Robust Inference in Locally Misspecified Bipartite Networks[*]

Luis E. Candelaria[†]   Yichong Zhang[‡]

March 2024

Abstract

This paper introduces a methodology to conduct robust inference in bipartite networks under local misspecification. We focus on a class of dyadic network models with misspecified conditional moment restrictions. The framework of misspecification is local, as the effect of misspecification varies with the sample size. We utilize this local asymptotic approach to construct a robust estimator that is minimax optimal for the mean square error within a neighborhood of misspecification. Additionally, we introduce bias-aware confidence intervals that account for the effect of the local misspecification. These confidence intervals have the correct asymptotic coverage for the true parameter of interest under sparse network asymptotics. Monte Carlo experiments demonstrate that the robust estimator performs well in finite samples and sparse networks. As an empirical illustration, we study the formation of a scientific collaboration network among economists.

1 Introduction

Bipartite networks are embedded in a wide range of economic interactions that entail bilateral relations, for example, those between exporters and importers, buyers and sellers, and scientists and research teams. Dyadic regression with two-sided heterogeneity represents a leading approach for analyzing bipartite networks as it can adequately account for dyad dependence and link sparsity (Graham 2020). While under correct model specification, it yields valid econometric inference for a structural parameter of interest, its performance under local misspecification has not been explored. To the best of our knowledge, this is the first paper to study the effects of local misspecification on a dyadic regression model for bipartite networks.

In this paper, we develop a methodology to conduct robust inference in bipartite networks under local misspecification. We focus on a class of dyadic models characterized by conditional moment restrictions when model misspecification prevents these restrictions from holding exactly. Some situations in dyadic network models that lead to the failure of these moment restrictions include incorrect specification of parametric assumptions, ignoring preferences for latent homophily, and excluding network externalities. Our framework is one of local misspecification, as the magnitude of misspecification vanishes as the sample size grows. This modelling approach is intended as an asymptotic device, which is particularly useful to provide a tractable approximation for the effect of the model misspecification. Moreover, it enables us to perform inference on the structural parameter of interest rather than on a pseudo-true parameter.

We propose a robust estimator that accounts for the local misspecification on the conditional moment restrictions. To construct this estimator, we consider a sieve approximation to form unconditional moment restrictions that grow in number with the sample size. The robust estimator is then computed as a one-step refinement, which improves on an initial estimator that ignores the model misspecification by incorporating an adjustment term that minimizes the sensitivity of the inferential procedure to local deviations in the unconditional moment restrictions. In particular, this adjustment term is chosen optimally to minimize the worst-case mean squared error (MSE). As a consequence, the robust estimator is minimax optimal in the sense that given the worst-case bias in a neighbourhood prespecified by the researcher, the robust estimator attains the minimum MSE within a class of regular estimators (Newey 1990).

The resulting confidence intervals are 'honest', as they are widened to account for the worst-case bias within a neighborhood of misspecification (Donoho 1994; Armstrong and Kolesár 2021; Bonhomme and Weidner 2022). This construction ensures that the bias-aware confidence intervals contain the structural parameter of interest under both correct specification and local misspecification with a prespecified coverage probability. The neighborhood of misspecification is determined ex-ante by the researcher but can be adjusted to conduct sensitivity analysis on the structural parameter of interest (Conley, Hansen, and Rossi 2012). Additionally, we show that in our setting,

the misspecification neighborhood takes on a simple form, which enables us to obtain a closed-form expression for the adjustment term. In Section 2, we describe in detail the construction of the robust estimator and bias-aware confidence intervals.

To illustrate the performance of our methodology in an empirical application, we study the formation of a scientific collaboration network among economists. While collaboration among scientists is pivotal for research productivity and innovation, the factors that determine the creation of these connections have remained underexplored (Goyal, Van Der Leij, and Moraga-González 2006; Anderson and Richards-Shubik 2022; Hsieh, Konig, Liu, and Zimmermann 2022). In addition, the existing approaches used to analyze these collaborations rely on strong parametric assumptions or limit the degree of unobserved heterogeneity, making them susceptible to model misspecification. To address this gap, we implement our robust methodology to study the factors driving the formation of scientific collaborations among economists. The web of collaborations is described as a bipartite network connecting scientists and research projects, and it is constructed using the universe of published papers in the top American economic journals of general interest during the period of 2000 to 2006. We describe in detail the dataset and network construction in Section 6.

As parameters of interest, we focus on those capturing assortative matching between scientists and the attributes of a research project. Additionally, we estimate global features of the network, such as the average out-degree. Our analysis documents that homophily in the research field of expertise is a strong predictor for link formation. Moreover, highly productive scientists are, on average, more likely to collaborate on higher-impact projects, measured by aggregate citations. The results also show that, during the sample period, female authors are more likely to sort in teams with a higher share of female authors. Finally, the evidence shows that our methodology leads to a reduction on the MSEs as predicted by out theoretical results.

Our paper is related to the literature on robust statistics (Huber 1964; Huber and Ronchetti 2009; Rieder 1994), sensitivity analysis (Leamer 1985; Imbens 2003; Chen, Tamer, and Torgovitsky 2011; Nevo and Rosen 2012; Masten and Poirier 2020, 2021), and local misspecification (Newey 1985; Conley et al. 2012; Kitamura, Otsu, and Evdokimov 2013; Bugni, Canay, and Guggenberger 2012; Andrews, Gentzkow, and Shapiro 2017, 2020; Bugni and Ura 2019; Armstrong and Kolesár 2021; Bonhomme and Weidner 2022; Armstrong, Weidner, and Zeleneev 2022; Armstrong, Kline, and Sun 2023; Christensen and Connault 2023). The most closely related papers to ours are those by Andrews et al. (2017); Armstrong and Kolesár (2021); Bonhomme and Weidner (2022) and Christensen and Connault (2023), who rely on local misspecification to conduct sensitivity analysis. In contrast to these papers, this work focuses on a class of models characterized by *conditional* moment restrictions, which is not nested within their settings. Moreover, the object of study is entirely different as we analyze dyadic models for bipartite networks. On a technical side, bipartite network exhibit patterns of dyadic dependence that preclude us from using efficiency results invoked by the previous papers (cf. Kitamura et al. 2013; Bonhomme and Weidner 2022).

3

The paper also contributes to the literature on dyadic regression (Graham 2017; Tabord-Meehan 2019; Davezies, D'Haultfœuille, and Guyonvarch 2021; Menzel 2021; Graham, Niu, and Powell 2021, 2019; Graham 2022). The closest paper to ours is the one by Graham (2022). In this paper, the author derives the asymptotic properties of a logistic regression under sparse network asymptotics and shows that under global misspecification, the pseudo-true parameter minimizes a Kullback–Leibler Information Criterion. In contrast, we construct a one-step robust estimator for a dyadic regression under local misspecification. Our methodology differs substantially as we make use of local approximations to conduct inference on the true structural parameter rather than on a pseudo-true parameter. Notably, our estimator provides robust inference in dyadic models with degree heterogeneity, even in the presence of network externalities (cf. Pelican and Graham 2020, and see de Paula (2020) for a survey on network models). On a technical note, our paper relies on the Yurinskii-type coupling under dyadic data with an increasing dimension of sieve bases. Our paper complements the kernel-based nonparametric estimation methods considered by Graham et al. (2021, 2019).

