On the event that $\|S_i - S_j\| > 0$, we have that

$$\arg\min_{\alpha} \{|\alpha| : S_j + \alpha(S_i - S_j) \in \mathcal{S}^{(i)}(\delta)\} = 1 - \frac{r^\star}{\|S_i - S_j\|} \; , \tag{B.9}$$

where

$$r^\star \in \arg\min_{r \in [r_-, r_+]} \left| 1 - \frac{r}{\|S_i - S_j\|} \right| = r^\star \in \arg\min_{r \in [r_-, r_+]} |r - \|S_i - S_j\|| \; . \tag{B.10}$$

Observe that $r^\star$ is defined as the projection of the point $\|S_i - S_j\|$ onto the closed interval $[r_-, r_+]$, which is unique. Thus, the minimizer of the problem (B.9) is unique, and so the counterfactual coordinate

$$S^{(i)}_j(\delta) = S_j + \alpha^{(i)}_j(\delta)(S_i - S_j) \tag{B.11}$$

is uniquely defined. Hence, the potential outcome

$$Y_{i,j}(\delta, w) = F(Z_i, Z_j(S^{(i)}_j(\delta), w), Z_{-i,j}) \; , \tag{B.12}$$

is uniquely defined.

As a consequence, for each (δ, w) in $\mathcal{F}$, Assumption 4.4 implies that

$$\mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, H_{i,j}, \overline{S}_j] = \mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid H_{i,j}, \overline{S}_j] \; . \tag{B.13}$$

Hence, we can evaluate

$$\theta^\star_n = \int_{\mathcal{D}} \int_{\mathcal{W}} \mathbb{E}[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j) \partial_\delta \mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, H_{i,j}, \overline{S}_j]] \, \mathrm{d}w \, \mathrm{d}\delta$$

$$= \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j)\partial_\delta \, \mathbb{E}[\partial_w \, Y_{i,j}(\delta, w) \mid H_{i,j}, \overline{S}_j]] \, \mathrm{d}w \, \mathrm{d}\delta$$
$$= \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j)\mathbb{E}[\partial^2_{\delta,w} \, Y_{i,j}(\delta, w) \mid H_{i,j}, \overline{S}_j]] \, \mathrm{d}w \, \mathrm{d}\delta$$
$$= \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}[\lambda(\delta, w \mid H_{i,j}, \overline{S}_j)\partial^2_{\delta,w} \, Y_{i,j}(\delta, w)] \, \mathrm{d}w \, \mathrm{d}\delta \ , \tag{B.14}$$

where the first equality follows by Corollary 4.1, the second equality follows by (B.13), the third equality follows by the Leibniz integral rule, and the final equality follows by the law of iterated expectations. Again, by the definition of the weight function $\lambda(\delta, w \mid H_{i,j}, \overline{S}_j)$ and the operators ∂_δ and ∂_w, the third equality in (B.14) only requires the application of the equality (B.13) at points δ in $\mathcal{D}(H_{i,j}, S_j)$. ■

Appendix C. Proofs for Asymptotic Results

In this Appendix, we give proofs for the theorems stated in Section 5 and Section 6, respectively. These results hold under additional regularity conditions, stated in Appendix C.1. Proofs are given in the subsequent subsections. Proofs for all supporting Lemmas are given in Appendix D. We use the notation

$$\tilde{D}^\star_{i,l} = \tilde{D}^\star(S_i, S_l) = D_{i,l} - \mathbb{E}[D_{i,l} \mid H_{i,l}, S_l] \tag{C.1}$$

throughout.

C.1 Additional Regularity Conditions

C.1.1 Theorems 5.1 and 5.2. We begin by giving the additional moment bounds needed for Theorems 5.1 and 5.2. Throughout, we let $\mathcal{A}_{i,j}$ denote the event that $A_{i,j} = 1$ and let $\tilde{A}_{i,j}$ denote its complement. Likewise, define the indicator

$$B_{i,j} = \mathbb{I}\{P\{\overline{S}_k \in L^{(1)}_n(\overline{S}_i, \overline{S}_j) \mid \overline{S}_i, \overline{S}_j\} > 0\} \tag{C.2}$$

where the set $L^{(1)}_n(\overline{S}_i, \overline{S}_j)$ is defined in Assumption 5.2. Define the events $\mathcal{B}_{i,j}$ and $\tilde{\mathcal{B}}_{i,j}$, accordingly.

Assumption C.1. *Define the conditional expectations*

$$\overline{F}(Z_i, Z_j) = \mathbb{E}[Y_i \mid Z_i, Z_j] \quad \text{and} \quad \tilde{F}(Z_i, Z_j) = \mathbb{E}[Y_i \mid Z_i, Z_j] - \mathbb{E}[Y_i \mid Z_i] \ . \tag{C.3}$$

The means and variances of each of the conditional moments (C.3) are bounded by a constant almost surely. Moreover, for each $a, b \in \{0, 1\}$, the moments

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[\overline{F}(Z_i, Z_j)^2 \mid A_{i,j} = a, Z_i]] \ , && \mathbb{E}[\mathbb{E}[\overline{F}(Z_i, Z_j)^2 \mid A_{i,j} = a, Z_j]] \ , \\
&\mathbb{E}[\mathbb{E}[\overline{F}(Z_i, Z_j)^2 \mid A_{k,j} = a, Z_k]] \ , && \mathbb{E}[\mathbb{E}[\tilde{F}(Z_i, Z_j)^2 \mid A_{i,j} = a, Z_j]] \ , \\
&\mathbb{E}[\mathbb{E}[\tilde{F}(Z_i, Z_j)^2 \mid A_{k,j} = a, Z_k]] \ , && \mathbb{E}[\tilde{F}(Z_i, Z_k)^2 \mid A_{i,j} = a, A_{i,k} = 1] \ , \\
\text{and} \quad &\mathbb{E}[\mathbb{E}[\tilde{F}(Z_i, Z_k)^2 \mid A_{i,j} = a, Z_j, Z_k] \mid B_{j,k} = b]
\end{aligned} \tag{C.4}$$

are each bounded by a constant almost surely.

This condition is relatively innocuous, and states that, in the asymptotic sequences under consideration, the moments (C.4) do not grow without bound. We note that this condition does not necessarily imply that the variance of the outcomes is bounded by a constant. This is important in settings in which, for instance, each unit has an increasing number of neighbors whose influence on a given unit's outcome is on the order of a constant. We go through the trouble of enumerating the moment bounds enforced by Assumption C.1 because the proof of Theorem 5.2 is based on constructing distributions that satisfy the conditions of Theorem 5.1, but where the distributions of the quantities

$$\mathbb{E}[Y_i \mid Z_i, Z_j] \quad \text{and} \quad \mathbb{E}[Y_i \mid Z_i, Z_j] - \mathbb{E}[Y_i \mid Z_i] \tag{C.5}$$

are unbounded.

C.1.2 Theorems 6.1 and 6.2. Next, we state the three additional conditions needed for Theorems 6.1 and 6.2. The first condition is a mild strengthening of Assumption 5.2.

Assumption C.2 (Extended Latent Spatial Structure)**.** *Define the set*

$$L_n^{(2)}(\overline{S}_i, \overline{S}_j) = L_n^{(1)}(\overline{S}_i) \cap L_n^{(1)}(\overline{S}_j) \ . \tag{C.6}$$

It holds that

$$P\{P\{\overline{S}_k \in L_n^{(2)}(\overline{S}_i, \overline{S}_j) \mid \overline{S}_i, \overline{S}_j\} > 0\} = O(\rho_n) \ . \tag{C.7}$$

Recall that the second part of the relation (5.6), stated in Assumption 5.2, ensures that the fraction of pairs of units i, j that could feasibly be proximate to a third unit k is of order $O(\rho_n)$. Assumption C.2 is a further iteration of this idea. That is, the fraction of pairs of units i, j that *could* each be proximate to units that are proximate to k is of order $O(\rho_n)$. As before, this condition is satisfied in settings where the latent proximity indicators are generated by the specification (5.10), and the measure of the latent factors is doubling.

The second condition is necessary for ensuring that the leading term in our asymptotic approximation to the estimator $\hat{\theta}_n^\star$ is non-degenerate.

Assumption C.3 (Non-Degeneracy)**.** *Define the conditional expectation*

$$\Psi(Z_i, Z_j) = \mathbb{E}[\tilde{F}(Z_k, Z_i)\tilde{D}^\star_{k,j} \mid Z_i, Z_j] \ . \tag{C.8}$$

The lower bound

$$\mathrm{Var}\left(\begin{pmatrix} \Psi(Z_i, Z_j)W_j^\star + \Psi(Z_j, Z_i)W_i^\star \\ 2\mathbb{E}[\tilde{D}_{k,i}\tilde{D}^\star_{k,j} \mid Z_i, Z_j]W_i^\star W_j^\star \end{pmatrix} \mid A_{i,j} = 1 \right) \succeq c\rho_n^2 \mathbf{I}_2 \tag{C.9}$$

holds, where $\preceq$ denotes the Loewner order on matrices.

To unpack Assumption C.3, consider the case that the outcomes are mean independent of the states of non-proximate units, in the sense that

$$\text{if} \quad A_{i,j} = 0 \quad \text{then} \quad \mathbb{E}[Y_i \mid Z_i, Z_j] - \mathbb{E}[Y_i \mid Z_i] = 0 \tag{C.10}$$

almost surely. In this case, Assumption C.3 reduces to the lower bound

$$\mathrm{Var}\left(\tilde{\Psi}(Z_j, Z_i)W_j^\star + \tilde{\Psi}(Z_i, Z_j)W_i^\star \mid A_{i,j}\right) \succeq c\mathbf{I}_2 \ , \tag{C.11}$$

where

$$\tilde{\Psi}(Z_i, Z_j) = \mathbb{E}\left[\begin{pmatrix} \tilde{F}_{k,i} \\ \tilde{D}_{k,i}^\star W_i^\star \end{pmatrix} \tilde{D}_{k,j}^\star \mid A_{k,j}, Z_i, Z_j\right] \ . \tag{C.12}$$

Consider, in turn, the closely related, but even simpler, lower bound

$$\mathrm{Var}(\mathbb{E}[\tilde{F}_{k,i} D_{k,j} \mid A_{k,j}, Z_i, Z_j] \mid A_{i,j} = 1) \geq c \ . \tag{C.13}$$

In effect, the lower bound (C.13) asks that, when i is proximate to j *and* k, the probability that k is also proximate to j is on the order of a constant. To see this, observe that in the latter event, i.e., when k is proximate to both i and j, the term $\tilde{F}(Z_k, Z_i)D_{k,j}$ can be on the order of a constant.

In other words, Assumption C.3 can be interpreted as stipulating that when i and j are proximate, the proportion of j's neighbors that are also proximate to i is on the order of a constant. Clustering, in this sense, is widely observed in social network data and arises in models where connections are generated by the specifications analogous to (5.10), as well as many other standard models of network formation (see e.g., Jackson and Rogers 2007 and Graham 2016 for further discussion). However, Assumption C.3 also asks that there no cancellations that cause the eigenvalues of the variance covariance matrix in (C.11) to be much smaller than the variance in (C.13).

The final condition enforces a slightly stronger set of moment bounds. Define the indicator

$$E_{i,j} = \mathbb{I}\{P\{\overline{S}_k \in L_n^{(2)}(\overline{S}_i, \overline{S}_j) \mid \overline{S}_i, \overline{S}_j\} > 0\} \ ,$$

where the sets $L_n^{(2)}(S_i, S_j)$ is defined in Assumption C.2. Define the events $\mathcal{E}_{i,j}$, and $\tilde{\mathcal{E}}_{i,j}$, accordingly.

Assumption C.4. *For each* $b, b', e \in \{0, 1\}$*, the moment*

$$\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4, Z_1)^4 \mid \mathcal{A}_{4,3}, Z_1, Z_2, Z_3] \mid B_{3,1} = b, B_{3,2} = b', Z_1, Z_2] \mid E_{1,2} = e] \tag{C.14}$$

is bounded by a constant.

Assumption C.4 again holds in the setting where the quantities (C.5) are bounded by constants almost surely.

C.2 Proof of Theorem 5.1

We continue to use the notation introduced in the proof of Theorem 4.1. For the sake of simplicity, we restrict attention to the case that data W_i, $D_{i,j}$, X_i, and $H_{i,j}$ have been unconditionally centered. The general case, that accounts for the inclusion of intercepts in the estimators $\hat{\pi}_n$ and $\hat{\gamma}_n$, follows from the same argument, but the expansions applied below are more tedious to verify. Throughout, we let $\overline{M} = M - \mathbb{E}[M]$ denote the unconditionally centered version of the random variable M.

Our objective is to give an asymptotic approximation to the estimator

$$\hat{\theta}_n^\star = \arg\min_\theta \min_\alpha \left\{\sum_{i=1}^n (Y_i - \alpha - \theta \cdot \hat{\Delta}_i^\star)^2\right\} \ . \tag{C.15}$$

By the Frisch-Waugh-Lovell and Projection Theorems, we can write

$$\hat{\theta}_n^{\star} = \frac{\sum_{i=1}^n Y_i \tilde{\Delta}_i^{\star}}{\sum_{i=1}^n (\tilde{\Delta}_i^{\star})^2} \; , \tag{C.16}$$

where

$$\tilde{\Delta}_i^{\star} = \sum_{j \neq i} (\hat{D}_{i,j}^{\star} - \frac{1}{n} \sum_{k \neq j} \hat{D}_{k,j}^{\star}) \hat{W}_j^{\star} - \frac{1}{n} \sum_{j \neq i} \hat{D}_{j,i}^{\star} \hat{W}_i^{\star} \; . \tag{C.17}$$

We successively evaluate the denominator and numerator of the representation (C.16).

First, we evaluate the denominator. With a mild abuse of notation, we let

$$\pi_n = \arg\min_{\pi} \mathbb{E}\left[(\overline{W}_i - \pi^{\top} \overline{X}_i)^2\right] \quad \text{and} \quad \gamma_n = \arg\min_{\gamma} \mathbb{E}\left[(\overline{D}_{i,l} - \gamma^{\top} \overline{H}_{i,l})^2\right] \tag{C.18}$$

denote the population best linear predictors associated with the two nuisance parameters and

$$\hat{\pi}_{0,n} = \pi_n - \hat{\pi}_n \quad \text{and} \quad \hat{\gamma}_{0,n} = \gamma_n - \hat{\gamma}_n \tag{C.19}$$

denote their respective estimation errors. Consider the decomposition

$$\begin{aligned} \tilde{\Delta}_i^{\star} &= \sum_{j \neq i} (D_{i,j}^{\star} - \frac{1}{n} \sum_{k \neq j} D_{k,j}^{\star}) W_j^{\star} - \frac{1}{n} \sum_{j \neq i} D_{j,i}^{\star} W_i^{\star} \\ &+ \hat{\gamma}_{0,n}^{\top} \left(\sum_{j \neq i} (\overline{H}_{i,j} - \frac{1}{n} \sum_{k \neq j} \overline{H}_{k,j}) W_j^{\star} - \frac{1}{n} \sum_{j \neq i} \overline{H}_{j,i} W_i^{\star} \right) \\ &+ \hat{\gamma}_{0,n}^{\top} \left(\sum_{j \neq i} (\overline{H}_{i,j} - \frac{1}{n} \sum_{k \neq j} \overline{H}_{k,j}) \overline{X}_j^{\top} - \frac{1}{n} \sum_{j \neq i} \overline{H}_{j,i} \tilde{X}_i^{\top} \right) \hat{\pi}_{0,n} \; . \end{aligned} \tag{C.20}$$

The leading order behavior of the denominator of (C.16) is determined by the first term in (C.20). To evaluate this term, consider the further decomposition

$$\begin{aligned} &\sum_{j \neq i} (D_{i,j}^{\star} - \frac{1}{n} \sum_{k \neq j} D_{k,j}^{\star}) W_j^{\star} - \frac{1}{n} \sum_{j \neq i} D_{j,i}^{\star} W_i^{\star} \\ &= \sum_{j \neq i} \tilde{D}_{i,j}^{\star} W_j^{\star} - \frac{1}{n-1} \sum_{j \neq i} \sum_{k \neq j} \tilde{D}_{k,j}^{\star} W_j^{\star} + \frac{1}{n(n-1)} \sum_{j \neq i} \sum_{k \neq j} D_{k,j}^{\star} W_j^{\star} - \frac{1}{n} \sum_{j \neq i} D_{j,i}^{\star} W_i^{\star} \; , \end{aligned} \tag{C.21}$$

where we have used the fact that Assumption 4.2 implies that

$$\tilde{D}_{i,j}^{\star} = D_{i,j} - \mathbb{E}[D_{i,j} \mid H_{i,j}, S_j] = D_{i,j}^{\star} - \mathbb{E}[D_{i,j}^{\star} \mid S_j] \; . \tag{C.22}$$

The following Lemma quantifies the contributions of the various terms in (C.21).

Lemma C.1. *If the conditions of Theorem 5.1 hold, then:*

(i) *It holds that*

$$\frac{1}{n^2} \sum_{i=1}^n \left(\sum_{j \neq i} \tilde{D}_{i,j}^{\star} W_j^{\star} \right)^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} \sum_{k \neq i,j} \tilde{D}_{i,j}^{\star} \tilde{D}_{i,k}^{\star} W_j^{\star} W_k^{\star}$$

$$+ \mathbb{E}[\mathrm{Var}(D_{i,l} \mid H_{i,j}, S_l)\,\mathrm{Var}(W_l \mid \overline{S}_l)] + O_p(n^{-1/2}\rho_n)$$
$$= \mathbb{E}[\mathrm{Var}(D_{i,l} \mid H_{i,j}, S_l)\,\mathrm{Var}(W_l \mid \overline{S}_l)] + O_p(\rho_n^{3/2}) \tag{C.23}$$

and

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}\sum_{k\neq j}\tilde{D}^\star_{k,j}W^\star_j\right)^2 = O_p(n^{-1}\rho_n)\ ,$$

respectively.

***(ii)** It holds that*

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{n(n-1)}\sum_{j\neq i}\sum_{k\neq j}D^\star_{k,j}W^\star_j\right)^2 = O(n^{-2})\ , \quad \text{and} \tag{C.24}$$
$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j\neq i}D^\star_{j,i}W^\star_i\right)^2 = O(n^{-1}\rho_n)\ ,$$

respectively.

As a consequence, we obtain the representation

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j\neq i}(D^\star_{i,j} - \frac{1}{n}\sum_{k\neq j}D^\star_{k,j})W^\star_j - \frac{1}{n}\sum_{j\neq i}D^\star_{j,i}W^\star_i\right)^2 \tag{C.25}$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}\tilde{D}^\star_{i,j}\tilde{D}^\star_{i,k}W^\star_jW^\star_k + \mathbb{E}[\mathrm{Var}(D_{i,l} \mid H_{i,j}, S_l)\,\mathrm{Var}(W_l \mid \overline{S}_l)] + O_p(n^{-1/2}\rho_n)$$
$$= \mathbb{E}[\mathrm{Var}(D_{i,l} \mid H_{i,l}, S_l)\,\mathrm{Var}(W_l \mid \overline{S}_l)] + O(\rho_n^{3/2})\ ,$$

by plugging the bounds stated in Lemma C.1 into the decomposition (C.21) and applying Cauchy-Schwarz.

In turn, we handle the second two terms in (C.20) through the application of the following Lemma.

Lemma C.2. *Assume that the covariates X_i and $H_{i,j}$ have lengths q and p, respectively. Let $\hat{\pi}_{n,k}$ and $\hat{\gamma}_{n,k}$ denote the kth components of the nuisance parameter estimators $\hat{\pi}_n$ and $\hat{\gamma}_n$. Define $\pi_{n,k}$ and $\gamma_{n,k}$ accordingly. If the conditions of Theorem 5.1 hold, then:*

***(i)** It holds that*

$$\hat{\pi}_{n,k} - \pi_{n,k} = O_p(n^{-1/2}) \quad \text{and} \quad \hat{\gamma}_{n,k} - \gamma_{n,k} = O_p(n^{-1/2}\kappa^{-1}_{n,k}\rho_n)\ , \tag{C.26}$$

respectively.

***(ii)** Define the quantities*

$$J_i = \sum_{j\neq i}(\overline{H}_{i,j} - \frac{1}{n}\sum_{k\neq j}\overline{H}_{k,j})W^\star_j - \frac{1}{n}\sum_{j\neq i}\overline{H}_{j,i}W^\star_i \quad \text{and} \tag{C.27}$$
$$M_i = \sum_{j\neq i}(\overline{H}_{i,j} - \frac{1}{n}\sum_{k\neq j}\overline{H}_{k,j})\overline{X}_j^\top - \frac{1}{n}\sum_{j\neq i}\overline{H}_{j,i}\overline{X}_i^\top\ .$$

Let $J_{i,r}$ denote the rth component of J_i and $M_{i,r} = (M_{i,r,s})_{s=1}^q$ denote the rth column of M_i. It holds that

$$\frac{1}{n^2}\sum_{i=1}^n J_{i,s}J_{i,r} = O_p(\kappa_{n,s}^{1/2}\kappa_{n,r}^{1/2}) \quad \text{and} \quad \frac{1}{n^2}\sum_{i=1}^n M_{i,s,r}M_{i,s',r'} = O_p(n\kappa_{n,s}^{1/2}\kappa_{n,s'}^{1/2}) \tag{C.28}$$

respectively.

In particular, observe that Lemma C.2, a union bound, and Assumption 5.3 imply that

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^n \left(\hat{\gamma}_{0,n}^\top \left(\sum_{j\neq i}(\overline{H}_{i,j} - \frac{1}{n}\sum_{k\neq j}\overline{H}_{k,j})W_j^\star - \frac{1}{n}\sum_{j\neq i}\overline{H}_{j,i}W_i^\star\right)\right)^2 \\
&= \sum_{s=1}^p\sum_{r=1}^p (\hat{\gamma}_{n,s}-\gamma_{n,s})\left(\frac{1}{n^2}\sum_{i=1}^n J_{i,s}J_{i,r}\right)(\hat{\gamma}_{n,r}-\gamma_{n,r}) \\
&= \sum_{s=1}^p\sum_{r=1}^p O_p(n^{-1}\kappa_{n,s}^{-1/2}\kappa_{n,r}^{-1/2}\rho_n^2) = O_p(n^{-1}\rho_n)\ .
\end{aligned} \tag{C.29}$$

Likewise, by an analogous argument, we can evaluate

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^n \left(\hat{\gamma}_{0,n}^\top \left(\sum_{j\neq i}(\overline{H}_{i,j} - \frac{1}{n}\sum_{k\neq j}H_{k,j})\overline{X}_j^\top - \frac{1}{n}\sum_{j\neq i}\overline{H}_{j,i}\overline{X}_i^\top\right)\hat{\pi}_{0,n}\right)^2 \\
&= \sum_{s=1}^p\sum_{s'=1}^p\sum_{r=1}^q\sum_{r'=1}^q (\hat{\gamma}_{n,s}-\gamma_{n,s})(\hat{\gamma}_{n,s'}-\gamma_{n,s'}) \\
&\qquad\qquad (\hat{\pi}_{n,r}-\pi_{n,r})(\hat{\pi}_{n,r'}-\pi_{n,r'})\left(\frac{1}{n^2}\sum_{i=1}^n M_{i,s,r}M_{i,s',r'}\right) \\
&= \sum_{s=1}^p\sum_{s'=1}^p\sum_{r=1}^q\sum_{r'=1}^q O_p(n^{-1}(\kappa_{n,s}\kappa_{n,s'})^{-1/2}\rho_n^2) = O_p(n^{-1}\rho_n)\ .
\end{aligned} \tag{C.30}$$

Thus, we can conclude that

$$\begin{aligned}
\frac{1}{n^2}\sum_{i=1}^n(\tilde{\Delta}_i^\star)^2 &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\neq i,j}\tilde{D}_{i,j}^\star\tilde{D}_{i,k}^\star W_jW_k + \mathbb{E}[\operatorname{Var}(D_{i,l}\mid H_{i,j},S_l)\operatorname{Var}(W_l\mid \overline{S}_l)] + O_p(n^{-1/2}\rho_n) \\
&= \mathbb{E}[\operatorname{Var}(D_{i,l}\mid H_{i,l},S_l)\operatorname{Var}(W_l\mid \overline{S}_l)] + O_p(\rho_n^{3/2})\ ,
\end{aligned} \tag{C.31}$$

by squaring the decomposition (C.20) within an expectation, applying Cauchy-Schwarz, and plugging in the bounds (C.25), (C.29), and (C.30).

Next, we evaluate the numerator of (C.16). Here, we consider the decomposition

$$\begin{aligned}
\frac{1}{n^2}\sum_{i=1}^n Y_i\tilde{\Delta}_i^\star &= \frac{1}{n^2}\sum_{i=1}^n Y_i\left(\sum_{j\neq i}(D_{i,j}^\star - \frac{1}{n}\sum_{k\neq j}D_{k,j}^\star)W_j^\star - \frac{1}{n}\sum_{j\neq i}D_{j,i}^\star W_i^\star\right) \\
&\quad + \hat{\gamma}_{0,n}^\top\left(\frac{1}{n^2}\sum_{i=1}^n Y_i\left(\sum_{j\neq i}(\overline{H}_{i,j} - \frac{1}{n}\sum_{k\neq j}\overline{H}_{k,j})W_j^\star - \frac{1}{n}\sum_{j\neq i}\overline{H}_{j,i}W_i^\star\right)\right)
\end{aligned} \tag{C.32}$$

$$+ \hat{\gamma}_{0,n}^\top \left(\frac{1}{n^2} \sum_{i=1}^n Y_i \left(\sum_{j \neq i} (\overline{H}_{i,j} - \frac{1}{n} \sum_{k \neq j} \overline{H}_{k,j}) \overline{X}_j^\top - \frac{1}{n} \sum_{j \neq i} \overline{H}_{j,i} \overline{X}_i^\top \right) \right) \hat{\pi}_{0,n} \, .$$

The leading order behavior of this quantity is determined by the first term in (C.32). To evaluate this term, we again consider the further decomposition

$$\begin{aligned} &\frac{1}{n^2} \sum_{i=1}^n Y_i \left(\sum_{j \neq i} (D^\star_{i,j} - \frac{1}{n} \sum_{k \neq j} D^\star_{k,j}) W^\star_j - \frac{1}{n} \sum_{j \neq i} D^\star_{j,i} W^\star_i \right) \\ &= \frac{n-2}{n-1} \frac{1}{n^2} \sum_{i=1}^n Y_i \check{\Delta}^\star_i - \frac{1}{n-1} \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} Y_i \check{\Delta}^\star_{i,j} \\ &+ \frac{1}{n^2} \sum_{i=1}^n Y_i (Q_i^{(1)} - Q_i^{(2)} + Q_i^{(3)}) \, , \end{aligned} \tag{C.33}$$

where

$$\check{\Delta}^\star_i = \sum_{j \neq i} \tilde{D}^\star_{i,j} W^\star_j \, , \qquad \check{\Delta}^\star_{i,j} = \sum_{k \notin \{i,j\}} \tilde{D}^\star_{k,j} W^\star_j \, , \tag{C.34}$$

$$Q_i^{(1)} = \frac{1}{n(n-1)} \sum_{j \neq i} \sum_{k \notin \{i,j\}} D^\star_{k,j} W^\star_j \, , \qquad Q_i^{(2)} = \frac{1}{n-1} \sum_{j \neq i} D^\star_{j,i} W^\star_i \, , \quad \text{and}$$

$$Q_i^{(3)} = \frac{1}{n(n-1)} \sum_{j \neq i} D^\star_{i,j} W^\star_j \, ,$$

respectively. The following Lemma quantifies the contributions of the various terms in (C.33).

Lemma C.3. *Define the conditional expectation*

$$\tilde{F}(Z_i, Z_k) = \mathbb{E}[Y_i \mid Z_k, Z_i] - \mathbb{E}[Y_i \mid Z_i] \, . \tag{C.35}$$

If the conditions of Theorem 4.1 hold, then:

***(i)** It holds that*

$$\frac{1}{n^2} \sum_{i=1}^n Y_i \check{\Delta}^\star_i = \mathbb{E}[Y_i \tilde{D}^\star_{i,j} W_j] + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} \sum_{k \notin \{i,j\}} \tilde{F}(Z_i, Z_k) \tilde{D}^\star_{i,j} W^\star_j + O_p(n^{-1/2} \rho_n) \quad \textit{and} \tag{C.36}$$

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} \sum_{k \notin \{i,j\}} \tilde{F}(Z_i, Z_k) \tilde{D}^\star_{i,j} W^\star_j = O_p(\rho_n^{3/2}) \, , \tag{C.37}$$

respectively.

***(ii)** It holds that*

$$\frac{1}{n-1} \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} Y_i \check{\Delta}^\star_{i,j} = \mathbb{E}[Y_i \tilde{D}^\star_{l,j} W^\star_j] + O_p(n^{-1/2} \rho_n) \quad \textit{and} \tag{C.38}$$

$$\frac{1}{n^2} \sum_{i=1}^n Y_i (Q_i^{(1)} - Q_i^{(2)} + Q_i^{(3)}) = \frac{1}{n} \mathbb{E}[Y_i D^\star_{q,l} W_l] - \frac{1}{n} \mathbb{E}[Y_i D^\star_{l,i} W_i] \tag{C.39}$$

$$+ \frac{1}{n^2}\mathbb{E}[Y_i D^\star_{i,l} W_l]) + O_p(n^{-1}\rho_n^{1/2}) \ ,$$

respectively.

Consequently, Lemma C.3, as well as the bounds (A.48), (A.49), and (A.53), imply that

$$\frac{1}{n^2}\sum_{i=1}^n Y_i \left(\sum_{j\neq i}(D^\star_{i,j} - \frac{1}{n}\sum_{k\neq j} D^\star_{k,j})W^\star_j - \frac{1}{n}\sum_{j\neq i} D^\star_{j,i} W^\star_i\right) = \mathbb{E}[Y_i \tilde{D}^\star_{i,l} W^\star_l] + O_p(\rho_n^{3/2}) \ . \tag{C.40}$$

In turn, we handle the second two terms in (C.32) through the application of the following Lemma.

Lemma C.4. *If the conditions of Theorem 4.1 hold, then*

$$\frac{1}{n^2}\sum_{i=1}^n Y_i J_{i,s} = O_p(\kappa_{n,s}) \quad \textit{and} \quad \frac{1}{n^2}\sum_{i=1}^n Y_i M_{i,s,r} = O_p(n^{1/2}\kappa_{n,s}) \ , \tag{C.41}$$

respectively.

In particular, Lemma C.2 and Lemma C.4 imply that

$$\hat{\gamma}_{0,n}^\top \left(\frac{1}{n^2}\sum_{i=1}^n Y_i \left(\sum_{j\neq i}(\overline{H}_{i,j} - \frac{1}{n}\sum_{k\neq j}\overline{H}_{k,j})W^\star_j - \frac{1}{n}\sum_{j\neq i}\overline{H}_{j,i} W^\star_i\right)\right) = O_p(n^{-1/2}\rho_n) \quad \text{and} \tag{C.42}$$

$$\hat{\gamma}_{0,n}^\top \left(\frac{1}{n^2}\sum_{i=1}^n Y_i \left(\sum_{j\neq i}(\overline{H}_{i,j} - \frac{1}{n}\sum_{k\neq j}\overline{H}_{k,j})\overline{X}_j^\top - \frac{1}{n}\sum_{j\neq i}\overline{H}_{j,i}\overline{X}_i^\top\right)\right)\hat{\pi}_{0,n} = O_p(n^{-1/2}\rho_n) \ ,$$

respectively. Thus, we obtain

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n Y_i \tilde{\Delta}^\star_i &= \mathbb{E}[Y_i \tilde{D}^\star_{i,l} W^\star_l] + \frac{1}{n^2}\sum_{i=1}^n \sum_{j\neq i}\sum_{k\notin\{i,j\}} \tilde{F}(Z_i, Z_k)\tilde{D}^\star_{i,j} W^\star_j + O_p(n^{-1/2}\rho_n)\\ &= \mathbb{E}[Y_i \tilde{D}^\star_{i,l} W^\star_l] + O_p(\rho_n^{3/2}) \ .\end{aligned} \tag{C.43}$$

by plugging the bounds (C.40) and (C.42) into the decomposition (C.32).

To conclude the proof, observe that Assumption A.4 and Assumption 5.2 imply that

$$\begin{aligned}&\mathbb{E}[\operatorname{Var}(D_{i,l} \mid H_{i,l}, S_l)\operatorname{Var}(W_l \mid \overline{S}_l)]\\ &\gtrsim \mathbb{E}[\mathbb{E}[D_{i,l} \mid H_{i,l}, S_l]\operatorname{Var}(W_l \mid \overline{S}_l)]\\ &= \mathbb{E}[\mathbb{E}[D_{i,l} \mid Z_l]\operatorname{Var}(W_l \mid \overline{S}_l)] \gtrsim \rho_n \ .\end{aligned} \tag{C.44}$$

Thus, the representations (A.42), (C.31), (C.43), and a Taylor expansion, imply that

$$\begin{aligned}\hat{\theta}^\star_n &= \frac{\sum_{i=1}^n Y_i \tilde{\Delta}^\star_i}{\sum_{i=1}^n (\tilde{\Delta}^\star_i)^2}\\ &= \overline{\theta}^\star_n + O_p(\rho_n^{3/2}\mathbb{E}[\operatorname{Var}(D_{i,l} \mid G_{i,l}, S_l)\operatorname{Var}(W_i \mid X_l, S_l)]^{-1})\\ &= \int_{\mathcal{D}}\int_{\mathcal{W}} \mathbb{E}[\lambda(\delta, w \mid G_{i,j}, X_j, \overline{S}_j)\partial^2_{\delta,w}\,\mu(\delta, w \mid G_{i,j}, X_j, \overline{S}_j)]\,\mathrm{d}w\,\mathrm{d}\delta + O_p(\rho_n^{1/2}) \ ,\end{aligned} \tag{C.45}$$

where the final equality follows by the bound (C.44) and Theorem 4.1, completing the proof. ■

C.3 Proof of Theorem 5.2

To ease exposition, we write $\mathcal{P}_n = \mathcal{P}_n(\rho_n)$. Moreover, for the sake of simplicity, we restrict attention to the case that ρ_n^{-1} is an integer that divides n. The more general result can be obtained by the same argument, at the cost of more involved notation. The desired bound follows from an application of Le Cam's two-point method, stated as follows.

Lemma C.5 (Equation 15.4, Wainwright, 2019)**.** *Let X be a random variable, valued on the domain $\mathcal{X}$ and drawn according to a distribution P in the class $\mathcal{P}$. Let $\theta(P)$ denote a real-valued parameter on $\mathcal{P}$. Let the distributions P_0 and P_1 in $\mathcal{P}$ satisfy $|\theta(P_0) - \theta(P_1)| \geq 2\delta$ for some positive constant δ. For any measurable function $\hat{\theta} : \mathcal{X} \mapsto \mathbb{R}$, it holds that*

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[(\hat{\theta}(X) - \theta(P))^2 \right] \geq \frac{\delta^2}{2} (1 - \|P_0 - P_1\|_{\mathsf{TV}}) ,$$

where $\|\cdot\|_{\mathsf{TV}}$ denotes the total variation norm on probability measures.

To apply Lemma C.5, we require some further notation. Observe that each distribution P^n in the class $\mathcal{P}_n$ induces a distribution $\hat{P}^n$ for the observable data $(Y_i, W_i, D_i)_{i=1}^n$. In turn, define the collection $J_n = (W_i, D_i)_{i=1}^n$ and let $\hat{P}^n_{|J_n}$ denote the distribution of the outcomes $(Y_i)_{i=1}^n$, conditioned on J_n, under the base distribution $\hat{P}^n$.