The rest of the paper is organized as follows. In Section 2, we introduce the class of dyadic regressions with local misspecification. We also show how to construct the robust estimator and the bias-aware confidence intervals. In Section 3, we prove that the bias-aware confidence intervals have the correct asymptotic size control. In Section 4, we show that the robust estimator is minimax-MSE optimal. Section 5 provides Monte Carlo experiments. Section 6 presents an empirical application. Additional results and proofs are collected in the appendices.

Notation. Throughout the paper, we use $||\cdot||_2$, $||\cdot||_{op}$, $||\cdot||_F$ to denote the ℓ_2-norm of a vector, the operator norm of a matrix, and the Frobenius norm of a matrix, respectively. For an positive integer N, we denote $[N]$ as $1, \cdots, N$.

2 Model

We consider a bipartite network of scientific collaborations with N scientists and M projects. For any $i \in [N]$, let X_i and A_i denote vectors of scientist i-specific observed and unobserved attributes. Similarly, for any $j \in [M]$, let W_j and B_j denote vectors of project j-specific observed and unobserved attributes. For any pair $i \in [N]$ and $j \in [M]$, U_{ij} denotes a scientist-project-specific unobserved attribute. The total sample size is denoted by $n = N + M$.

For any finite n, the bipartite network $[Y_{ij}]_{i\in[N],j\in[M]}$ is determined according to the X-W-exchangeable graphon

$$Y_{ij} = h_n\left(v_{ij}, X_i, W_j, A_i, B_j, U_{ij}\right), \tag{2.1}$$

4

where h_n is a measurable function that maps $(v_{ij}, X_i, W_j, A_i, B_j, U_{ij}) \mapsto \{0, 1\}$, and v_{ij} represents an i.i.d. latent mixing factor that is unidentified, as discussed in Graham (2022). In our framework, v_{ij} represents a misspecification component that will affect the conditional probability of establishing a collaboration link as specified by (2.2). The graphon in (2.1) is used as a nonparametric data generating process for the bipartite network of scientific collaborations with N scientists and M projects. Intuitively, (2.1) states that $Y_{ij} = 1$ if scientist i collaborates in project j, and 0 otherwise.

We assume that the conditional probability of establishing a collaboration link is given by

$$\mathbb{P}\left[Y_{ij} = 1 \mid X_i, W_j\right] = \int \Lambda\left(\alpha_{0,n} + Z_{ij}^\top \beta_0 + n^{-1/2}v\right) \pi_{ij}(v)dv, \tag{2.2}$$

where $\Lambda(u) = \exp(u)/(1 + \exp(u))$ is the standard logistic CDF, $Z_{ij} = z(X_i, W_j) \in \Re^{d_z}$ is a vector-valued distance of the observed attributes X_i and W_j, and v_{ij} is a random variable with PDF given by $\pi_{ij}(\cdot)$. The distribution $\pi_{ij}(\cdot)$ can depend on (X_i, W_j), it integrates to one $\int \pi_{ij}(v)dv = 1$, and it has mean equal to $\int v\pi_{ij}(v)dv = \eta_{ij}$. Note that η_{ij} is a function of (X_i, W_j), and thus, it is indexed by (i, j).

The coefficient $\alpha_{0,n}$ depends on the sample size n to accommodate for sparse network asymptotics. As in Graham (2022), we set $\alpha_{0,n} \equiv \log(\alpha_0/n)$, which jointly with the specification of the conditional distribution in (2.2) and Assumption 1 below, ensure that in the limit the network is sparse. Consequently, the average in-degree and out-degree of this network will be bounded. That is, for any $i \in [N]$ and $j \in [M]$, let $\rho_n \equiv \mathbb{E}[Y_{ij}]$ denote the marginal probability of forming a link, then $\lim_{n\to\infty} M\rho_n = \lambda_0^a < \infty$ and $\lim_{n\to\infty} N\rho_n = \lambda_0^p < \infty$.

The linear index $Z_{ij}^\top \beta_0$ in (2.2) captures the contribution that sharing similar observed attributes has on a scientist i's decision to collaborate in the project j. In other words, this component represents an assortative matching mechanism for the formation of a scientific collaboration network. Consequently, the coefficient β_0 is interpreted as a homophily parameter.

Finally, notice that the latent component v_{ij} affects the probability of establishing scientific collaboration. In this setting, the component $n^{-1/2}v$ represents a source of misspecification. The misspecification design is local as it is indexed by n and vanishes away as the sample size increases at a $\sqrt{n}$-rate. In Section 3, we show that at this rate, the local misspecification induces a bias term into the asymptotic distribution of interest. Moreover, the effect of misspecification is determined by the first moment of the distribution $\pi_{ij}(\cdot)$ of the latent factor v_{ij}.

Remark 1 (Conditional probability). *Alternatively, we also consider the equivalent specification for the true conditional probability*

$$\mathbb{P}\left[Y_{ij} = 1 \mid X_i, W_j\right] = \Lambda\left(\alpha_{0,n} + Z_{ij}^\top \beta_0 + n^{-1/2}\eta_{ij}\right), \tag{2.3}$$

where $\Lambda(\cdot)$ and $Z_{ij} \in \Re^{d_z}$ are as defined above. The factor $\eta_{ij} = \eta(X_i, Z_j)$ is a scalar random

5

variable. In fact, (2.2) and (2.3) lead to the same local misspecification in our setting because in both cases, we have

$$\sqrt{n}(\mathbb{P}\left[Y_{ij} = 1 \mid X_i, W_j\right] - \Lambda\left(\alpha_{0,n} + Z_{ij}^\top \beta_0\right)) = \eta_{ij} + o_p(1).$$

Therefore, all results in this paper hold without any change if the local misspecification is formulated as (2.3).

We now provide four empirical illustrations that can be analyzed using the local misspecification framework characterized by (2.1) and (2.2).

Example 1 (Latent homophily). *Suppose the bipartite network $[Y_{ij}]_{i\in[N],j\in[M]}$ is generated according to*

$$Y_{ij} = \mathbf{1}\left[\alpha_{0,n} + Z_{ij}^\top \beta_0 + n^{-1/2} v_{ij} + U_{ij} \geq 0\right] \tag{2.4}$$

where $v_{ij} = \psi(v_i, v_j)$, $v_i, v_j, \in \Re^{d_v}$ are unobserved fixed effects that are potentially correlated with X_i and W_j, and $\psi(\cdot, \cdot)$ is a known function. Common choices of $\psi(\cdot, \cdot)$ include (i) nonparametric homophily, $\psi(v_i, v_j) = -\|v_i - v_j\|_2$, and (ii) interactive fixed effects, $\psi(v_i, v_j) = v_i^\top v_j$. Note that if U_{ij} has a Logistic CDF, then the conditional probability of establishing a link is equal to (2.2). In this setting, the local misspecification in (2.2) is induced by the presence of latent homophily v_{ij}.