With this in place, we begin by specifying two distributions, P_0^n and P_1^n, in $\mathcal{P}_n$. In both cases, the treatments W_i are i.i.d. Bernoulli with success rate $1/2$. Likewise, in both cases, the coordinates S_i are mutually independent, independent of the treatments, and uniformly distributed on the set $\{1, \dots, \rho_n^{-1}\}$. Again, in both cases, the distances $D_{i,j}$ and latent proximity indicators are given by

$$A_{i,j} = D_{i,j} = \mathbb{I}\{S_i = S_j\} . \tag{C.46}$$

Now, again in both cases, the collection $(U_i)_{i=1}^n$ are real-valued, i.i.d., independent of the treatments W_i and coordinates ξ_i, and have a standard Gaussian distribution. Consequently, it holds that

$$P\{A_{i,j} = 1 \mid S_i\} = \rho_n . \tag{C.47}$$

It remains to specify the distribution of the outcomes. Under P_0^n, the outcomes Y_i are generated by

$$Y_i = U_i + \sum_{j \neq i} D_{i,j} U_j . \tag{C.48}$$

In turn, under P_1^n, the outcomes are generated by

$$Y_i = \psi_n \left(\left(W_i - \frac{1}{2} \right) + \sum_{j \neq i} D_{i,j} \left(W_j - \frac{1}{2} \right) \right) + \left(U_i + \sum_{j \neq i} D_{i,j} U_j \right) , \tag{C.49}$$

for a sequence ψ_n that will be specified at a later point in the proof. The specification of P_0^n and P_1^n is now complete. It can be verified that both distributions are elements of $\mathcal{P}_n$.

Observe that, in both cases, as W_j and $D_{i,j}$ are binary, and their variances, conditioned on S_j, are constant, the weight function $\lambda(\delta, w \mid \overline{S}_j)$ is equal to zero almost surely if δ or w is equal to zero, and is equal to one otherwise. Thus, the parameter $\hat{\theta}^\star(\hat{P}^n)$ can be expressed as the double difference

$$\hat{\theta}^\star(\hat{P}^n) = \mathbb{E}_{\hat{P}^n}[(\mu(1,1 \mid \overline{S}_j) - \mu(1,0 \mid \overline{S}_j)) - (\mu(0,1 \mid \overline{S}_j) - \mu(0,0 \mid \overline{S}_j))] \,, \tag{C.50}$$

where we recall that $\mu(\delta, w \mid \overline{S}_j) = \mathbb{E}[Y_i \mid D_{i,j} = \delta, W_j = w, \overline{S}_j]$ for each $\delta, w \in \{0,1\}$. Hence, we can evaluate

$$\hat{\theta}^\star(\hat{P}_0^n) = 0 \quad \text{and} \quad \hat{\theta}^\star(\hat{P}_1^n) = \psi_n \,, \tag{C.51}$$

respectively

Thus, it remains to bound the total variation distance between P_0^n and P_1^n. We require some further notation. The random variables

$$B_k = \sum_{i=1}^{n} \mathbb{I}\{S_i = k\} \quad \text{and} \quad V_k = \sum_{i=1}^{n} \mathbb{I}\{S_i = k\} \left(W_j - \frac{1}{2}\right) \tag{C.52}$$

measure the number of units whose coordinate is equal to k and the aggregate, centered treatments associated with those units, respectively. In turn, we let $\mathsf{N}^b(m,v)$ denote the distribution of the b-vector $(X,\dots,X)$, where X has the distribution $\mathsf{N}(m,v)$. With this in place, observe that, conditional on J_n, both P_0^n and P_1^n can be expressed as product distributions though

$$\hat{P}^n_{0|J_n} = \prod_{k=1}^{\rho_n^{-1}} \mathsf{N}^{B_k}(0, B_k) \quad \text{and} \quad \hat{P}^n_{1|J_n} = \prod_{k=1}^{\rho_n^{-1}} \mathsf{N}^{B_k}(\psi_n V_k, B_k) \,, \tag{C.53}$$

respectively. Consequently, we obtain the bound

$$\begin{aligned} \|\hat{P}_0^n - \hat{P}_1^n\|_{\mathsf{TV}}^2 &\leq \frac{1}{2}\mathsf{D}_{\mathsf{KL}}\left(\hat{P}_0^n, \hat{P}_1^n\right) \\ &\leq \frac{1}{2}\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\hat{P}^n_{0|J_n}, \hat{P}^n_{1|J_n}\right)\right] \\ &\leq \frac{1}{2}\sum_{k=1}^{\rho_n^{-1}} \mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\mathsf{N}^{B_k}(0,B_k), \mathsf{N}^{B_k}(\psi_n V_k, B_k)\right) \mid B_k \neq 0\right] P\{B_k \neq 0\} \\ &\leq \frac{1}{2}\sum_{k=1}^{\rho_n^{-1}} \mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\mathsf{N}(0,B_k), \mathsf{N}(\psi_n V_k, B_k)\right) \mid B_k \neq 0\right] \,, \end{aligned} \tag{C.54}$$

where $\mathsf{D}_{\mathsf{KL}}(\cdot,\cdot)$ denotes the KL-divergence between probability measures, the first inequality follows from Pinsker's inequality (e.g., Lemma 15.2 of Wainwright, 2019), the second inequality follows from the joint-convexity of the KL-divergence (e.g., Proposition 2.2.11 of Duchi, 2024), the third inequality follows from the tensorization of the KL-divergence over product distributions (e.g., Equation 15.11 of Wainwright, 2019), and the final inequality follows from the fact that the KL divergence is invariant under invertible transformations. Now, observe that, we can evaluate

$$\begin{aligned} &\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\mathsf{N}(0,B_k), \mathsf{N}(\psi_n V_k, B_k)\right) \mid B_k \neq 0\right] \\ &= \psi_n^2 \mathbb{E}[V_k^2 B_k^{-1} \mid B_k \neq 0] \end{aligned}$$

$$= \psi_n^2 \mathbb{E}\left[\mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{I}\{S_i = k\}\left(W_j - \frac{1}{2}\right)\right)^2 \mid S_1, ..., S_n\right] B_k^{-1} \mid B_k \neq 0\right]$$

$$= \frac{\psi_n^2}{4}\mathbb{E}\left[B_k B_k^{-1} \mid B_k \neq 0\right] = \frac{\psi_n^2}{4} \tag{C.55}$$

where the first equality follows from the standard characterization of the KL-divergence between two Gaussian distributions with the same variance, the second equality follows by the assumed independence between the treatments and the coordinates. Thus, we find that

$$\|\hat{P}_0^n - \hat{P}_1^n\|_{\mathsf{TV}}^2 \leq \frac{1}{8}\rho_n^{-1}\psi_n^2 \tag{C.56}$$

by plugging the bound (C.55) into (C.54). Hence, by making the choice $\psi_n = \rho_n^{1/2}$, we can conclude that

$$\mathfrak{M}_n(\rho_n) \gtrsim \rho_n^{1/2} \tag{C.57}$$

by applying Lemma C.5 with the distributions P_0^n and P_1^n, as required. ∎

C.4 Proof of Theorem 6.1

We apply the following Lemma at several points in the argument.

Lemma C.6. *Define the sequences*

$$\sigma_{n,1}^2 = \mathbb{E}\left[\left(\mathbb{E}[\tilde{F}(Z_1, Z_3)\tilde{D}_{1,2}^\star \mid Z_2, Z_3]W_2^\star + \mathbb{E}[\tilde{F}(Z_1, Z_2)\tilde{D}_{1,3}^\star \mid Z_2, Z_3]W_3^\star\right)^2\right], \tag{C.58}$$

$$\sigma_{n,2}^2 = 4\mathbb{E}\left[\left(\mathbb{E}[\tilde{D}_{1,3}\tilde{D}_{1,2}^\star \mid Z_2, Z_3]W_3^\star W_2^\star\right)^2\right], \quad \textit{and}$$

$$\varsigma_n = 4\mathbb{E}\left[\mathbb{E}[\tilde{F}(Z_1, Z_3)\tilde{D}_{1,2}^\star \mid Z_2, Z_3]\mathbb{E}[\tilde{D}_{1,2}^\star W_2^\star \tilde{D}_{1,3}^\star \mid Z_2, Z_3]W_2^\star W_3^\star\right], \tag{C.59}$$

and the matrix

$$\Sigma_n = \begin{pmatrix} \sigma_{n,1}^2 & \varsigma_n \\ \varsigma_n & \sigma_{n,2}^2 \end{pmatrix}. \tag{C.60}$$

Under the conditions of Theorem 6.1, it holds that

$$c\rho_n^3 \mathbf{I}_2 \preceq \Sigma_n \preceq C\rho_n^3 \mathbf{I}_2, \tag{C.61}$$

where $\preceq$ denotes the Loewner order on matrices.

Proof. The sequence of inequalities (C.61) is equivalent to the assertion that

$$\xi^\top \Sigma_n \xi \asymp \rho_n^3 \tag{C.62}$$

uniformly over the collection of real-valued vectors $\xi = (\xi_1, \xi_2)^\top$ that satisfy $\xi_1^2 + \xi_2^2 = 1$. The upper bound $\xi^\top \hat{\Sigma}_n \xi \lesssim \rho_n^3$ follows immediately from the the bounds (D.134) and (D.185) stated in the Proof of Lemma D.5 and the Proof of Part (iii) of Lemma D.7, respectively. The lower bound $\xi^\top \hat{\Sigma}_n \xi \gtrsim \rho_n^3$ follows immediately from Assumptions 5.2 and C.3, respectively. ∎

The result is an application of the following central limit theorem, obtained by applying a general result concerning the Gaussian approximation of degenerate U-statistics, due to Liu et al. (2025).

Lemma C.7. *Under the conditions of Theorem 6.2, it holds that*

$$\frac{\sqrt{2}}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\Sigma_n^{-1/2}\begin{pmatrix}\tilde{F}(Z_i,Z_k)\tilde{D}^{\star}_{i,j}W^{\star}_j\\ \tilde{D}^{\star}_{i,k}\tilde{D}^{\star}_{i,j}W^{\star}_jW^{\star}_k\end{pmatrix}\xrightarrow{d}\mathsf{N}(\mathbf{0}_2,\mathbf{I}_2)\,, \tag{C.63}$$

where Σ_n is defined in the statement of Lemma C.6.

Recall the expansions

$$\frac{1}{n^2}\sum_{i=1}^{n}Y_i\tilde{\Delta}^{\star}_i=\mathbb{E}[Y_i\tilde{D}_{i,l}W^{\star}_l]+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{F}(Z_i,Z_k)\tilde{D}^{\star}_{i,j}W^{\star}_j+O_p(n^{-1/2}\rho_n)\,, \tag{C.64}$$

$$\frac{1}{n^2}\sum_{i=1}^{n}(\tilde{\Delta}^{\star}_i)^2=\mathbb{E}[(\tilde{D}^{\star}_{i,l}W^{\star}_l)^2]+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{D}^{\star}_{i,k}W^{\star}_k\tilde{D}^{\star}_{i,j}W^{\star}_j+O_p(n^{-1/2}\rho_n)\,, \tag{C.65}$$

stated as (C.43) and (C.31) in the Proof of Theorem 5.1. Thus, we obtain

$$\frac{\sqrt{2}}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\Sigma_n^{-1/2}\begin{pmatrix}Y_i\tilde{\Delta}^{\star}_i-\mathbb{E}[\tilde{D}_{i,l}W^{\star}_l]\\ (\tilde{\Delta}^{\star}_i)^2-\mathbb{E}[(\tilde{D}^{\star}_{i,l}W^{\star}_l)^2]\end{pmatrix}\xrightarrow{d}\mathsf{N}(\mathbf{0}_2,\mathbf{I}_2)\,, \tag{C.66}$$

by Lemmas C.6 and C.7 and the continuous mapping theorem.

Now, recall the representation

$$\hat{\theta}^{\star}_n=\frac{\frac{1}{n^2}\sum_{i=1}^{n}Y_i\tilde{\Delta}^{\star}_i}{\frac{1}{n^2}\sum_{i=1}^{n}(\tilde{\Delta}^{\star}_i)^2}\,, \tag{C.67}$$

obtained in the proof of Theorem 5.1. By an application of the Delta method, the approximation (C.66) implies that

$$\sigma_n^{-1}(\hat{\theta}^{\star}_n-\theta^{\star}_n)\xrightarrow{d}\mathsf{N}(0,1) \tag{C.68}$$

where

$$\begin{aligned}\sigma_n^2&=\frac{1}{2\mathbb{E}[(\tilde{D}^{\star}_{i,l}W^{\star}_l)^2]^2}\left(\sigma^2_{n,1}-2\varsigma_n\theta^{\star}_n+\sigma^2_{n,2}(\theta^{\star}_n)^2\right)\\&=\frac{1}{2\mathbb{E}[(\tilde{D}^{\star}_{i,j}W^{\star}_j)^2]^2}\operatorname{Var}(\mathbb{E}[\tilde{\varepsilon}_{k,i}\tilde{D}^{\star}_{k,j}\mid Z_i,Z_j]W^{\star}_j+\mathbb{E}[\tilde{\varepsilon}_{k,j}\tilde{D}^{\star}_{k,i}\mid Z_j,Z_i]W^{\star}_i)\,.\end{aligned} \tag{C.69}$$

We conclude the proof by noting that the relation $\sigma_n^{-1}\asymp\rho_n^{1/2}$ follows immediately from Lemma C.6 and the bound

$$\mathbb{E}[(\tilde{D}^{\star}_{i,j}W^{\star}_j)^2]\gtrsim\rho_n\,, \tag{C.70}$$

implied by Assumptions 5.2 and A.4, as required. ∎

C.5 Proof of Theorem 6.2

The result follows from an application of the following lemma, due to Hoeffding (1952).

Lemma C.8 (Theorem 17.2.3, Lehmann and Romano, 2022)**.** *Suppose that $\mathcal{X}^n$ has distribution P_n in $\mathcal{X}_n$ and $\mathbf{G}_n$ is a finite group of transformations from $\mathcal{X}_n$ to $\mathcal{X}_n$. Let G_n be a random variable that is uniform on $\mathbf{G}_n$. Also, let G_n' have the same distribution as G_n, with X^n, G_n, and G_n' mutually independent. Define the randomization distribution*

$$\hat{R}_n(t) = M_n^{-1} \sum_{g \in \mathbf{G}_n} \mathbb{I}\{T_n(gX^n) \le t\} , \tag{C.71}$$

where M_n is the cardinality of $\mathbf{G}_n$. Suppose, under P_n, it holds that

$$(T_n(G_n X^n), T_n(G_n' X_n)) \xrightarrow{d} (T, T') , \tag{C.72}$$

where T and T' are independent, each with common c.d.f. $R(\cdot)$. Then, under P_n, it holds that

$$\hat{R}_n(t) \xrightarrow{p} R(t) , \tag{C.73}$$

for every t that is a continuity point of $R(\cdot)$.

To apply this strategy to our setting, we require some intermediate results. First, we use the following central limit theorem. We omit the proof of this result, as it follows from an argument that is identical to the proof of Lemma C.7.

Lemma C.9. *Fix any scalar θ. Define the random variable*

$$\tilde{\varepsilon}_\theta(Z_i, Z_j) = \tilde{F}(Z_i, Z_j) - \theta \tilde{D}_{i,j} W_j^\star \tag{C.74}$$

and the sequences

$$\varphi_{n,1}^2(\theta) = 2 \operatorname{Var}\left(\mathbb{E}[\tilde{\varepsilon}_\theta(Z_i, Z_j) \tilde{D}_{1,2}^\star \mid Z_2, Z_3] W_2^\star\right) \quad \text{and} \quad \tilde{\Sigma}_n(\theta) = \begin{pmatrix} \varphi_{n,1}^2(\theta) & 0 \\ 0 & \sigma_{n,2}^2 \end{pmatrix} \tag{C.75}$$

respectively. Let $V = (V_i)_{i=1}^n$ and $V' = (V_i')_{i=1}^n$ denote two independent collections of i.i.d. Rademacher random variables. Under the conditions of Theorem 6.2, it holds that

$$\frac{\sqrt{2}}{n^2} \sum_{i=1}^n \sum_{j \ne i} \sum_{k \notin \{i,j\}} \begin{pmatrix} \tilde{\Sigma}_n^{-1/2}(\theta) & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \tilde{\Sigma}_n^{-1/2}(\theta) \end{pmatrix} \begin{pmatrix} \tilde{\varepsilon}_\theta(Z_i, Z_k) \tilde{D}_{i,j}^\star W_j^\star V_j \\ \tilde{D}_{i,k}^\star \tilde{D}_{i,j}^\star W_j^\star W_k^\star V_j V_k \\ \tilde{\varepsilon}_\theta(Z_i, Z_k) \tilde{D}_{i,j}^\star W_j^\star V_j' \\ \tilde{D}_{i,k}^\star \tilde{D}_{i,j}^\star W_j^\star W_k^\star V_j' V_k' \end{pmatrix} \xrightarrow{d} \mathsf{N}(\mathbf{0}_4, \mathbf{I}_4) , \tag{C.76}$$

respectively.

Moreover, we make use of the following expansions, which follow from arguments identical to the proofs of Lemmas C.1 to C.4, respectively. Again, we have omitted the details for reasons of space.

Lemma C.10. *Let $V = (V_i)_{i=1}^n$ denote an independent collection of i.i.d. Rademacher random variables. Define the variable*

$$\tilde{\Delta}_i^\star(V) = \sum_{j\neq i}(\hat{D}_{i,j}^\star - \frac{1}{n}\hat{D}_{k,j}^\star)\hat{W}_j^\star V_j - \frac{1}{n}\sum_{j\neq i}\hat{D}_{j,i}^\star \hat{W}_i^\star \ . \tag{C.77}$$

Fix any scalar θ. Under the conditions of Theorem 5.1, it holds that

$$\frac{1}{n^2}\sum_{i=1}^n (Y_i - \theta\tilde{\Delta}_i^\star)\tilde{\Delta}_i^\star(V) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\notin\{i,j\}} \varepsilon_\theta(Z_i, Z_k)\tilde{D}_{i,j}^\star W_j^\star V_j + O_p(n^{-1/2}\rho_n) \quad \textit{and} \tag{C.78}$$

$$\frac{1}{n^2}\sum_{i=1}^n (\tilde{\Delta}_i^\star(V))^2 = \mathbb{E}[(\tilde{D}_{i,l}^\star W_l^\star)^2] + \frac{1}{n^2}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\notin\{i,j\}} \tilde{D}_{i,k}^\star W_k^\star V_k \tilde{D}_{i,j}^\star W_j^\star V_j + O_p(n^{-1/2}\rho_n) \ ,$$

respectively.

With these results in place, consider the statistic

$$\hat{\phi}_n^\star(V,\theta) = \frac{\frac{1}{n^2}\sum_{i=1}^n (Y_i - \theta\tilde{\Delta}_i^\star)\tilde{\Delta}_i^\star(V)}{\frac{1}{n^2}\sum_{i=1}^n(\tilde{\Delta}_i^\star(V))^2} \tag{C.79}$$

and observe that Lemmas C.9 and C.10, the continuous mapping theorem, and the Delta method imply that

$$\begin{pmatrix} \varphi_n^{-1/2}(\theta)\hat{\phi}_n^\star(V,\theta) \\ \varphi_n^{-1/2}(\theta)\hat{\phi}_n^\star(V',\theta) \end{pmatrix} \xrightarrow{d} \mathsf{N}(\mathbf{0}_2, \mathbf{I}_2) \ , \tag{C.80}$$

where

$$\varphi_n^{-1/2}(\theta) = \frac{1}{\mathbb{E}[(\tilde{D}_{i,j}^\star W_j^\star)^2]^2}\operatorname{Var}(\mathbb{E}[\tilde{\varepsilon}_\theta(Z_k, Z_i)\tilde{D}_{k,j}^\star \mid Z_i, Z_j]W_j^\star) \ . \tag{C.81}$$

Thus, Lemma C.8 implies that

$$\frac{1}{2^n}\sum_{v\in\{-1,1\}^n} \mathbb{I}\{\varphi_n^{-1/2}(\theta)\hat{\phi}_n^\star(v,\theta) \leq x\} \xrightarrow{p} \Phi(x) \ , \tag{C.82}$$

where $\Phi(x)$ denotes the standard normal cumulative distribution function. Now, observe that

$$\hat{\phi}_n^\star(V) = \hat{\phi}_n^\star(V, \hat{\theta}_n^\star) \tag{C.83}$$

by the Frisch-Waugh-Lovell Theorem, and so

$$\hat{R}_n(x) = \frac{1}{2^n}\sum_{v\in\{-1,1\}^n} \mathbb{I}\{\varphi_n^{-1/2}\hat{\phi}_n^\star(v) \leq x\} = \frac{1}{2^n}\sum_{v\in\{-1,1\}^n} \mathbb{I}\{\varphi_n^{-1/2}(\theta_n^\star)\hat{\phi}_n^\star(v,\hat{\theta}_n^\star) \leq x\} \xrightarrow{p} \Phi(x) \tag{C.84}$$

by Theorem 5.1 and Slutsky's Lemma. Observe that the relation

$$\rho_n \leq \sigma_n^2 \leq 2\varphi_n^2 \lesssim \rho_n^2 \tag{C.85}$$

follows immediately from Lemma C.6, the bound (C.70), and Cauchy-Schwarz, as required. ■

Appendix D. Proofs for Lemmas Supporting Appendix C

Throughout the proofs stated in this section, we apply results concerning the decomposition and asymptotic approximation of symmetric statistics. In particular, at several points, our characterization of the limiting behavior of various statistics is facilitated by Hoeffding's decomposition (Hoeffding, 1948; Efron and Stein, 1981). We adopt the following somewhat nonstandard formulation, considered in Lachièze-Rey and Peccati (2017), based on the use of stochastic differences.

Lemma D.1 (Theorem 2.2, Lachièze-Rey and Peccati, 2017)**.** *Let $X = (X_i)_{i=1}^n$ denote a sequence of independent random variables valued in $\mathcal{X}$. Let X' denote an independent copy of X. Construct $X^{(i)}$ by replacing X_i with X_i' in X, leaving all other entries unchanged. For each i, define the difference operator*

$$\nabla_i f(X) = f(X) - f(X^{(i)}) \tag{D.1}$$

on the set of all measurable functions $f : \mathcal{X}^n \to \mathbb{R}$. If the moment $\mathbb{E}[f(X)^2]$ is finite, then the representation

$$f(X) = \mathbb{E}[f(X)] + \sum_{m=1}^{n} \sum_{1\le i_1<\cdots<i_m<n} f^{(m)}(X_{i_1}, ..., X_{i_m}) \,, \quad \text{where} \tag{D.2}$$

$$f^{(m)}(X_{i_1}, ..., X_{i_m}) = (-1)^m \mathbb{E}[\nabla_{i_1} \cdots \nabla_{i_m} f(X', X) \mid X_{i_1}, ..., X_{i_m}] \,,$$

holds. Moreover, all 2^n terms appearing on the right-hand side of (D.2) are mean-zero and uncorrelated.

When applied to U-statistics, Lemma D.1 has the following, widely-known effect.

Lemma D.2 (Lemmas 5.1.5 A and 5.2.1 A, Serfling, 1980)**.** *Let $X = (X_i)_{i=1}^n$ denote an i.i.d. sequence of random variables valued in $\mathcal{X}$. Suppose that the function $f : \mathcal{X}^b \to \mathbb{R}$ is symmetric in its arguments and that the statistic $f(X_1, \dots, X_b)$ is mean-zero.*

(i) *It holds that*

$$F_n = \frac{1}{\binom{n}{b}} \sum_{1\le i_1\le\cdots\le i_b\le n} f(X_{i_1}, \dots, X_{i_b}) = \sum_{m=1}^{b} \binom{b}{m} F_{n,m} \,, \quad \text{where} \tag{D.3}$$

$$F_{n,m} = \frac{1}{\binom{n}{m}} \sum_{1\le i_1\le\cdots\le i_m\le n} f^{(m)}(X_{i_1}, \dots, X_{i_m}) \tag{D.4}$$

and the quantities $f^{(m)}(\cdot)$ are defined in Lemma D.1.

(ii) *The decomposition*

$$\mathrm{Var}\,(F_n) = \binom{n}{b}^{-1} \sum_{m=1}^{b} \binom{b}{m}\binom{n-b}{b-m} \mathrm{Var}(E[f(X_1, \dots, X_n \mid X_1, \dots, X_m]) \tag{D.5}$$

holds.

At some points, we will make use of the following two, more exotic results. The first is a version of a second order Efron-Stein inequality, due to Bobkov et al. (2019).

Lemma D.3 (Theorem 1.8, Bobkov et al., 2019)**.** *Let $X = (X_i)_{i=1}^n$ denote a collection of independent random variances. Let X' denote an independent copy of X. Construct $X^{(i)}$ by replacing X_i with X_i' in X,* *leaving all other entries unchanged. Construct* $X^{(i,j)}$ *analogously. If* $f(X)$ *has a finite variance and satisfies the equality*

$$\mathbb{E}[f(X) \mid X_i] = \mathbb{E}[f(X)] \tag{D.6}$$

almost surely for each i *in* $[n]$*, then the inequality*

$$\operatorname{Var}(f(X)) \leq \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}[\mathbb{E}[(f(X)-f(X^{(i)}))-(f(X^{(j)})-f(X^{(i,j)}))\mid X]^2] \tag{D.7}$$

holds

The second is a Berry-Esseen type central limit theorem for degenerate U-statistics, due to Liu et al. (2025). The result stated in Liu et al. (2025) is given for U-statistics with kernels and degeneracies of arbitrary orders. The version that follows has been adapted to suit our purposes.

Lemma D.4 (Theorem 2.2, Liu et al., 2025)**.** *Let* $X = (X_i)_{i=1}^n$ *denote an i.i.d. sequence of random variables valued in* $\mathcal{X}$*. Let* X' *denote an independent copy of* X*. Suppose that the symmetric function* $f : \mathcal{X}^3 \to \mathbb{R}$ *satisfies the equality*

$$\mathbb{E}[f(X_1, X_2, X_3) \mid X_1] = 0 \tag{D.8}$$

almost surely. Then, it holds that

$$\sup_{x\in\mathbb{R}}\left|P\left\{\frac{1}{\sigma_n}\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n}(f(X_i,X_j,X_k)-\mathbb{E}[f(X_1,X_2,X_2)])\leq x\right\}-\Phi(x)\right|$$
$$\lesssim \sqrt{\frac{1}{n}+\frac{\mathbb{E}[(f^{(3)}(X_1,X_2,X_3))^2]}{n\mathbb{E}[(f^{(2)}(X_1,X_2))^2]}}+\frac{\sqrt{\frac{1}{n}\mathbb{E}[(f^{(2)}(X_1,X_2))^4]+\mathbb{E}[(g^{(2)}(X_1,X_2))^2]}}{\mathbb{E}[(f^{(2)}(X_1,X_2))^2]}\,, \tag{D.9}$$

where

$$\sigma_n^2 = \frac{\binom{3}{2}^2}{\binom{n}{2}}\mathbb{E}[(f^{(2)}(X_1,X_2))^2]\,, \quad g^{(2)}(x_1,x_2) = \mathbb{E}[f^{(2)}(X_2,x_1)f^{(2)}(X_1,x_2)]\,, \tag{D.10}$$

the quantities $f^{(m)}(\cdot)$ *are defined in Lemma D.1, and* $\Phi(\cdot)$ *denotes the standard normal c.d.f.*

D.1 Proof of Lemma C.1

D.1.1 Part (i). We begin by verifying the first bound stated in (C.23). Consider the decomposition

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j\neq i}\tilde{D}^\star_{i,j}W^\star_j\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}(\tilde{D}^\star_{i,j})^2(W^\star_j)^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{D}^\star_{i,j}\tilde{D}^\star_{i,k}W^\star_jW^\star_k\,. \tag{D.11}$$

We bound the two terms in (D.11) through the application of the following Lemma.

Lemma D.5. *If the conditions of Theorem 5.1 hold, then*

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}(\tilde{D}^\star_{i,j})^2(W^\star_j) - \mathbb{E}[\operatorname{Var}(D_{i,l}\mid H_{i,l},S_l)\operatorname{Var}(W_l\mid \overline{S}_l)] = O_p(n^{-1/2}\rho_n) \quad \textit{and} \tag{D.12}$$

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{D}^{\star}_{i,j}\tilde{D}^{\star}_{i,k}W^{\star}_{j}W^{\star}_{k} = O_p(\rho_n^{3/2}) \tag{D.13}$$

hold, respectively.

Consequently, applying Lemma D.5 to the representation (D.11), we find that

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j\neq i}\tilde{D}^{\star}_{i,j}W^{\star}_{j}\right)^2 - \mathbb{E}[\mathrm{Var}(D_{i,l} \mid H_{i,l}, S_l)\,\mathrm{Var}(W_l \mid \overline{S}_l)] = O_p(\rho_n^{3/2})\ . \tag{D.14}$$

as required. Next, we consider second bound stated in (C.23). Here, we consider the decomposition

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}\sum_{k\neq j}\tilde{D}^{\star}_{k,j}W^{\star}_{j}\right)^2 \qquad\qquad (\mathrm{D.15})\\
&= \frac{1}{n^2(n-1)}\sum_{j=1}^{n}\sum_{k\neq j}\sum_{k'\neq j}\tilde{D}^{\star}_{k,j}\tilde{D}^{\star}_{k',j}(W^{\star}_{j})^2 + \frac{n-2}{n^2(n-1)^2}\sum_{j=1}^{n}\sum_{j'\neq j}\sum_{k\neq j}\sum_{k'\neq j'}\tilde{D}^{\star}_{k,j}\tilde{D}^{\star}_{k',j'}W^{\star}_{j}W^{\star}_{j'}\\
&= \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}(\tilde{D}^{\star}_{i,j})^2(W^{\star}_{j})^2 + \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{D}^{\star}_{i,k}\tilde{D}^{\star}_{j,k}(W^{\star}_{k})^2\\
&+ \frac{n-2}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{D}^{\star}_{j,i}\tilde{D}^{\star}_{i,j}W^{\star}_{j}W^{\star}_{i} + \frac{2(n-2)}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{D}^{\star}_{i,k}\tilde{D}^{\star}_{j,i}W^{\star}_{k}W^{\star}_{i}\\
&+ \frac{n-2}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{k\neq i}\sum_{l\notin\{i,k\}}\tilde{D}^{\star}_{i,k}\tilde{D}^{\star}_{i,l}W^{\star}_{k}W^{\star}_{l}\\
&+ \frac{n-2}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j,l\}}\tilde{D}^{\star}_{i,k}\tilde{D}^{\star}_{j,l}W^{\star}_{k}W^{\star}_{l}\ .
\end{aligned}$$

Observe that Lemma D.5 implies that

$$\begin{aligned}
\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}(\tilde{D}^{\star}_{i,j})^2(W^{\star}_{j})^2 &= n^{-1}\mathbb{E}[\mathrm{Var}(D_{i,l} \mid H_{i,l}, S_l)\,\mathrm{Var}(W_l \mid \overline{S}_l)] + O_p(n^{-3/2}\rho_n) \qquad (\mathrm{D.16})\\
&= O_p(n^{-1}\rho_n^{1/2})
\end{aligned}$$

and

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{k\neq i}\sum_{l\notin\{i,k\}}\tilde{D}^{\star}_{i,k}\tilde{D}^{\star}_{i,l}W^{\star}_{k}W^{\star}_{l} = O_p(n^{-1}\rho_n^{3/2})\ , \tag{D.17}$$

respectively. The remaining four terms are then handled through the following Lemma.

Lemma D.6. *If the conditions of Theorem 5.1 hold, then*

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{D}^{\star}_{j,i}\tilde{D}^{\star}_{i,j}W^{\star}_{j}W^{\star}_{i} = O_p(n^{-3/2}\rho_n^{1/2})\ , \tag{D.18}$$

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{D}^\star_{i,k}\tilde{D}^\star_{j,k}(W^\star_k)^2 = O_p(n^{-1}\rho_n^{1/2})\ , \tag{D.19}$$

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{D}^\star_{i,k}\tilde{D}^\star_{j,i}W^\star_k W^\star_i = O_p(n^{-3/2}\rho_n^{1/2})\ , \quad \textit{and} \tag{D.20}$$

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j,l\}}\tilde{D}^\star_{i,k}\tilde{D}^\star_{j,l}W^\star_k W^\star_l = O(n^{-1}\rho_n^{1/2})\ , \tag{D.21}$$

hold, respectively.

Thus, combining (D.16), (D.17), and Lemma D.6, we find that

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}\sum_{k\neq j}\tilde{D}^\star_{k,j}W^\star_j\right)^2 = O_p(n^{-1}\rho_n^{1/2})\ , \tag{D.22}$$

as required. ■

D.1.2 Part (ii). We begin by verifying the first bound in (C.24). By a decomposition analogous to (D.15), we have that

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{n(n-1)}\sum_{j\neq i}\sum_{k\neq j}D^\star_{k,j}W^\star_j\right)^2 \\
&= \frac{1}{n^4(n-1)}\sum_{j=1}^{n}\sum_{k\neq j}\sum_{k'\neq j}D^\star_{k,j}D^\star_{k',j}(W^\star_j)^2 \\
&+ \frac{n-2}{n^4(n-1)^2}\sum_{j=1}^{n}\sum_{j'\neq j}\sum_{k\neq j}\sum_{k'\neq j'}D^\star_{k,j}D^\star_{k',j'}W^\star_j W^\star_{j'} = O(n^{-2})\ ,
\end{aligned} \tag{D.23}$$

where the bound follows from the fact that the random variables $D_{i,j}$ and W_j are bounded by constants almost surely. Next, we verify the second bound in (C.24). Observe that, by writing

$$g(Z_1,Z_2,Z_3) = \sum_{\pi\in\Pi_3}D^\star_{\pi(2),\pi(1)}D^\star_{\pi(3),\pi(1)}(W^\star_{\pi(1)})^2\ , \tag{D.24}$$

the term

$$\begin{aligned}
\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j\neq i}D^\star_{j,i}W^\star_i\right)^2 &= \frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i}D^\star_{j,i}D^\star_{k,i}(W^\star_i)^2 \\
&= \frac{\binom{n}{3}}{n^4}\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n}g(Z_i,Z_j,Z_k)
\end{aligned} \tag{D.25}$$

can be recognized as a scaled U-statistic of order 3. Thus, Lemma D.1 implies that

$$\operatorname{Var}\left(\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n}g(Z_i,Z_j,Z_k)\right) \lesssim n^{-1}\operatorname{Var}(g(Z_i,Z_j,Z_k))\ . \tag{D.26}$$

We can evaluate

$$\operatorname{Var}(g(Z_i, Z_j, Z_k)) \lesssim \mathbb{E}[(D^\star_{2,1} D^\star_{3,1})^2 (W^\star_1)^4] \tag{D.27}$$

$$\lesssim \mathbb{E}[(D^\star_{2,1})^4] \lesssim \mathbb{E}[D^4_{2,1}] = O(\rho_n) \tag{D.28}$$

by Cauchy-Schwarz and Assumption 5.2. Hence, we obtain

$$\left| \frac{1}{n^2} \sum_{i=1}^n \left(\frac{1}{n} \sum_{j \neq i} D^\star_{j,i} W^\star_i \right)^2 \right| \leq \left| \frac{1}{n} \mathbb{E}[D^\star_{j,i} D^\star_{k,i} (W^\star_i)^2] \right| + O_p(n^{-3/2} \rho_n^{1/2}) \tag{D.29}$$

$$\lesssim n^{-1} \mathbb{E}[(D^\star_{j,i})^2] + O_p(n^{-3/2} \rho_n^{1/2}) = O(n^{-1} \rho_n) , \tag{D.30}$$

by Chebyshev's inequality, Cauchy-Schwarz, and Assumption 5.2, as required. ■

D.2 Proof of Lemma C.2

D.2.1 Part (i). We being by evaluating the estimator $\hat{\pi}_n$. Observe that we can express

$$\hat{\pi}_n = \left(\frac{1}{n} \sum_{i=1}^n \overline{X}_i \overline{X}_i^\top \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \overline{X}_i \overline{W}_i \right) \tag{D.31}$$

by the Projection Theorem. As the dimension of covariate vector X_j is fixed, and each component is bounded by a constant, Chebyshev's inequality and a union bound imply that

$$\left\| \frac{1}{n} \sum_{i=1}^n \overline{X}_i \overline{X}_i^\top - \operatorname{Var}(X_i) \right\|_\infty = O_p(n^{-1/2}) \quad \text{and} \tag{D.32}$$

$$\left\| \frac{1}{n} \sum_{i=1}^n \overline{X}_i \overline{W}_i - \operatorname{Cov}(X_i, W_i) \right\|_\infty = O_p(n^{-1/2})$$

respectively. Consequently, as the Projection Theorem implies that $\pi_n = \operatorname{Var}(X_i)^{-1} \operatorname{Cov}(X_i, W_i)$, a Taylor expansion, the fact that variance of each component of X_i is bounded below by a constant, and Assumption 5.1 give the bound

$$\begin{aligned} \|\hat{\pi}_n - \pi_n\|_\infty &= O_p(n^{-1/2} \| \operatorname{Var}(X_i)^{-1} \|_{\text{op}}) \\ &\quad + O_p(n^{-1/2} \| \operatorname{Var}(X_i)^{-1} \|_{\text{op}}^2 \| \operatorname{Cov}(X_i W_i) \|_2) = O_p(n^{-1/2}) , \end{aligned}$$

as required.