Example 2 (Semiparametric distribution). *Consider the graphon $[Y_{ij}]_{i\in[N],j\in[M]}$ is generated according to (2.4), where U_{ij} is an i.i.d. random term with Logistic CDF and v_{ij} is a latent mixing component with unknown distribution $\pi_{ij}(\cdot)$ that has mean η_{ij}, which can depend on (X_i, W_j). In this case, $n^{-1/2} v_{ij} + U_{ij}$ represents a composite error term with an unknown distribution that renders the estimation of $(\alpha_{0,n}, \beta_0)$ nonparametric. In this setting, the presence of local misspecification is due to specifying incorrectly the distribution of U_{ij}.*

Example 3 (Functional form misspecification). *Suppose the graphon $[Y_{ij}]_{i\in[N],j\in[M]}$ is generated according to*

$$Y_{ij} = \mathbf{1}\left[U_{ij} \leq f(X_i, W_j)\right]$$

where U_{ij} is an i.i.d. random variable with Logistic CDF and $f(\cdot, \cdot)$ is an unknown measurable function of (X_i, W_j). Assuming that $U_{i,j}$ has a Logistic distribution is without loss of generality if we allow the functional form of $f(\cdot, \cdot)$ to be flexible. The latent component v_{ij} in (2.2) can be interpreted as a local specification error from specifying a linear index model. That is, $f(X_i, W_j) = \alpha_{0,n} + Z_{ij}^\top \beta_0 + n^{-1/2} v_{ij}$, where the distribution of v_{ij} given by π_{ij} depends on (X_i, W_j).

Example 4 (Network externalities). *Suppose that the graphon $[Y_{ij}]_{i\in[N],j\in[M]}$ is an equilibrium outcome from a strategic network formation model with incomplete information as in Leung (2015) or Ridder and Sheng (2020). Under Assumption 1, and regularity conditions in Leung (2015), there exists a symmetric Bayesian Nash equilibrium (BNE) under which edges form independently*

6

conditionally on the observed attributes $\mathcal{X}_N$ and $\mathcal{W}_M$, and the conditional linking probabilities are X-W- exchangeable. In particular, the symmetric BNE yields the following network formation equation

$$Y_{ij} = \mathbf{1}\left[\alpha_{0,n} + Z_{ij}^\top \beta_0 + \gamma_n\left(\mathbb{E}\left[\mathcal{E}_{ij} \mid \mathcal{X}_N, \mathcal{W}_M, \tau^*\right]\right) \geq U_{ij}\right]$$

where $\mathcal{E}_{ij} \equiv \mathcal{E}_{ij}\left(Y_n, \mathcal{X}_N, \mathcal{W}_M\right)$, $Y_n = [Y_{ij}]_{i\in[N],j\in[M]}$, $\mathcal{X}_N = (X_1, \cdots, X_N)$, $\mathcal{W}_M = (W_1, \cdots, W_M)$, and τ^* denotes the equilibrium belief profile. Here, $\mathcal{E}_{ij}$ represents an endogenous network statistic such as (i) in-degree of project j, $\frac{1}{N}\sum_{k\neq i} Y_{kj}$; (ii) out-degree of scientist i, $\frac{1}{M}\sum_{k\neq j} Y_{ik}$; or (iii) average transitivity, $\frac{1}{NM}\sum_{k\neq i}\sum_{l\neq j} Y_{kj}Y_{kl}$. In this setting, by assuming γ_n scales with $n^{-1/2}$ (i.e., $\gamma_n = n^{-1/2}\gamma$), the misspecification component $v_{ij} \equiv \gamma\mathbb{E}\left[\mathcal{E}_{ij} \mid \mathcal{X}_N, \mathcal{W}_M, \tau^*\right]$, and we assume it has a conditional density $\pi_{ij}(\cdot)$ given (X_i, W_j). Then, we have

$$\begin{aligned}
\mathbb{P}\left(Y_{ij} = 1 \mid X_i, W_j\right) &= \mathbb{E}\left[\mathbb{P}\left(Y_{ij} = 1 \mid \mathcal{X}_N, \mathcal{W}_M\right) \mid X_i, W_j\right] \\
&= \mathbb{E}\left[\Lambda(\alpha_{0,n} + Z_{ij}^\top \beta_0 + n^{-1/2}v_{ij}) \mid X_i, W_j\right] \\
&= \int \Lambda(\alpha_{0,n} + Z_{ij}^\top \beta_0 + n^{-1/2}v)\pi_{ij}(v)dv,
\end{aligned}$$

which is just (2.2).

As parameter of interest, we focus on the scalar $\Psi_{0,n}$ that is defined as

$$\Psi_{0,n} = \mathbb{E}\int \gamma_n\left(D_{ij}, n^{-1/2}v, \theta_{0,n}\right)\pi_{ij}(v)dv, \tag{2.5}$$

where γ_n is a known measurable function that maps $\left(D_{ij}, n^{-1/2}v, \theta_{0,n}\right) \mapsto \Re$, with $D_{ij} \equiv (X_i^\top, W_j^\top)^\top$ and $\theta_{0,n} \equiv (\alpha_{0,n}, \beta_0)$. Notice that the misspecification term $n^{-1/2}v$ can affect $\Psi_{0,n}$ directly, and thus, $\Psi_{0,n}$ is allowed to be misspecified itself (cf. Armstrong and Kolesár 2021). Next, we provide some examples of the parameter of interest as described by (2.5).

Example 5 (Individual Parameters)**.** *If the research interest is to make inference on the homophily parameters β_0 or the intercept $\alpha_{0,n}$, the parameter of interest can be defined as $\Psi_{0,n} = \theta_{0,n}^{(k)}$, where $\theta_{0,n}^{(k)}$ denotes the kth element in $\theta_{0,n} \in \Re^{d_z+1}$.*

Example 6 (Network Statistics)**.** *The parameter of interest $\Psi_{0,n}$ can be defined to capture global features of the network, such as the average in-degree of the network. In this case,*

$$\Psi_{0,n} = N\mathbb{E}\int\left[\Lambda\left(\alpha_{0,n} + Z_{ij}^\top \beta_0 + n^{-1/2}v\right)\right]\pi_{ij}(v)dv,$$

which represents the average number of scientists that participate in a randomly selected project. Similarly, the parameter of interest could be the average out-degree of the network, i.e,

$$\Psi_{0,n} = M\mathbb{E}\int\left[\Lambda\left(\alpha_{0,n} + Z_{ij}^\top \beta_0 + n^{-1/2}v\right)\right]\pi_{ij}(v)dv,$$

which captures the average number of projects in which a randomly selected scientist participates.