Next, we evaluate the estimator $\hat{\gamma}_n$. Again, by Assumption 5.1 and the Projection Theorem, we can write

$$\hat{\gamma}_n = \hat{\Sigma}^{-1}_{H,H} \hat{\Sigma}_{H,D} \quad \text{and} \quad \gamma_n = \Sigma^{-1}_{H,H} \Sigma_{H,D} \tag{D.33}$$

where

$$\hat{\Sigma}_{H,H} = \frac{1}{n-1} \sum_{i \neq j} \overline{H}_{i,j} \overline{H}_{i,j}^\top , \quad \Sigma_{H,H} = \mathbb{E}[\overline{H}_{i,j} \overline{H}_{i,j}^\top] , \tag{D.34}$$

and

$$\hat{\Sigma}_{H,D} = \frac{1}{n-1} \sum_{i \neq j} \overline{H}_{i,j} \overline{D}_{i,j} , \quad \Sigma_{H,D} = \mathbb{E}[\overline{H}_{i,j} \overline{D}_{i,j}] , \tag{D.35}$$

respectively. Define the perturbations

$$\Gamma_{H,H} = \hat{\Sigma}_{H,H} - \Sigma_{H,H} \quad \text{and} \quad \Gamma_{H,D} = \hat{\Sigma}_{H,D} - \Sigma_{H,D} \, . \tag{D.36}$$

Now, observe that, for any invertible matrices B and C, it holds that

$$(B+C)^{-1} = B^{-1} - B^{-1}C(B+C)^{-1} \, . \tag{D.37}$$

Consequently, we have that

$$\hat{\Sigma}_{H,H}^{-1} = \Sigma_{H,H}^{-1} - \Sigma_{H,H}^{-1}\Gamma_{H,H}\hat{\Sigma}_{H,H}^{-1} \tag{D.38}$$

and thereby

$$\begin{aligned}
\hat{\gamma}_n - \gamma_n &= \Sigma_{H,H}^{-1}\Gamma_{H,D} - \Sigma_{H,H}^{-1}\Gamma_{H,H}\hat{\Sigma}_{H,H}^{-1}\Sigma_{H,D} - \Sigma_{H,H}^{-1}\Gamma_{H,H}\hat{\Sigma}_{H,H}^{-1}\Gamma_{H,D} \\
&= \Sigma_{H,H}^{-1}\Gamma_{H,D} - \Sigma_{H,H}^{-1}\Gamma_{H,H}\gamma_n \\
&+ \Sigma_{H,H}^{-1}\Gamma_{H,H}(\hat{\Sigma}_{H,H}^{-1} - \Sigma_{H,H}^{-1})\Sigma_{H,D} - \Sigma_{H,H}^{-1}\Gamma_{H,H}\hat{\Sigma}_{H,H}^{-1}\Gamma_{H,D} \\
&= \Sigma_{H,H}^{-1}\Gamma_{H,D} - \Sigma_{H,H}^{-1}\Gamma_{H,H}\gamma_n \\
&+ \Sigma_{H,H}^{-1}\Gamma_{H,H}\Sigma_{H,H}^{-1}\Gamma_{H,H}\hat{\Sigma}_{H,H}^{-1}\Sigma_{H,D} - \Sigma_{H,H}^{-1}\Gamma_{H,H}\hat{\Sigma}_{H,H}^{-1}\Gamma_{H,D} \, ,
\end{aligned} \tag{D.39}$$

where the final equality follows from the identity

$$\hat{\Sigma}_{H,H}^{-1} - \Sigma_{H,H}^{-1} = -\Sigma_{H,H}^{-1}\Gamma_{H,H}\Sigma_{H,H}^{-1} \, . \tag{D.40}$$

The result follows by giving suitable bounds for the kth component of each of the four terms in (D.39).

In order to do this, we make repeated use of the inequalities

$$\Gamma_{H,H}^{k,l} = O_p(n^{-1/2}\min\{\kappa_{n,k}, \kappa_{n,l}\}) \quad \text{and} \quad \Gamma_{H,D}^{k} = O_p(n^{-1/2}\rho_n) \tag{D.41}$$

where $\Gamma_{H,H}^{k,l}$ and $\Gamma_{H,D}^{k}$ are the k,lth and kth components of $\Gamma_{H,H}$ and $\Gamma_{H,D}$, respectively. To verify the first inequality in (D.41), observe that Lemma D.1 implies that

$$\operatorname{Var}\left(\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\overline{H}_{i,j}^{k}\overline{H}_{i,j}^{l}\right) \tag{D.42}$$

$$\begin{aligned}
&\lesssim n^{-1}\operatorname{Var}(\mathbb{E}[\overline{H}_{i,j}^{(k)}\overline{H}_{i,j}^{(l)} \mid \overline{S}_j] + \mathbb{E}[\overline{H}_{i,j}^{(k)}\overline{H}_{i,j}^{(l)} \mid \overline{S}_i]) + n^{-2}\operatorname{Var}(\overline{H}_{i,j}^{(k)}\overline{H}_{i,j}^{(l)}) \\
&\lesssim n^{-1}\min\left\{P\{H_{i,j}^{(k)} \neq 0 \mid \overline{S}_j\}, P\{H_{i,j}^{(l)} \neq 0 \mid \overline{S}_j\}\right\}^2 \\
&+ n^{-1}\min\left\{P\{H_{i,j}^{(k)} \neq 0 \mid \overline{S}_i\}, P\{H_{i,j}^{(l)} \neq 0 \mid \overline{S}_i\}\right\}^2 \lesssim n^{-1}\min\{\kappa_{n,k}, \kappa_{n,l}\}^2 \, ,
\end{aligned} \tag{D.43}$$

where the last inequality follows by Assumption 5.3. Thus, the first inequality in (D.41) follows from Chebyshev's inequality. By an analogous argument, Lemma D.1 implies that

$$\operatorname{Var}\left(\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\overline{H}_{i,j}^{k}\overline{D}_{i,j}\right)$$

$$\lesssim n^{-1}\min\left\{P\{H_{i,j}^{(k)}\neq 0\mid \overline{S}_i\}, P\{D_{i,j}\neq 0\mid \overline{S}_i\}\right\}^2$$
$$+\, n^{-1}\min\left\{P\{H_{i,j}^{(k)}\neq 0\mid \overline{S}_j\}, P\{D_{i,j}\neq 0\mid \overline{S}_j\}\right\}^2 = O(n^{-1}\rho_n^2)\ , \tag{D.44}$$

and so the second inequality in (D.41) also follows from Chebyshev's inequality.

With this in place, we now turn to bounding the kth component of the first term in (D.39). Let e_k denote the kth basis vector in $\mathbb{R}^p$. Observe that

$$\|e_k^\top \Sigma_{H,H}^{-1}\|_1 = \sum_{l=1}^p |(\Sigma_{H,H}^{-1})_{k,l}| = \sum_{l=1}^p \frac{|K_{k,l}^{-1}|}{\sqrt{\mathrm{Var}(H_{i,j}^k)\,\mathrm{Var}(H_{i,j}^l)}}$$
$$\lesssim \frac{1}{\kappa_{n,k}}\sum_{l=1}^p \sqrt{\frac{\kappa_{n,l}}{\kappa_{n,l}}}|K_{k,l}^{-1}| = O(\kappa_{n,k}^{-1}) \tag{D.45}$$

by Assumption 5.3. Thus, we have that

$$|e_k^\top \Sigma_{H,H}^{-1}\Gamma_{H,D}| \leq \|e_k^\top \Sigma_{H,H}^{-1}\|_1\|\Gamma_{H,D}\|_\infty = O_p(n^{-1/2}\kappa_{n,k}^{-1}\rho_n)\ . \tag{D.46}$$

by the second inequality in (D.41) and (D.45).

Next, we consider the second term in (D.39). To this end, observe that

$$|\gamma_{n,k}| = |e_k^\top \Sigma_{H,H}^{-1}\Sigma_{H,D}| \leq \|e_k^\top \Sigma_{H,H}^{-1}\|\|\Sigma_{H,D}\|_\infty = O(\kappa_{n,k}^{-1}\rho_n) \tag{D.47}$$

for each k. Likewise, we can evaluate

$$|(\Gamma_{H,H}\gamma_n)_l| = \sum_{k=1}^p |\Gamma_{H,H}^{l,k}\gamma_{n,l}| \leq \sum_{k=1}^p |\Gamma_{H,H}^{l,k}||\gamma_{n,l}|$$
$$= \sum_{k=1}^p O_p(n^{-1/2}\min\{\kappa_{n,k},\kappa_{n,l}\}\kappa_{n,k}^{-1}\rho_n) = O_p(n^{-1/2}\rho_n)\ , \tag{D.48}$$

by the first inequality in (D.41) and the inequality (D.47). Consequently, we have that

$$|e_k^\top \Sigma_{H,H}^{-1}\Gamma_{H,H}\gamma_n| \leq \|e_k^\top \Sigma_{H,H}^{-1}\|_1\|\Gamma_{H,H}\gamma_n\|_\infty = O_p(n^{-1/2}\kappa_{n,k}^{-1}\rho_n)\ , \tag{D.49}$$

by the bound (D.45).

In turn, we consider the third term in (D.39), given by

$$R_n^{(1)} = \Sigma_{H,H}^{-1}\Gamma_{H,H}\Sigma_{H,H}^{-1}\Gamma_{H,H}\hat{\Sigma}_{H,H}^{-1}\Sigma_{H,D}\ . \tag{D.50}$$

Define the matrix

$$\sigma_n = \mathsf{diag}(\sqrt{\mathrm{Var}(H_{i,j}^1)},\ldots,\sqrt{\mathrm{Var}(H_{i,j}^p)})\ . \tag{D.51}$$

and observe that we can write

$$R_n^{(1)} = \sigma_n^{-1}K^{-1}\tilde{\Gamma}_{H,H}K^{-1}\tilde{\Gamma}_{H,H}(K+\tilde{\Gamma}_{H,H})^{-1}\tilde{\Gamma}_{H,D}\ , \tag{D.52}$$

where

$$\tilde{\Gamma}_{H,H} = \sigma_n^{-1}\Gamma_{H,H} \quad \text{and} \quad \tilde{\Sigma}_{H,G} = \sigma_n^{-1}\Sigma_{H,D}\ . \tag{D.53}$$

Observe that Assumption 5.3 and the first inequality in (D.41) imply that

$$\|\tilde{\Gamma}_{H,H}\|_\infty = O_p(n^{-1/2}) \,, \quad \|K^{-1}\|_\infty = O_p(1) \,, \quad \text{and} \quad \|(K^{-1} + \tilde{\Gamma}_{H,H})^{-1}\|_\infty = O_p(1) \,, \tag{D.54}$$

respectively. In turn, we have that

$$\|e_k^\top \sigma_n^{-1} K^{-1}\|_1 = O(\kappa_{n,k}^{-1}) \quad \text{and} \quad \|\tilde{\Sigma}_{H,D}\|_\infty = O(\rho_n^{1/2}) \tag{D.55}$$

by the same steps used to establish (D.48) and the fact that $\rho_n \lesssim \kappa_{n,k}$, respectively. Hence, we obtain

$$\begin{aligned} \|e_k^\top R_n^{(1)}\|_1 &= \|e_k^\top \sigma_n^{-1} K^{-1}\|_1 \|\tilde{\Gamma}_{H,H}\|_\infty \|K^{-1}\|_\infty \|\tilde{\Gamma}_{H,H}\|_\infty \|(K^{-1} + \tilde{\Gamma}_{H,H})^{-1}\|_\infty \|\tilde{\Sigma}_{H,D}\|_\infty \\ &= O_p(n^{-1}\kappa_{n,k}^{-1}\rho_n^{1/2}) = o_p(n^{-1/2}\kappa_{n,k}^{-1}\rho_n) \end{aligned} \tag{D.56}$$

by plugging the bounds (D.54) and (D.55) into (D.52).

Finally, we consider the fourth term in (D.39), given by

$$R_n^{(2)} = \Sigma_{H,H}^{-1}\Gamma_{H,H}\hat{\Sigma}_{H,H}^{-1}\Gamma_{H,D} \,. \tag{D.57}$$

Observe that we can write

$$R_n^{(2)} = \sigma_n^{-1} K^{-1} \tilde{\Gamma}_{H,H} (K + \tilde{\Gamma}_{H,H})^{-1} \tilde{\Gamma}_{H,D} \tag{D.58}$$

where

$$\tilde{\Gamma}_{H,D} = \sigma_n^{-1}\Gamma_{H,D} \tag{D.59}$$

Observe that the second inequality in (D.41) and the fact that $\rho_n \lesssim \kappa_{n,k}$ imply that

$$\|\tilde{\Gamma}_{H,D}\|_\infty = O_p(n^{-1/2}\rho_n^{1/2}) \,. \tag{D.60}$$

Consequently, we obtain the bound

$$\begin{aligned} \|e_k^\top R_n^{(1)}\|_1 &= \|e_k^\top \sigma_n^{-1} K^{-1}\|_1 \|\tilde{\Gamma}_{H,H}\|_\infty \|(K + \tilde{\Gamma}_{H,H})^{-1}\|_\infty \|\tilde{\Gamma}_{H,D}\|_\infty \\ &= O_p(n^{-1}\kappa_{n,k}^{-1}\rho_n^{1/2}) = o_p(n^{-1/2}\kappa_{n,k}^{-1}\rho_n) \end{aligned} \tag{D.61}$$

by plugging the bounds (D.54), (D.55) and (D.55) into (D.58). Hence, we obtain the bound

$$\hat{\gamma}_{n,k} - \gamma_{n,k} = O_p(n^{-1}\kappa_{n,k}^{-1}\rho_n^{1/2}) \tag{D.62}$$

by plugging the bounds (D.46), (D.49), (D.56), and (D.61) into the decomposition (D.39), as required. ■

D.2.2 Part (ii). Observe that Cauchy-Schwarz implies that

$$\left| \frac{1}{n^2} \sum_{i=1}^n J_{i,s} J_{i,r} \right| \leq \sqrt{\left(\frac{1}{n^2} \sum_{i=1}^n J_{i,s}^2 \right) \left(\frac{1}{n^2} \sum_{i=1}^n J_{i,r}^2 \right)} \,. \tag{D.63}$$

Thus, to verify the first bound, it suffices to show that

$$\mathbb{E}[J_{i,s}^2] = O(n\kappa_{n,s}) \,, \tag{D.64}$$

for each s, as this implies that

$$\frac{1}{n^2}\sum_{i=1}^{n} J_{i,s}^2 = O_p(\kappa_{n,s}) \tag{D.65}$$

by Markov's inequality. Thus, to ease notation, and without loss, we can drop the subscript s from $A_{i,s}$ and $\kappa_{n,s}$ and treat the quantity $\overline{H}_{i,j}$ as a scalar-valued random variable. Consider the decomposition

$$J_i = \sum_{j\neq i} \tilde{H}_{i,j} W_j^\star - \frac{1}{n-1}\sum_{j\neq i}\sum_{k\neq j} \tilde{H}_{k,j} W_j^\star + \frac{1}{n(n-1)}\sum_{j\neq i}\sum_{k\neq j} \overline{H}_{k,j} W_j^\star - \frac{1}{n}\sum_{j\neq i} \overline{H}_{k,i} W_i^\star \tag{D.66}$$

where

$$\tilde{G}_{i,j} = \tilde{H}_{i,j} - \mathbb{E}[\tilde{H}_{i,j} \mid \overline{S}_j] \; . \tag{D.67}$$

Observe that

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{j\neq i} \tilde{H}_{i,j} W_j^\star\right)^2\right] &= \sum_{j\neq i}\sum_{j'\neq i} \mathbb{E}[\tilde{H}_{i,j}\tilde{H}_{i,j'} W_j^\star W_{j'}^\star] \\ &= \sum_{j\neq i} \mathbb{E}[\tilde{H}_{i,j}^2 (W_j^\star)^2] \lesssim n\mathbb{E}[H_{i,j}^2] = O(n\kappa_n) \; , \end{aligned} \tag{D.68}$$

where the second equality follows from Assumption 4.2 and the final bound follows from Assumption 5.1. Analogously, we can evaluate

$$\begin{aligned} \mathbb{E}\left[\left(\frac{1}{n-1}\sum_{j\neq i}\sum_{k\neq j} \tilde{H}_{k,j} W_j^\star\right)^2\right] &= \frac{1}{(n-1)^2}\sum_{j\neq i}\sum_{k\neq j}\sum_{j'\neq i}\sum_{k'\neq j'} \mathbb{E}[\tilde{H}_{k,j}\tilde{H}_{k',j'} W_j^\star W_{j'}^\star] \\ &= \frac{1}{(n-1)^2}\sum_{j\neq i}\sum_{k\neq j} \mathbb{E}[\tilde{H}_{k,j}^2 (W_j^\star)^2] = O(\kappa_n) \; , \end{aligned} \tag{D.69}$$

as well as

$$\begin{aligned} &\mathbb{E}\left[\left(\frac{1}{n(n-1)}\sum_{j\neq i}\sum_{k\neq j} \overline{H}_{k,j} W_j^\star\right)^2\right] \\ &\qquad = \frac{1}{n^2(n-1)^2}\sum_{j\neq i}\sum_{k\neq j}\sum_{k'\neq j} \mathbb{E}[\overline{H}_{k,j}\overline{H}_{k',j}(W_j^\star)^2] = O(n^{-1}\kappa_n) \end{aligned} \tag{D.70}$$

and

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j\neq i} \overline{H}_{j,i} W_i^\star\right)^2\right] = \frac{1}{n^2}\sum_{j\neq i}\sum_{j'\neq i} \mathbb{E}[\overline{H}_{j,i}\overline{H}_{j',i}(W_i^\star)^2] = O(\kappa_n) \; , \tag{D.71}$$

respectively. Hence, the moment bound (D.64) follows by squaring the decomposition (D.66) within an expectation, applying Cauchy-Schwarz, and plugging in the bounds (D.68) through (D.71).

The proof of the second bound has the same structure. In particular, Cauchy-Schwarz implies that

$$\left| \frac{1}{n^2} \sum_{i=1}^{n} M_{i,s,r} M_{i,s',r'} \right| \lesssim \sqrt{\left(\frac{1}{n^2} \sum_{i=1}^{n} M_{i,s,r}^2 \right) \left(\frac{1}{n^2} \sum_{i=1}^{n} M_{i,s',r'}^2 \right)} \,. \tag{D.72}$$

Thus, it suffices to show that

$$\mathbb{E}\left[M_{i,s,r}^2\right] = O(n^2 \kappa_{n,s}) \tag{D.73}$$

for each s, r, as this implies that

$$\frac{1}{n^2} \sum_{i=1}^{n} M_{i,s,r}^2 = O_p(n \kappa_{n,s}) \tag{D.74}$$

by Markov's inequality. Thus, again, to ease notation, we can drop the subscripts s and r and treat the covariates $\overline{H}_{i,j}$ and $\overline{X}_j$ as scalar-valued random variables, without loss. Here, it suffices to consider the decomposition

$$M_i = \sum_{j \neq i} \overline{H}_{i,j} \overline{X}_j - \frac{1}{n} \sum_{j=1}^{n} \sum_{k \neq j} \overline{H}_{k,j} \overline{X}_j \,. \tag{D.75}$$

Observe that we can evaluate

$$\mathbb{E}\left[\left(\sum_{j \neq i} \overline{H}_{i,j} \overline{X}_j \right)^2 \right] = \sum_{j \neq i} \sum_{j' \neq i} \mathbb{E}[\overline{H}_{i,j} \overline{H}_{i,j'} \overline{X}_j \overline{X}_{j'}] \lesssim n^2 \mathbb{E}[\overline{H}_{i,j}^2 \overline{X}_j^2] = O(n^2 \kappa_n) \,, \tag{D.76}$$

and

$$\begin{aligned} &\mathbb{E}\left[\left(\frac{1}{n} \sum_{j=1}^{n} \sum_{k \neq j} \overline{H}_{k,j} \overline{X}_j \right)^2 \right] \\ &= \frac{1}{n^2} \sum_{j=1}^{n} \sum_{j'=1}^{n} \sum_{k \neq j} \sum_{k' \neq j'} \mathbb{E}[\overline{H}_{k,j} \overline{X}_j \overline{H}_{k',j'} \overline{X}_{j'}] \lesssim n^2 \mathbb{E}[\overline{H}_{i,j}^2 \overline{X}_j^2] = O(n^2 \kappa_n) \,, \end{aligned} \tag{D.77}$$

by Cauchy-Schwarz and Assumption 5.1. Hence, the moment bound (D.73) follows by squaring the decomposition (D.75) within an expectation, applying Cauchy-Schwarz to remove the cross-terms, and plugging in the bounds (D.76) and (D.77). ■

D.3 Proof of Lemma C.3

D.3.1 Part (i). Define the conditional expectations

$$\tilde{F}(Z_i) = \mathbb{E}[Y_i \mid Z_i] \quad \text{and} \quad \tilde{F}(Z_i, Z_j) = \mathbb{E}[Y_i \mid Z_i, Z_j] - \tilde{F}(Z_i) \tag{D.78}$$

and the associated residual

$$R_i = Y_i - \tilde{F}(Z_i) - \sum_{j \neq i} \tilde{F}(Z_i, Z_j) \,, \tag{D.79}$$

for each unit i in $[n]$. Consider the decomposition

$$\frac{1}{n^2}\sum_{i=1}^{n} Y_i \check{\Delta}_i = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} Y_i \tilde{D}^{\star}_{i,j} W^{\star}_j = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} R_i \tilde{D}^{\star}_{i,j} W^{\star}_j$$
$$+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} \mathbb{E}[Y_i \mid Z_i, Z_j]\tilde{D}^{\star}_{i,j} W^{\star}_j$$
$$+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}} \tilde{F}(Z_i, Z_k)\tilde{D}^{\star}_{i,j} W^{\star}_j \ . \qquad \text{(D.80)}$$

We evaluate each term, in succession, through the application of the following Lemma.

Lemma D.7. *If the conditions of Theorem 5.1 hold, then:*
(i) *It holds that*

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} R_i \tilde{D}^{\star}_{i,j} W^{\star}_j = O_p(n^{-1/2}\rho_n) \ . \qquad \text{(D.81)}$$

(ii) *It holds that*

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} \mathbb{E}[Y_i \mid Z_i, Z_j]\tilde{D}^{\star}_{i,j} W^{\star}_j = \mathbb{E}[Y_i \tilde{D}^{\star}_{i,j} W^{\star}_j] + O_p(\rho_n n^{-1/2}) \ . \qquad \text{(D.82)}$$

(iii) *It holds that*

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}} \tilde{F}(Z_i, Z_k)\tilde{D}^{\star}_{i,j} W^{\star}_j = O_p(\rho_n^{3/2}) \ . \qquad \text{(D.83)}$$

Consequently, by applying each part of Lemma D.7 to the representation (D.80), we find that

$$\frac{1}{n^2}\sum_{i=1}^{n} Y_i \check{\Delta}_i - \mathbb{E}[Y_i \tilde{D}^{\star}_{i,j} W^{\star}_j] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}} \tilde{F}(Z_i, Z_k)\tilde{D}^{\star}_{i,j} W^{\star}_j + O_p(n^{-1/2}\rho_n) \qquad \text{(D.84)}$$

and

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}} \tilde{F}(Z_i, Z_k)\tilde{D}^{\star}_{i,j} W^{\star}_j = O_p(\rho_n^{3/2}) \ . \qquad \text{(D.85)}$$

as required. ■

D.3.2 Part (ii). We now turn to the terms

$$\frac{1}{n-1}\frac{1}{n^2}\sum_{i=1}\sum_{j\neq i} Y_i \check{\Delta}^{\star}_{i,j} \quad \text{and} \quad \frac{1}{n^2}\sum_{i=1}\sum_{j\neq i} Y_i (Q_i^{(1)} - Q_i^{(2)} + Q_i^{(3)}) \ . \qquad \text{(D.86)}$$

Observe that each term can be decomposed analogously to (D.80). That is, we have that

$$\frac{1}{n^3}\sum_{i=1}\sum_{j\neq i} Y_i \check{\Delta}^{\star}_{i,j} = \frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}} R_i \tilde{D}^{\star}_{k,j} W^{\star}_j + \frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}} \mathbb{E}[Y_i \mid Z_i, Z_j]\tilde{D}^{\star}_{k,j} W^{\star}_j$$

$$+\frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)\tilde{D}^\star_{k,j}W^\star_j\ . \tag{D.87}$$

Likewise, we can write

$$\frac{n-1}{n^3}\sum_{i=1}\sum_{j\neq i}Y_iQ^{(1)}_i=\frac{1}{n^4}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\notin\{i,j\}}R_iD^\star_{k,j}W^\star_j+\frac{1}{n^4}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\notin\{i,j\}}\mathbb{E}[Y_i\mid Z_i,Z_j]D^\star_{k,j}W^\star_j$$
$$+\frac{1}{n^4}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)D^\star_{k,j}W^\star_j \tag{D.88}$$

as well as

$$\frac{n-1}{n^3}\sum_{i=1}^n\sum_{j\neq i}Y_iQ^{(2)}_i=\frac{1}{n^3}\sum_{i=1}^n\sum_{j\neq i}R_iD^\star_{j,i}W^\star_i+\frac{1}{n^3}\sum_{i=1}^n\sum_{j\neq i}\mathbb{E}[Y_i\mid Z_i,Z_j]D^\star_{j,i}W^\star_i$$
$$+\frac{1}{n^3}\sum_{i=1}^n\sum_{j\neq i}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)D^\star_{j,i}W^\star_i \tag{D.89}$$

and

$$\frac{n-1}{n^3}\sum_{i=1}\sum_{j\neq i}Y_iQ^{(3)}_i=\frac{1}{n^4}\sum_{i=1}^n\sum_{j\neq i}R_iD^\star_{i,j}W^\star_j+\frac{1}{n^4}\sum_{i=1}^n\sum_{j\neq i}\mathbb{E}[Y_i\mid Z_i,Z_j]D^\star_{i,j}W^\star_j$$
$$+\frac{1}{n^4}\sum_{i=1}^n\sum_{j\neq i}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)D^\star_{i,j}W^\star_j\ , \tag{D.90}$$

respectively. We evaluate each term through the application of the following lemma.

Lemma D.8. *If the conditions of Theorem 5.1 hold, then:*
(i) *It holds that*

$$\frac{1}{n^3}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\notin\{i,j\}}R_i\tilde{D}^\star_{k,j}W^\star_j=O_p(n^{-1}\rho_n)\ ,\quad \frac{1}{n^4}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\notin\{i,j\}}R_iD^\star_{k,j}W^\star_j=O_p(n^{-3/2}\rho_n) \tag{D.91}$$
$$\frac{1}{n^3}\sum_{i=1}^n\sum_{j\neq i}R_iD^\star_{j,i}W^\star_i=O_p(n^{-1}\rho_n)\ ,\quad \text{and}\quad \frac{1}{n^4}\sum_{i=1}^n\sum_{j\neq i}R_iD^\star_{i,j}W^\star_j=O_p(n^{-5/2}\rho_n)\ ,$$

respectively.
(ii) *It holds that*

$$\frac{1}{n^2(n-1)}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\mathbb{E}[Y_i\mid Z_i,Z_j]\tilde{D}^\star_{k,j}W^\star_j=\mathbb{E}[Y_i\tilde{D}_{l,q}W^\star_q]+O_p(n^{-1/2}\rho_n)\ , \tag{D.92}$$
$$\frac{1}{n^3(n-1)}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\notin\{i,j\}}\mathbb{E}[Y_i\mid Z_i,Z_j]D^\star_{k,j}W^\star_j=\frac{1}{n}\mathbb{E}[Y_iD^\star_{l,q}W^\star_q]+O_p(n^{-3/2}\rho_n^{1/2})\ ,$$
$$\frac{1}{n^2(n-1)}\sum_{i=1}^n\sum_{j\neq i}\mathbb{E}[Y_i\mid Z_i,Z_j]D^\star_{j,i}W^\star_i=\frac{1}{n}\mathbb{E}[Y_iD^\star_{l,i}W^\star_i]+O_p(n^{-3/2}\rho_n^{1/2})\ ,$$

$$\frac{1}{n^3(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}[Y_i \mid Z_i, Z_j]D^{\star}_{i,j}W^{\star}_j = \frac{1}{n^2}\mathbb{E}[Y_i D^{\star}_{i,l}W^{\star}_l] + O_p(n^{-5/2}\rho_n^{1/2})$$

respectively.

(iii) *It holds that*

$$\frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)\tilde{D}^{\star}_{k,j}W^{\star}_j = O_p(n^{-1/2}\rho_n)\ , \tag{D.93}$$

$$\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)D^{\star}_{k,j}W^{\star}_j = O_p(n^{-1}\rho_n^{1/2})\ ,$$

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)D^{\star}_{j,i}W^{\star}_i = O_p(n^{-1}\rho_n^{1/2})\ , \quad \textit{and}$$

$$\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)D^{\star}_{i,j}W^{\star}_j = O_p(n^{-3/2}\rho_n^{1/2})$$

respectively.

By plugging each of the bounds from Lemma D.8 into the decompositions (D.87) though (D.90), we obtain

$$\frac{1}{n-1}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}Y_i\check{\Delta}_{i,j} = \mathbb{E}[Y_i\tilde{D}_{l,j}W_j] + O_p(n^{-1/2}\rho_n)\ , \quad \text{and} \tag{D.94}$$

$$\frac{1}{n^2}\sum_{i=1}^{n}Y_i(Q_i^{(1)} - Q_i^{(2)} + Q_i^{(3)}) = \frac{1}{n}\mathbb{E}[Y_i\overline{D}_{q,l}W_l] - \frac{1}{n}\mathbb{E}[Y_i\overline{D}_{l,i}W_i] + \frac{1}{n^2}\mathbb{E}[Y_i\overline{D}_{i,l}W_l] + O_p(n^{-1}\rho_n^{1/2})\ , \tag{D.95}$$

verifying the representations (C.38) and (C.39), respectively. ■

D.4 Proof of Lemma C.4

We continue to use the notation introduced in the proof of Lemma C.3. To verify the bounds, we consider the decompositions

$$\frac{1}{n^2}\sum_{i=1}^{n}Y_iJ_{i,s} = \frac{1}{n^2}\sum_{i=1}^{n}R_iJ_{i,s} + \frac{1}{n^2}\sum_{i=1}^{n}\tilde{F}(Z_i)J_{i,s} + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{F}(Z_i,Z_j)J_{i,s} \quad \text{and} \tag{D.96}$$

$$\frac{1}{n^2}\sum_{i=1}^{n}Y_iM_{i,s,r} = \frac{1}{n^2}\sum_{i=1}^{n}R_iM_{i,s,r} + \frac{1}{n^2}\sum_{i=1}^{n}\tilde{F}(Z_i)M_{i,s,r} + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{F}(Z_i,Z_j)M_{i,s,r}\ .$$

Bounds for each term are given in the following Lemma.

Lemma D.9. *If the conditions of Theorem 5.1 hold, then:*

(i) *It holds that*

$$\frac{1}{n^2}\sum_{i=1}^{n}R_iJ_{i,s} = O_p((\rho_n\kappa_{n,s})^{1/2}n^{-1/2}) \quad \textit{and} \quad \frac{1}{n^2}\sum_{i=1}^{n}R_iM_{i,s,r} = O_p((\rho_n\kappa_{n,s})^{1/2})\ . \tag{D.97}$$

(ii) *It holds that*

$$\frac{1}{n^2}\sum_{i=1}^{n}\tilde{F}(Z_i)J_{i,s} = O_p(\kappa_{n,s}) \quad \textit{and} \quad \frac{1}{n^2}\sum_{i=1}^{n}\tilde{F}(Z_i)M_{i,s,r} = O_p(n^{1/2}\kappa_{n,s}) \, . \tag{D.98}$$

(iii) *It holds that*

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{F}(Z_i,Z_j)J_{i,s} = O_p(\kappa_n) \quad \textit{and} \quad \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{F}(Z_i,Z_j)M_{i,s,r} = O_p(n^{1/2}\kappa_n) \, . \tag{D.99}$$

Thus, we can conclude that

$$\frac{1}{n^2}\sum_{i=1}^{n}Y_iJ_{i,s} = O_p(\kappa_{n,s}) \quad \text{and} \quad \frac{1}{n^2}\sum_{i=1}^{n}Y_iM_{i,s,r} = O_p(n^{1/2}\kappa_{n,s}) \, , \tag{D.100}$$

by plugging each of the bounds from Lemma D.9 into (D.96). ■

D.5 Proof of Lemma C.7

Let $\xi = (\xi_1, \xi_2)$ be any real-valued vector that satisfies $\xi_1^2 + \xi_2^2 = 1$. By the Cramér-Wold device, it suffices to show that

$$\frac{\sqrt{2}}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\overline{\sigma}_n^{-1}(\xi)(\xi_1\tilde{F}(Z_i,Z_k) + \xi_2\tilde{D}^\star_{i,k}W^\star_k)\tilde{D}^\star_{i,j}W^\star_j \xrightarrow{d} \mathsf{N}(0,1) \, , \tag{D.101}$$

where

$$\overline{\sigma}_n^2(\xi) = \xi_1^2\sigma_{n,1}^2 + 2\xi_1\xi_2\varsigma_n + \xi_2^2\sigma_{n,2}^2 \, . \tag{D.102}$$

To this end, observe that, by writing

$$f_\xi(Z_1,Z_2,Z_3) = \sum_{\pi\in\Pi_3}(\xi_1\tilde{F}(Z_{\pi(1)},Z_{\pi(3)}) + \xi_2\tilde{D}^\star(Z_{\pi(1)},Z_{\pi(3)})W^\star_{\pi(3)})\tilde{D}^\star(S_{\pi(1)},S_{\pi(2)})W^\star_{\pi(2)} \, , \tag{D.103}$$

the quantity

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}(\xi_1\tilde{F}(Z_i,Z_k) + \xi_2\tilde{D}^\star_{i,k}W^\star_k)\tilde{D}^\star_{i,j}W^\star_j \\ &= \frac{(n-1)(n-2)}{6n}\left(\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n}f_\xi(Z_i,Z_j,Z_k)\right)\end{aligned} \tag{D.104}$$

can be recognized as scaled U-statistic of order 3. Moreover, by the equalities (D.121), (D.177), and (D.178), the U-statistic (D.104) is degenerate of order one. In turn, by the equalities (D.129) and (D.180), the second term in Hoeffding decomposition of (D.104) is given by

$$\begin{aligned}f_\xi^{(2)}(Z_2,Z_3) &= \mathbb{E}[(\xi_1\tilde{F}(Z_1,Z_3) + \xi_2\tilde{D}^\star_{1,3}W^\star_3)\tilde{D}^\star_{1,2} \mid Z_2,Z_3]W^\star_2 \\ &\quad + \mathbb{E}[(\xi_1\tilde{F}(Z_1,Z_2) + \xi_2\tilde{D}^\star_{1,2}W^\star_2)\tilde{D}^\star_{1,3} \mid Z_2,Z_3]W^\star_3 \, .\end{aligned} \tag{D.105}$$

Thus, Lemma D.4 implies that

$$\sup_{x\in\mathbb{R}}\left|P\left\{\frac{1}{\overline{\sigma}_n(\xi)}\frac{\sqrt{\binom{n}{2}}}{\binom{3}{2}}\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n} f_\xi(Z_i,Z_j,Z_k)\le x\right\}-\Phi(x)\right|$$
$$\lesssim \sqrt{\frac{1}{n}+\frac{\mathbb{E}[(f_\xi^{(3)}(Z_1,Z_2,Z_3))^2]}{n\mathbb{E}[(f_\xi^{(2)}(Z_1,Z_2))^2]}}+\frac{\sqrt{\frac{1}{n}\mathbb{E}[(f_\xi^{(2)}(Z_1,Z_2))^4]+\mathbb{E}[(g_\xi^{(2)}(Z_1,Z_2))^2]}}{\mathbb{E}[(f_\xi^{(2)}(Z_1,Z_2))^2]}\ , \tag{D.106}$$

where $f_\xi^{(3)}(Z_1,Z_2,Z_3)$ is the third term in Hoeffding decomposition of (D.104), and

$$g_\xi^{(2)}(z_1,z_2)=\mathbb{E}[f_\xi^{(2)}(Z_3,z_1)f_\xi^{(2)}(Z_3,z_2)]\ , \tag{D.107}$$

respectively. Observe that Lemma C.6 implies that

$$\overline{\sigma}_n^2(\xi)=\mathbb{E}[(f_\xi^{(2)}(X_1,X_2))^2]\gtrsim \rho_n^3 \tag{D.108}$$

Upper bounds for each of the other three quantities appearing in (D.106) are given by the following Lemma.