Example 7 (Average Effects)**.** *The parameter of interest $\Psi_{0,n}$ could capture the average marginal effect. For example, if $Z_{ij}^{(1)}$ is assumed to be continuous, then $\Psi_{0,n}$ can be defined as*

$$\Psi_{0,n} = n\mathbb{E}\int \frac{\partial}{\partial Z_{ij}^{(1)}}\Lambda\left(\alpha_{0,n} + Z_{ij}^\top\beta_0 + n^{-1/2}v\right)\pi_{ij}(v)dv.$$

Alternatively, if the object of interest is to do inference on an average partial effect, then $\Psi_{0,n}$ can be defined as

$$\Psi_{0,n} = n\,\mathbb{E}\int \frac{\partial}{\partial\beta}\Lambda\left(\alpha_{0,n} + Z_{ij}^\top\beta_0 + n^{-1/2}v\right)\Big|_{\beta=\beta_0}\pi_{ij}(v)dv.$$

2.1 Minimum-MSE Estimator

The robust estimator $\hat{\Psi}(\hat{\kappa})$ of $\Psi_{0,n}$ is defined as the following one-step regular estimator

$$\hat{\Psi}(\hat{\kappa}) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\gamma_n(D_{ij}, 0, \hat{\theta}_{\text{initial}}) + \hat{\kappa}^\top\left[\frac{n}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(Y_{ij} - \Lambda(R_{ij}^\top\hat{\theta}_{\text{initial}}))H_{ij}\right], \qquad (2.6)$$

where $R_{ij} \equiv (1, Z_{ij}^\top)^\top \in \Re^{d_z+1}$, $\hat{\theta}_{\text{initial}}$ is a $\sqrt{n}$-consistent initial estimator of $\theta_{0,n}$, i.e., $\hat{\theta}_{\text{initial}} - \theta_{0,n} = O_P(n^{-1/2})$, H_{ij} is a k_n-dimensional vector of sieve basis functions of X_i and W_j, and $\hat{\kappa}$ is a k_n-dimensional estimator of the parameter κ. The dimension k_n can be either fixed or increasing with n, given that its growth rate satisfies Assumption 4 below. We note that restricting H_{ij} to be a function of (X_i, W_j) only is necessary because we focus on the conditional (and locally misspecified) moment[1]

$$\mathbb{E}(Y_{ij} - \Lambda(R_{ij}^\top\theta) \mid X_i, W_j) = 0. \qquad (2.7)$$

The robust estimator $\hat{\Psi}(\hat{\kappa})$ is composed of two components. The first term in the right-hand side of (2.6) represents a plug-in estimator of $\Psi_{0,n}$, which ignores the effect of the local misspecification. The plug-in estimator is evaluated at the initial estimator $\hat{\theta}_{\text{initial}}$. As initial estimator, we consider the logistic maximum likelihood estimator (MLE), i.e.,

$$\hat{\theta}_{\text{initial}} = (\hat{\alpha}_n, \hat{\beta}_n) \equiv \arg\max_\theta L_n(\theta), \qquad (2.8)$$

---

[1]Moreover, it follows from the random sampling of nodes in Assumption 1 that the conditional moment satisfies $\mathbb{E}(Y_{ij} - \Lambda(R_{ij}^\top\theta) \mid \mathcal{X}_N, \mathcal{W}_M) = \mathbb{E}(Y_{ij} - \Lambda(R_{ij}^\top\theta) \mid X_i, W_j)$. Hence, it is sufficient to define H_{ij} as a function of (X_i, W_j).

where

$$L_n(\theta) = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \ell_{ij}(\theta), \quad \ell_{ij}(\theta) = Y_{ij} R_{ij}^\top \theta - \log(1 + \exp(R_{ij}^\top \theta)).$$

Notice that under local misspecification, the asymptotic distribution of the plug-in estimator will be centered at $\mathbb{E}\gamma_n(D_{ij}, 0, \theta_{0,n})$ rather than at $\Psi_{0,n}$ as it assumes away the presence of the local misspecification.

The second component in (2.6) corresponds to a one-step adjustment that accounts for the asymptotic bias generated by the local misspecification. The term in brackets is the sample analogue of the unconditional moments $\mathbb{E}\left[\left(Y_{ij} - \Lambda\left(R_{ij}^\top \theta\right)\right) H_{ij}\right]$, which are scaled by n to account for sparse network asymptotics (cf. Graham 2022). Notice that the following unconditional moment restrictions will fail to hold exactly

$$\mathbb{E}\left[\left(Y_{ij} - \Lambda\left(R_{ij}^\top \theta\right)\right) H_{ij}\right] = 0, \tag{2.9}$$

due to the effect of the local misspecification in the conditional probability given by (2.2). The vector $\hat{\kappa}$ is an estimate of κ, which captures the response of the estimator $\hat{\Psi}(\hat{\kappa})$ to the misspecification of the $k_n \times 1$ moment conditions in (2.9). Thus, following Andrews et al. (2020); Armstrong and Kolesár (2021); Bonhomme and Weidner (2022), we refer to κ as a sensitivity parameter. Below, we discuss how to optimally select κ to minimize the worst-mean squared error (MSE).

In contrast to the existing literature on sensitivity analysis that relies on maximum likelihood methods or method of moments, the dimension of the sensitivity parameter κ grows with the dimension of the sieve space. This is an implication that the underlying moment restrictions in (2.9) are constructed based on the locally misspecified conditional moments in (2.7). The growing sieve space will determine the limiting distribution of the one-step robust estimator.

Next, we describe the limiting distribution of the robust estimator, which will be central to performing robust inference on $\Psi_{0,n}$. Following Andrews et al. (2017); Armstrong and Kolesár (2021); Bonhomme and Weidner (2022), we focus on a class of regular estimators. As in Newey (1990), a necessary and sufficient condition for the one-step robust estimator to be regular is

$$G_0^\top \kappa = \Gamma, \tag{2.10}$$

where $\Gamma \equiv \Gamma(0)$ with $\Gamma(t)$ being defined as $\sup_{t \in \mathcal{N}(0)} \|\partial_t \mathbb{E}\gamma_n(D_{ij}, 0, \theta_{0,n} + t) - \Gamma(t)\|_2 \to 0$ and $G_0 \equiv \mathbb{E}\alpha_0 \exp(Z_{ij}^\top \beta_0) H_{ij} R_{ij}^\top$. Under this requirement and provided that $\hat{\kappa}$ is a consistent estimator of κ, i.e., $\|\hat{\kappa} - \kappa\|_2 = o_P(1)$ and the initial estimator is $\sqrt{n}$-consistent, i.e., $\hat{\theta}_{\text{initial}} - \theta_{0,n} = O_P(n^{-1/2})$, we have the next strong approximation of the distribution of the robust estimator $\hat{\Psi}(\hat{\kappa})$ which does

not depend on the asymptotic distribution of the initial estimator:

$$\sqrt{n}(\hat{\Psi}(\hat{\kappa}) - \Psi_{0,n}) - \omega_n(\kappa) = o_P(1), \tag{2.11}$$

where

$$\omega_n(\kappa) \overset{d}{=} \mathcal{N}\left(\mathbb{E}\left[\eta_{ij}(\kappa^\top \alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij})\right], \Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa + \kappa^\top \Sigma_H \kappa\right), \tag{2.12}$$