Lemma D.10. *Under the conditions of Theorem 6.2, the bounds*

$$\begin{aligned}\mathbb{E}[(f_\xi^{(3)}(Z_1,Z_2,Z_3))^2]&=O(\rho_n^2)\ ,\\ \mathbb{E}[(f_\xi^{(2)}(Z_1,Z_2))^4]&=O(\rho_n^5)\ ,\quad \textit{and}\\ \mathbb{E}[(g_\xi^{(2)}(Z_1,Z_2))^2]&=o(\rho_n^6)\end{aligned} \tag{D.109}$$

hold.

Plugging (D.108) and each of the bounds from Lemma D.10 into (D.106), we obtain

$$\frac{1}{\overline{\varphi}_n(\xi)}\frac{\sqrt{\binom{n}{2}}}{\binom{3}{2}}\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n} f_\xi(Z_i,Z_j,Z_k)\xrightarrow{d}\mathsf{N}(0,1)\ . \tag{D.110}$$

Consequently, as

$$\frac{\sqrt{\binom{n}{2}}}{\binom{3}{2}}\frac{1}{\binom{n}{3}}=\frac{\sqrt{2}}{(n-2)\sqrt{n(n-1)}}=\frac{\sqrt{2}}{n^2}+O(n^{-3}) \tag{D.111}$$

we can conclude that

$$\frac{\sqrt{2}}{\overline{\sigma}_n(\xi)}\frac{1}{n^2}\sum_{i=1}^n\sum_{j\ne i}\sum_{k\notin\{i,j\}}(\xi_1\tilde{F}(Z_i,Z_k)+\xi_2\tilde{D}^\star_{i,k}W^\star_k)\tilde{D}^\star_{i,j}W^\star_j\xrightarrow{d}\mathsf{N}(0,1)\ , \tag{D.112}$$

by the continuous mapping theorem, as required. ■

D.6 Proof of Lemma D.5

First, we verify the bound (D.12). By writing

$$g(Z_1,Z_2)=\sum_{\pi\in\Pi_2}\tilde{D}^\star(S_{\pi(1)},S_{\pi(2)})^2(W^\star_{\pi(2)})^2\ , \tag{D.113}$$

where we recall that Π_m denotes the set of permutations on $[m]$, the term

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}(\tilde{D}^\star_{i,j})^2(W^\star_j) = \frac{n-1}{2n}\left(\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j>i} g(Z_i, Z_j)\right) \tag{D.114}$$

can be recognized as a scaled U-statistic of order 2. Thus, Part (ii) of Lemma D.2 implies that

$$\mathrm{Var}\left(\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j>i} g(Z_i, Z_j)\right) \lesssim n^{-1}\,\mathrm{Var}(\mathbb{E}[g(Z_1, Z_2) \mid Z_1]) + n^{-2}\,\mathrm{Var}(g(Z_1, Z_2))\;. \tag{D.115}$$

Observe that

$$\begin{aligned}
\mathrm{Var}(g(Z_1, Z_2)) &= \mathrm{Var}((\tilde{D}^\star_{2,1})^2(W^\star_1)^2 + (\tilde{D}^\star_{1,2})^2(W^\star_2)^2)\\
&\lesssim \mathbb{E}[(\tilde{D}^\star_{2,1})^4(W^\star_1)^4]\\
&\lesssim \mathbb{E}[D^4_{2,1} + \mathbb{E}[D(S_2, S_1) \mid H_{2,1}, S_1]^4] \lesssim \rho_n\mathbb{E}[D^4_{2,1} \mid \mathcal{A}_{2,1}] = O(\rho_n)\;,
\end{aligned} \tag{D.116}$$

where the first two inequalities follow from Cauchy-Schwarz, the third inequality follows from Assumption 5.2, and the final bound follows from the fact that the treatments and distances are bounded buy a constant. In turn, we can evaluate

$$\begin{aligned}
&\mathrm{Var}(\mathbb{E}[g(Z_1, Z_2) \mid Z_1])\\
&\lesssim \mathbb{E}[\mathbb{E}[\tilde{D}^\star(S_2, S_1)^2 \mid Z_1]^2] + \mathbb{E}[\mathbb{E}[\tilde{D}^\star(S_1, S_2)^2] \mid Z_1]^2]\\
&\lesssim \mathbb{E}[\mathbb{E}[D(S_2, S_1) \mid Z_1]^2 + \mathbb{E}[\mathbb{E}[D(S_2, S_1) \mid H_{2,1}, S_1] \mid Z_1]^2]\\
&+ \mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid Z_1]^2 + \mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2] \mid Z_1]^2]\\
&\lesssim \rho_n^2(\mathbb{E}[\mathbb{E}[D(S_2, S_1) \mid \mathcal{A}_{2,1}, Z_1]^2 + \mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid \mathcal{A}_{2,1}, Z_1]^2])\\
&+ \rho_n^2\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2] \mid \mathcal{A}_{2,1}, Z_1]^2]\\
&+ \mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2] \mid \tilde{\mathcal{A}}_{2,1}, Z_1]^2] \lesssim O(\rho_n^2)\;,
\end{aligned} \tag{D.117}$$

by Cauchy-Schwarz and Assumption 5.2. Thus, the bound (D.115) and Chebyshev's inequality imply that

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}(\tilde{D}^\star_{i,j})^2(W^\star_j)^2 - \frac{n-1}{n}\mathbb{E}[\mathrm{Var}(D_{i,j} \mid H_{i,j}, S_j)\,\mathrm{Var}(W_i \mid S_j)]\\
&\qquad = O_p(\rho_n n^{-1/2} + \rho_n^{1/2}n^{-1}) = O_p(\rho_n n^{-1/2})\;,
\end{aligned} \tag{D.118}$$

as required.

Next, we verify the bound (D.13). Proceeding in the same way, by writing

$$h(Z_1, Z_2, Z_3) = \sum_{\pi\in\Pi_3}\tilde{D}^\star(S_{\pi(1)}, S_{\pi(2)})\tilde{D}^\star(S_{\pi(1)}, S_{\pi(3)})W^\star_{\pi(2)}W^\star_{\pi(3)}\;, \tag{D.119}$$

the term

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{D}^\star_{i,j}\tilde{D}^\star_{i,k}W^\star_jW^\star_k = \frac{(n-1)(n-2)}{6n}\left(\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n} h(Z_i, Z_j, Z_k)\right) \tag{D.120}$$

can be recognized as a scaled U-statistic order 3. Now, by Assumption 4.2, it holds that

$$\mathbb{E}[\tilde{D}^\star_{1,2}\tilde{D}^\star_{1,3}W^\star_2 W^\star_3 \mid Z_i] = 0 \tag{D.121}$$

for each i in $\{1, 2, 3\}$, and so

$$\mathrm{Var}(\mathbb{E}[h(Z_1, Z_2, Z_3) \mid Z_1]) = 0 \ . \tag{D.122}$$

Consequently, Part (ii) of Lemma D.2 implies that

$$\begin{aligned}&\mathrm{Var}\left(\frac{(n-1)(n-2)}{6n}\left(\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n} h(X_i, X_j, X_k)\right)\right)\\&\lesssim \mathrm{Var}(\mathbb{E}[h(Z_1, Z_2, Z_3) \mid Z_2, Z_3]) + n^{-1}\,\mathrm{Var}(h(Z_1, Z_2, Z_3)) \ .\end{aligned} \tag{D.123}$$

Observe that

$$\begin{aligned}&\mathrm{Var}(h(Z_1, Z_2, Z_3))\\&\lesssim \mathbb{E}[(\tilde{D}^\star(S_1, S_2)\tilde{D}^\star(S_1, S_3))^2]\\&\lesssim \mathbb{E}[(D(S_1, S_2)D(S_1, S_3))^2] + \mathbb{E}[D(S_1, S_2)^2\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3]^2]\\&+ \mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]^2\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3]^2] \ ,\end{aligned} \tag{D.124}$$

by Cauchy-Schwarz. We can evaluate

$$\mathbb{E}[(D(S_1, S_2)D(S_1, S_3))^2] \lesssim \rho_n^2\mathbb{E}[(D(S_1, S_2)D(S_1, S_3))^2 \mid \mathcal{A}_{1,2}, \mathcal{A}_{1,3}] = O(\rho_n^2) \tag{D.125}$$

by Assumption 5.2. Likewise, Assumption 5.2 implies that

$$\begin{aligned}&\mathbb{E}[D(S_1, S_2)^2\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3]^2]\\&\lesssim \rho_n^2\mathbb{E}[D(S_1, S_2)^2\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3]^2 \mid \mathcal{A}_{1,2}, \mathcal{A}_{1,3}]\\&+ \rho_n\mathbb{E}[D(S_1, S_2)^2\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3]^2 \mid \mathcal{A}_{1,2}, \tilde{\mathcal{A}}_{1,3}] = O(\rho_n^2) \ ,\end{aligned} \tag{D.126}$$

and analogously

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]^2\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3]^2]\\&\lesssim \rho_n^2\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]^2\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3]^2 \mid \mathcal{A}_{1,2}, \mathcal{A}_{1,3}]\\&+ \rho_n\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]^2\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3]^2 \mid \mathcal{A}_{1,2}, \tilde{\mathcal{A}}_{1,3}]\\&+ \mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]^2\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3]^2 \mid \tilde{\mathcal{A}}_{1,2}, \tilde{\mathcal{A}}_{1,3}] = O(\rho_n^2) \ .\end{aligned} \tag{D.127}$$

Thus, we have that

$$\mathrm{Var}(h(Z_1, Z_2, Z_3)) = O(\rho_n^2) \ . \tag{D.128}$$

In turn, observe that

$$\mathbb{E}[\tilde{D}^\star_{1,2}\tilde{D}^\star_{1,3}W^\star_2 W^\star_3 \mid Z_1, Z_2] = 0 \quad \text{and} \quad \mathbb{E}[\tilde{D}^\star_{1,2}\tilde{D}^\star_{1,3}W^\star_2 W^\star_3 \mid Z_1, Z_3] = 0 \ , \tag{D.129}$$

respectively. As a consequence, we can evaluate

$$\begin{aligned}
&\mathrm{Var}(\mathbb{E}[h(Z_1, Z_2, Z_3) \mid Z_2, Z_3]) \\
&\lesssim \mathrm{Var}(\mathbb{E}[\tilde{D}^\star(S_1, S_2)\tilde{D}^\star(S_1, S_3) \mid Z_2, Z_3]W_2 W_3) \\
&\leq \mathbb{E}[\mathbb{E}[\tilde{D}^\star(S_1, S_2)\tilde{D}^\star(S_1, S_3) \mid Z_2, Z_3]^2] \\
&\lesssim \mathbb{E}[\mathbb{E}[D(S_1, S_2)D(S_1, S_3) \mid S_2, S_3]^2] \\
&+ \mathbb{E}[\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]D(S_1, S_3) \mid S_2, S_3]^2] \\
&+ \mathbb{E}[\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3] \mid S_2, S_3]^2] \ . \qquad \text{(D.130)}
\end{aligned}$$

Observe that

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[D(S_1, S_2)D(S_1, S_3) \mid S_2, S_3]^2] \\
&\leq P\{\mathcal{B}_{2,3}\}\mathbb{E}[(P\{\mathcal{A}_{1,2}\}\mathbb{E}[D(S_1, S_2)D(S_1, S_3) \mid S_2, S_3, \mathcal{A}_{1,2}])^2 \mid \mathcal{B}_{2,3}] = O(\rho_n^3) \ , \qquad \text{(D.131)}
\end{aligned}$$

by Assumption 5.2. In turn, as $\tilde{\mathcal{A}}_{1,2} \subseteq \tilde{\mathcal{B}}_{2,3} \cap \mathcal{A}_{1,3}$ we have that

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid G_{1,2}, S_2]D(S_1, S_3) \mid S_2, S_3]^2] \\
&\lesssim \rho_n^3\mathbb{E}[\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]D(S_1, S_3) \mid \mathcal{A}_{1,3}, S_2, S_3]^2 \mid \mathcal{B}_{2,3}] \\
&+ \rho_n^2\mathbb{E}[\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]D(S_1, S_3) \mid \mathcal{A}_{1,3}, S_2, S_3]^2 \mid \tilde{\mathcal{B}}_{2,3}] = O(\rho_n^3) \qquad \text{(D.132)}
\end{aligned}$$

by Assumption 5.2. Likewise, we have that

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3] \mid S_2, S_3]^2] \\
&\lesssim \rho_n^4\mathbb{E}[\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3] \mid \mathcal{A}_{1,3}, \mathcal{A}_{2,3}, S_2, S_3]^2] \\
&+ \rho_n^2\mathbb{E}[\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3] \mid \tilde{\mathcal{A}}_{1,3}, \mathcal{A}_{2,3}, S_2, S_3]^2] \\
&+ \mathbb{E}[\mathbb{E}[\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]\mathbb{E}[D(S_1, S_3) \mid H_{1,3}, S_3] \mid \tilde{\mathcal{A}}_{1,3}, \tilde{\mathcal{A}}_{2,3}, S_2, S_3]^2] = O(\rho_n^4) \ . \qquad \text{(D.133)}
\end{aligned}$$

Thus, we obtain the bound

$$\mathrm{Var}(\mathbb{E}[h(Z_1, Z_2, Z_3) \mid Z_2, Z_3]) = O(\rho_n^3) \ . \qquad \text{(D.134)}$$

Consequently, plugging (D.128) and (D.134) into (D.123), we find that

$$\mathrm{Var}\left(\frac{(n-1)(n-2)}{6n}\left(\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n} h(Z_i, Z_j, Z_k)\right)\right) = O(\rho_n^3) \ . \qquad \text{(D.135)}$$

The representation (D.120) and Chebyshev's inequality, in turn, imply that

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}} \tilde{D}^\star_{i,j}\tilde{D}^\star_{i,k}W^\star_j W^\star_k = O(\rho_n^{3/2}) \ , \qquad \text{(D.136)}$$

as required. ■

D.7 Proof of Lemma D.6

We continue to use the notation introduced in the proofs of Lemma C.3 and Lemma D.5. Each bound follows by demonstrating that each term can be expressed as a U-statistic, and applying Lemma D.2. In particular, observe that we can write

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{D}^{\star}_{j,i}\tilde{D}^{\star}_{i,j}W^{\star}_{j}W^{\star}_{i} = \frac{\binom{n}{2}}{n^3}\frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n}h_0(Z_i,Z_j) \tag{D.137}$$

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{D}^{\star}_{i,k}\tilde{D}^{\star}_{j,k}(W^{\star}_{k})^2 = \frac{\binom{n}{3}}{n^3}\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n}h_1(Z_i,Z_j,Z_k)\ , \tag{D.138}$$

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\notin\{j,k\}}\tilde{D}^{\star}_{i,j}\tilde{D}^{\star}_{j,l}W^{\star}_{j}W^{\star}_{l} = \frac{\binom{n}{3}}{n^3}\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n}h_2(Z_i,Z_j,Z_k)\ , \quad \text{and} \tag{D.139}$$

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j,l\}}\tilde{D}^{\star}_{i,k}\tilde{D}^{\star}_{j,l}W^{\star}_{k}W^{\star}_{l} = \frac{\binom{n}{4}}{n^3}\frac{1}{\binom{n}{4}}\sum_{1\le i<j<k<l\le n}h_3(Z_i,Z_j,Z_k,Z_l)\ , \tag{D.140}$$

where

$$h_0(Z_1,Z_2) = \sum_{\pi\in\Pi_2}\tilde{D}^{\star}(S_{\pi(2)},S_{\pi(1)})\tilde{D}^{\star}(S_{\pi(1)},S_{\pi(2)})W^{\star}_{\pi(2)}W^{\star}_{\pi(1)} \tag{D.141}$$

$$h_1(Z_1,Z_2,Z_3) = \sum_{\pi\in\Pi_3}\tilde{D}^{\star}(S_{\pi(1)},S_{\pi(3)})\tilde{D}^{\star}(S_{\pi(2)},S_{\pi(3)})(W^{\star}_{\pi(3)})^2\ ,$$

$$h_2(Z_1,Z_2,Z_3) = \sum_{\pi\in\Pi_3}\tilde{D}^{\star}(S_{\pi(1)},S_{\pi(2)})\tilde{D}^{\star}(S_{\pi(2)},S_{\pi(3)})W^{\star}_{\pi(2)}W^{\star}_{\pi(3)}\ , \quad \text{and}$$

$$h_3(Z_1,Z_2,Z_3,Z_4) = \sum_{\pi\in\Pi_4}\tilde{D}^{\star}(S_{\pi(1)},S_{\pi(3)})\tilde{D}^{\star}(S_{\pi(2)},S_{\pi(4)})W^{\star}_{\pi(3)}W^{\star}_{\pi(4)}\ ,$$

respectively. We begin by considering the term (D.137). Part (ii) of Lemma D.1 and the law of total variance imply that

$$\operatorname{Var}\left(\frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n}h_0(Z_i,Z_j)\right) \lesssim n^{-1}\operatorname{Var}(h_0(Z_i,Z_j))\ . \tag{D.142}$$

We can evaluate

$$\begin{aligned}\operatorname{Var}(h_0(Z_i,Z_j)) &\lesssim \mathbb{E}[\tilde{D}^{\star}(S_2,S_1)^2\tilde{D}^{\star}(S_1,S_2)^2(W^{\star}_2)^2(W^{\star}_1)^2] \\ &\lesssim \mathbb{E}[\tilde{D}^{\star}(S_2,S_1)^4] \lesssim \mathbb{E}[D(S_2,S_1)^4] \lesssim \rho_n\end{aligned} \tag{D.143}$$

by Cauchy-Schwarz and Assumption 5.2. Consequently, we find that

$$\operatorname{Var}\left(\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{D}^{\star}_{j,i}\tilde{D}^{\star}_{i,j}W^{\star}_{j}W^{\star}_{i}\right) = O(n^{-3}\rho_n) \tag{D.144}$$

and so Chebyshev's inequality gives

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{D}^{\star}_{j,i}\tilde{D}^{\star}_{i,j}W^{\star}_{j}W^{\star}_{i} = O_p(n^{-3/2}\rho_n^{1/2})\ , \tag{D.145}$$

which verifies the bound (D.19).

Next, we consider the term (D.138). Observe that

$$\mathbb{E}[\tilde{D}^\star_{1,3}\tilde{D}^\star_{2,3}(W^\star_3)^2 \mid Z_i] = 0\ , \tag{D.146}$$

holds for each i in $\{1,2,3\}$, as the treatments are independent of the latent coordinates and the latent coordinates are i.i.d. Thus, Part (ii) of Lemma D.2 and the law of total variance imply that

$$\mathrm{Var}\left(\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n} h_1(Z_1,Z_2,Z_3)\right) \lesssim n^{-2}\,\mathrm{Var}(h_1(Z_1,Z_2,Z_3))\ . \tag{D.147}$$

Observe that the bound

$$\mathrm{Var}(h_1(X_1,X_2,X_3)) \lesssim \mathbb{E}[(\tilde{D}^\star_{1,3}\tilde{D}^\star_{2,3})^2 W_3^4] \lesssim \sqrt{\mathbb{E}[(\tilde{D}^\star_{1,3})^4]\mathbb{E}[(\tilde{D}^\star_{2,3})^4]} = O(\rho_n) \tag{D.148}$$

follows from Cauchy-Schwarz and Assumption 5.2. Consequently, we find that

$$\mathrm{Var}\left(\frac{1}{n^3}\sum_{i=1}^n\sum_{j\ne i}\sum_{k\notin\{i,j\}} \tilde{D}^\star_{i,k}\tilde{D}^\star_{j,k}(W^\star_k)^2\right) = O(n^{-2}\rho_n) \tag{D.149}$$

and so Chebyshev's inequality gives

$$\frac{1}{n^3}\sum_{i=1}^n\sum_{j\ne i}\sum_{k\notin\{i,j\}} \tilde{D}^\star_{i,k}\tilde{D}^\star_{j,k}(W^\star_k)^2 = O_p(n^{-1}\rho_n^{1/2})\ , \tag{D.150}$$

which verifies the bound (D.19).

Next, we consider the term (D.139). In this case, observe that both

$$\mathbb{E}[\tilde{D}^\star_{1,2}\tilde{D}^\star_{2,3}W^\star_2W^\star_3 \mid Z_i] = 0 \quad \text{and} \quad \mathbb{E}[\tilde{D}^\star_{1,2}\tilde{D}^\star_{2,3}W^\star_2W^\star_3 \mid Z_i, Z_j] = 0 \tag{D.151}$$

hold for each i and j in $\{1,2,3\}$, by the same reasoning as before. Thus, Part (ii) of Lemma D.2 again implies

$$\mathrm{Var}\left(\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n} h_2(Z_1,Z_2,Z_3)\right) \lesssim n^{-3}\,\mathrm{Var}(h_2(Z_1,Z_2,Z_3))\ . \tag{D.152}$$

We can evaluate

$$\begin{aligned}\mathrm{Var}(h_2(X_1,X_2,X_3)) &\lesssim \mathbb{E}[(\tilde{D}^\star_{1,2}\tilde{D}^\star{}_{2,3})^2(W^\star_2)^2(W^\star_3)^2] \\ &\lesssim \sqrt{\mathbb{E}[(\tilde{D}^\star_{1,2})^4]\mathbb{E}[(\tilde{D}^\star_{2,3})^4]} = O(\rho_n)\end{aligned} \tag{D.153}$$

by Cauchy=Schwarz and Assumption 5.2. Thus, by plugging (D.153) into (D.152) and applying Chebyshev's inequality, we find that

$$\frac{1}{n^3}\sum_{i=1}^n\sum_{j\ne i}\sum_{l\notin\{j,k\}} \tilde{D}^\star_{i,j}\tilde{D}^\star_{j,l}W^\star_jW^\star_l = O_p(n^{3/2}\rho_n^{1/2})\ , \tag{D.154}$$

which verifies the bound (D.20).

Finally, we consider the term (D.140). Here, we have that

$$\mathbb{E}[\tilde{D}^\star_{1,3}\tilde{D}^\star_{2,4}W^\star_3 W^\star_4 \mid Z_i, Z_j, Z_k] = 0 \tag{D.155}$$

holds for any i, j, and k in $\{1,2,3,4\}$. Thus, Part (ii) of Lemma D.2 implies that

$$\operatorname{Var}\left(\frac{1}{\binom{n}{4}} \sum_{1\le i<j<k<l\le n} h_2(Z_1, Z_2, Z_3, Z_4)\right) \lesssim n^{-4} \operatorname{Var}(h_3(Z_1, Z_2, Z_3, Z_4)) \,. \tag{D.156}$$

We can again evaluate

$$\begin{aligned}\operatorname{Var}(h_3(Z_1, Z_2, Z_3, XZ4)) &\lesssim \mathbb{E}[(\tilde{D}^\star_{1,3}\tilde{D}^\star_{2,4})^2 (W^\star_3)^2 (W^\star_4)^2] \\ &\lesssim \sqrt{\mathbb{E}[(\tilde{D}^\star_{1,3})^4]\mathbb{E}[(\tilde{D}^\star_{2,4})^4]} = O(\rho_n) \,. \end{aligned}\tag{D.157}$$

Hence, by plugging (D.157) into (D.156), and applying the representation (D.140) and Chebyshev's inequality, we find that

$$\frac{1}{n^3}\sum_{i=1}^n \sum_{j\ne i} \sum_{k\notin\{i,j\}} \sum_{l\notin\{i,j,l\}} \tilde{D}^\star_{i,k}\tilde{D}^\star_{j,l} W^\star_k W^\star_l = O(n^{-1}\rho_n^{1/2}) \,, \tag{D.158}$$

which verifies the bound (D.21) and completes the proof. ∎

D.8 Proof of Lemma D.7

We continue to use the notation introduced in the proofs of Lemma C.3 and Lemma D.5. Moreover, we adopt the short-hand

$$F_i(Z) = F_i(Z_j, Z_{-j}) = F(Z_i, Z_{-i}) \quad \text{and} \quad \overline{F}(Z_i, Z_j) = \mathbb{E}[Y_i \mid Z_i, Z_j] \tag{D.159}$$

throughout.

D.8.1 Part (i). The result is obtained by applying the following bound

Lemma D.11. *If the conditions of Theorem 5.1 hold, then* $\mathbb{E}[R_i^2] = O_p(\rho_n)$.

In particular, Cauchy-Schwarz implies that

$$\left|\frac{1}{n}\sum_{i=1}^n \left(\frac{1}{n}\sum_{j\ne i}\tilde{D}^\star_{i,j}W^\star_j\right) R_i\right| \le \frac{1}{\sqrt{n}}\sqrt{\left(\frac{1}{n^2}\sum_{i=1}^n\left(\sum_{j\ne i}\tilde{D}^\star_{i,j}W^\star_j\right)^2\right)\left(\frac{1}{n}\sum_{i=1}^n R_i^2\right)} \,. \tag{D.160}$$

In turn, Lemma D.11 and Markov's inequality imply that

$$\frac{1}{n}\sum_{i=1}^n R_i^2 = O_p(\rho_n) \tag{D.161}$$

and Lemma D.5 and Assumption 5.2 imply that

$$\frac{1}{n^2}\sum_{i=1}^n\left(\sum_{j\ne i}\tilde{D}^\star_{i,j}W^\star_j\right)^2 = \mathbb{E}[\operatorname{Var}(D_{i,j} \mid H_{i,j}, S_j)\operatorname{Var}(W_j \mid \overline{S}_j)] + O_p(\rho_n^{3/2}) = O_p(\rho_n) \,. \tag{D.162}$$

Consequently, the bound (D.160) implies that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j\neq i}\tilde{D}^{\star}_{i,j}W^{\star}_{j}\right)R_i = O_p(n^{-1/2}\rho_n) \tag{D.163}$$

as required. ∎

D.8.2 Part (ii). Observe that, by writing

$$f_1(Z_1,Z_2) = \sum_{\pi\in\Pi_2}\overline{F}(Z_{\pi(1)},Z_{\pi(2)})\tilde{D}^{\star}(S_{\pi(1)},S_{\pi(2)})W^{\star}_{\pi(2)}\ , \tag{D.164}$$

the quantity

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\overline{F}(Z_i,Z_j)\tilde{D}^{\star}_{i,j}W^{\star}_{j} = \frac{n-1}{2n}\left(\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j>i}f_1(Z_i,Z_j)\right) \tag{D.165}$$

can be recognized as scaled U-statistic of order 2. Thus, Part (ii) of Lemma D.2 implies that

$$\begin{aligned}&\mathrm{Var}\left(\frac{n-1}{2n}\left(\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j>i}f_1(Z_i,Z_j)\right)\right)\\&\lesssim n^{-1}\,\mathrm{Var}(\mathbb{E}[f_1(Z_1,Z_2)\mid Z_1]) + n^{-2}\,\mathrm{Var}(f_1(Z_1,Z_2))\ .\end{aligned} \tag{D.166}$$

We successively bound the two variances appearing in (D.166).

First, observe that

$$\begin{aligned}\mathrm{Var}(\mathbb{E}[f_1(Z_1,Z_2)\mid Z_1]) &= \mathrm{Var}(\mathbb{E}[\overline{F}(Z_1,Z_2)\tilde{D}^{\star}_{1,2}W^{\star}_{2} + \overline{F}(Z_2,Z_1)\tilde{D}^{\star}_{2,1}W^{\star}_{1}\mid Z_1])\\&\lesssim \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)\tilde{D}^{\star}_{1,2}W^{\star}_{2}\mid Z_1]^2] + \mathbb{E}[\mathbb{E}[\overline{F}(Z_2,Z_1)\tilde{D}^{\star}_{2,1}W^{\star}_{1}\mid Z_1]^2]\ ,\end{aligned} \tag{D.167}$$

by Cauchy-Schwarz. We can evaluate

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)\tilde{D}^{\star}_{1,2}W^{\star}_{2}\mid Z_1]^2]\\&\lesssim \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)D_{1,2}W^{\star}_{2}\mid Z_1]^2] + \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)\mathbb{E}[D_{1,2}\mid H_{1,2},S_2]W^{\star}_{2}\mid Z_1]^2]\\&\lesssim \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)D_{1,2}W^{\star}_{2}\mid \mathcal{A}_{2,1},Z_1]^2] + \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)\mathbb{E}[D_{1,2}\mid H_{1,2},S_2]W^{\star}_{2}\mid \mathcal{A}_{2,1},Z_1]^2]\\&+\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)\mathbb{E}[D_{1,2}\mid H_{1,2},S_2]W^{\star}_{2}\mid \tilde{\mathcal{A}}_{2,1},Z_1]^2]\\&\lesssim \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2\mid \mathcal{A}_{2,1},Z_1]] + \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2\mid \tilde{\mathcal{A}}_{2,1},Z_1]] = O(\rho_n^2)\ ,\end{aligned} \tag{D.168}$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. We can evaluate

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[\overline{F}(Z_2,Z_1)\tilde{D}^{\star}_{2,1}W^{\star}_{1}\mid Z_1]^2]\\&\lesssim \mathbb{E}[\mathbb{E}[\overline{F}(Z_2,Z_1)D_{2,1}W^{\star}_{1}\mid Z_1]^2] + \mathbb{E}[\mathbb{E}[\overline{F}(Z_2,Z_1)\mathbb{E}[D_{2,1}\mid H_{2,1},S_1]W^{\star}_{1}\mid Z_1]^2]\\&\lesssim \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_2,Z_1)D_{2,1}W^{\star}_{1}\mid \mathcal{A}_{2,1},Z_1]^2] + \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_2,Z_1)\mathbb{E}[D_{2,1}\mid H_{2,1},S_1]W^{\star}_{1}\mid \mathcal{A}_{2,1},Z_1]^2]\\&+\mathbb{E}[\mathbb{E}[\overline{F}(Z_2,Z_1)\mathbb{E}[D_{2,1}\mid H_{2,1},S_1]W^{\star}_{1}\mid \tilde{\mathcal{A}}_{2,1},Z_1]^2]\\&\lesssim \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_2,Z_1)^2\mid \mathcal{A}_{2,1},Z_1]] + \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_2,Z_1)^2\mid \tilde{\mathcal{A}}_{2,1},Z_1]] = O(\rho_n^2)\ ,\end{aligned} \tag{D.169}$$

again by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. Consequently, we find that

$$\mathrm{Var}(\mathbb{E}[f_1(Z_1,Z_2) \mid Z_1]) = O(\rho_n^2) \tag{D.170}$$

by plugging (D.168) and (D.169) into (D.167). In turn, observe that

$$\begin{aligned}
&\mathrm{Var}(f_1(Z_1,Z_2)) \\
&\lesssim \mathbb{E}[\overline{F}(Z_1,Z_2)^2(\tilde{D}^\star_{1,2})^2(W_2^\star)^2] \\
&\lesssim \mathbb{E}[\overline{F}(Z_1,Z_2)^2 D_{1,2}^2(W_2^\star)^2] \\
&+ \mathbb{E}[\overline{F}(Z_1,Z_2)^2 D_{1,2}\mathbb{E}[D_{1,2} \mid H_{1,2},S_2](W_2^\star)^2] \\
&+ \mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mathbb{E}[D_{1,2} \mid G_{1,2},S_2]^2(W_2^\star)^2] \\
&\lesssim \rho_n \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mid \mathcal{A}_{1,2},Z_2]] + \rho_n \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mid \tilde{\mathcal{A}}_{1,2},Z_2]] = O(\rho_n)\,,
\end{aligned} \tag{D.171}$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. Thus, by plugging (D.170) and (D.171) into (D.166), we obtain

$$\mathrm{Var}\left(\frac{n-1}{2n}\left(\frac{1}{\binom{n}{2}}\sum_{i=1}^n\sum_{j>i} f_1(Z_i,Z_j)\right)\right) = O(\rho_n^2 n^{-1} + \rho_n n^{-2}) \tag{D.172}$$

Hence, as

$$\mathbb{E}[\overline{F}(Z_i,Z_j)\tilde{D}^\star_{i,j}W_j^\star] = \mathbb{E}[Y_i\tilde{D}^\star_{i,j}W_j^\star] \tag{D.173}$$

we can conclude that

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j\neq i}\overline{F}(Z_i,Z_j)\tilde{D}^\star_{i,j}W_j^\star - \mathbb{E}[Y_i\tilde{D}_{i,j}W_j] = O_p(\rho_n n^{-1/2})\,, \tag{D.174}$$

by applying Chebyshev's inequality and the representation (D.165). ■

D.8.3 Part (iii). Observe that, by writing

$$f_2(Z_1,Z_2,Z_3) = \sum_{\pi\in\Pi_3}\tilde{F}(Z_{\pi(1)},Z_{\pi(3)})\tilde{D}^\star(S_{\pi(1)},S_{\pi(2)})W^\star_{\pi(2)}\,, \tag{D.175}$$

the quantity

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{F}(Z_i,Z_k)\tilde{D}^\star_{i,j}W_j^\star = \frac{(n-1)(n-2)}{6n}\left(\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n} f_2(Z_i,Z_j,Z_k)\right) \tag{D.176}$$

can be recognized as scaled U-statistic of order 3. We can evaluate

$$\mathbb{E}[\tilde{F}(Z_i,X_k)\tilde{D}^\star_{i,j}W_j^\star \mid Z_i] = 0 \quad\text{and}\quad \mathbb{E}[\tilde{F}(Z_i,X_k)\tilde{D}^\star_{i,j}W_j^\star \mid Z_k] = 0\,, \tag{D.177}$$

by Assumption 4.2. Moreover, it holds that

$$\begin{aligned}
\mathbb{E}[\tilde{F}(Z_i,Z_k)\tilde{D}^\star_{i,j}W_j^\star \mid Z_j] &= \mathbb{E}[\mathbb{E}[\tilde{F}(Z_i,Z_k) \mid Z_i,Z_j]\tilde{D}^\star_{i,j} \mid Z_j]W_j^\star \\
&= \mathbb{E}[\mathbb{E}[\tilde{F}(Z_i,Z_k) \mid Z_i]\tilde{D}^\star_{i,j} \mid Z_j]W_j^\star = 0\,,
\end{aligned} \tag{D.178}$$

by the definition (D.78). Consequently, Part (ii) of Lemma D.2 implies that

$$\begin{aligned}&\operatorname{Var}\left(\frac{(n-1)(n-2)}{6n}\left(\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n} f_2(Z_i,Z_j,Z_k)\right)\right)\\&\lesssim \operatorname{Var}(\mathbb{E}[f_2(Z_1,Z_2,Z_3)\mid Z_2,Z_3]) + n^{-1}\operatorname{Var}(f_2(Z_1,Z_2,Z_3))\ .\end{aligned} \tag{D.179}$$

We successively bound the two variances appearing in (D.179).