$\Delta_{v,ij} \equiv \partial_v \gamma_n(D_{ij}, v, \theta_{0,n})|_{v=0}$ and $(\Sigma_H, \Sigma_{H,\gamma}, \Sigma_\gamma)$ are three submatrices of a $(k_n + 1) \times (k_n + 1)$ covariance matrix

$$\Sigma_n = \begin{pmatrix} \Sigma_H & \Sigma_{H,\gamma} \\ \Sigma_{H,\gamma}^\top & \Sigma_\gamma \end{pmatrix}.$$

The covariance matrix Σ_n is defined and characterized in Section 2.3. Here, $\omega_n(\kappa)$ is referred to as the strong approximation of $\sqrt{n}(\hat{\Psi}(\hat{\kappa}) - \Psi_{0,n})$ rather than the limit under weak convergence because the dimensions of κ, Σ_H, $\Sigma_{H,\gamma}$, and thus themselves, implicitly depend on the sample size n.

The strong approximation result in (2.11) and (2.12) provides two main insights. First, the one-step robust estimator $\hat{\Psi}(\hat{\kappa})$ is asymptotically biased. This asymptotic bias is multifaceted: the term $-\mathbb{E}\left[\eta_{ij}\Delta_{v,ij}\right]$ reflects the bias induced by the plug-in estimator in (2.6) which is the result of having a limiting distribution centered around $\mathbb{E}\gamma_n(D_{ij}, 0, \theta_{0,n})$ rather than at $\Psi_{0,n}$. While the term $\mathbb{E}\left[\eta_{ij}\kappa^\top \alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij}\right]$ stems from the adjustment term in (2.6) and represents the asymptotic bias due to using $e(R_{ij}^\top \theta)$ as the identification condition in (2.9) as opposed to $\mathbb{P}[Y_{ij} = 1 \mid X_i, W_j]$. Second, the strong approximation result underscores that the local misspecification does not have a higher-order effect, as its complete impact concentrates on the asymptotic bias. In other words, the asymptotic variance does not depend on the local misspecification parameter η_{ij}.

Importantly, notice that the direction of the asymptotic bias is determined by η_{ij}, which corresponds to the mean of the latent distribution. Because this direction is unidentified, it renders the asymptotic distribution in (2.11) not immediately useful to conduct inference on $\Psi_{0,n}$. Nevertheless, the magnitude of the asymptotic bias can be bounded if an upper bound on $\mathbb{E}(\eta_{ij}^2)$ is imposed. Let $\bar{\sigma}^2$ denote this upper bound, which is specified by the researcher. It follows then that

$$\mathbb{E}\left[\eta_{ij}(\kappa^\top \alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij})\right] \leq \bar{\sigma}(\mathbb{E}(\kappa^\top \alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij})^2)^{1/2}.$$

We can use this inequality to define the worst-case bias within the neighborhood of misspecification characterized by $\mathbb{E}(\eta_{ij}^2) \leq \bar{\sigma}^2$ as

$$\mathcal{B}(\kappa) \equiv \bar{\sigma}(\mathbb{E}(\kappa^\top \alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij})^2)^{1/2}$$

10

$$= \left\{ \overline{\sigma}^2 \left[\kappa^\top B_H \kappa - 2 B_{H,\gamma}^\top \kappa + B_\gamma \right] \right\}^{1/2},$$

where

$$B_H = \alpha_0^2 \mathbb{E} \exp(2 Z_{ij}^\top \beta_0) H_{ij} H_{ij}^\top, \quad B_{H,\gamma} = \alpha_0 \mathbb{E} \exp(Z_{ij}^\top \beta_0) H_{ij} \Delta_{v,ij}, \quad \text{and} \quad B_\gamma = \mathbb{E} \Delta_{v,ij}^2.$$

The optimal choice of the sensitivity parameter κ is set to minimize the worst-case MSE defined as

$$MSE_H(\kappa) = \mathcal{B}(\kappa)^2 + (\Sigma_\gamma + 2 \Sigma_{H,\gamma}^\top \kappa + \kappa^\top \Sigma_H \kappa) \tag{2.13}$$

subject to the regularity condition (2.10). It follows from the method of Lagrange multipliers that the optimal κ, denoted by κ^*, is defined as the solution to this convex optimization

$$\kappa^* = \arg \min_\kappa \left\{ MSE(\kappa) + \lambda^\top (G_0^\top \kappa - \Gamma) \right\},$$

where λ denotes the Lagrange multiplier associated to regularity condition (2.10). Moreover, κ^* can be characterized in closed-form, which is given by

$$\kappa^* = -\Phi_1^{-1} \left\{ \Phi_2 - G_0 \left[G_0^\top \Phi_1^{-1} G_0 \right]^{-1} \left[G_0^\top \Phi_1^{-1} \Phi_2 + \Gamma \right] \right\}, \tag{2.14}$$

where

$$\Phi_1 = \Sigma_H + \overline{\sigma}^2 B_H \quad \text{and} \quad \Phi_2 = \Sigma_{H,\gamma} - \overline{\sigma}^2 B_{H,\gamma}.$$

In Section 2.3, we provide consistent estimators $\hat{G}$, $\hat{\Gamma}$, $\hat{B}_H$, $\hat{B}_{H,\gamma}$, $\hat{\Sigma}_H$, $\hat{\Sigma}_{H,\gamma}$, $\hat{\Sigma}_\gamma$ for their population counterparts. Then, the optimal sensitivity parameter κ^* is estimated by

$$\hat{\kappa}^* = -\hat{\Phi}_1^{-1} \left\{ \hat{\Phi}_2 - \hat{G} \left[\hat{G}^\top \hat{\Phi}_1^{-1} \hat{G} \right]^{-1} \left[\hat{G}^\top \hat{\Phi}_1^{-1} \hat{\Phi}_2 + \hat{\Gamma} \right] \right\},$$

where

$$\hat{\Phi}_1 = \hat{\Sigma}_H + \overline{\sigma}^2 \hat{B}_H \quad \text{and} \quad \hat{\Phi}_2 = \hat{\Sigma}_{H,\gamma} - \overline{\sigma}^2 \hat{B}_{H,\gamma}.$$

The approach of selecting κ^* to minimize the worst-case MSE is similar in spirit to that in Bonhomme and Weidner (2022) and has a number of antecedents in robust statistics (Huber and Ronchetti 2009; Rieder 1994). It is important to emphasize that a unique feature of our setting is that the dimension κ^* is determined by the sieve basis and can increase with the sample size. Consequently, our methodology to select and consistently estimate κ^* complements the existing

approaches in the literature.

2.2 Robust Confidence Interval

The bias-aware confidence interval for $\Psi_{0,n}$ is designed to account for the effect of the local misspecification. This is obtained by enlarging the $1 - \alpha$ quantile value by the worst-case bias in the strong approximation of $\sqrt{n}(\hat{\Psi}(\hat{\kappa}) - \Psi_{0,n})$ (cf. Armstrong and Kolesár 2021). In particular, we define the feasible optimal bias-aware confidence interval for $\Psi_{0,n}$ as

$$
\widehat{CI}_{1-\alpha}(\hat{\kappa}^*) \equiv \hat{\Psi}(\hat{\kappa}^*) \pm CV_\alpha \left(\frac{\hat{\mathcal{B}}(\hat{\kappa}^*)}{(\hat{\Sigma}_\gamma + 2\hat{\Sigma}_{H,\gamma}^\top \hat{\kappa}^* + \hat{\kappa}^{*,\top} \hat{\Sigma}_H \hat{\kappa}^*)^{1/2}} \right) \frac{(\hat{\Sigma}_\gamma + 2\hat{\Sigma}_{H,\gamma}^\top \hat{\kappa}^* + \hat{\kappa}^{*,\top} \hat{\Sigma}_H \hat{\kappa}^*)^{1/2}}{\sqrt{n}},
$$

$$(2.15)$$

where feasible worst-case bias $\hat{\mathcal{B}}(\hat{\kappa}^*)$ is defined as

$$
\hat{\mathcal{B}}(\hat{\kappa}^*) = \left\{ \overline{\sigma}^2 \left[\hat{\kappa}^{*,\top} \hat{B}_H \hat{\kappa}^* - 2\hat{B}_{H,\gamma}^\top \hat{\kappa}^* + \hat{B}_\gamma \right] \right\}^{1/2}
$$

and $CV_\alpha(t)$ denotes the $1 - \alpha$ quantile of $\mathcal{Z}$ where $\mathcal{Z} \sim \mathcal{N}(t, 1)$.

Remark 2 (Neighborhood of misspecification). *The choice of $\overline{\sigma}^2$ cannot be determined in a data-driven way. However, this upper bound can be varied to conduct sensitivity analysis on the magnitude of the local misspecification (cf. Conley et al. 2012). The values of $\overline{\sigma}^2$ should reflect the potential forms of misspecification that are admissible in a given application and thus should be established on an application-specific basis by the researcher. In likelihood-based models, Bonhomme and Weidner (2022) offer an interpretation of $\overline{\sigma}^2$ as a lower bound on the local power of a likelihood ratio specification test. In a setting of overidentified unconditional moments, Armstrong and Kolesár 2021 suggest using a J-test of overidentifying restrictions to obtain a lower bound on $\overline{\sigma}^2$. For additional discussions on the interpretation of the neighborhood of misspecification, we refer the reader to Andrews et al. (2017); Noack and Rothe (2019); Armstrong et al. (2022).*

Remark 3 (Other performance criteria). *The researcher might be interested in a robust estimate for $\Psi_{0,n}$ that minimizes the length of the bias-aware CIs as in Armstrong and Kolesár (2021). The optimal sensitivity parameter κ^* can be set to minimize the asymptotic variance in the strong approximation (2.12) subject to an upper bound on the worst-case bias and condition (2.10). In extended simulation experiments, we verify that the resulting bias-aware CIs also have the correct asymptotic coverage.*

2.3 Variance and Bias Components

In this section, we propose consistent estimators for G, Γ, B_H, $B_{H,\gamma}$, Σ_H, $\Sigma_{H,\gamma}$, and Σ_γ. Recall that the initial estimator $\hat{\theta}_{\text{initial}} = (\hat{\alpha}_n, \hat{\beta}_n)$ is defined in (2.8). Then, we define the estimators

$$\hat{G} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n \exp(R_{ij}^\top \hat{\theta}_{\text{initial}}) H_{ij} R_{ij}^\top,$$

$$\hat{\Gamma} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \partial_\theta \gamma_n(D_{ij}, 0, \hat{\theta}_{\text{initial}}),$$

$$\hat{B}_H = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n^2 \exp(2R_{ij}^\top \hat{\theta}_{\text{initial}}) H_{ij} H_{ij}^\top,$$

$$\hat{B}_{H,\gamma} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n^2 \exp(2R_{ij}^\top \hat{\theta}_{\text{initial}}) H_{ij} \hat{\Delta}_{v,ij}, \quad \text{and} \quad \hat{B}_\gamma = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \hat{\Delta}_{v,ij}^2,$$

where

$$\hat{\Delta}_{v,ij} = \partial_v \gamma_n(D_{ij}, v, \hat{\theta}_{\text{initial}})|_{v=0}.$$

To estimate the components of the covariance matrix in (2.12), recall that

$$\Sigma_n = \begin{pmatrix} \Sigma_H & \Sigma_{H,\gamma} \\ \Sigma_{H,\gamma}^\top & \Sigma_\gamma \end{pmatrix},$$

where Σ_n is the covariance matrix of the joint scores

$$s_{ij} = \begin{pmatrix} n \left(Y_{ij} - \int \Lambda(R_{ij}^\top \theta_{0,n} + n^{-1/2} v) \pi_{ij}(v) dv \right) H_{ij} \\ \gamma_n(D_{ij}, 0, \theta_{0,n}) - \mathbb{E}\gamma_n(D_{ij}, 0, \theta_{0,n}) \end{pmatrix} \equiv (s_{Y,ij} H_{ij}^\top, s_{\gamma,ij})^\top.$$

Let $\bar{s}_{ij} = \mathbb{E}(s_{ij}|X_i, A_i, W_j, B_j)$, $\bar{s}_{1i}^a = \mathbb{E}(\bar{s}_{ij}|X_i, A_i)$, $\bar{s}_{1j}^p = \mathbb{E}(\bar{s}_{ij}|W_j, B_j)$, $\Sigma_n^a = \mathbb{E}\bar{s}_{1i}^a(\bar{s}_{1i}^a)^\top$, $\Sigma_n^p = \mathbb{E}\bar{s}_{1j}^p(\bar{s}_{1j}^p)^\top$, and $\Sigma_{2n} = n^{-1}\mathbb{E}(s_{ij} - \bar{s}_{ij})(s_{ij} - \bar{s}_{ij})^\top$. Because $D_{ij} = (X_i^\top, W_j^\top)$, we have

$$\bar{s}_{\gamma,ij} \equiv \mathbb{E}(s_{\gamma,ij}|X_i, A_i, W_j, B_j) = s_{\gamma,ij}.$$