First, observe that

$$\begin{aligned}&\mathbb{E}[\tilde{F}(Z_1,Z_3)\tilde{D}^\star(S_1,S_2)W_2^\star \mid Z_1,Z_3] = \mathbb{E}[\tilde{D}^\star(S_1,S_2)W_2^\star \mid Z_1]\tilde{F}(Z_1,Z_3) = 0 \quad \text{and} \\ &\mathbb{E}[\tilde{F}(Z_1,Z_3)\tilde{D}^\star(S_1,S_2)W_2^\star \mid Z_1,Z_2] = \mathbb{E}[\tilde{F}(Z_1,Z_3)\mid Z_1]\tilde{D}^\star(S_1,S_2)W_2^\star = 0\ .\end{aligned} \tag{D.180}$$

Thus, we obtain the bound

$$\begin{aligned}\operatorname{Var}(\mathbb{E}[f_2(Z_1,Z_2,Z_3)\mid Z_2,Z_3]) &\lesssim \mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W_2^\star \mid Z_2,Z_3]^2] \\ &\quad + \mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)\mathbb{E}[D(S_1,S_2)\mid H_{1,2},S_2]W_2^\star \mid Z_2,Z_3]^2]\ .\end{aligned} \tag{D.181}$$

Observe that $\tilde{\mathcal{A}}_{1,3} \subseteq \tilde{B}_{2,3} \cap \mathcal{A}_{1,2}$ and that on the event $\tilde{\mathcal{A}}_{1,3}$, we have that

$$\tilde{F}(Z_1,Z_3) = O(\rho_n) \tag{D.182}$$

uniformly almost surely, by Assumption 5.2. Thus, we can evaluate

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W_2^\star \mid Z_2,Z_3]^2] \\ &\lesssim \rho_n\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W_2^\star \mid Z_2,Z_3]^2 \mid \mathcal{B}_{2,3}] \\ &+ \mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W_2^\star \mid Z_2,Z_3]^2 \mid \tilde{\mathcal{B}}_{2,3}] \\ &\lesssim \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W_2^\star \mid \mathcal{A}_{1,2},Z_2,Z_3]^2 \mid \mathcal{B}_{2,3}] \\ &+ \rho_n^2\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W_2^\star \mid \mathcal{A}_{1,2},Z_2,Z_3]^2 \mid \tilde{\mathcal{B}}_{2,3}] \\ &\lesssim \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)^2 \mid \mathcal{A}_{1,2},Z_2,Z_3] \mid \mathcal{B}_{2,3}] + O(\rho_n^3) = O(\rho_n^3)\ ,\end{aligned} \tag{D.183}$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. Similarly, we can evaluate

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)\mathbb{E}[D(S_1,S_2)\mid H_{1,2},S_2]W_2^\star \mid Z_2,Z_3]^2] \\ &\lesssim \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)\mathbb{E}[D(S_1,S_2)\mid H_{1,2},S_2]W_2^\star \mid \mathcal{A}_{1,2},Z_2,Z_3]^2 \mid \mathcal{B}_{2,3}] \\ &+ \rho_n\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)\mathbb{E}[D(S_1,S_2)\mid H_{1,2},S_2]W_2^\star \mid \tilde{\mathcal{A}}_{1,2},Z_2,Z_3]^2 \mid \mathcal{B}_{2,3}] \\ &+ \rho_n^2\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)\mathbb{E}[D(S_1,S_2)\mid H_{1,2},S_2]W_2^\star \mid \mathcal{A}_{1,2},Z_2,Z_3]^2 \mid \tilde{\mathcal{B}}_{2,3}] \\ &+ \rho_n^2\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)\mathbb{E}[D(S_1,S_2)\mid H_{1,2},S_2]W_2^\star \mid \mathcal{A}_{1,3},Z_2,Z_3]^2 \mid \tilde{\mathcal{B}}_{2,3}] \\ &+ \mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)\mathbb{E}[D(S_1,S_2)\mid H_{1,2},S_2]W_2^\star \mid \tilde{\mathcal{A}}_{1,3},\tilde{\mathcal{A}}_{1,2},Z_2,Z_3]^2 \mid \tilde{\mathcal{B}}_{2,3}] \\ &\lesssim \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)^2 \mid \mathcal{A}_{1,2},Z_2,Z_3] \mid \mathcal{B}_{2,3}] + \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)^2 \mid \tilde{\mathcal{A}}_{1,2},Z_2,Z_3] \mid \mathcal{B}_{2,3}] \\ &+ \rho_n^4\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)^2 \mid \mathcal{A}_{1,3},Z_2,Z_3] \mid \tilde{\mathcal{B}}_{2,3}] + O(\rho_n^3) = O(\rho_n^3)\ ,\end{aligned} \tag{D.184}$$

by Assumption 5.2 and Assumption C.1. Consequently, we find that

$$\mathrm{Var}(\mathbb{E}[f_2(Z_1, Z_2, Z_3) \mid Z_2, Z_3]) = O(\rho_n^3) \,, \tag{D.185}$$

by plugging (D.183) and (D.184) into (D.181). In turn, we can evaluate

$$\begin{aligned}
&\mathrm{Var}(f(Z_1, Z_2, Z_3)) \\
&\lesssim \mathbb{E}[(\tilde{F}(Z_1, Z_3)\tilde{D}^\star(S_1, S_2)W_2^\star)^2] \\
&\lesssim \mathbb{E}[(\tilde{F}(Z_1, Z_3)D(S_1, S_2)W_2^\star)^2] \\
&+ \mathbb{E}[\tilde{F}(Z_1, Z_3)^2 D(S_1, S_2)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2](W_2^\star)^2] \\
&+ \mathbb{E}[(\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]W_2^\star)^2] \\
&\lesssim \rho_n^2\mathbb{E}[(\tilde{F}(Z_1, Z_3)D(S_1, S_2)W_2^\star)^2 \mid \mathcal{A}_{1,3}, \mathcal{A}_{1,2}] \\
&+ \rho_n\mathbb{E}[(\tilde{F}(Z_1, Z_3)D(S_1, S_2)W_2^\star)^2 \mid \tilde{\mathcal{A}}_{1,3}, \mathcal{A}_{1,2}] \\
&+ \rho_n^2\mathbb{E}[\overline{F}^{(1)}(Z_1, Z_3)^2 D(S_1, S_2)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2](W_2^\star)^2 \mid \mathcal{A}_{1,3}, \mathcal{A}_{1,2}] \\
&+ \rho_n\mathbb{E}[\overline{F}^{(1)}(Z_1, Z_3)^2 D(S_1, S_2)\mathbb{E}[D(S_1, S_2| \mid H_{1,2}, S_2](W_2^\star)^2 \mid \tilde{\mathcal{A}}_{1,3}, \mathcal{A}_{1,2}] \\
&+ \rho_n^2\mathbb{E}[(\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]W_2^\star)^2 \mid \mathcal{A}_{1,3}, \mathcal{A}_{1,2}] \\
&+ \rho_n\mathbb{E}[(\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]W_2^\star)^2 \mid \tilde{\mathcal{A}}_{1,3}, \mathcal{A}_{1,2}] \\
&+ \mathbb{E}[(\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]W_2^\star)^2 \mid \tilde{\mathcal{A}}_{1,2}] \\
&\lesssim \rho_n^2\mathbb{E}[\tilde{F}(Z_1, Z_3)^2 \mid \mathcal{A}_{1,3}, \mathcal{A}_{1,2}] + \rho_n^3\mathbb{E}[\tilde{F}(Z_1, Z_3)^2 \mid \mathcal{A}_{1,3}, \tilde{\mathcal{A}}_{1,2}] + O(\rho_n^2) = O(\rho_n^2) \,,
\end{aligned} \tag{D.186}$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. By plugging (D.184) and (D.186) into (D.179), we find that

$$\mathrm{Var}\left(\frac{(n-1)(n-2)}{6n}\left(\frac{1}{\binom{n}{3}}\sum_{1 \le i < j < k \le n} f_2(Z_i, Z_j, Z_k)\right)\right) = O(\rho_n^3) \,. \tag{D.187}$$

Hence, Chebyshev's inequality and the representation (D.176) imply that

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j \neq i}\sum_{k \notin \{i,j\}} \tilde{F}(Z_i, Z_k)\tilde{D}_{i,j}^\star W_j^\star = O_p(\rho_n^{3/2}) \,, \tag{D.188}$$

as required. ■

D.9 Proof of Lemma D.8

We continue to use the notation introduced in the proofs of Lemmas C.3, D.5, and D.7.

D.9.1 Part (i). Each bound follows from an application of Lemma D.11. In particular, we can evaluate

$$\left|\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j \neq i}\sum_{k \notin \{i,j\}} R_i\tilde{D}_{k,j}^\star W_j^\star\right| \le \frac{1}{n}\sqrt{\left(\frac{1}{n^3}\sum_{i=1}^{n}\left(\sum_{j \neq i}\sum_{k \notin \{i,j\}} \tilde{D}_{k,j}^\star W_j^\star\right)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} R_i^2\right)} \,, \tag{D.189}$$

$$\left|\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}} R_i D^{\star}_{k,j} W^{\star}_j\right| \leq \frac{1}{n^{3/2}}\sqrt{\left(\frac{1}{n^4}\sum_{i=1}^{n}\left(\sum_{j\neq i}\sum_{k\notin\{i,j\}} D^{\star}_{k,j}W^{\star}_j\right)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}R_i^2\right)}\,,$$

$$\left|\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i} R_i D^{\star}_{j,i} W^{\star}_i\right| \leq \frac{1}{n}\sqrt{\left(\frac{1}{n^3}\sum_{i=1}^{n}\left(\sum_{j\neq i} D^{\star}_{j,i}W^{\star}_i\right)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}R_i^2\right)}\,, \quad \text{and}$$

$$\left|\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i} R_i D^{\star}_{i,j} W^{\star}_j\right| \leq \frac{1}{n^{5/2}}\sqrt{\left(\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j\neq i} D^{\star}_{i,j}W^{\star}_j\right)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}R_i^2\right)}$$

by Cauchy-Schwarz. Recall that

$$\frac{1}{n}\sum_{i=1}^{n} R_i^2 = O_p(\rho_n) \tag{D.190}$$

by Markov's inequality and Lemma D.11. In turn, we can evaluate

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n^3}\sum_{i=1}^{n}\left(\sum_{j\neq i}\sum_{k\notin\{i,j\}} \tilde{D}^{\star}_{k,j}W^{\star}_j\right)^2\right] &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{j'\neq i}\sum_{k\notin\{i,j\}}\sum_{k'\notin\{i,j\}} \mathbb{E}[\tilde{D}^{\star}_{k,j}\tilde{D}^{\star}_{k',j'}W^{\star}_j W^{\star}_{j'}] \\
&= \frac{1}{n^2}\sum_{k\notin\{i,j\}}\sum_{k'\notin\{i,j\}} \mathbb{E}[\tilde{D}^{\star}_{k,j}\tilde{D}^{\star}_{k',j}(W^{\star}_j)^2] = O(\rho_n)\,, \qquad \text{(D.191)} \\
\mathbb{E}\left[\frac{1}{n^4}\sum_{i=1}^{n}\left(\sum_{j\neq i}\sum_{k\notin\{i,j\}} D^{\star}_{k,j}W^{\star}_j\right)^2\right] &= \frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{j'\neq i}\sum_{k\notin\{i,j\}}\sum_{k'\notin\{i,j\}} \mathbb{E}[D^{\star}_{k,j}D^{\star}_{k',j'}W^{\star}_j W^{\star}_{j'}] \\
&= \frac{1}{n^2}\sum_{k\notin\{i,j\}}\sum_{k'\notin\{i,j\}} \mathbb{E}[D^{\star}_{k,j}D^{\star}_{k',j}(W^{\star}_j)^2] = O(\rho_n)\,, \\
\mathbb{E}\left[\frac{1}{n^3}\sum_{i=1}^{n}\left(\sum_{j\neq i} D^{\star}_{j,i}W^{\star}_i\right)^2\right] &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{j'\neq i}\mathbb{E}[D^{\star}_{j,i}D^{\star}_{j',i}(W^{\star}_i)^2] = O(\rho_n)\,, \quad \text{and} \\
\mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j\neq i} D^{\star}_{i,j}W^{\star}_j\right)^2\right] &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}[(D^{\star}_{i,j})^2(W^{\star}_j)^2] = O(\rho_n)\,.
\end{aligned}$$

Thus, the desired bounds follow by applying Markov's inequality, using the bounds (D.191), and by plugging the resultant bounds and (D.190) into the inequalities (D.189). ■

D.9.2 Part (ii). We continue to use the notation adopted in the proof of Lemma D.7. Observe that we can write

$$\frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}} \overline{F}(Z_i,Z_j)\tilde{D}^{\star}_{k,j}W^{\star}_j = \frac{\binom{n}{3}}{n^3}\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n} f_1(Z_i,Z_j,Z_k)\,, \tag{D.192}$$

$$\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}} \overline{F}(Z_i,Z_j)D^{\star}_{k,j}W^{\star}_j = \frac{\binom{n}{3}}{n^4}\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n} f_3(Z_i,Z_j,Z_k)\,, \tag{D.193}$$

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\overline{F}(Z_i,Z_j)D^\star_{j,i}W^\star_i = \frac{\binom{n}{2}}{n^3}\frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n} f_3(Z_i,Z_j)\ , \quad \text{and} \tag{D.194}$$

$$\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\overline{F}(Z_i,Z_j)D^\star_{i,j}W^\star_j = \frac{\binom{n}{2}}{n^4}\frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n} f_4(Z_i,Z_j) \tag{D.195}$$

where

$$\begin{aligned}
f_1(Z_i,Z_j,Z_k) &= \sum_{\pi\in\Pi_3}\overline{F}(Z_{\pi(1)},Z_{\pi(2)})\tilde{D}^\star(S_{\pi(3)},S_{\pi(2)})W^\star_{\pi(2)}\ ,\\
f_2(Z_i,Z_j,Z_k) &= \sum_{\pi\in\Pi_3}\overline{F}(Z_{\pi(1)},Z_{\pi(2)})D^\star(S_{\pi(3)},S_{\pi(2)})W^\star_{\pi(2)}\ ,\\
f_3(Z_i,Z_j) &= \sum_{\pi\in\Pi_2}\overline{F}(Z_{\pi(1)},Z_{\pi(2)})D^\star(S_{\pi(2)},S_{\pi(1)})W^\star_{\pi(1)}\ , \quad \text{and}\\
f_4(Z_i,Z_j) &= \sum_{\pi\in\Pi_2}\overline{F}(Z_{\pi(1)},Z_{\pi(2)})D^\star(S_{\pi(1)},S_{\pi(2)})W^\star_{\pi(2)}\ .
\end{aligned} \tag{D.196}$$

We successively apply Lemma D.2 to bound the variance of each term.

We begin by considering the term (D.192). Lemma D.2 and the law of total variance imply that

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n} f_1(Z_i,Z_j,Z_k)\right)\\
&\lesssim n^{-1}\,\mathrm{Var}(\mathbb{E}[f_1(Z_1,Z_2,Z_3)\mid Z_1]) + n^{-2}\,\mathrm{Var}(f_1(Z_1,Z_2,Z_3))\ .
\end{aligned} \tag{D.197}$$

Observe that

$$\mathbb{E}[\overline{F}(Z_i,Z_j)\tilde{D}^\star_{k,j}W^\star_j \mid Z_i] = 0 \quad \text{and} \quad \mathbb{E}[\overline{F}(Z_i,Z_j)\tilde{D}^\star_{k,j}W^\star_j \mid Z_j] = 0 \tag{D.198}$$

respectively. We can evaluate

$$\begin{aligned}
&\mathrm{Var}(\mathbb{E}[f_1(Z_1,Z_2,Z_3)\mid Z_3])\\
&\lesssim \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)\tilde{D}^\star(S_3,S_2)W^\star_2 \mid Z_3]^2]\\
&\lesssim \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)D(S_3,S_2)W^\star_2 \mid Z_3]^2]\\
&+ \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)\mathbb{E}[D(S_3,S_2)\mid H_{3,2},S_2]W^\star_2 \mid Z_3]^2]\\
&\lesssim \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)D(S_3,S_2)W^\star_2 \mid \mathcal{A}_{3,2},Z_3]^2]\\
&+ \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)\mathbb{E}[D(S_3,S_2)\mid H_{3,2},S_2]W^\star_2 \mid \mathcal{A}_{3,2},Z_3]^2]\\
&+ \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)\mathbb{E}[D(S_3,S_2)\mid H_{3,2},S_2]W^\star_2 \mid \tilde{\mathcal{A}}_{3,2},Z_3]^2]\\
&\lesssim \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mid \mathcal{A}_{3,2},Z_3]] + \rho_n^2\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mid \tilde{\mathcal{A}}_{3,2},Z_3]] = O(\rho_n^2)
\end{aligned} \tag{D.199}$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. Likewise, we can evaluate

$$\begin{aligned}
&\mathrm{Var}(f_1(Z_1,Z_2,Z_3))\\
&\lesssim \mathbb{E}[\overline{F}(Z_1,Z_2)^2\tilde{D}^\star(S_3,S_2)^2(W^\star_2)^2]
\end{aligned}$$

$$
\begin{aligned}
&\lesssim \mathbb{E}[\overline{F}(Z_1,Z_2)^2 D(S_3,S_2)^2 (W_2^\star)^2] \\
&+ \mathbb{E}[\overline{F}(Z_1,Z_2)^2 D(S_3,S_2)\mathbb{E}[D(S_3,S_2) \mid H_{3,2},S_2](W_2^\star)^2] \\
&+ \mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mathbb{E}[D(S_3,S_2) \mid H_{3,2},S_2]^2 (W_2^\star)^2] \\
&\lesssim \rho_n \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mid \mathcal{A}_{2,3},Z_3]] + \rho_n^2 \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mid \tilde{\mathcal{A}}_{2,3},Z_3]] = O(\rho_n) \, , \qquad \text{(D.200)}
\end{aligned}
$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. Thus, by plugging (D.199) and (D.200) into the bound (D.197), and applying Chebyshev's inequality, we obtain

$$
\frac{1}{n^3} \sum_{i=1} \sum_{j \neq i} \sum_{k \notin \{i,j\}} \overline{F}(Z_i,Z_j) \tilde{D}_{k,j}^\star W_j^\star = O_p(n^{-1/2}\rho_n) \, , \qquad \text{(D.201)}
$$

verifying the first representation in (D.92).

Next, we consider the terms (D.193) through (D.195). Lemma D.2 and the law of total variance imply that

$$
\mathrm{Var}\left(\frac{1}{\binom{n}{3}} \sum_{1 \le i < j < k \le n} f_2(Z_i,Z_j,Z_k) \right) \lesssim n^{-1} \mathrm{Var}(f_1(Z_1,Z_2,Z_3)) \, , \qquad \text{(D.202)}
$$

$$
\mathrm{Var}\left(\frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} f_3(Z_i,Z_j) \right) \lesssim n^{-1} \mathrm{Var}(f_3(Z_1,Z_2)) \, , \quad \text{and} \qquad \text{(D.203)}
$$

$$
\mathrm{Var}\left(\frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} f_4(Z_i,Z_j) \right) \lesssim n^{-1} \mathrm{Var}(f_4(Z_1,Z_2)) \, , \qquad \text{(D.204)}
$$

respectively. We can evaluate

$$
\begin{aligned}
&\mathrm{Var}(f_1(Z_1,Z_2,Z_3) \\
&\lesssim \mathbb{E}[\overline{F}(Z_1,Z_2)^2 D^\star(S_3,S_2)^2 (W_2^\star)^2] \\
&\lesssim \mathbb{E}[\overline{F}(Z_1,Z_2)^2 D(S_3,S_2)^2 (W_2^\star)^2] \\
&+ \mathbb{E}[\overline{F}(Z_1,Z_2)^2 D(S_3,S_2)\mathbb{E}[D(S_3,S_2) \mid G_{3,2}](W_2^\star)^2] \\
&+ \mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mathbb{E}[D(S_3,S_2) \mid G_{3,2}]^2 (W_2^\star)^2] \\
&\lesssim \rho_n \mathbb{E}[\mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mid \mathcal{A}_{2,3},Z_3]] \\
&+ \rho_n \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mid \mathcal{A}_{1,2},Z_2]] + \mathbb{E}[\mathbb{E}[\overline{F}(Z_1,Z_2)^2 \mid \tilde{\mathcal{A}}_{1,2},Z_2]] = O(\rho_n) \, , \qquad \text{(D.205)}
\end{aligned}
$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. The bounds

$$
\mathrm{Var}(f_3(Z_1,Z_2)) = O(\rho_n) \quad \text{and} \quad \mathrm{Var}(f_4(Z_1,Z_2)) = O(\rho_n) \, , \qquad \text{(D.206)}
$$

follow from identical arguments. Thus, Chebyshev's inequality implies that

$$
\frac{1}{n^4} \sum_{i=1} \sum_{j \neq i} \sum_{k \notin \{i,j\}} \overline{F}(Z_i,Z_j) D_{k,j}^\star W_j^\star = \frac{1}{n} \mathbb{E}[Y_i D_{l,q}^\star W_q^\star] + O_p(n^{-3/2}\rho_n^{1/2}) \, , \qquad \text{(D.207)}
$$

$$
\frac{1}{n^3} \sum_{i=1}^{n} \sum_{j \neq i} \overline{F}(Z_i,Z_j) D_{j,i}^\star W_i^\star = \frac{1}{n} \mathbb{E}[Y_i D_{l,i}^\star W_i^\star] + O_p(n^{-3/2}\rho_n^{1/2}) \, , \quad \text{and} \qquad \text{(D.208)}
$$

$$\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\overline{F}(Z_i,Z_j)D^{\star}_{j,i}W^{\star}_i = \frac{1}{n^2}\mathbb{E}[Y_iD^{\star}_{i,l}W^{\star}_l] + O_p(n^{-5/2}\rho_n^{1/2})\ , \tag{D.209}$$

verifying the second though fourth representations in (D.92) and completing the proof. ∎

D.9.3 Part (iii). Observe that each of the terms on the left-hand side of (D.93) has expectation equal to zero. Thus, it will suffice to bound their variances and apply Chebyshev's inequality. We begin by considering the term

$$\frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)\tilde{D}^{\star}_{k,j}W^{\star}_j$$
$$= \frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{F}(Z_i,Z_k)\tilde{D}^{\star}_{k,j}W^{\star}_j + \frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j,k\}}\tilde{F}(Z_i,Z_l)\tilde{D}^{\star}_{k,j}W^{\star}_j\ . \tag{D.210}$$

Observe that we can write

$$\frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{F}(Z_i,Z_k)\tilde{D}^{\star}_{k,j}W^{\star}_j = \frac{\binom{n}{3}}{n^3}\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n}g_1(Z_i,Z_j,Z_k) \quad\text{and} \tag{D.211}$$

$$\frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j,k\}}\tilde{F}(Z_i,Z_l)\tilde{D}^{\star}_{k,j}W^{\star}_j = \frac{\binom{n}{4}}{n^3}\frac{1}{\binom{n}{4}}\sum_{1\leq i<j<k<l\leq n}g_2(Z_i,Z_j,Z_k,Z_l)\ , \tag{D.212}$$

where

$$g_1(Z_i,Z_j,Z_k) = \sum_{\pi\in\Pi_3}\tilde{F}(Z_{\pi(1)},Z_{\pi(3)})\tilde{D}^{\star}(S_{\pi(3)},S_{\pi(2)})W^{\star}_{\pi(2)} \quad\text{and} \tag{D.213}$$
$$g_2(Z_i,Z_j,Z_k,Z_l) = \sum_{\pi\in\Pi_4}\tilde{F}(Z_{\pi(1)},Z_{\pi(4)})\tilde{D}^{\star}(S_{\pi(3)},S_{\pi(2)})W^{\star}_{\pi(2)}\ ,$$

respectively.

First, we evaluate the term (D.211). Observe that Lemma D.2 and the law of total variance imply

$$\operatorname{Var}\left(\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n}g_1(Z_i,Z_j,Z_k)\right)$$
$$\lesssim n^{-1}\operatorname{Var}(\mathbb{E}[g_1(Z_i,Z_j,Z_k)\mid Z_i]) + n^{-2}\operatorname{Var}(g_1(Z_i,Z_j,Z_k))\ . \tag{D.214}$$

Observe that

$$\mathbb{E}[\tilde{F}(Z_i,Z_k)\tilde{D}^{\star}_{k,j}W^{\star}_j\mid Z_i] = 0 \quad\text{and}\quad \mathbb{E}[\tilde{F}(Z_i,Z_k)\tilde{D}^{\star}_{k,j}W^{\star}_j\mid Z_k] = 0\ , \tag{D.215}$$

respectively. Thus, we can evaluate

$$\operatorname{Var}(\mathbb{E}[g_1(Z_i,Z_j,Z_k)\mid Z_i])$$
$$\lesssim \mathbb{E}[\mathbb{E}[\tilde{F}(Z_i,Z_k)\tilde{D}^{\star}_{k,j}W^{\star}_j\mid Z_j]^2]$$
$$\lesssim \mathbb{E}[\mathbb{E}[\tilde{F}(Z_i,Z_k)D_{k,j}W^{\star}_j\mid Z_j]^2] + \mathbb{E}[\mathbb{E}[\tilde{F}(Z_i,Z_k)\mathbb{E}[D_{k,j}\mid G_{k,j},S_j]W^{\star}_j\mid Z_j]^2]$$
$$\lesssim \rho_n^2\mathbb{E}[\mathbb{E}[\tilde{F}(Z_i,Z_k)^2\mid \mathcal{A}_{k,j},Z_j]] + \rho_n^2\mathbb{E}[\mathbb{E}[\tilde{F}(Z_i,Z_k)^2\mid \tilde{\mathcal{A}}_{k,j},Z_j]] = O(\rho_n^2) \tag{D.216}$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. Similarly, we can evaluate

$$\begin{aligned}\operatorname{Var}(g_1(Z_i,Z_j,Z_k)) &\lesssim \rho_n \mathbb{E}[(\tilde{F}(Z_i,Z_k)\tilde{D}^\star_{k,j}W^\star_j)^2 \mid \mathcal{A}_{i,k}] \\ &\quad + \mathbb{E}[(\tilde{F}(Z_i,Z_k)\tilde{D}^\star_{k,j}W^\star_j)^2 \mid \tilde{\mathcal{A}}_{i,k}] \\ &\lesssim \rho_n \mathbb{E}[\tilde{F}(Z_i,Z_k)^2 \mid \mathcal{A}_{i,k}] + O(\rho_n) = O(\rho_n) \end{aligned} \tag{D.217}$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. Consequently, Chebyshev's inequality implies

$$\frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}} \tilde{F}(Z_i,Z_k)\tilde{D}^\star_{k,j}W^\star_j = O_p(n^{-1/2}\rho_n)\,. \tag{D.218}$$

Second, we evaluate the term (D.212). Observe that

$$\mathbb{E}[\tilde{F}(Z_1,Z_4)\tilde{D}^\star(S_3,S_2)W^\star_2 \mid Z_i,Z_j] = 0 \tag{D.219}$$

for any pair i,j in $\{1,2,3,4\}$. Thus, Lemma D.2 and the law of total variance imply that

$$\begin{aligned}&\operatorname{Var}\left(\frac{1}{\binom{n}{4}}\sum_{1\leq i<j<k<l\leq n} g_2(Z_i,Z_j,Z_k,Z_l)\right) \\ &\lesssim n^{-3}\operatorname{Var}(\mathbb{E}[g_2(Z_i,Z_j,Z_k,Z_l) \mid Z_i,Z_j,Z_k]) + n^{-4}\operatorname{Var}(g_2(Z_i,Z_j,Z_k,Z_l))\,. \end{aligned} \tag{D.220}$$

Observe that

$$\mathbb{E}[\tilde{F}(Z_1,Z_4)\tilde{D}^\star(S_3,S_2)W^\star_2 \mid Z_i,Z_j,Z_k] = 0 \tag{D.221}$$

for each collection i, j, and k in $\{1,2,3,4\}$ other than $\{2,3,4\}$. Hence, we can evaluate

$$\begin{aligned}&\operatorname{Var}(\mathbb{E}[g_2(Z_i,Z_j,Z_k,Z_l) \mid Z_i,Z_j,Z_k]) \\ &\lesssim \mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_4) \mid Z_4]^2\tilde{D}^\star_{3,2})^2(W^\star_2)^2] \\ &\lesssim \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_4) \mid \mathcal{A}_{1,4},Z_4]^2 D^2_{3,2}(W^\star_2)^2 \mid \mathcal{A}_{2,3}] \\ &+ \rho_n^2\mathbb{E}[D^2_{3,2}(W^\star_2)^2 \mid \mathcal{A}_{2,3}] \\ &+ \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_4) \mid \mathcal{A}_{1,4},Z_4]^2 D_{3,2}\mathbb{E}[D_{3,2} \mid H_{3,2},S_2](W^\star_2)^2 \mid \mathcal{A}_{2,3}] \\ &+ \rho_n^2\mathbb{E}[\mathbb{E}[D_{3,2}\mathbb{E}[D_{3,2} \mid H_{3,2},S_2](W^\star_2)^2 \mid \mathcal{A}_{2,3}] \\ &+ \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_4) \mid \mathcal{A}_{1,4},Z_4]^2\mathbb{E}[D_{3,2} \mid H_{3,2},S_2]^2(W^\star_2)^2 \mid \mathcal{A}_{2,3}] \\ &+ \rho_n^2\mathbb{E}[\mathbb{E}[\mathbb{E}[D_{3,2} \mid H_{3,2},S_2]^2(W^\star_2)^2 \mid \mathcal{A}_{2,3}] \\ &+ \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_4) \mid Z_4]^2\mathbb{E}[D_{3,2} \mid H_{3,2},S_2]^2(W^\star_2)^2 \mid \tilde{\mathcal{A}}_{2,3}] \\ &\lesssim \rho_n^3\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_4)^2 \mid \mathcal{A}_{1,4},Z_4]] + O(\rho_n^2) = O(\rho_n^2) \end{aligned} \tag{D.222}$$

by Cauchy-Schwarz, Assumption 5.2, Assumption C.1. Likewise, we can evaluate

$$\begin{aligned}&\operatorname{Var}(g_2(Z_i,Z_j,Z_k,Z_l)) \\ &\lesssim \rho_n\mathbb{E}[\tilde{F}(Z_1,Z_4)^2\tilde{D}^\star(S_3,S_2)^2(W^\star_2)^2 \mid \mathcal{A}_{1,4}] + O(\rho_n) = O(\rho_n) \end{aligned} \tag{D.223}$$

by Assumption 5.2 and Assumption C.1. Consequently, Chebyshev's inequality implies that

$$\frac{1}{n^3}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j,k\}}\tilde{F}(Z_i,Z_l)\tilde{D}^{\star}_{k,j}W^{\star}_j = O_p(n^{-1/2}\rho_n)\ . \tag{D.224}$$

Hence, the first representation in (D.93) follows by plugging (D.218) and (D.224) into (D.210).

Next, we consider the term

$$\begin{aligned}&\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j\}}\tilde{F}(Z_i,Z_l)D^{\star}_{k,j}W^{\star}_j\\ &=\frac{1}{n^4}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{F}(Z_i,Z_k)D^{\star}_{k,j}W^{\star}_j+\frac{1}{n^4}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j,k\}}\tilde{F}(Z_i,Z_l)D^{\star}_{k,j}W^{\star}_j\ .\end{aligned} \tag{D.225}$$

As before, we can write

$$\frac{1}{n^4}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{F}(Z_i,Z_k)D^{\star}_{k,j}W^{\star}_j=\frac{\binom{n}{3}}{n^4}\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n}\overline{g}_1(Z_i,Z_j,Z_k)\quad\text{and} \tag{D.226}$$

$$\frac{1}{n^4}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\sum_{l\notin\{i,j,k\}}\tilde{F}(Z_i,Z_l)D^{\star}_{k,j}W^{\star}_j=\frac{\binom{n}{4}}{n^4}\frac{1}{\binom{n}{4}}\sum_{1\leq i<j<k<l\leq n}\overline{g}_2(Z_i,Z_j,Z_k,Z_l)\ , \tag{D.227}$$

where

$$\overline{g}_1(Z_i,Z_j,Z_k)=\sum_{\pi\in\Pi_3}\tilde{F}(Z_{\pi(1)},Z_{\pi(3)})D^{\star}(S_{\pi(3)},S_{\pi(2)})W^{\star}_{\pi(2)}\quad\text{and} \tag{D.228}$$

$$\overline{g}_2(Z_i,Z_j,Z_k,Z_l)=\sum_{\pi\in\Pi_4}\tilde{F}(Z_{\pi(1)},Z_{\pi(4)})D^{\star}(S_{\pi(3)},S_{\pi(2)})W^{\star}_{\pi(2)}\ ,$$

respectively. First, we consider the term (D.226). Lemma D.2 and the law of total variance imply that

$$\operatorname{Var}\left(\frac{1}{\binom{n}{3}}\sum_{1\leq i<j<k\leq n}\overline{g}_1(Z_i,Z_j,Z_k)\right)\lesssim n^{-1}\operatorname{Var}(\overline{g}_1(Z_i,Z_j,Z_k))\ . \tag{D.229}$$

We can evaluate

$$\operatorname{Var}(\overline{g}_1(Z_i,Z_j,Z_k))\lesssim\rho_n\mathbb{E}[\tilde{F}(Z_1,Z_3)^2D^{\star}(S_3,S_2)^2(W^{\star}_2)^2\mid\mathcal{A}_{1,3}]+O(\rho_n)=O(\rho_n) \tag{D.230}$$

by Assumption 5.2 and Assumption C.1. Thus, Chebyshev's inequality and (D.226) imply that

$$\frac{1}{n^4}\sum_{i=1}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\tilde{F}(Z_i,Z_k)D^{\star}_{k,j}W^{\star}_j=O_p(n^{-1}\rho_n^{-1/2})\ . \tag{D.231}$$

Second, we consider the term (D.227). Observe that

$$\mathbb{E}[\tilde{F}(Z_1,Z_4)D^{\star}(S_3,S_2)W_2\mid X_i]=0 \tag{D.232}$$

for each i in $\{1,2,3,4\}$. Thus, Lemma D.2 and the law of total variance imply that

$$\operatorname{Var}\left(\frac{1}{\binom{n}{4}}\sum_{1\leq i<j<k<l\leq n}\overline{g}_2(Z_i,Z_j,Z_k,Z_l)\right)\lesssim n^{-2}\operatorname{Var}(\overline{g}_2(Z_i,Z_j,Z_k,Z_l))\ . \tag{D.233}$$

We can evaluate

$$\operatorname{Var}(\overline{g}_2(Z_i, Z_j, Z_k, Z_l)) \lesssim \rho_n \mathbb{E}[\tilde{F}(Z_1, Z_4)^2 D^\star(S_3, S_2)^2 (W_2^\star)^2 \mid \mathcal{A}_{1,4}] + O(\rho_n) = O(\rho_n) \tag{D.234}$$

by Assumption 5.2 and Assumption C.1. Thus, Chebyshev's inequality and (D.226) imply that

$$\frac{1}{n^4} \sum_{i=1} \sum_{j \neq i} \sum_{k \notin \{i,j\}} \sum_{l \notin \{i,j,k\}} \tilde{F}(Z_i, Z_l) D_{k,j}^\star W_j^\star = O_p(n^{-1} \rho_n^{-1/2}) \; . \tag{D.235}$$

Hence, the second representation in (D.93) follows by plugging (D.231) and (D.235) into (D.225).

Next, we consider the term

$$\frac{1}{n^3} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{l \notin \{i,j\}} \tilde{F}(Z_i, Z_l) D_{j,i}^\star W_i^\star \; . \tag{D.236}$$

Observe that we can write

$$\frac{1}{n^3} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{l \notin \{i,j\}} \tilde{F}(Z_i, Z_l) D_{j,i}^\star W_i^\star = \frac{\binom{n}{3}}{n^3} \frac{1}{\binom{n}{3}} \sum_{1 \leq i < j < k \leq n} g_3(Z_1, Z_j, Z_k) \tag{D.237}$$

where

$$g_3(Z_1, Z_j, Z_k) = \sum_{\pi \in \Pi_3} \tilde{F}(Z_{\pi(1)}, Z_{\pi(3)}) D^\star(S_{\pi(2)}, S_{\pi(1)}) W_{\pi(1)}^\star \; . \tag{D.238}$$

It holds that that

$$\mathbb{E}[\tilde{F}(Z_1, Z_3) D^\star(S_2, S_1) W_1^\star \mid X_i] = 0 \tag{D.239}$$

for each i in $\{1, 2, 3\}$. Thus, Lemma D.2 and the law of total variance imply that

$$\operatorname{Var}\left(\frac{1}{\binom{n}{3}} \sum_{1 \leq i < j < k \leq n} g_3(Z_1, Z_j, Z_k) \right) \lesssim n^{-2} \operatorname{Var}(g_3(Z_1, Z_j, Z_k)) \; . \tag{D.240}$$

We can evaluate

$$\operatorname{Var}(g_3(Z_1, Z_j, Z_k)) \lesssim \rho_n \mathbb{E}[\tilde{F}(Z_1, Z_3)^2 D^\star(S_2, S_1)^2 (W_1^\star)^2 \mid \mathcal{A}_{1,3}] + O(\rho_n) = O(\rho_n) \tag{D.241}$$

by Assumption 5.2 and Assumption C.1. Hence, Chebyshev's inequality implies that

$$\frac{1}{n^3} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{l \notin \{i,j\}} \tilde{F}(Z_i, Z_l) D_{j,i}^\star W_i^\star = O_p(n^{-1} \rho_n^{1/2}) \; , \tag{D.242}$$

verifying the third representation in (D.93).

Finally, we consider the term

$$\frac{1}{n^4} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{l \notin \{i,j\}} \tilde{F}(Z_i, Z_l) D_{i,j}^\star W_j^\star \; . \tag{D.243}$$

Observe that we can write

$$\frac{1}{n^4} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{l \notin \{i,j\}} \tilde{F}(Z_i, Z_l) D_{i,j}^\star W_j^\star = \frac{\binom{n}{3}}{n^4} \frac{1}{\binom{n}{3}} \sum_{1 \leq i < j < k \leq n} g_4(Z_1, Z_j, Z_k) \tag{D.244}$$

where

$$g_4(Z_1, Z_j, Z_k) = \sum_{\pi \in \Pi_3} \tilde{F}(Z_{\pi(1)}, Z_{\pi(3)}) \overline{D}(S_{\pi(1)}, S_{\pi(2)}) W^\star_{\pi(2)} \,. \tag{D.245}$$

Lemma D.2 and the law of total variance imply that

$$\operatorname{Var}\left(\frac{1}{\binom{n}{3}} \sum_{1 \le i < j < k \le n} g_4(Z_1, Z_j, Z_k) \right) \lesssim n^{-1} \operatorname{Var}(g_4(Z_1, Z_j, Z_k)) \,. \tag{D.246}$$

We can evaluate

$$\operatorname{Var}(g_4(Z_1, Z_j, Z_k)) \lesssim \rho_n \mathbb{E}[\tilde{F}(Z_1, Z_3)^2 D^\star(S_1, S_2)^2 (W_2^\star)^2 \mid \mathcal{A}_{1,3}] + O(\rho_n) = O(\rho_n) \tag{D.247}$$

by Assumption 5.2 and Assumption C.1. Hence, Chebyshev's inequality implies that

$$\frac{1}{n^4} \sum_{i=1}^n \sum_{j \ne i} \sum_{l \notin \{i,j\}} \tilde{F}(Z_i, Z_l) D^\star_{i,j} W^\star_j = O_p(n^{-3/2} \rho_n^{1/2}) \,, \tag{D.248}$$

verifying the fourth representation in (D.93) and completing the proof. ■

D.10 Proof of Lemma D.9

We continue to use the notation introduced in the proofs of Lemmas C.3, D.5, and D.7.

D.10.1 Part (i). Observe that

$$\left| \frac{1}{n^2} \sum_{i=1}^n R_i J_{i,s} \right| \le \frac{1}{\sqrt{n}} \sqrt{ \left(\frac{1}{n^2} \sum_{i=1}^n J_{i,s}^2 \right) \left(\frac{1}{n} \sum_{i=1}^n R_i^2 \right) } \quad \text{and} \tag{D.249}$$

$$\left| \frac{1}{n^2} \sum_{i=1}^n R_i M_{i,s,r} \right| \le \frac{1}{\sqrt{n}} \sqrt{ \left(\frac{1}{n^2} \sum_{i=1}^n M_{i,s,r}^2 \right) \left(\frac{1}{n} \sum_{i=1}^n R_i^2 \right) }$$

by Cauchy-Schwarz. Recall that

$$\frac{1}{n} \sum_{i=1}^n R_i^2 = O_p(\rho_n) \tag{D.250}$$

by Markov's inequality and Lemma D.11. Moreover, we have that

$$\frac{1}{n^2} \sum_{i=1}^n J_{i,s}^2 = O_p(\kappa_{n,s}) \quad \text{and} \quad \frac{1}{n^2} \sum_{i=1}^n M_{i,r,s}^2 = O_p(n \kappa_{n,s}) \tag{D.251}$$

by Lemma C.2. Hence, we can conclude that

$$\frac{1}{n^2} \sum_{i=1}^n R_i J_{i,s} = O_p((\rho_n \kappa_{n,s})^{1/2} n^{-1/2}) \quad \text{and} \quad \frac{1}{n^2} \sum_{i=1}^n R_i M_{i,r,s} = O_p((\rho_n \kappa_{n,s})^{1/2}) \,, \tag{D.252}$$

as required. ■

D.10.2 Part (ii). Observe that

$$\left| \frac{1}{n^2} \sum_{i=1}^n \tilde{F}(Z_i) J_{i,s} \right| \le \frac{1}{\sqrt{n}} \sqrt{ \left(\frac{1}{n^2} \sum_{i=1}^n J_{i,s}^2 \right) \left(\frac{1}{n} \sum_{i=1}^n \tilde{F}(Z_i)^2 \right) } \quad \text{and} \tag{D.253}$$

$$\left|\frac{1}{n^2}\sum_{i=1}^{n}\tilde{F}(Z_i)M_{i,s,r}\right| \leq \frac{1}{\sqrt{n}}\sqrt{\left(\frac{1}{n^2}\sum_{i=1}^{n}M_{i,s,r}^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{F}(Z_i)^2\right)}$$

by Cauchy-Schwarz. Moreover, as Assumption C.1 implies that

$$\mathbb{E}[\tilde{F}(Z_i)^2] = O(1) \;, \tag{D.254}$$

Markov's inequality implies that

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{F}(Z_i)^2 = O(1) \;. \tag{D.255}$$

Thus, by applying the bounds (D.251), we can conclude

$$\frac{1}{n^2}\sum_{i=1}^{n}\tilde{F}(Z_i)J_{i,s} = O_p(n^{-1/2}\kappa_{n,s}^{1/2}) = O_p(\kappa_{n,s}) \;. \tag{D.256}$$

and

$$\frac{1}{n^2}\sum_{i=1}^{n}\tilde{F}(Z_i)M_{i,s,r} = O_p(\kappa_{n,s}^{1/2}) = O_p(n^{1/2}\kappa_{n,s}) \;, \tag{D.257}$$

as required. ■

D.10.3 Part (iii). Observe that

$$\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{F}(Z_i,Z_j)J_{i,s}\right| \leq \sqrt{\left(\frac{1}{n^2}\sum_{i=1}^{n}J_{i,s}^2\right)\left(\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j\neq i}\tilde{F}(Z_i,Z_j)\right)^2\right)} \tag{D.258}$$

and

$$\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{F}(Z_i,Z_j)M_{i,s,r}\right| \leq \sqrt{\left(\frac{1}{n^2}\sum_{i=1}^{n}M_{i,s,r}^2\right)\left(\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j\neq i}\tilde{F}(Z_i,Z_j)\right)^2\right)} \tag{D.259}$$

by Cauchy-Schwarz. We can evaluate

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j\neq i}\tilde{F}(Z_i,Z_j)\right)^2\right] &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{j'\neq i}\mathbb{E}[\tilde{F}(Z_i,Z_j)\tilde{F}(Z_i,Z_{j'})] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}[\tilde{F}(Z_i,Z_j)^2] \\ &= \rho_n\mathbb{E}[\tilde{F}(Z_i,Z_j)^2 \mid \mathcal{A}_{i,j}] + O(\rho_n) = O(\rho_n) \;, \end{aligned} \tag{D.260}$$

where the second equality follows from the definition (D.78) and the bound follows from Assumptions 5.2 and C.1. Consequently, Markov's inequality implies that

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j\neq i}\tilde{F}(Z_i,Z_j)\right)^2 = O_p(\rho_n) \;. \tag{D.261}$$

Thus, by applying the bounds (D.251), we can conclude

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\tilde{F}(Z_i,Z_j)J_{i,s}=O_p(\rho_n^{1/2}\kappa_n^{1/2})=O_p(\kappa_{n,s})\ . \tag{D.262}$$

and

$$\frac{1}{n^2}\sum_{i=1}^{n}\tilde{F}(Z_i)M_{i,s,r}=O_p(\rho_n^{1/2}n^{1/2}\kappa_{n,s}^{1/2})=O_p(n^{1/2}\kappa_{n,s})\ , \tag{D.263}$$

as required. ■

D.11 Proof of Lemma D.10

To verify the first bound, observe that

$$\begin{aligned}\mathrm{Var}(f_\xi^{(3)}(Z_1,Z_2,Z_3)) &\leq \mathrm{Var}(f_\xi(Z_1,Z_2,Z_3))\\ &\leq \mathrm{Var}(\tilde{F}(Z_1,Z_3)\tilde{D}^\star_{1,2}W^\star_2)+\mathrm{Var}(\tilde{D}^\star\tilde{D}^\star_{1,2}W^\star_2)=O(\rho_n^2)\ ,\end{aligned} \tag{D.264}$$

where the first inequality follows from Lemma D.1, the second inequality follows from the definition of $f_\xi(Z_1,Z_2,Z_3)$, and the third inequality follows from the bounds (D.186) and (D.128) stated in Part (iii) of the Proof of Lemma D.7 and the Proof of Lemma D.5, respectively.

The second inequality follows from an argument similar to the bounds (D.134) and (D.185), respectively. In particular, observe that

$$\mathbb{E}[(f^{(2)}(Z_2,Z_2))^4]\lesssim\mathbb{E}[(\mathbb{E}[\tilde{F}(Z_1,Z_3)\tilde{D}^\star_{1,2}\mid Z_2,Z_3]W^\star_2)^4] \tag{D.265}$$

$$+\,\mathbb{E}[(\mathbb{E}[\tilde{D}^\star_{1,3}W^\star_3\mid Z_2,Z_3]W^\star_2)^4]\ . \tag{D.266}$$

We provide the details for the verification of the bound

$$\mathbb{E}[(\mathbb{E}[\tilde{F}(Z_1,Z_3)\tilde{D}^\star_{1,2}\mid Z_2,Z_3]W^\star_2)^4]=O(\rho_n^5) \tag{D.267}$$

as the obtaining the same bound for the second term in (D.266) follows from ana analogous argument, in the same sense that the bound (D.185) follows from an argument analogous to the bound (D.134). To this end, observe that

$$\begin{aligned}\mathbb{E}[(\mathbb{E}[\tilde{F}(Z_1,Z_3)\tilde{D}^\star_{1,2}\mid Z_2,Z_3]W^\star_2)^4] &\lesssim \mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W^\star_2\mid Z_2,Z_3]^4]\\ &\quad+\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)\mathbb{E}[D(S_1,S_2)\mid H_{1,2},S_2]W^\star_2\mid Z_2,Z_3]^4]\ .\end{aligned} \tag{D.268}$$

We can evaluate

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W^\star_2\mid Z_2,Z_3]^4]\\ &\lesssim \rho_n^5\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W^\star_2\mid \mathcal{A}_{1,2},Z_2,Z_3]^4\mid \mathcal{B}_{2,3}]\\ &+\rho_n^4\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)D(S_1,S_2)W^\star_2\mid \mathcal{A}_{1,2},Z_2,Z_3]^4\mid \tilde{\mathcal{B}}_{2,3}]\\ &\lesssim \rho_n^5\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1,Z_3)^4\mid \mathcal{A}_{1,2},Z_2,Z_3]\mid \mathcal{B}_{2,3}]+O(\rho_n^6)=O(\rho_n^5)\ ,\end{aligned} \tag{D.269}$$

by Cauchy-Schwarz, Assumption 5.2, and Assumption C.1. Similarly, we can evaluate

$$
\begin{aligned}
&\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid G_{1,2}, S_2]W_2^\star \mid Z_2, Z_3]^4] \\
&\lesssim \rho_n^5\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]W_2^\star \mid \mathcal{A}_{1,2}, Z_2, Z_3]^4 \mid \mathcal{B}_{2,3}] \\
&+ \rho_n\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]W_2^\star \mid \tilde{\mathcal{A}}_{1,2}, Z_2, Z_3]^4 \mid \mathcal{B}_{2,3}] \\
&+ \rho_n^4\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]W_2^\star \mid \mathcal{A}_{1,2}, Z_2, Z_3]^4 \mid \tilde{\mathcal{B}}_{2,3}] \\
&+ \rho_n^4\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]W_2^\star \mid \mathcal{A}_{1,3}, Z_2, Z_3]^4 \mid \tilde{\mathcal{B}}_{2,3}] \\
&+ \mathbb{E}[\mathbb{E}[\tilde{F}(Z_1, Z_3)\mathbb{E}[D(S_1, S_2) \mid H_{1,2}, S_2]W_2^\star \mid \tilde{\mathcal{A}}_{1,3}, \tilde{\mathcal{A}}_{1,2}, Z_2, Z_3]^4 \mid \tilde{\mathcal{B}}_{2,3}] \\
&\lesssim \rho_n^5\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1, Z_3)^4 \mid \mathcal{A}_{1,2}, Z_2, Z_3] \mid \mathcal{B}_{2,3}] \\
&+ \rho_n^5\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1, Z_3)^4 \mid \tilde{\mathcal{A}}_{1,2}, Z_2, Z_3] \mid \mathcal{B}_{2,3}] \\
&+ \rho_n^8\mathbb{E}[\mathbb{E}[\tilde{F}(Z_1, Z_3)^4 \mid \mathcal{A}_{1,3}, Z_2, Z_3] \mid \tilde{\mathcal{B}}_{2,3}] + O(\rho_n^5) = O(\rho_n^5)\,,
\end{aligned}
\tag{D.270}
$$

by Assumption 5.2 and Assumption C.4. Consequently, we find that

$$
\mathbb{E}[(\mathbb{E}[\tilde{F}(Z_1, Z_3)\tilde{D}^\star_{1,2} \mid Z_2, Z_3]W_2^\star)^4] = O(\rho_n^5)\,,
\tag{D.271}
$$

by plugging (D.269) and (D.270) into (D.268).

Finally, we verify the third bound in (D.109). Observe that

$$
\begin{aligned}
&\mathbb{E}[g_\xi^{(2)}(Z_1, Z_2)^2] \\
&= \mathbb{E}[\mathbb{E}[f_\xi^{(2)}(Z_3, Z_1)f_\xi^{(2)}(Z_3, Z_2) \mid Z_1, Z_2]^2] \\
&\lesssim \mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4, Z_1)\tilde{D}^\star_{4,3} \mid S_3, Z_1]\mathbb{E}[\tilde{F}(Z_4, Z_2)\tilde{D}^\star_{4,3} \mid S_3, Z_2](W_3^\star)^2 \mid Z_1, Z_2]^2] \\
&+ \mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{D}^\star_{4,1}W_1^\star\tilde{D}^\star_{4,3} \mid S_3, Z_1]\mathbb{E}[\tilde{F}(Z_4, Z_2)\tilde{D}^\star_{4,3} \mid S_3, Z_2](W_3^\star)^2 \mid Z_1, Z_2]^2] \\
&+ \mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{D}^\star_{4,1}W_1^\star\tilde{D}^\star_{4,3} \mid S_3, Z_1]\mathbb{E}[\tilde{D}^\star_{4,2}W_2^\star\tilde{D}^\star_{4,3} \mid S_3, Z_2](W_3^\star)^2 \mid Z_1, Z_2]^2]\,,
\end{aligned}
\tag{D.272}
$$

by Cauchy-Schwarz and Assumption 4.2. We give the details for verifying the bound

$$
\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4, Z_1)\tilde{D}^\star_{4,3} \mid S_3, Z_1]\mathbb{E}[\tilde{F}(Z_4, Z_2)\tilde{D}^\star_{4,3} \mid S_3, Z_2](W_3^\star)^2 \mid Z_1, Z_2]^2] = o(\rho_n^6)
\tag{D.273}
$$

as the argument supporting the same bound for the other two terms is identical. To this end, we have that

$$
\begin{aligned}
&\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4, Z_1)\tilde{D}^\star_{4,3} \mid S_3, Z_1]\mathbb{E}[\tilde{F}(Z_4, Z_2)\tilde{D}^\star_{4,3} \mid S_3, Z_2](W_3^\star)^2 \mid Z_1, Z_2]^2] \\
&\lesssim \mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4, Z_1)D_{4,3} \mid S_3, Z_1]\mathbb{E}[\tilde{F}(Z_4, Z_2)D_{4,3} \mid S_3, Z_2](W_3^\star)^2 \mid Z_1, Z_2]^2] \\
&+ \mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4, Z_1)D_{4,3} \mid S_3, Z_1] \\
&\qquad \mathbb{E}[\tilde{F}(Z_4, Z_2)\mathbb{E}[D_{4,3} \mid H_{4,3}, S_3] \mid S_3, Z_2](W_3^\star)^2 \mid Z_1, Z_2]^2] \\
&+ \mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4, Z_1)\mathbb{E}[D_{4,3} \mid H_{4,3}, S_3] \mid S_3, Z_1] \\
&\qquad \mathbb{E}[\tilde{F}(Z_4, Z_2)\mathbb{E}[D_{4,3} \mid H_{4,3}, S_3] \mid S_3, Z_2](W_3^\star)^2 \mid Z_1, Z_2]^2]\,,
\end{aligned}
\tag{D.274}
$$

by Cauchy-Schwarz. In turn, we give the details for the verification that

$$\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)D_{4,3} \mid S_3,Z_1]\mathbb{E}[\tilde{F}(Z_4,Z_2)D_{4,3} \mid S_3,Z_2](W_3^\star)^2 \mid Z_1,Z_2]^2] = o(\rho_n^6)\ . \tag{D.275}$$

An argument with an identical structure will verify the same bound for the second two terms in (D.274), in the same manner that the bound (D.270) is obtained from an argument with the same structure as the bound (D.269). To do this, recall the definition of the event $\mathcal{E}_{i,j}$, given at the beginning of Appendix C, and observe that $P\{\mathcal{E}_{i,j}\} = O(\rho_n)$ by Assumption C.2. We can evaluate

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)D_{4,3} \mid S_3,Z_1]\mathbb{E}[\tilde{F}(Z_4,Z_2)D_{4,3} \mid S_3,Z_2](W_3^\star)^2 \mid Z_1,Z_2]^2] \\
&\lesssim \rho_n\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)D_{4,3} \mid S_3,Z_1]\mathbb{E}[\tilde{F}(Z_4,Z_2)D_{4,3} \mid S_3,Z_2](W_3^\star)^2 \mid Z_1,Z_2]^2 \mid \mathcal{E}_{1,2}] \\
&+ \mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)D_{4,3} \mid S_3,Z_1]\mathbb{E}[\tilde{F}(Z_4,Z_2)D_{4,3} \mid S_3,Z_2](W_3^\star)^2 \mid Z_1,Z_2]^2 \mid \tilde{\mathcal{E}}_{1,2}] \\
&\lesssim \rho_n^7\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^2 \mid \mathcal{A}_{4,3},Z_3,Z_1] \\
&\qquad\qquad \mathbb{E}[\tilde{F}(Z_4,Z_2)^2 \mid \mathcal{A}_{4,3},Z_3,Z_2] \mid \mathcal{B}_{3,1},\mathcal{B}_{3,2},Z_1,Z_2] \mid \mathcal{E}_{1,2}] \\
&+ \rho_n^7\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^2 \mid \mathcal{A}_{4,3},Z_3,Z_1] \\
&\qquad\qquad \mathbb{E}[\tilde{F}(Z_4,Z_2)^2 \mid \mathcal{A}_{4,3},Z_3,Z_2] \mid \mathcal{B}_{3,1},\tilde{\mathcal{B}}_{3,2},Z_1,Z_2]^2 \mid \mathcal{E}_{1,2}] \\
&+ \rho_n^5\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^2 \mid \mathcal{A}_{4,3},Z_3,Z_1] \\
&\qquad\qquad \mathbb{E}[\tilde{F}(Z_4,Z_2)^2 \mid \mathcal{A}_{4,3},Z_3,Z_2] \mid \tilde{\mathcal{B}}_{3,1},\tilde{\mathcal{B}}_{3,2},Z_1,Z_2] \mid \mathcal{E}_{1,2}] \\
&+ \rho_n^6\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^2 \mid \mathcal{A}_{4,3},Z_3,Z_1] \\
&\qquad\qquad \mathbb{E}[\tilde{F}(Z_4,Z_2)^2 \mid \mathcal{A}_{4,3},Z_3,Z_2] \mid \tilde{\mathcal{B}}_{3,1},\mathcal{B}_{3,2},Z_1,Z_2] \mid \tilde{\mathcal{E}}_{1,2}] \\
&+ \rho_n^4\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^2 \mid \mathcal{A}_{4,3},Z_3,Z_1] \\
&\qquad\qquad \mathbb{E}[\tilde{F}(Z_4,Z_2)^2 \mid \mathcal{A}_{4,3},Z_3,Z_2] \mid \tilde{\mathcal{B}}_{3,1},\tilde{\mathcal{B}}_{3,2},Z_1,Z_2] \mid \tilde{\mathcal{E}}_{1,2}] \\
&\lesssim \rho_n^7\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^4 \mid \mathcal{A}_{4,3},Z_3,Z_1,Z_2] \mid \mathcal{B}_{3,1},\mathcal{B}_{3,2},Z_1,Z_2] \mid \mathcal{E}_{1,2}] \\
&+ \rho_n^8\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^2 \mid \mathcal{A}_{4,3},Z_3,Z_1] \mid \mathcal{B}_{3,1},\tilde{\mathcal{B}}_{3,2},Z_1,Z_2]^2 \mid \mathcal{E}_{1,2}] \\
&+ \rho_n^7\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_2)^2 \mid \mathcal{A}_{4,3},Z_3,Z_2] \mid \tilde{\mathcal{B}}_{3,1},\mathcal{B}_{3,2},Z_1,Z_2] \mid \tilde{\mathcal{E}}_{1,2}] \\
&+ \rho_n^4\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^2 \mid \mathcal{A}_{4,3},Z_3,Z_1] \\
&\qquad\qquad \mathbb{E}[\tilde{F}(Z_4,Z_2)^2 \mid \mathcal{A}_{4,3},Z_3,Z_2] \mid \tilde{\mathcal{B}}_{3,1},\tilde{\mathcal{B}}_{3,2},Z_1,Z_2] \mid \tilde{\mathcal{E}}_{1,2}] + O(\rho_n^7) \\
&= \rho_n^4\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^2 \mid \mathcal{A}_{4,3},Z_3,Z_1] \\
&\qquad\qquad \mathbb{E}[\tilde{F}(Z_4,Z_2)^2 \mid \mathcal{A}_{4,3},Z_3,Z_2] \mid \tilde{\mathcal{B}}_{3,1},\tilde{\mathcal{B}}_{3,2},Z_1,Z_2] \mid \tilde{\mathcal{E}}_{1,2}] + O(\rho_n^7)\ ,
\end{aligned} \tag{D.276}$$

through several applications of Assumption 5.2 and Assumption C.4. In turn, we can evaluate

$$\begin{aligned}
&\rho_n^4\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^2 \mid \mathcal{A}_{4,3},Z_3,Z_1] \\
&\qquad\qquad \mathbb{E}[\tilde{F}(Z_4,Z_2)^2 \mid \mathcal{A}_{4,3},Z_3,Z_2] \mid \tilde{\mathcal{B}}_{3,1},\tilde{\mathcal{B}}_{3,2},Z_1,Z_2] \mid \tilde{\mathcal{E}}_{1,2}] \\
&\lesssim \rho_n^4\mathbb{E}[\mathbb{E}[\mathbb{E}[\tilde{F}(Z_4,Z_1)^4 \mid \mathcal{A}_{4,3},Z_3,Z_1] \mid \tilde{\mathcal{B}}_{3,1},\tilde{\mathcal{B}}_{3,2},Z_1,Z_2] \mid \tilde{\mathcal{E}}_{1,2}] = o(\rho_n^6)
\end{aligned} \tag{D.277}$$

by Assumption 5.2 and Assumption C.4. The bound (D.275) is thus verified by plugging (D.275) into (D.276), completing the proof. ■

D.12 Proof of Lemma D.11

The result follows from an application of Lemma D.3. Throughout, we adopt the short-hand

$$F_i(Z) = F_i(Z_j, Z_{-j}) = F(Z_i, Z_{-i}) \; . \tag{D.278}$$

Now, observe that as

$$\mathbb{E}[R_i \mid Z_i] = 0 \tag{D.279}$$

we have that

$$\mathbb{E}[R_i^2] = \mathbb{E}[\mathbb{E}[R_i^2 \mid Z_i]] = \mathbb{E}[\mathrm{Var}(R_i \mid Z_i)] \; . \tag{D.280}$$

Thus, as the equality

$$\mathbb{E}[R_i \mid Z_i, Z_j] = 0 \tag{D.281}$$

holds almost surely, by definition, Lemma D.3 implies that the inequality

$$\begin{aligned}
&\mathbb{E}[\mathrm{Var}(R_i \mid Z_i)] \\
&\leq \frac{1}{2} \sum_{j \neq i}^{n} \sum_{k \notin \{i,j\}} \mathbb{E}[\mathbb{E}[(F(Z_i, Z_{-i}) - F(Z_i, Z_{-i}^{(j)})) - (F(Z_i, Z_{-i}^{(k)}) - F(Z_i, Z_{-i}^{(j,k)})) \mid Z]^2] \\
&\lesssim n^2 \mathbb{E}[\mathbb{E}[\nabla_{j,k}^2 F(Z_i, Z_{-i}) \mid Z^{(j,k)}]^2]
\end{aligned} \tag{D.282}$$

holds, where we recall the notation introduced in Assumption C.1. Now, observe that we can evaluate

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[\nabla_{j,k}^2 F(Z_i, Z_{-i}) \mid Z^{(j,k)}]^2] \\
&\lesssim \rho_n^4 \mathbb{E}[\mathbb{E}[\nabla_{j,k} F(Z_i, Z_{-i}) \mid A_{i,j} = 1, A_{i,k} = 1, Z^{(j,k)}]^2] \\
&+ \rho_n^2 \mathbb{E}[\mathbb{E}[\nabla_{j,k} F(Z_i, Z_{-i}) \mid A_{i,j} = 1, A_{i,k} = 0, Z^{(j,k)}]^2] \\
&+ \mathbb{E}[\mathbb{E}[\nabla_{j,k} F(Z_i, Z_{-i}) \mid A_{i,j} = 1, A_{i,k} = 0, Z^{(j,k)}]^2] = O(n^{-2} \rho_n)
\end{aligned} \tag{D.283}$$

by Assumption C.1. By plugging (D.283) into (D.282), we find that

$$\mathbb{E}[R_i^2] = O(\rho_n) \; , \tag{D.284}$$

completing the proof. ■

Appendix E. Additional Results and Discussion

E.1 Extensions to Theorem 2.1

In this appendix, we outline several extensions to the baseline decomposition given in Section 2.

E.1.1 Direct Effects. In many applications, the regression specification (2.1) is augmented to additionally include the covariate W_i; that is, through the specification

$$Y_i = \alpha + \tau \cdot W_i + \theta \cdot \Delta_i + \varepsilon_i \ . \tag{E.1}$$

The coefficient on the covariate W_i is often interpreted as an estimate of the "direct effect" of the treatment W_i on Y_i.[37] The decomposition stated in Theorem 2.1 holds for this augmented specification.

In particular, consider the augmented loss function

$$L_n(\tau, \theta) = \min_{\alpha} \sum_{i=1}^{n} (Y_i - \alpha - \tau \cdot W_i - \theta \cdot \Delta_i)^2 \tag{E.2}$$

associated with the specification (E.1). By the Frisch-Waugh-Lovell Theorem, we can write

$$\mathbb{E}[L_n(\tau,\theta)] = \mathbb{E}\left[\sum_{i=1}^{n}(Y_i - \tau\tilde{W}_i - \theta \cdot \tilde{\Delta}_i)^2\right] ,$$

$$\text{where} \quad \tilde{W}_i = W_i - \frac{1}{n}\sum_{j=1}^{n} W_j \quad \text{and}$$

$$\tilde{\Delta}_i = \sum_{j \neq i}\left(D_{i,j} - \frac{1}{n}\sum_{k \neq j} D_{k,j}\right) W_j - \frac{1}{n}\sum_{k \neq i} D_{k,i} W_i \ . \tag{E.3}$$

Observe that

$$\mathbb{E}[\tilde{\Delta}_i \tilde{W}_i] = -\frac{n-1}{n}\mathbb{E}[D_{i,j} W_j^2] \tag{E.4}$$

and that

$$\mathrm{Var}(\tilde{W}_i) = \frac{n-1}{n}\,\mathrm{Var}(W_i) \ . \tag{E.5}$$

Thus, by a second application of the Frisch-Waugh-Lovell Theorem, it holds that

$$\min_{\tau} \mathbb{E}[L_n(\tau,\theta)] = \mathbb{E}\left[\sum_{i=1}^{n}(Y_i - \theta \cdot \tilde{\Delta}_i')^2\right] ,$$

$$\text{where} \quad \tilde{\Delta}_i' = \sum_{j \neq i}\left(D_{i,j} - \frac{1}{n}\sum_{k \neq j} D_{k,j}\right) W_j - \left(\frac{1}{n}\sum_{k \neq i} D_{k,i} - \frac{\mathbb{E}[D_{i,j}W_j^2]}{\mathrm{Var}(W_i)}\right) W_i \ . \tag{E.6}$$

Hence, by Projection Theorem, it holds that

$$\bar{\theta}_n = \frac{\sum_{i=1}^{n} \mathbb{E}[Y_i \tilde{\Delta}_i']}{\sum_{i=1}^{n} \mathbb{E}[(\tilde{\Delta}_i')^2]} \ . \tag{E.7}$$

The representation (E.7) is very similar to the representation (A.22) considered in the proof of Theorem 2.1. The only difference is through the second term in (E.6), which will contribute slightly different lower order terms when expressed as a representation analogous to (A.23). It can be shown that these lower order terms are smaller than $o(n^{-1/2})$ under the same regularity conditions used for Theorem 2.1.

[37] See Li and Wager (2022) for discussion of estimators of direct effects in models similar to the settings considered in this paper.

E.1.2 Monotone Transformations. In some settings, the proximity measure of interest enters through a monotonic function $Q(D_{i,j})$; that is, thorough the specification

$$Y_i = \alpha + \theta \cdot Q_i + \varepsilon_i \quad \text{where} \quad Q_i = \sum_{j \neq i} Q(D_{i,j}) W_j \ . \tag{E.8}$$

For instance, when $D_{i,j}$ is a function of geographic distance, researchers often set $Q(\cdot)$ as a binary indicator that $D_{i,j}$ is an element of some pre-defined interval; see, e.g, Miguel and Kremer (2004), Feyrer et al. (2017), Egger et al. (2022), Myers and Lanahan (2022), and Muralidharan et al. (2023). In other cases, the function $Q(\cdot)$ often depends on various elasticities fixed at values chosen by referencing pre-existing literature; see e.g., Kovak (2013), Autor et al. (2013), Donaldson and Hornbeck (2016), and Adao et al. (2025).

It follows immediately from Theorem 2.1 that, in this case, we have that

$$\overline{\theta}_n = \int_{\overline{\mathcal{Q}}} \int_{\overline{\mathcal{W}}} \mathbb{E}\Big[\lambda(q, w \mid S_j) \partial^2_{q,w} \mathbb{E}[Y_i \mid Q_i = q, W_j = w, S_j]\Big] \, \mathrm{d}w \, \mathrm{d}q + o(n^{-1/2}) \ , \tag{E.9}$$

where the weights are defined analogously to (2.8) and $\overline{\mathcal{Q}}$ is some interval containing the support of $Q(D_{i,j})$. In cases where the function $Q(\cdot)$ is invertible (e.g., $Q(D_{i,j})$ is a smooth decreasing function of a distance $D_{i,j}$), it follows immediately that

$$\overline{\theta}_n = \int_{\overline{\mathcal{Q}}} \int_{\overline{\mathcal{W}}} \mathbb{E}\Big[\lambda(q, w \mid S_j) \partial^2_{q,w} \mu(Q^{-1}(q), w \mid S_j)\Big] \, \mathrm{d}w \, \mathrm{d}q + o(n^{-1/2}) \ . \tag{E.10}$$

That is, in this case, the parameter $\overline{\theta}_n$ is still a convex average of the parameter $\partial^2_{q,w}\mu(\delta, w \mid S_j)$, but now the weights in this average depend on the function $Q(\cdot)$.

E.1.3 Dyadic Information. Network linkages are often modeled through specifications of the form

$$D_{i,j} = D_n(S_i, S_j, \varepsilon_{i,j}) \ , \tag{E.11}$$

where the variables $\varepsilon_{i,j}$ are i.i.d. and distributed independently of the coordinates S_i. See, for instance, the latent position or graphon network models considered by, e.g., Hoff et al. (2002) and Li and Wager (2022), respectively. Indeed, by the Aldous-Hoover representation theorem, any exchangeable array admits such a representation (Aldous, 1981; Hoover, 1979). See Menzel (2021) for discussion.

Under the assumption that the treatments W_i are independent of the error terms $\varepsilon_{i,j}$, the decomposition Theorem 2.1 holds for this setting, unchanged. In particular, Theorem 2.1 only requires that the treatment W_j is independent of the proximity measure $D_{i,j}$, conditional on the coordinate S_j. So long as this continues to hold, the theorem will be unchanged.

We impose Assumption 2.1 because the variables $\varepsilon_{i,j}$ would complicate the construction of potential outcomes. In particular, when considering interventions on the proximity measure $D_{i,j}$, while holding S_i fixed, we would need to specify whether we are changing the value of the coordinate S_j or the error term $\varepsilon_{i,j}$. Likewise, incorporating the variables $\varepsilon_{i,j}$ would require augmenting the Assumption 3.1 to accommodate dyadically indexed variables, which would substantially complicate our theoretical analysis.