Then, the covariance matrix Σ_n is given by

$$\Sigma_n \equiv \begin{pmatrix} \Sigma_H & \Sigma_{H,\gamma} \\ \Sigma_{H,\gamma}^\top & \Sigma_\gamma \end{pmatrix} = \frac{\Sigma_n^a}{1-\phi} + \frac{\Sigma_n^p}{\phi} + \frac{\Sigma_{2n}}{\phi(1-\phi)} \tag{2.16}$$

where $\phi_n \equiv M/n \to \phi \in (0,1)$,

$$\Sigma_n^a = \mathbb{E}\begin{pmatrix} [\mathbb{E}(s_{Y,ij}H_{ij}|X_i,A_i)][\mathbb{E}(s_{Y,ij}H_{ij}|X_i,A_i)]^\top & \mathbb{E}(s_{Y,ij}H_{ij}|X_i,A_i)\mathbb{E}(s_{\gamma,ij}|X_i,A_i) \\ \mathbb{E}(s_{Y,ij}H_{ij}^\top|X_i,A_i)\mathbb{E}(s_{\gamma,ij}|X_i,A_i) & [\mathbb{E}(s_{\gamma,ij}|X_i,A_i)]^2 \end{pmatrix},$$

$$\Sigma_n^p = \mathbb{E}\begin{pmatrix} [\mathbb{E}(s_{Y,ij}H_{ij}|W_j,B_j)][\mathbb{E}(s_{Y,ij}H_{ij}|W_j,B_j)]^\top & \mathbb{E}(s_{Y,ij}H_{ij}|W_j,B_j)\mathbb{E}(s_{\gamma,ij}|W_j,B_j) \\ \mathbb{E}(s_{Y,ij}H_{ij}^\top|W_j,B_j)\mathbb{E}(s_{\gamma,ij}|W_j,B_j) & [\mathbb{E}(s_{\gamma,ij}|W_j,B_j)]^2 \end{pmatrix},$$

and

$$\Sigma_{2n} = \mathbb{E}\begin{pmatrix} n^{-1}Var(s_{Y,ij}|X_i,A_i,W_j,B_j)H_{ij}H_{ij}^\top & 0 \\ 0 & 0 \end{pmatrix}.$$

Let $\hat{s}_{Y,ij} = n(Y_{ij} - \exp(R_{ij}^\top \hat{\theta}_{\text{initial}}))$ and $\hat{s}_{\gamma,ij} = \gamma_n(D_{ij},0,\hat{\theta}_{\text{initial}}) - \frac{1}{NM}\sum_{i=1}^N \sum_{j=1}^M \gamma_n(D_{ij},0,\hat{\theta}_{\text{initial}})$. Following Graham (2022), we define

$$\hat{\Sigma}_n \equiv \begin{pmatrix} \hat{\Sigma}_H & \hat{\Sigma}_{H,\gamma} \\ \hat{\Sigma}_{H,\gamma}^\top & \hat{\Sigma}_\gamma \end{pmatrix} = \frac{\hat{\Sigma}_n^a}{1-\phi_n} + \frac{\hat{\Sigma}_n^p}{\phi_n} + \frac{\hat{\Sigma}_{2n}}{\phi_n(1-\phi_n)},$$

where

$$\hat{\Sigma}_n^a = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} \frac{2}{M(M-1)}\sum_{j=1}^{M-1}\sum_{k=j+1}^M (\hat{s}_{Y,ij}\hat{s}_{Y,ik}H_{ij}H_{ik}^\top) & \frac{2}{M(M-1)}\sum_{j=1}^{M-1}\sum_{k=j+1}^M (\hat{s}_{Y,ij}\hat{s}_{\gamma,ik}H_{ij}) \\ \frac{2}{M(M-1)}\sum_{j=1}^{M-1}\sum_{k=j+1}^M (\hat{s}_{Y,ij}\hat{s}_{\gamma,ik}H_{ij}^\top) & \frac{2}{M(M-1)}\sum_{j=1}^{M-1}\sum_{k=j+1}^M \hat{s}_{\gamma,ij}\hat{s}_{\gamma,ik} \end{pmatrix},$$

$$\hat{\Sigma}_n^p = \frac{1}{M}\sum_{j=1}^M \begin{pmatrix} \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{k=i+1}^N (\hat{s}_{Y,ij}\hat{s}_{Y,ik}H_{ij}H_{ik}^\top) & \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{k=i+1}^N (\hat{s}_{Y,ij}\hat{s}_{\gamma,ik}H_{ij}) \\ \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{k=i+1}^N (\hat{s}_{Y,ij}\hat{s}_{\gamma,ik}H_{ij}^\top) & \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{k=i+1}^N \hat{s}_{\gamma,ij}\hat{s}_{\gamma,ik} \end{pmatrix},$$

$$\hat{\Sigma}_{2n} = \frac{1}{n}\begin{pmatrix} \frac{1}{NM}\sum_{i=1}^N \sum_{j=1}^M \hat{s}_{Y,ij}^2 H_{ij}H_{ij}^\top - \hat{\Sigma}_{YY,n}^a - \hat{\Sigma}_{YY,n}^p & 0 \\ 0 & 0 \end{pmatrix},$$

and $\hat{\Sigma}_{YY,n}^a$ and $\hat{\Sigma}_{YY,n}^p$ are the upper-left $k_n \times k_n$ submatrices of $\hat{\Sigma}_n^a$ and $\hat{\Sigma}_n^p$, respectively. In Lemma A.1 of the Appendix, we show that the estimators $\hat{G}$, $\hat{\Gamma}$, $\hat{B}_H$, $\hat{B}_{H,\gamma}$, $\hat{\Sigma}_H$, $\hat{\Sigma}_{H,\gamma}$, and $\hat{\Sigma}_\gamma$ are consistent.

3 Asymptotic Size Control

In this section, we show the bias-aware CI proposed in (2.15) has the correct asymptotic size control, which relies on the following assumptions.

Assumption 1. *1. The graphon $\{Y_{ij}\}_{i\in[N],j\in[M]}$ is generated according to (2.1) with $\{X_i, A_i\}_{i=1}^N$, $\{W_j, B_j\}_{j=1}^M$, $\{U_{ij}\}_{i\in[N],j\in[M]}$, $\{v_{ij}\}_{i\in[N],j\in[M]}$ be sequences of i.i.d. random variables that are independent, $M/n \to \phi \in (0,1)$ and $n = M + N$.*

2. *The conditional distribution $\mathbb{P}(Y_{ij} = 1|W_i, X_j)$ is as defined in (2.2) with $\eta_{ij} = \int v\pi_{ij}(v)dv$. In addition, suppose $\mathbb{E}\eta_{ij}^2 \leq \bar{\sigma}^2$ for some known upper bound $\bar{\sigma}^2$ and $\max_{i\in[N],j\in[M]} \int |v| \exp(v)\pi_{ij}dv \leq C$ for some constant $C \in (0,\infty)$.*

3. *Let $\theta_0 = (\alpha_0, \beta_0^\top)^\top \in \mathbb{A} \times \mathbb{B} = \Theta$, the parameter spaces $\mathbb{A}$ and $\mathbb{B}$ are compact.*

4. *Suppose that Z_{ij} is bounded, and*

$$\Lambda(R_{ij}^\top \theta_{0,n}) \in \left[\frac{c_1}{n}, \frac{c_2}{n}\right]$$

for some constants $0 < c_1 \leq c_2 < \infty$, where $R_{ij} = (1, Z_{ij}^\top)^\top$.