E.1.4 Bipartite Data. In many applications, treatments and outcomes are measured for different types of units. In particular, suppose that there are n units and m treatment nodes. Each unit i is associated with an outcome Y_i. Each node k is associated with a treatment W_k. Each unit i and node k is associated with the measure $D_{i,k}$. A prominent instance of this setting are cases where treatments are assigned to industries and outcomes are measured at geographic units of aggregation. See, for instance, Kovak (2013), Autor et al. (2013), Adao et al. (2019), and Borusyak et al. (2022b). Here, the treatments W_k might be changes to the productivity of industry k and $D_{i,k}$ might be the share of employment of city i in industry k.

Researchers often report coefficients from regression specifications of the form

$$Y_i = \alpha + \theta \cdot \Delta_i + \varepsilon_i \quad \text{and} \quad \Delta_i = \sum_{j=1}^{m} D_{j,k} W_k \tag{E.12}$$

A result very similar to Theorem 2.1 holds for this setting. That is, such regressions identify convex averages of the parameters

$$\partial^2_{\delta,w} \mathbb{E}[Y_i \mid D_{i,k} = \delta, W_k = w, R_k] \tag{E.13}$$

Thus, "shift-share" specifications of the form (E.12) can be interpreted as measuring the effect of changing the measure $D_{i,k}$ on the effect of the treatment W_k on the outcome Y_i. We conclude this subsection by giving a formal statement and proof for this result.

Causal interpretations of the parameter (E.13) are susceptible to the same sort of confounding considered in the main text. For instance, suppose that the variable $D_{i,k}$ denotes the share of employment in location i that belongs to industry k and that the variable $G_{i,k}$ denotes the share of imports to location i from industry k. If the productivity of industry k increases, wages in location i may increase through $D_{i,k}$, due to increased local labor demand, or through $G_{i,k}$, due to cheaper intermediates.

Approaches analogous to those considered in the main text can be used to disentangle these channels. In particular, the variable of interest $D_{i,k}$ can be residualized on the auxiliary channel $G_{i,k}$, and the residuals can be used in the place of $D_{i,k}$ in the specification (E.12). Estimators with this structure can be shown to be consistent through arguments analogous to those used to establish Theorem 5.1. We note, however, that obtaining a result analogous to the Gaussian approximation result given in Theorem 6.1 would be a more significant undertaking, as the central limit theorem from Liu et al. (2025) is not immediately applicable.

We now state and prove a version of Theorem 2.1 for settings with bipartite data. We consider the following assumptions, which mirror Assumptions 2.1 to 2.3.

Assumption E.1. *Each unit i and node k are associated with coordinates S_i and R_k, respectively. There exists a function $D_n(\cdot,\cdot)$, valued on the non-negative, real-valued set $\mathcal{D}$, such that the representation*

$$D_{i,k} = D_n(S_i, R_k) \ , \tag{E.14}$$

holds almost surely for each unit i and node k.

Assumption E.2. *The replicates $(R_k, W_k)_{k=1}^m$ and $(S_i)_{i=1}^n$ are i.i.d., and are independent of each other. Moreover, the orthogonality condition*

$$\mathbb{E}[W_i \mid R_i] = 0 \tag{E.15}$$

holds almost surely.

Assumption E.3. *For any units i and j, and node k, the map*

$$(\delta, w) \mapsto \mathbb{E}[Y_i \mid D_{j,k} = \delta, W_j = w, R_j] \tag{E.16}$$

is twice continuously differentiable, almost surely. Moreover, it holds that

$$\partial^2_{\delta,w} \mathbb{E}[Y_i \mid D_{j,k} = \delta, W_k = w, R_k = s] = o(n^{-1/2}) \tag{E.17}$$

uniformly over each δ, w, and s in their respective domains.

Theorem E.1. *Assume that the treatments and proximity measures have support contained in intervals $\overline{\mathcal{W}}$ and $\overline{\mathcal{D}}$ of constant length and that the outcomes are identically distributed. Consider the loss function*

$$L_n(\theta) = \min_{\alpha} \sum_{i=1}^{n} (Y_i - \alpha - \theta \cdot \Delta_i)^2, \quad \text{where} \quad \Delta_i = \sum_{j=1}^{m} D_{i,k} W_k \; . \tag{E.18}$$

Under Assumptions E.1 to E.3, the parameter $\overline{\theta}_n$ that minimizes $\mathbb{E}[L_n(\theta)]$ admits the representation

$$\overline{\theta}_n = \int_{\overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} \mathbb{E}\Big[\lambda(\delta, w \mid R_j)\, \partial^2_{\delta,w} \mathbb{E}[Y_i \mid D_{i,k} = \delta, W_k = w, R_k]\Big] \, \mathrm{d}w \, \mathrm{d}\delta + o(n^{-1/2}) \; , \tag{E.19}$$

where the weights

$$\lambda(\delta, w \mid R_j) = \frac{\operatorname{Cov}(\mathbb{I}\{D_{i,j} \geq \delta\}, D_{i,j} \mid R_j) \operatorname{Cov}(\mathbb{I}\{W_j \geq w\}, W_j \mid R_j)}{\mathbb{E}[\operatorname{Var}(D_{i,j} \mid R_j) \operatorname{Var}(W_j \mid R_j)]} \tag{E.20}$$

are convex.

Proof. By the Frisch-Waugh-Lovell and Projection Theorems, we can write

$$\overline{\theta}_n = \frac{\sum_{i=1}^{n} \mathbb{E}[Y_i \tilde{\Delta}_i]}{\sum_{i=1}^{n} \mathbb{E}[(\tilde{\Delta}_i)^2]} \; , \tag{E.21}$$

where

$$\tilde{\Delta}_i = \sum_{k=1}^{m} D_{i,k} W_k - \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{m} D_{j,k} W_k = \sum_{k=1}^{m} (D_{i,k} - \frac{1}{n} \sum_{j=1}^{n} D_{j,k}) W_k \; . \tag{E.22}$$

Observe that

$$\mathbb{E}[Y_i \tilde{\Delta}_i] = m \left(1 - \frac{1}{n}\right) \left(\mathbb{E}[Y_i \tilde{D}_{i,k} W_k] - \mathbb{E}[Y_i \tilde{D}_{j,k} W_k]\right) \tag{E.23}$$

and

$$\mathbb{E}[(\tilde{\Delta}_i)^2] = m \left(1 - \frac{1}{n}\right) \mathbb{E}[\tilde{D}^2_{i,k} W^2_k] \; . \tag{E.24}$$

Thus, we obtain the representation

$$\overline{\theta}_n = \frac{\mathbb{E}[Y_i \tilde{D}_{i,k} W_k] - \mathbb{E}[Y_i \tilde{D}_{j,k} W_k]}{\mathbb{E}[\operatorname{Var}(D_{i,k} \mid R_k) \operatorname{Var}(R_k \mid W_k)]} \; . \tag{E.25}$$

Consequently, by Lemma A.2, we obtain the representation

$$\overline{\theta}_n = \int_{\overline{\mathcal{D}}} \int_{\mathcal{W}} \mathbb{E}\Big[\lambda(\delta, w \mid R_j) \left(\partial^2_{\delta,w}\mathbb{E}[Y_i \mid D_{i,k} = \delta, W_k = w, R_k] - \mathbb{E}[Y_i \mid D_{j,k} = \delta, W_k = w, R_k]\right)\Big] \,\mathrm{d}w\, \mathrm{d}\delta \tag{E.26}$$

where the weights $\lambda(\delta, w \mid R_j)$ are defined in (E.20). The weights are convex by Lemma A.3. Thus, the representation (E.19) follows from Assumption E.3, as required. ■

E.2 Extensions to Assumption 3.2

In this Appendix, we detail two extensions to Assumption 3.2.

E.2.1 Discontinuities. Assumption 3.2 stipulates that

$$D_{i,j} = D(S_i, S_j) = \overline{D}(\|S_i - S_j\|) \tag{E.27}$$

for some function $\overline{D}(\cdot)$ that is continuous and monotone. Here, we outline how to extend this assumption to treat cases where the function $\overline{D}(\cdot)$ has discontinuities. Observe that, in our formal arguments, Assumption 3.2 is only applied once, in Corollary 4.2 to ensure that the counterfactual coordinate $S_j^{(i)}(\delta)$ is uniquely defined. Here, the restriction that $\overline{D}(\cdot)$ is continuous is only used to ensure that the level set $\mathcal{S}^{(i)}(\delta)$ is closed.

Thus, to handle the discontinuous case, let

$$\mathcal{S}^{(i)}(\delta, \varepsilon) = \{s : \|s - s'\| \leq \varepsilon, s' \in \mathsf{cl}(\mathcal{S}^{(i)}(\delta))\}. \tag{E.28}$$

denote the ε enlargement of the closure of the set

$$\mathcal{S}^{(i)}(\delta) = \{s : D_n(S_i, s) = \delta\}\ . \tag{E.29}$$

Define the counterfactual coordinate

$$S_j^{(i)}(\delta, \varepsilon) = S_j + \alpha_j^{(i)}(\delta, \varepsilon)(S_i - S_j)\ , \quad \text{where}$$
$$\alpha_j^{(i)}(\delta, \varepsilon) = \arg\min_{\alpha} \{|\alpha| : S_j + \alpha(S_i - S_j) \in \mathcal{S}^{(i)}(\delta, \varepsilon)\}\ . \tag{E.30}$$

With this construction, we can define the potential outcome through the limit

$$Y_{i,j}(\delta, w) = \lim_{\varepsilon \downarrow 0} F(Z_i, Z_j(w, S_j^{(i)}(\delta, \varepsilon)), Z_{-i,j})\ . \tag{E.31}$$

Observe that the construction agrees with the construction stated in the main text for the continuous case, so long as the function F is continuous in S_j.

For the sake of concreteness, consider the case that

$$D_{i,j} = \mathbb{I}\{\|S_i - S_j\| \leq \phi\} \tag{E.32}$$

for some constant radius ϕ. The level set $\mathcal{S}^{(i)}(1)$ is closed, and so it suffices to consider the construction of the counterfactual coordinate $S_j^{(i)}(0)$. In this case, the coordinate $S_j^{(i)}(\delta, \varepsilon)$ moves toward S_i on the line $S_j + \alpha(S_i - S_j)$ as α increases toward one, achieving $\|S_i - S_j\| = \phi$ as $\varepsilon \to 0$.

E.2.2 Several Distances. In some applications, proximity measures of interest are functions of many features of the units under consideration. Commuting and migration probabilities, for instance, are often modeled as being functions of geographic distance, as well as wages and amenity values (Monte et al., 2018). We sketch three distinct approaches for handling settings with this structure.

Monotone Index: The first approach is based on the assumption that we can decompose the coordinate S_i into the components $S_i = (S_{i,0}, S_{i,1})$. For instance, suppose that $D_{i,j}$ denotes the probability of commuting from location i to location j. In this case, the component $S_{i,0}$ might denote the geographic coordinate of location i and $S_{i,1}$ might denote the average utility associated with working in location i.

We consider the following alternative to Assumption 3.2.

Assumption E.4 (Extended Index Structure)**.** *The coordinate $S_{i,1}$ is scalar-valued and the proximity measure $D_{i,j}$ admits the representation*

$$D_{i,j} = D(S_i, S_j) = \overline{D}(S_i, S_{j,0}, S_{j,1}) \tag{E.33}$$

where the function $\overline{D}(S_i, S_{j,0}, \cdot)$ is monotone.

That is, the proximity measure of interest $D_{i,j}$ is monotone, almost surely, in some component $S_{j,1}$ of the coordinate S_j. For instance, the probability of commuting from S_i to S_j might be decreasing in the utility associated with working in location j, conditional on location j's geographic location and the features of the location i. In this case, we can define

$$\begin{aligned} &\mathcal{S}_j^{(i)}(\delta) = \{s : D_n(S_i, (S_{j,1}, s)) = \delta\} \,, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(E.34)} \\ &S_{j,1}^{(i)}(\delta) = S_{j,1} + \alpha_j^{(i)}(\delta) \,, \quad \text{and} \quad \alpha_j^{(i)}(\delta) = \arg\min_{\alpha} \{|\alpha| : S_{j,0} + \alpha \in \mathcal{S}_j^{(i)}(\delta)\} \,, \end{aligned}$$

as before. Potential outcomes can then be defined by

$$Y_{i,j}(\delta, w) = F(Z_i, Z_j(w, S_{j,1}^{(i)}(\delta)), Z_{-i,j}) \,, \quad \text{where} \quad Z_j(w, s) = (w, S_{j,0}, s, U_j) \,. \tag{E.35}$$

it is clear that, so long as (E.34) is non-empty, the potential outcome (E.35) is uniquely defined.

In this setting, identification of averages of spillover proximity gradients is most credible in the case that the coordinates $S_{i,0}$ are observed. In this case, Assumption 4.4 should be replaced by the condition that

$$\mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, H_{i,j}, S_{i,0}, \overline{S}_j] = \mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid H_{i,j}, S_{i,0}, \overline{S}_j] \,. \tag{E.36}$$

That is, we can more transparently compare changes in spillovers induced by changes to the index $S_{1,i}$ by holding constant the features $S_{i,0}$ that might otherwise change the value of the proximity measure $D_{i,j}$.

Monotone Norm: The second approach is closely related and again based on the decomposition $S_i = (S_{i,0}, S_{i,1})$. Here, we consider the following extension to Assumption 3.2.

Assumption E.5 (Conditional Monotone Factor Structure)**.** *There exists a norm $\| \cdot \|$ such that*

$$D_{i,j} = D(S_i, S_j) = \overline{D}(\|S_{0,i} - S_{0,j}\|, S_{i,1}, S_{j,1}) \tag{E.37}$$

for some function $\overline{D}(\cdot)$ that is continuous and non-increasing in its first entry.

In other words, we assume that the proximity measures of interest still depend monotonically on a distance $\|S_{0,i} - S_{0,j}\|$. For instance, the probability of commuting from S_i to S_j might be decreasing in the commuting time between location i and location j, conditional on all other features of the two locations. In this case, we can define

$$\begin{aligned} \mathcal{S}^{(i)}(\delta) &= \{s : D_n(S_i, (s, S_{j,1})) = \delta\} , \\ S_{j,0}^{(i)}(\delta) &= S_{j,0} + \alpha_j^{(i)}(\delta)(S_{i,0} - S_{j,0}) , \quad \text{and} \\ \alpha_j^{(i)}(\delta) &= \arg\min_{\alpha} \{|\alpha| : S_{j,0} + \alpha(S_{i,0} - S_{j,0}) \in \mathcal{S}^{(i)}\} . \end{aligned} \tag{E.38}$$

and, thus, potential outcomes can be defined by

$$Y_{i,j}(\delta, w) = F(Z_i, Z_j(w, S_{j,0}^{(i)}(\delta)), Z_{-i,j}) , \quad \text{where} \quad Z_j(w, s) = (w, s, S_{j,1}U_j) . \tag{E.39}$$

By arguments analogous to those used in the proof of Corollary 4.2, the potential outcome (E.39) is uniquely defined whenever $\mathcal{S}^{(i)}(\delta)$ is non-empty.

As before, identification is most credible when the features $S_{i,1}$ are observed. In this case, Assumption 4.4 should be replaced by the condition that

$$\mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, H_{i,j}, \overline{S}_{i,1}, \overline{S}_j] = \mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid H_{i,j}, \overline{S}_{i,1}, \overline{S}_j] . \tag{E.40}$$

Here, we compare changes in spillovers induced by changes to the distance $\|S_{0,i} - S_{0,j}\|$ by holding constant the features $\overline{S}_{i,1}$ of the unit i that might otherwise change the value of the proximity measure $D_{i,j}$.

Invariance: The final approach is best suited for settings where the researcher is not willing to pose a concrete interpretation for the coordinate S_i. For instance, in Examples 2 and 3, although the coordinate S_i is probably best understood as collecting the geographic coordinates, wages, and amenity values of the location i, at this level of abstraction, this interpretation is contestable. This ambiguity can be resolved by imposing the following restriction.

Assumption E.6 (Stable Pairwise Proximity Spillovers)**.** *It holds that*

$$Y_{i,j}(\delta, w) = F(Z_i, Z_j(w, s), Z_{-i,j}), \quad \textit{for all} \quad s \in \mathcal{S}_j^{(i)}(\delta) . \tag{E.41}$$

This assumption is the analogue of the classical "Stable Unit Treatment Value Assumption" (SUTVA), due to Cox (1958) and Rubin (1980), adapted to our setting. In words, Assumption E.6 imposes the restriction that unit j's location S_j only modifies the influence of the treatment W_j on the outcome Y_i though $D_{i,j}$. That is, the "direction" of the spillover from j into i makes no difference, all that matters is the "distance" $D_{i,j}$.

Like SUTVA, Assumption E.6 places strong restrictions on the structure of interference. In particular, Assumption E.6 effectively rules out endogenous spillovers, in the sense of Manski (1993). That is, as the coordinate S_j varies over the level set $\mathcal{S}^{(i)}(\delta)$, unit j's proximity to units other than i changes. Assumption E.6 stipulates that these changes have no influence on the spillover effect of j on i. Of course, standard models of economic geography and trade feature endogenous spillovers (Allen and Arkolakis, 2014; Monte et al., 2018). On the other hand, approaches to estimating spillover effects based on well-posed exposure mappings rule out endogenous spillovers as well (Hudgens and Halloran, 2008; Manski, 2013; Li and Wager, 2022).

In sum, Assumption E.6 provides a clear, albeit strong, tool for reasoning about the potential outcomes $Y_{i,j}(\delta, w)$ in settings where the the coordinates S_i lack a concrete interpretation or the proximity measure of interest $D_{i,j}$ cannot be represented in a way that is amenable to the application of Assumption 3.2, Assumption E.4, or Assumption E.5. When applying Assumption E.6 it is important make it plain that endogenous spillover effects are assumed to be negligible.

E.3 Differences Between Residualized and Long Spillover Regressions

In this appendix, we compare regressions

$$Y_i = \alpha + \theta \cdot \Delta_i^\star + \varepsilon_i \quad \text{and} \tag{E.42}$$

$$Y_i = \alpha + \theta \cdot \Delta_i + \xi \cdot \Gamma_i + \varepsilon_i \ , \tag{E.43}$$

where

$$\Delta_i = \sum_{j \neq i} D_{i,j} W_j \ , \quad \Delta_i^\star = \sum_{j \neq i} (D_{i,j} - \mathbb{E}[D_{i,j} \mid G_{i,j}]) W_j \ , \quad \text{and} \quad \Gamma_i = \sum_{j \neq i} G_{i,j} W_j \ . \tag{E.44}$$

In particular, we outline the conditions under which the population coefficient on the covariate Δ_i in the "long" regression (E.43) is equal to the population coefficient on $\Delta_i^\star$ in the "short" regression (E.42).

To this end, consider the loss function

$$L_n^{\text{long}}(\theta, \xi) = \min_\alpha \sum_{i=1}^n (Y_i - \alpha - \theta \cdot \Delta_i - \xi \cdot \Gamma_i) \tag{E.45}$$

associated with the specification (E.42). By the Frisch-Waugh-Lovell Theorem, we can write

$$\mathbb{E}[L_n^{\text{long}}(\theta, \xi)] = \mathbb{E}\left[\sum_{i=1}^n \left((Y_i - \tilde{Y}_n) - \theta \cdot \tilde{\Delta}_i - \xi \cdot \tilde{\Delta}_i\right)^2\right] \ , \tag{E.46}$$

where $\tilde{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$ denotes the centered outcome,

$$\tilde{\Delta}_i = \sum_{j \neq i} \left(D_{i,j} - \frac{1}{n}\sum_{k \neq j} D_{k,j}\right) W_j - \frac{1}{n}\sum_{k \neq i} D_{k,i} W_i \ , \quad \text{and} \tag{E.47}$$

$$\tilde{\Gamma}_i = \sum_{j \neq i} \left(G_{i,j} - \frac{1}{n}\sum_{k \neq j} G_{k,j}\right) W_j - \frac{1}{n}\sum_{k \neq i} G_{k,i} W_i \ .$$

Thus, by the Projection Theorem and a second application of the Frisch-Waugh-Lovell Theorem, the parameter $\overline{\theta}_n^{\text{long}}$ that minimizes the expected loss (E.46) admits the representation

$$\overline{\theta}_n^{\text{long}} = \frac{\mathbb{E}[Y_i \tilde{\Delta}_i']}{\mathbb{E}[(\tilde{\Delta}_i')^2]} \quad \text{where} \quad \tilde{\Delta}_i' = \sum_{j \neq i} \left(D_{i,j}' - \frac{1}{n}\sum_{k \neq j} D_{k,j}'\right) W_j - \frac{1}{n}\sum_{k \neq i} D_{k,i}' W_i \ , \tag{E.48}$$

$$D_{i,j}' = D_{i,j} - \xi' \cdot G_{i,j} \ , \quad \text{and} \quad \xi' = \frac{\text{Cov}(\tilde{\Delta}_i, \tilde{\Gamma}_i)}{\text{Var}(\tilde{\Gamma}_i)} \ .$$

By comparing the representation (E.48) to the representation (A.34) stated in the Proof of Theorem 4.1, we find that the condition

$$\xi' \cdot G_{i,j} = \mathbb{E}[D_{i,j} \mid G_{i,j}] \tag{E.49}$$

is necessary and sufficient for the population coefficient on Δ_i in the regression (E.43) to coincide with the population coefficient on $\Delta_i^\star$ in the regression (E.42).

We conclude this appendix by showing that the condition (E.49) holds, up to a small error term, if and only if the conditional expectation $\mathbb{E}[D_{i,j} \mid G_{i,j}]$ is linear and the treatments W_j are homoskedastic, in the sense that their variances conditional on the variables $\overline{S}_j$ are constant. To see this, observe that

$$\begin{aligned}
\operatorname{Cov}(\tilde{\Delta}_i, \tilde{\Gamma}_i) &= \sum_{j \neq i} \operatorname{Cov}\left(\left(D_{i,j} - \frac{1}{n}\sum_{k \neq j} D_{k,j}\right) W_j, \left(G_{i,j} - \frac{1}{n}\sum_{k \neq j} G_{k,j}\right) W_j\right) \\
&\quad + \frac{1}{n}\sum_{k \neq i}\sum_{q \neq i} \operatorname{Cov}(D_{k,i} W_i, G_{k,i} W_i) \\
&= (n-1)\mathbb{E}[\operatorname{Cov}(D_{i,j}, G_{i,j} \mid S_j) \operatorname{Var}(W_j \mid \overline{S}_j)] \; ,
\end{aligned} \tag{E.50}$$

where we have applied Assumptions 2.1, 4.1, and 4.2. Likewise, we can evaluate

$$\operatorname{Var}(\tilde{\Gamma}_i) = (n-2)\mathbb{E}[\operatorname{Var}(G_{i,j} \mid \overline{S}_j) \operatorname{Var}(W_j \mid \overline{S}_j)] + O(n^{-1}) \tag{E.51}$$

by Lemma A.1. Thus, we have that

$$\xi' \cdot G_{i,j} = \frac{\mathbb{E}[\operatorname{Cov}(D_{i,j}, G_{i,j} \mid S_j) \operatorname{Var}(W_j \mid \overline{S}_j)]}{\mathbb{E}[\operatorname{Var}(G_{i,j} \mid \overline{S}_j) \operatorname{Var}(W_j \mid \overline{S}_j)]} + O(n^{-2}) \tag{E.52}$$

by (E.50) and (E.51). Now, observe that Assumption 4.3 and the assumption that $\mathbb{E}[D_{i,j} \mid G_{i,j}]$ is linear imply that we can write

$$\mathbb{E}[D_{i,j} \mid G_{i,j}] = \frac{\mathbb{E}[\operatorname{Cov}(D_{i,j}, G_{i,j} \mid S_j)]}{\mathbb{E}[\operatorname{Var}(G_{i,j} \mid \overline{S}_j)]} G_{i,j} \; . \tag{E.53}$$

Hence, so long as the quantities $\operatorname{Cov}(D_{i,j}, G_{i,j} \mid S_j)]$ and $\operatorname{Var}(G_{i,j} \mid \overline{S}_j)]$ are non-constant, the condition (E.49) holds if any only if $\operatorname{Var}(W_j \mid \overline{S}_j)$ is constant almost surely.

E.4 Alternative Approaches for Addressing Treatment Endogeneity

In this Appendix, we illustrate how to augment the framework proposed in the main text to accommodate settings where endogeneity in treatment assignment is addressed with instrumental variable or difference-in-differences research designs.

E.4.1 Instrumental Variables. First, we demonstrate how to adapt our estimation framework to treat settings based on instrumental variable research designs. For the sake of simplicity, we assume that we observe an instrumental variable M_i, that is also valued on $\{0, 1\}$ and is some deterministic function of the latent variable $\overline{S}_i$. For the sake of simplicity, we focus attention on the case without covariates X_i or auxiliary measures of proximity $G_{i,j}$, as the results stated here will generalize to that setting in the same way that Theorem 2.1 generalizes to Theorem 4.1.

We consider a distinct two-step procedure, analogous to the standard cross-sectional two-stage least-squares estimator. In the first stage, we compute an estimate $\hat{\pi}_n(m)$ of the conditional expectation

$$\pi_n(m) = \mathbb{E}[W_i \mid M_i] \; . \tag{E.54}$$

In the second stage, we consider the regression specification

$$Y_i = \alpha + \theta \cdot \hat{\Delta}_i^{\text{IV}} + \varepsilon_i \; , \quad \text{where} \quad \hat{\Delta}_i^{\text{IV}} = \sum_{j \neq i} D_{i,j} \hat{W}_j^{\text{IV}} \quad \text{and} \quad \hat{W}_j^{\text{IV}} = \hat{\pi}_n(M_j) - \frac{1}{n}\sum_{i=1}^{n} W_j \tag{E.55}$$

denotes the centered, predicted value of the treatment W_j conditional on the instrument M_j. We impose the condition

$$M_j \perp\!\!\!\perp S_j \; . \tag{E.56}$$

Observe that this condition can be interpreted as (or implied by) an instrumental variable exogeneity condition.

In this section, we sketch the derivation of a result analogous to Corollary 4.1 for this two-stage estimator. For the sake of space, we omit many of the formal details of this argument. Moreover, we leave considerations concerning consistency and inference to further work. We consider the infeasible loss function

$$L_n^{\text{IV}}(\theta) = \min_{\alpha} \sum_{i=1}^{n} (Y_i - \alpha - \theta \cdot \Delta_i^{\text{IV}})^2 \; , \tag{E.57}$$

where

$$\Delta_i^{\text{IV}} = \sum_{j \neq i} D_{i,j} W_j^{\text{IV}} \quad \text{and} \quad W_j^{\text{IV}} = \pi_n(M_j) - \mathbb{E}[W_j] \; . \tag{E.58}$$

Under regularity conditions analogous to those imposed in Theorem 2.1, and by the same steps used in the proof of that result, the exogeneity condition (E.56) implies that the population minimizer $\overline{\theta}_n^{\text{IV}}$ of the risk $\mathbb{E}[L_n^{\text{IV}}(\theta)]$ can be expressed as

$$\begin{aligned}
\overline{\theta}_n^{\text{IV}} &= \frac{\mathbb{E}[Y_i(D_{i,j} - \mathbb{E}[D_{i,j} \mid S_j])(\mathbb{E}[W_j \mid M_j] - \mathbb{E}[W_j])]}{\mathbb{E}[\operatorname{Var}(D_{i,j} \mid S_j)(\mathbb{E}[W_j \mid M_j] - \mathbb{E}[W_j])^2]} + o(n^{-1/2}) \\
&= \frac{\operatorname{Var}(M_j)}{\operatorname{Cov}(W_j, M_j)} \frac{\mathbb{E}[Y_i(D_{i,j} - \mathbb{E}[D_{i,j} \mid S_j])(M_j - \mathbb{E}[M_j])]}{\mathbb{E}[\operatorname{Var}(D_{i,j} \mid S_j)(M_j - \mathbb{E}[M_j])^2]} + o(n^{-1/2}) \\
&= \frac{\mathbb{E}[Y_i(D_{i,j} - \mathbb{E}[D_{i,j} \mid S_j])(M_j - \mathbb{E}[M_j])]}{\mathbb{E}[\operatorname{Var}(D_{i,j} \mid S_j) \operatorname{Cov}(W_j, M_j)]} + o(n^{-1/2}) \; ,
\end{aligned} \tag{E.59}$$

where we have used the fact that the instrument M_j is binary in the second equality and the exogeneity condition (E.56) again in the third equality. By the argument applied in the proof of Lemma A.2, we can express the parameter (E.59) as

$$\overline{\theta}_n^{\text{IV}} = \int_{\overline{\mathcal{D}}} \mathbb{E}\left[\lambda(\delta \mid S_j) \left(\frac{(M_j - \mathbb{E}[M_j])\partial_\delta \mathbb{E}[Y_i \mid D_{i,j} = \delta, S_j]}{\operatorname{Cov}(M_j, W_j)}\right)\right] \mathrm{d}\delta \tag{E.60}$$

$$= \int_{\overline{\mathcal{D}}} \mathbb{E}\Bigg[\lambda(\delta \mid S_j)$$

$$\partial_\delta \left(\frac{\mathbb{E}[Y_i \mid D_{i,j} = \delta, M_j = 1, S_j] - \mathbb{E}[Y_i \mid D_{i,j} = \delta, M_j = 0, S_j]}{\mathbb{E}[W_j \mid M_j = 1] - \mathbb{E}[W_j \mid M_j = 0]}\Bigg]\right) \mathrm{d}\delta \; , \tag{E.61}$$

where

$$\lambda(\delta \mid S_j) = \frac{\mathrm{Cov}(\mathbb{I}\{D_{i,j} \geq \delta\}, D_{i,j})}{\mathrm{Var}(D_{i,j} \mid S_j)} \tag{E.62}$$

and we have again used the fact that the instruments are binary to obtain the second equality.

Now, observe that the parameter

$$\frac{\mathbb{E}[Y_i \mid D_{i,j} = \delta, M_j = 1, S_j] - \mathbb{E}[Y_i \mid D_{i,j} = \delta, M_j = 0, S_j]}{\mathbb{E}[W_j \mid M_j = 1] - \mathbb{E}[W_j \mid M_j = 0]} \tag{E.63}$$

can be interpreted as the usual Wald instrumental variables estimand, associated with the treatment W_j, the instrument M_j, and the outcome

$$\mathbb{E}[Y_i \mid D_{i,j} = \delta, S_j] \ . \tag{E.64}$$

Thus, under standard exogeneity, exclusion, relevance, and monotonicity conditions, the parameter (E.63) can be expressed as a convex average of the spillover effect

$$\mathbb{E}[\partial_w Y_{i,j}(\delta, w) \mid D_{i,j} = \delta, S_j, W_j(1) > W_j(0)] \tag{E.65}$$

on the population of compilers (see, for instance, Imbens and Angrist (1994) and Angrist et al. (2000)).

E.4.2 Parallel Trends. Next, we sketch how adapt our framework to settings based on difference-in-differences designs. For the sake of simplicity, we assume that there are two time periods and that the treatment is binary. We assume that all units are untreated at time zero. Let $Y_{i,t}$ denote the outcome observed at period t and

$$Y_i^{(1)} = Y_{i,1} - Y_{i,0} \tag{E.66}$$

denote the first difference in the outcome for unit i.

Suppose that each unit i is associated with an additional error term $U_i^{(1)}$ and that the outcomes and first differences are generated by the structural equations

$$\begin{aligned} Y_{i,1} &= F_1(Z_i, Z_{-i}) \ , \quad \text{where} \quad Z_i = (W_i, S_i, U_i) \ , \quad \text{and} \\ Y_i^{(1)} &= F^{(1)}(Z_i^{(1)}, Z_{-i}^{(1)}) \ , \quad \text{where} \quad Z_i^{(1)} = (W_i, S_i, U_i^{(1)}) \ . \end{aligned} \tag{E.67}$$

This is a natural extension of Assumption 3.1. Here, the unobservable terms U_i and $U_i^{(1)}$ determine differences in period one values $Y_{i,1}$ and trends $Y_i^{(1)}$, respectively. In general, these terms can be distinct, but correlated. We define the potential trend and outcome

$$\begin{aligned} Y_{i,j}^{(1)}(\delta, w) &= F^{(1)}(Z_i, Z_j(w, S_j^{(i)}(\delta)), Z_{-i,j}) \quad \text{and} \\ Y_{i,j,1}(\delta, w) &= F_1(Z_i^{(1)}, Z_j^{(1)}(w, S_j^{(i)}(\delta)), Z_{-i,j}^{(1)}) \ , \end{aligned} \tag{E.68}$$

as before.

In this setting, a suitable parallel trends condition is given by

$$\mathbb{E}[W_j \mid \overline{S}_j^{(1)}] = \mathbb{E}[W_j], \quad \text{where} \quad \overline{S}_j^{(1)} = (S_j, U_j^{(1)}) \ . \tag{E.69}$$

In particular, the restriction (E.69) says that any of the factors S_j and $U_j^{(1)}$ that might contribute to larger or smaller values of the potential trend $Y_i^{(1)}(\delta, 0)$ are mean-independent of the treatments W_j. Observe that this condition is directly analogous to the condition (4.2) considered in the main text, where the outcomes are replaced by the first differences.

Indeed, under this assumption, by Theorem 4.1, the regression

$$Y_i^{(1)} = \alpha + \theta \cdot \Delta_i^\star + \varepsilon_i \quad \text{where} \quad \Delta^\star = \sum_{j \neq i} D_{i,j}(W_j - \mathbb{E}[W_j]) \tag{E.70}$$

identifies a convex average of the parameters

$$\begin{aligned} &\partial_\delta(\mathbb{E}[Y_{i,j}^{(1)}(\delta, 1) \mid D_{i,j} = \delta, W_j = 1, \overline{S}_j] - \mathbb{E}[Y_{i,j}^{(1)}(\delta, 0) \mid D_{i,j} = \delta, W_j = 0, \overline{S}_j]) \\ &= \partial_\delta(\mathbb{E}[Y_{i,j}^{(1)}(\delta, 1) \mid D_{i,j} = \delta, \overline{S}_j] - \mathbb{E}[Y_{i,j}^{(1)}(\delta, 0) \mid D_{i,j} = \delta, \overline{S}_j]) \\ &= \partial_\delta(\mathbb{E}[Y_{i,j,1}(\delta, 1) \mid D_{i,j} = \delta, \overline{S}_j] - \mathbb{E}[Y_{i,j,1}(\delta, 0) \mid D_{i,j} = \delta, \overline{S}_j])\ , \end{aligned} \tag{E.71}$$

reproducing a version of Corollary 4.1. In sum, if the outcome in the class of regressions under consideration takes the form of a first difference, then the mean-independence condition (E.69) can be interpreted as a parallel trends assumption and the class of estimators proposed in the main text can be applied, unchanged.

E.5 Addressing Endogenous Measures of Proximity with Instrumental Variables

In this appendix, we illustrate how to adapt the framework proposed in the main text to accommodate settings where endogeneity in a measure of proximity is addressed with an instrumental variable. We assume that, for each pair of units i, j, we observe a proximity measure

$$M_{i,j} = M(\overline{S}_i, \overline{S}_j) \tag{E.72}$$

valued on $\{0, 1\}$. For the sake of simplicity, we focus attention on the case without covariates X_i or auxiliary measures of proximity $G_{i,j}$.

We consider a two-step procedure, analogous to the standard cross-sectional two-stage least-squares estimator. In the first stage, we compute an estimate $\hat{\gamma}_n^{\text{IV}}(m)$ of the conditional expectation

$$\gamma_n^{\text{IV}}(m) = \mathbb{E}[D_{i,j} \mid M_{i,j}]\ . \tag{E.73}$$

In the second stage, we consider the regression specification

$$Y_i = \alpha + \theta \cdot \hat{\Delta}_i^{\text{IV}} + \varepsilon_i\ , \quad \text{where} \quad \hat{\Delta}_i^{\text{IV}} = \sum_{j \neq i} \hat{D}_{i,j}^{\text{IV}} W_j \quad \text{and} \quad \hat{D}_{i,j}^{\text{IV}} = \hat{\gamma}_n^{\text{IV}}(M_{i,j}) \tag{E.74}$$

denotes the predicted value of the proximity measure treatment $D_{i,j}$ conditional on the instrument $M_{i,j}$. For the sake of simplicity, we assume that the treatments satisfy the condition

$$\mathbb{E}[W_j \mid \overline{S}_j] = 0 \tag{E.75}$$

almost surely. That is, in effect, they arise from a randomized experiment. Further extensions to settings where the treatments satisfy Assumption 4.2, or the conditions considered in Appendix E.4, follow from the same arguments.

To justify the specification (E.74), we require the following analogue to Assumption 4.3.

Assumption E.7 (Additively Separable Heterogeneity)**.** *There exists a function* $\psi_n^{\mathrm{IV}}(\cdot)$ *such that*

$$\mathbb{E}[D_{i,j} \mid M_{i,j}, S_j] = \gamma_n^{IV}(M_{i,j}) + \psi_n^{IV}(S_j) \tag{E.76}$$

almost surely.

Like Assumption 4.3, Assumption E.7 stipulates that any unobserved heterogeneity in the relationship between the proximity measure and the instrument is additively separable.

In this section, we sketch the derivation of a result analogous to Corollary 4.2 for the two-stage estimator (E.74). Like in Appendix E.4, we omit many of the formal details of this argument and leave considerations concerning consistency and inference to further work. We consider the infeasible loss function

$$L_n^{\mathrm{IV}}(\theta) = \min_{\alpha} \sum_{i=1}^{n} (Y_i - \alpha - \theta \cdot \Delta_i^{\mathrm{IV}})^2 \,, \tag{E.77}$$

where

$$\Delta_i^{\mathrm{IV}} = \sum_{j \neq i} D_{i,j}^{\mathrm{IV}} W_j \quad \text{and} \quad D_{i,j}^{\mathrm{IV}} = \pi_n(M_j) - \mathbb{E}[W_j] \,. \tag{E.78}$$

Observe that Assumption E.7 implies that

$$\mathbb{E}[D_{i,j} \mid M_{i,j}] - \mathbb{E}[\mathbb{E}[D_{i,j} \mid M_{i,j}] \mid S_j] = \mathbb{E}[D_{i,j} \mid M_{i,j}, S_j] - \mathbb{E}[D_{i,j} \mid S_j] \,. \tag{E.79}$$

Thus, under regularity conditions analogous to those imposed in Theorem 2.1, and by the same steps used in the proof of that result, the condition (E.75) implies that the population minimizer $\overline{\theta}_n^{\mathrm{IV}}$ of the risk $\mathbb{E}[L_n^{\mathrm{IV}}(\theta)]$ can be expressed as

$$\overline{\theta}_n^{\mathrm{IV}} = \frac{\mathbb{E}[Y_i(\mathbb{E}[D_{i,j} \mid M_{i,j}, \overline{S}_j] - \mathbb{E}[D_{i,j} \mid \overline{S}_j])W_j]}{\mathbb{E}[(\mathbb{E}[D_{i,j} \mid M_{i,j}, \overline{S}_j] - \mathbb{E}[D_{i,j} \mid \overline{S}_j])^2 \operatorname{Var}(W_j \mid \overline{S}_j)]} + o(n^{-1/2}) \,. \tag{E.80}$$

Observe that, because the instrument $M_{i,j}$ is binary, we can write

$$\begin{aligned} \mathbb{E}[D_{i,j} \mid M_{i,j}, \overline{S}_j] - \mathbb{E}[D_{i,j} \mid \overline{S}_j] &= \eta(\overline{S}_j)(M_{i,j} - \mathbb{E}[M_{i,j} \mid \overline{S}_j]) \,, \quad \text{where} \\ \eta(\overline{S}_j) &= \mathbb{E}[D_{i,j} \mid M_{i,j} = 1, \overline{S}_j] - \mathbb{E}[D_{i,j} \mid M_{i,j} = 1, \overline{S}_j] \,. \end{aligned} \tag{E.81}$$

Thus, we have that

$$\overline{\theta}_n^{\mathrm{IV}} = \frac{\mathbb{E}[\eta(\overline{S}_j)Y_i W_j (M_{i,j} - \mathbb{E}[M_{i,j} \mid \overline{S}_j])]}{\mathbb{E}[\eta(\overline{S}_j)^2 (M_{i,j} - \mathbb{E}[M_{i,j} \mid \overline{S}_j])^2 \operatorname{Var}(W_j \mid \overline{S}_j)]} + o(n^{-1/2}) \,. \tag{E.82}$$

Moreover, again by the fact that the instrument $M_{i,j}$ is binary, we can evaluate

$$\begin{aligned} \operatorname{Cov}(Y_i W_j, M_{i,j} \mid \overline{S}_j) &= \operatorname{Var}(M_{i,j} \mid \overline{S}_j)(\mathbb{E}[Y_i W_j \mid M_{i,j} = 1, \overline{S}_j] \\ &\qquad - \mathbb{E}[Y_i W_j \mid M_{i,j} = 0, \overline{S}_j]) \\ &= \operatorname{Var}(M_{i,j} \mid \overline{S}_j)\eta(\overline{S}_j)\beta(\overline{S}_j) \,, \end{aligned} \tag{E.83}$$

where

$$\beta(\overline{S}_j) = \frac{\mathbb{E}[Y_i W_j \mid M_{i,j} = 1, \overline{S}_j] - \mathbb{E}[Y_i W_j \mid M_{i,j} = 0, \overline{S}_j]}{\mathbb{E}[D_{i,j} \mid M_{i,j} = 1, \overline{S}_j] - \mathbb{E}[D_{i,j} \mid M_{i,j} = 0, \overline{S}_j]} \; . \tag{E.84}$$

Consequently, we obtain the representation

$$\overline{\theta}_n^{\mathrm{IV}} = \frac{\mathbb{E}[\eta^2(\overline{S}_j) \operatorname{Var}(M_{i,j} \mid \overline{S}_j) \beta(\overline{S}_j)]}{\mathbb{E}[\eta(\overline{S}_j)^2 \operatorname{Var}(M_{i,j} \mid \overline{S}_j) \operatorname{Var}(W_j \mid \overline{S}_j)]} + o(n^{-1/2}) \; . \tag{E.85}$$

Observe that the parameter (E.84) can be interpreted as the usual Wald instrumental variables estimand, associated with the treatment $D_{i,j}$, the instrument $M_{i,j}$, and the outcome $Y_i W_j$. Thus, under standard exogeneity, exclusion, relevance, and monotonicity conditions (Imbens and Angrist, 1994), we can write

$$\beta(\overline{S}_j) = \int_{\overline{D}} \lambda^{\mathrm{IV}}(\delta \mid \overline{S}_j) \mathbb{E}[\partial_\delta Y_{i,j}(\delta, W_j) W_j \mid \overline{S}_j, D_{i,j}(1) > D_{i,j}(0)] \, \mathrm{d}\delta \; , \tag{E.86}$$

where the weights $\lambda^{\mathrm{IV}}(\delta \mid \overline{S}_j)$ are convex (see Angrist et al. (2000) for details). Hence, by plugging (E.86) into (E.85) and applying an argument very similar to the proof of Lemma A.2, we can conclude that

$$\overline{\theta}_n^{\mathrm{IV}} = \int_{\overline{D}} \int_{\overline{W}} \lambda^{\mathrm{IV}}(\delta, w \mid \overline{S}_j) \mathbb{E}[\partial^2_{\delta,w} Y_{i,j}(\delta, W_j) \mid \overline{S}_j, D_{i,j}(1) > D_{i,j}(0)] \, \mathrm{d}\delta \, \mathrm{d}w + o(n^{-1/2}) \; . \tag{E.87}$$

where the weights $\lambda^{\mathrm{IV}}(\delta, w \mid \overline{S}_j)$ are convex.

E.6 Additional Discussion Concerning Assumption 5.4

In this Appendix, we give addition details concerning Assumption 5.4. Recall that Assumption 5.4 stipulates that the bounds

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[(\nabla^2_{j,k} F(Z_i, Z_{-i}))^2 \mid A_{i,j} = 1, A_{i,k} = 1, Z^{(j,k)}]] = O(n^{-2} \rho_n^{-3}) \; , \\
&\mathbb{E}[\mathbb{E}[(\nabla^2_{j,k} F(Z_i, Z_{-i}))^2 \mid A_{i,j} = 1, A_{i,k} = 0, Z^{(j,k)}]] = O(n^{-2} \rho_n^{-1}) \; , \quad \text{and} \\
&\mathbb{E}[\mathbb{E}[(\nabla^2_{j,k} F(Z_i, Z_{-i}))^2 \mid A_{i,j} = 0, A_{i,k} = 0, Z^{(j,k)}]] = O(n^{-2} \rho_n)
\end{aligned} \tag{E.88}$$

hold, where

$$\nabla_j f(Z) = f(Z) - f(Z^{(j)}) \quad \text{and} \quad \nabla^2_{j,k} f(Z) = \nabla_j \nabla_k f(Z) \; , \tag{E.89}$$

denote difference operators, Z' denotes an independent copy of Z, $Z^{(j)}$ is constructed by replacing Z_j with Z'_j in Z, and $Z^{(j,k)}$ is constructed analogously, by replacing Z_j and Z_k with Z'_j and Z'_k in Z.

To unpack the restrictions implied by these conditions, observe that the quantity

$$P\{|\nabla^2_{j,k} F(Z_i, Z_{-i})| \geq C \mid A_{i,j} = 1, A_{i,k} = 1, Z^{(j,k)}\} \tag{E.90}$$

can be interpreted as the measuring the proportion of pairs of units j, k, that are proximate to a unit i, for which the curvature

$$\nabla^2_{j,k} F(Z_i, Z_{-i}) \; , \tag{E.91}$$

i.e., the change in dependence of i on j, induced by changes in k, is larger than a constant C. The bound (E.88) and Chebyshev's inequality imply that

$$\begin{aligned}&\mathbb{E}\left[P\{|\nabla^2_{j,k}F(Z_i, Z_{-i})| \geq C \mid A_{i,j}=1, A_{i,k}=1, Z^{(j,k)}\}\right]\\ &\lesssim \mathbb{E}[\mathbb{E}[(\nabla^2_{j,k}F(Z_i, Z_{-i}))^2 \mid A_{i,j}=1, A_{i,k}=1, Z^{(j,k)}]] = O(n^{-2}\rho_n^{-3})\end{aligned} \tag{E.92}$$

In other words, if $\rho_n \lesssim n^{-2/3}$, then this proportion can be on the order of a constant for all pairs of units proximate to each unit i. For larger values of ρ_n, the proportion decreases. In particular, in general, each unit i will have $n^2\rho_n^2$ pairs of neighbors. The bound (E.92) says that ρ_n^{-1} of these pairs can be associated with curvature on the order of a constant.

We conclude showing that Assumption 5.4 holds in settings where

$$Y_i = F_i\left(W_i, \frac{1}{n\rho_n}\sum_{j\neq i} D_{i,j}W_j\right), \tag{E.93}$$

where F_i is some unit-specific function. That is, each unit's outcome depends on its own treatment as well as the proportion of its "neighbors" that are treated. We have used the denominator $n\rho_n$, rather than $\sum_{j=1}^n D_{i,j}$, for the sake of simplicity. We assume that $F_i(\cdot)$ is twice continuously differentiable, with second derivative uniformly bounded by the sequence B_n. A closely related model is considered in Li and Wager (2022).

In particular, define the differences $R_j = D_{i,j}W_j - D'_{i,j}W'_j$, where $D'_{i,j}$ and W'_j are constructed with Z'_j. We have that

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[(\nabla^2_{j,k}F(Z_i, Z_{-i}))^2 \mid A_{i,j}=1, A_{i,k}=1, Z^{(j,k)}]]\\ &\leq n^{-4}\rho_n^{-4}B_n^2\mathbb{E}[\mathbb{E}[(R_jR_k)^2 \mid A_{i,j}=1, A_{i,k}=1, Z^{(j,k)}]]\end{aligned} \tag{E.94}$$

by two applications of the mean-value theorem. Thus, the bound (E.88) holds so long as the curvature bound

$$B_n \lesssim n\rho_n^{1/2} \tag{E.95}$$

holds. The second two bounds in (E.88) hold by construction.

Appendix F. Data and Additional Applications

In this Appendix, we detail our treatment of the data associated with Example 1 and give additional applications to data from Examples 2 and 3.

F.1 Bloom et al. (2013)

First, we give details of our treatment of the data associated with Example 1. Results are given in Section 7. We use data obtained from the replication package associated with Lucking et al. (2019), who increase the coverage of the data considered in Bloom et al. (2013).[38]

We follow the steps used in Bloom et al. (2013) and Lucking et al. (2019) to construct the inputs used in our analysis. There are two, main raw data components. The first components are the Compustat Fundamentals

[38] These data are available on Nicholas Bloom's website, at the link https://www.dropbox.com/scl/fi/fmtqtpuv41l6c0p30ybop/spillovers_rep.zip?rlkey=pu0bsc3scesewb3yt9y7juh9m&dl=0.

Annual and Compustat Segments datasets. From these sources, we assemble a cleaned panel of measurements of R&D expenditure, sales, market value, and product market composition for a sample of U.S. public companies. All measures of sales, value, and R&D expenditure are deflated to 1996 U.S. dollars. All sample restrictions and data cleaning steps follow Lucking et al. (2019). For each firm, we aggregate these measures over five year periods, spanning 1986 to 2005. We use the observations of the shares of sales across four digit manufacturing industries to construct our measure of product market proximity for each five year time period.

The second component is data from the NBER patent data project (Hall et al., 2001). With these data, we measure each firm's patenting activity, across USPTO technology classes, and total patent citations. Again, all sample restrictions and data cleaning steps follow Lucking et al. (2019). We aggregate these measures to the same five-year periods. We use the observations of the shares of patents across USPTO technology classes to construct our measure of technological proximity for each five year time period. Additionally, following Bloom et al. (2013), use these data to construct a measure of geographic proximity, in an analogous way. In particular, we compute the share of each firm's patents filed by inventors across a set of locations. The geographic proximity measure is then given by the uncentered correlation between these shares.

For the estimates reported in Table 7.1, we restrict attention to the same subset of firms considered in Lucking et al. (2019). Moreover, for each outcome, we omit firms associated with observations of the treatment variable and outcome variable are equal to zero, as both measures enter in logs. For the resultant samples, all missing proximity measures are imputed to zero.

F.2 Hornbeck and Moretti (2024)

Next, we detail an additional application, associated with Example 2. Recall that Hornbeck and Moretti (2024) are interested in characterizing how the spillover effects of shocks to one cities' total factor productivity (TFP) on another cities' labor market outcomes depends on the propensity to migrate between the two cities. Hornbeck and Moretti (2024) use a simple structural model to address this objective. Here, we outline how to apply the methodology developed in this paper to this problem.

F.2.1 Data. We use two sources of data. First, we use data obtained from the replication package associated with Hornbeck and Moretti (2024).[39] We use four types of data from this source as inputs into our analysis. First, we obtain measurements of the number of workers in each United States Metropolitan Statistical Area (MSA) that are employed in each two digit industry in 1980. These data were originally obtained by Hornbeck and Moretti (2024) though an IPUMS census extract. Second, we use measurements of the pairwise distances between the centroids of each MSA. Third, we obtain measurements of the bilateral migration shares between each pair of MSAs for 1975 to 1980. That is, this is this the proportion, of individuals who lived in MSA i in 1975 and migrated to another MSA, that live in MSA j in 1980. Fourth, we use a set of MSA level variables, associated with the years 1980 and 1990. These variables are the TFP, total employment, average earnings, and average home value associated with each MSA.

[39] These data were obtained from Richard Hornbeck's website, at the link https://www.dropbox.com/scl/fi/66qqjxtlzg3xmgwxjx5mi/TFP_HM_May2022_ReplicationFiles.rar?rlkey=yf4dje4zhpw3erqb3es8pp99z&dl=0.

Second, we use data the 1997 Bureau of Economic Analysis Historical Benchmark Input-Output Table.[40] These data report input output tables between two-digit industries in the United States.

F.2.2 Input-Output Flows. To the best of our knowledge, no input-output tables across MSAs are publicly available for the relevant time period. Thus, we impute shares with a simple model. Our approach is based on first constructing industry-specific gravity matrices, whose bilateral terms combine exponential distance frictions with the destination-specific intermediate-input demand implied by an input-output table. For each origin–industry pair, we row-normalize the bilateral terms to obtain destination-choice shares (the fraction of shipments from origin i in industry j going to destination i'). We then average across industries using the origin's employment mix to form the origin–destination share matrix. This, highly stylized, approach is closely related to standard models of structural gravity and market access (Redding and Venables, 2004; Head and Mayer, 2011).

Formally, let $Q_{i,j}$ denote the number of workers in MSA i that are employed in industry j. Collect these data into the matrix Q. Let $I_{j,j'}$ denote the share of industry j' inputs that are produced by industry j. Collect these data into the matrix I. Let $\mathsf{Dist}_{i,i'}$ denote the distance between MSAs i and i' in miles. We use the Distance kernel

$$\tau_{i,i'} = \exp(-\lambda \mathsf{Dist}_{i,i'}) \,, \tag{F.1}$$

where we chose the parameter λ so that the 90th percentile of $\tau_{i,i'}$ over pairs of MSAs $i \neq i'$ is equal to 0.01. We estimate intermediate demand by

$$\mathsf{ID} = QI^{\top} \,. \tag{F.2}$$

That is, the i, j component of ID denotes destination j's demand for inputs from industry g. Industry-specific destination choice shares are then constructed by

$$\pi^{(j)}_{i,i} = \frac{\tau_{i,i'} \mathsf{ID}_{i',j}}{\sum_k \tau_{i,k} \mathsf{ID}_{k,j}} \,. \tag{F.3}$$

We then aggregate across industries by employment mix through

$$\tilde{\Pi}_{i,i'} = \sum_j S_{i,j} \pi^{(j)}_{i,i} \quad \text{where} \quad S_{i,j} = Q_{i,j} / \sum_k Q_{i,k} \,. \tag{F.4}$$

Our final estimate $\Pi_{i,i'}$ of the proportion of intermediates purchased by MSA i that are produced in MSA i' is obtained by row-normalizing the shares (F.4).

Figure F.1 displays the joint distributions of the trade flow, bilateral distance, and migration probability between each pair of distinct MSAs. By construction, pairs of MSAs that are geographically close are closely connected by trade. However, for pairs of MSAs that are relatively close, there is a considerable amount of variation in trade flows, driven by heterogeneity in intermediate demand and employment. Likewise, pairs of MSAs that are closely connected by trade have large migration flows and the migration flows for pairs of MSAs that have negligible trade are also negligible.

[40] These data were obtained from the BEA's webpage, at the link `https://www.bea.gov/industry/historical-benchmark-input-output-tables`, on May 12th, 2025.

TABLE F.1. Application to Productivity Spillover Estimation

		Employment	Wages	Home Value
Unadjusted		0.618	0.392	0.229
		(1.556)	(0.880)	(3.219)
Resid. Treatment		-0.619	-0.244	-0.336
		(1.186)	(0.878)	(3.087)
Resid. Proximity	Distance	-1.456	-0.325	-0.151
		(1.205)	(0.787)	(3.214)
	Distance + IO	-1.507	-0.311	-0.078
		(1.204)	(0.807)	(3.292)
Number of MSAs		193	193	193

Notes: Table F.1 displays values of the estimator $\hat{\theta}_n^{\star}$ for the application to data from Hornbeck and Moretti (2024), using the choices for the outcome variable. The standard error $\hat{\sigma}_n$, defined in Algorithm 1, is displayed in parentheses below each estimate. Further details are given in Appendix F.2. In each case, the treatment variable W_i denotes the total factor productivity growth for MSA i between 1980 and 1990 and the proximity measures $D_{i,j}$ are probability that a migrant from MSA i relocates to MSA j between 1975 and 1980. The outcomes are differences between 1980 and 1990, and enter in logs. In the first row, neither the treatments nor proximity measures have been residualized. In the second row, the treatments are residualized using the values of total factor productivity and the three outcomes in 1980. In the remaining rows, the proximity measure has been residualized using various measures of proximity, in addition to the covariates used to residualize the treatments. The estimators $\hat{\pi}_n(\cdot)$ and $\hat{\gamma}_n(\cdot)$ are both obtained from linear regressions.

F.2.3 Analysis and Results. In this application, the treatment W_i denotes the TFP growth in MSA i between 1980 and 1990. The proximity measure of interest $D_{i,j}$ is the share of people who live in MSA j in 1975 and live in a different MSA in 1980 that live in MSA j in 1980. That is, $D_{i,j}$ measures the probability that a migrant from MSA i relocates to MSA j. We consider three choices of outcomes Y_i: the differences between the total employment, averages wages, and average home values between 1980 and 1990. We use the logarithm of each outcome variable. Results are displayed in Table F.1.

The first row displays the realized value of the estimator $\hat{\theta}_n$ associated with each outcome, where neither the treatment nor proximity measures have been residualized. To address endogeneity in the treatments, in the second row, we residualize the treatments with the values of each MSA's total factor productivity, as well as each of the outcome variables, in 1980.[41] The sign of the coefficient changes for each outcome.

Now, our interest is in characterizing how changes to the propensity to migrate between two MSAs change the spillover effects of TFP growth. The hypothesis is that, when productivity growth increases labor demand in MSA j, migration from MSA i is encouraged, reducing labor supply and increasing wages. Migration probabilities $D_{i,j}$ depend on many features of both the MSAs i and j. As a consequence, Assumption 3.2 is

[41] Hornbeck and Moretti (2024) consider "shift-share" instruments, based on exposures to industry level changes to productivity, stock market returns, export markets, and patenting activity. We do not consider these instruments because they are spatially correlated. That is an analogue of Assumption 4.2 would be violated.

not immediately applicable. Instead, we apply Assumption E.4, stated in Appendix E.2. In particular, we assume that probability of migrating from i to j is a function of each MSA's geographic coordinates $S_{i,0}$ and $S_{j,0}$, as well as the utilities $S_{i,1}$ and $S_{j,1}$ associated with living and working in the respective MSAs. To apply Assumption E.4, we assume that the probability of migrating from i to j is monotonically decreasing in the utility $S_{j,1}$ associated with living in MSA j. Thus, the potential outcome $Y_{i,j}(\delta, w)$ denotes the value of MSA i's outcome associated with the intervention that sets TFP growth in MSA j to w and changes the probability of migrating from i to j to δ by changing the utility $S_{j,1}$.

To operationalize this assumption, it is essential to control for differences in other features of the unit i that might otherwise affect the value of the migration probability $D_{i,j}$. See Appendix E.2 for further discussion. Accordingly, the third row of Table F.1 displays revised estimates, where the migration probabilities have been residualized with flexible functions of the distance between MSAs.[42] This change decreases the coefficients associated with employment and wages and increases the coefficient associated with home values.

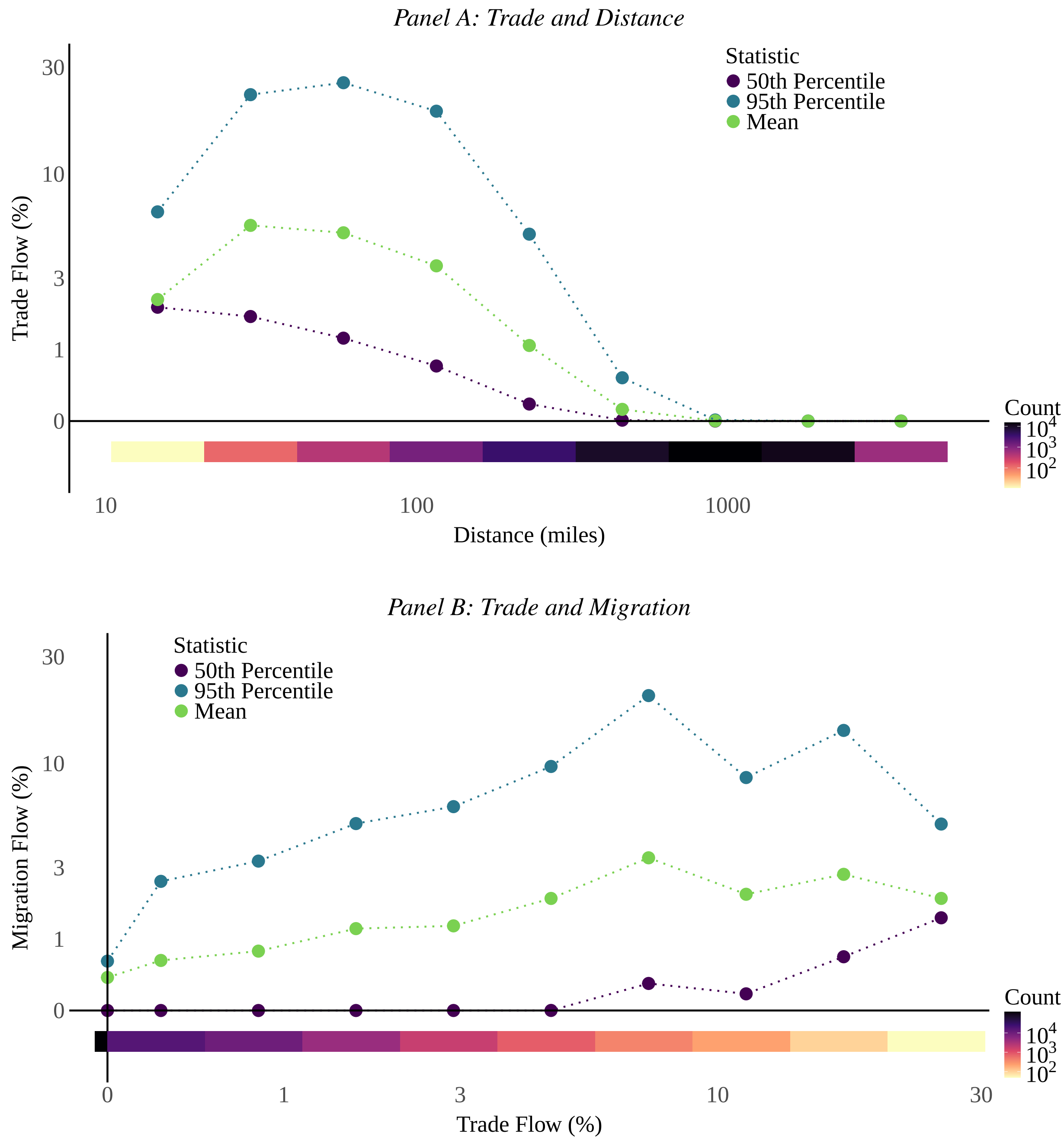
Moreover, migration probabilities are likely to be associated with other channels that otherwise mediate spillover effects. For instance, MSAs that are closely connected by migration are likely to be closely connected by trade, and TFP growth in MSA i might increase demand for intermediate inputs produced by city i, increasing labor demand and wages. To mitigate this source of confounding, in the fourth row of Table F.1 we additionally control for flexible functions of the origin-destination trade flows constructed in the previous section.[43] The changes in the estimates are more modest.

As would be expected, we find a negative association between migration probabilities and the spillover effect of TFP changes on employment. However, in contrast to estimates obtained in Hornbeck and Moretti (2024), we find negative associations with wages and home values as well. We caution that all estimates are statistically insignificant at conventional levels. Greater power might be attained if analogous exercises could be implemented at lower levels of geographic aggregation, such as counties or commuting zones.

[42] In particular, we residualize the migration probabilities with indicators that the distance between the two MSAs is less than the 0.5th, 1st, 2.5th, and 5th percentiles of the distribution of the distance between MSAs.

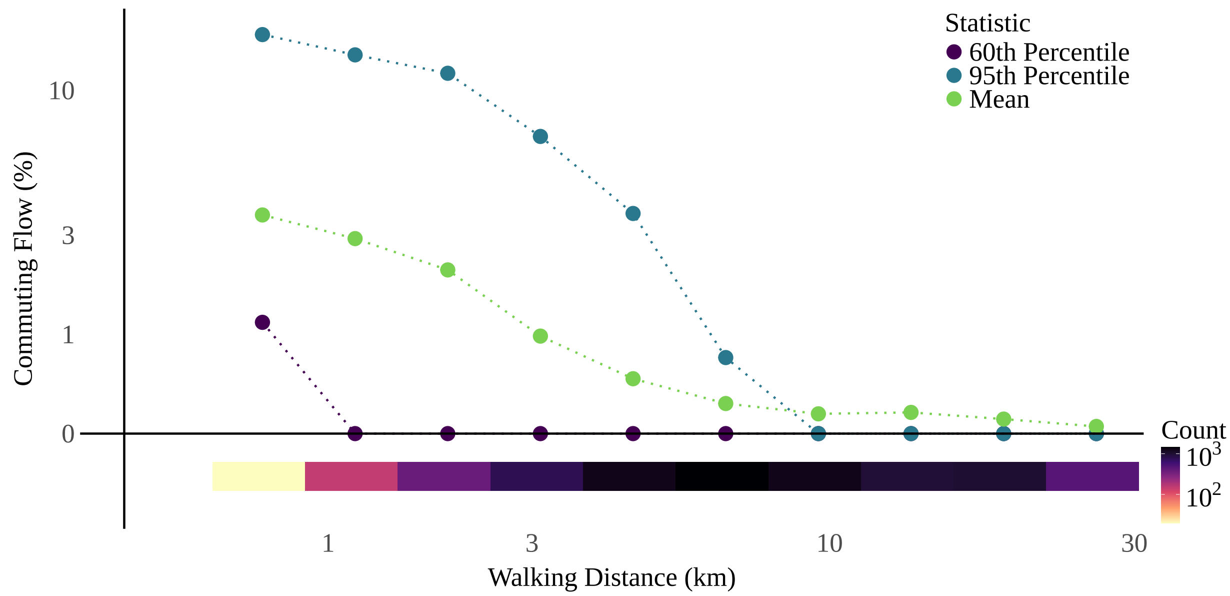
[43] We additionally residualize the migration probabilities with indicators that share of goods that used in MSA i are produced in MSA j greater than the 99.5th, 99th, 97.5th, and 95th percentiles of the distribution of this quantity.

FIGURE F.1. Trade, Migration, and Distance



Notes: Figure F.1 displays features of the joint distributions of the trade flow, bilateral distance, and migration probability between each pair of distinct MSAs. The measures are computed used data obtained from the replication package associated with Hornbeck and Moretti (2024). The x-axes measure distance between MSAs in miles and the proportion of intermediates purchased by MSA i that are produced in MSA i', respectively, and have been partitioned into bins. The y-axes display the proportion of intermediates purchased by MSA i that are produced in MSA i' and the proportion of workers from MSA i who migrate to MSA i', respectively. Details on the construction of these measures are given in Appendix F.2. The means, 50th percentiles, and 90th percentiles of each statistic within each bin are displayed using green, purple, and blue dots and are measured relative to the y-axis. A heatmap measuring the number of pairs of units in each bin is displayed below both the x-axis.

FIGURE F.2. Commuting and Distance



Notes: Figure F.2 displays features of the joint distributions of the commuting flows and bilateral distance for the sample of woredas considered in Franklin et al. (2024). The x-axis measures the walking distance between each pair of woredas, and has been partitioned into bins. The y-axis displays the probability that a worker who lives in woreda i commutes to woreda j. The means, 60th percentiles, and 90th percentiles of each statistic within each bin are displayed using green, purple, and blue dots and are measured relative to the y-axis. A heatmap measuring the number of pairs of units in each bin is displayed below both the x-axis.

F.3 Franklin et al. (2024)

Finally, we detail our application associated with Example 3, using data from Franklin et al. (2024). Franklin et al. (2024) develop a model of urban spatial equilibrium in which the labor market response to a public works program depends on commuting flows. They use this model to argue that the the indirect effects of the program are larger for pairs of neighborhoods that are closely connected by commuting. In this section, we detail how to apply the methodology developed in this paper to this problem.

F.3.1 Data. We use data obtained from the replication package associated with Franklin et al. (2024).[44] Franklin et al. (2024) consider a public works program implemented in Addis Ababa, Ethiopia. The program provided guaranteed public work to targeted households and was randomized across neighborhoods (referred to as "Woredas"). The treatment W_j indicates that neighborhood j has been assigned to the program. The outcome Y_i denotes the log average earnings in neighborhood i. The proximity measure of interest $D_{i,j}$ is the pretreatment probability that a worker who works in neighborhood i lives in neighborhood j. Figure F.2 displays the joint distribution between these commuting probabilities and the walking distance between each pair of neighborhoods.

[44]This replication package is available from ICPSR at the link `https://www.openicpsr.org/openicpsr/project/195483/version/V1/view`.

TABLE F.2. Application to Public Works Spillover Estimation

		Commuting	Dist. < 2 km
Unadjusted		0.396	0.011
		(0.439)	(0.062)
Resid. Treatment		0.542	0.018
		(0.539)	(0.057)
Resid. Proximity	Distance	0.510	
		(0.481)	
	Commuting		0.002
			(0.056)
Number of Woredas		90	90

Notes: Table F.2 displays values of the estimator $\hat{\theta}_n^\star$ for the application to data from Franklin et al. (2024). The standard error $\hat{\sigma}_n$, defined in Algorithm 1, is displayed in parentheses below each estimate. Further details are given in Appendix F.3. In each case, the treatment variable W_i indicates assignment to treatment and the outcome Y_i measures the log average wage. We consider two proximity measures of interest: the probability of commuting from Woreda j to Woreda i and an indicator that the walking distance between the two Woredas is less than two kilometers. In the first row, neither the treatments nor proximity measures have been residualized. In the second row, the treatments are residualized using sub-city indicators. In the remaining rows, the proximity measure of interest has been residualized using various additional measures of proximity, in addition to the covariates used to residualize the treatments. The estimators $\hat{\pi}_n(\cdot)$ and $\hat{\gamma}_n(\cdot)$ are both obtained from linear regressions.

F.3.2 Results. The first row of the first column of Table F.2 displays the coefficient $\hat{\theta}_n$ obtained from the regression specification

$$Y_i = \alpha + \theta \cdot \Delta_i + \varepsilon_i \ , \quad \text{where} \quad \Delta_i = \sum_{j \neq i} D_{i,j}(W_j - \mathbb{E}[W_j]) \ . \tag{F.5}$$

The second row of this column gives an analogous estimate, where the treatments have been residualized on indicators for a higher level of geographic aggregation (referred to as a "sub-city"). A very similar specification is reported in Franklin et al. (2024). The main difference is that they include each neighborhood's own treatment W_i in the covariate Δ_i with the weight $D_{i,i} = 1$. In Franklin et al. (2024), this specification approximates the reduced form of a model of spatial equilibrium.

We have shown that such specifications admit a nonparametric interpretation, as targeting the effect of the proximity measure $D_{i,j}$ on the intensity of the spillover effect of W_j on Y_i. The hypothesis is that individuals who participate in the program no longer commute to other neighborhoods, reducing their private sector labor supply and increasing wages.

Observe that, like the application considered in Appendix F.2, commuting probabilities will not satisfy Assumption 3.2. Instead, we again apply Assumption E.4. That is, we assume that commuting probabilities are functions of neighborhoods' locations and expected wages, and that the probability of commuting from j to i is decreasing in the expected wage in neighborhood j. Thus, the potential outcome $Y_{i,j}(\delta, w)$ denotes the value of neighborhood i's post-treatment wages associated with the intervention that sets j's treatment status to w and changes the probability of commuting from j to i by changing the pre-treatment wage in neighborhood j. To operationalize this assumption, we must control for other features of the unit i that might otherwise affect the commuting probability $D_{i,j}$. Thus, the third row of the first column displays estimates where the commuting probabilities have been residualized using flexible functions of the walking distance between neighborhoods. All three estimates are insignificant at conventional levels.

The second column of Table F.2 displays analogous results, where the proximity measure of interest is an indicator that the walking distance is smaller than two kilometers. Here, as well, all results are insignificant at conventional levels.