5. *There exist constants $0 < \lambda_1 < \lambda_2 < \infty$ such that, with probability approaching one,*

$$\lambda_1 \leq \lambda_{\min}\left(\frac{1}{NM}\sum_{i=1}^N\sum_{j=1}^M R_{ij}R_{ij}^\top\right) \leq \lambda_{\max}\left(\frac{1}{NM}\sum_{i=1}^N\sum_{j=1}^M R_{ij}R_{ij}^\top\right) \leq \lambda_2,$$

where for a symmetric matrix A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum singular values of A, respectively.

6. *Let $G(\alpha, \beta) = \mathbb{E}\alpha\exp(Z_{ij}^\top\beta)H_{ij}R_{ij}^\top$ and $G_0 = G(\alpha_0, \beta_0)$. Then,*

$$\sup_{(\alpha,\beta^\top)^\top\in\Theta} \frac{\lambda_{max}(G(\alpha,\beta) - G_0)}{|\alpha - \alpha_0| + ||\beta - \beta_0||_2} \leq C$$

for a positive constant $C < \infty$.

Condition 1 of Assumption 1 describes the graph formation and ensures that $\{Y_{ij}\}_{i\in[N],j\in[M]}$ is X-W exchangeable. Condition 2 restricts the conditional probability to satisfy (2.2). Condition 3 is a regularity condition on the parameter space Θ. Condition 4 bounds the largest magnitude of the covariates and the rate at which the probability of forming a link decays. This condition, jointly with 3, controls the rate of sparsity in the network. The boundedness of Z_{ij} is for theoretical simplicity and can be relaxed a sub-exponential tail of Z_{ij}, which is necessary for G_0 in Condition 6 to be well-defined. Condition 5 is an identification condition for $\theta_{0,n}$. Finally, Condition 6 is a continuity condition of $G(\alpha, \beta)$ at (α_0, β_0), which bounds the maximum eigenvalue difference. Conditions 1-5 of Assumption 1 has been used in Graham (2022). Condition 6 is specific to our setting.

The next assumption provides the smoothness conditions required for the asymptotic expansions.

Assumption 2.   *1. Recall $\theta_{0,n} \equiv (\alpha_{0,n}, \beta_0)$. Suppose $\gamma_n(D_{ij}, 0, \theta_{0,n} + t)$ is twice-differentiable in t,*

$$\mathbb{E}\,\|\partial_\theta \gamma_n(D_{ij}, 0, \theta_{0,n})\|_2^2 = O(1),$$

$$\sup_{t \in \mathcal{N}(0)} \|\partial_t \mathbb{E}\gamma_n(D_{ij}, 0, \theta_{0,n} + t) - \Gamma(t)\|_2 \to 0,$$

and

$$\sup_{t \in \mathcal{N}(0)} \left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \partial_{t, t^\top} \gamma_n(D_{ij}, 0, \theta_{0,n} + t) \right\|_{op} = O_P(1),$$

where $\mathcal{N}(0) \in \Re^{d_z + 1}$ is an arbitrary neighborhood of 0. Also, let $\Gamma(0)$ be denoted as Γ.

2. Let $\Delta_{v,ij} = \partial_v \gamma_n(D_{ij}, v, \theta_{0,n})|_{v=0}$. Then, we have

$$\sqrt{n}\mathbb{E}\left(\int \gamma_n(D_{ij}, n^{-1/2}v, \theta_{0,n})\pi_{ij}(v)dv - \gamma_n(D_{ij}, 0, \theta_{0,n}) \right) = \mathbb{E}\eta_{ij}\Delta_{v,ij} + o(1), \qquad (3.1)$$

$$\sup_{t \in \mathcal{N}(0)} \left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \partial_{v,t} \gamma_n(D_{ij}, 0, \theta_{0,n} + t) \right\|_2 = O_P(1)$$

and $\mathbb{E}\Delta_{v,ij}^2 = O(1)$.

Condition 1 of Assumption 2 states that, in the absence of local misspecification, the parameter of interest is twice differentiable at $\theta_{0,n}$ with bounded derivatives. Meanwhile, Condition 2 requires the parameter of interest to be Gateaux differentiable in the local misspecification component. This condition ensures that the bias in the parameter of interest induced by the local misspecification is non-vanishing at a rate $\sqrt{n}$. All the parameters in Examples 5–7 satisfy this assumption.

Assumption 3.   *1. Let $p_{ij} = \mathbb{E}(Y_{ij}|X_i, A_i, W_j, B_j)$. Then, we assume np_{ij} and $\gamma_n(D_{ij}, 0, \theta_{0,n})$ have bounded support.*

2. Suppose $\|G_0\|_{op} = O(1)$, $\|B_{H,\gamma}\|_2 = O(1)$, $\|\Sigma_{H,\gamma}\|_2 = O(1)$,

$$\lambda_1 \le \lambda_{\min}(\Sigma_H) \le \lambda_{\max}(\Sigma_H) \le \lambda_2, \quad \lambda_{\max}(B_H) \le \lambda_2,$$
$$\lambda_1 \le \lambda_{\min}(G_0^\top G_0) \le \lambda_{\max}(G_0^\top G_0) \le \lambda_2,$$

$$\lambda_{\max}\left(\frac{1}{N} \sum_{i=1}^{N} [\mathbb{E}(s_{Y,ij}H_{ij}|X_i, A_i)][\mathbb{E}(s_{Y,ij}H_{ij}|X_i, A_i)]^\top \right) = O_P(1),$$

and

$$\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^* \geq \lambda_1$$

for some constants $0 < \lambda_1 \leq \lambda_2 < \infty$.

Assumption 3 contains the regularity conditions that ensure $\kappa^* = O(1)$.

Assumption 4. *Suppose* $H_{ij} \in \Re^{k_n}$, $\max_{i \in [N], j \in [M]} ||H_{ij}||_\infty \leq \zeta_n$, *and*

$$\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}||H_{ij}||_2^2 = O_P(k_n),$$

such that $k_n = o(n^{1/5})$ *and* $k_n^3 \zeta_n^2 \log^2(n) = o(n)$.

Assumption 4 controls the complexity of the sieve basis by imposing a rate condition on the dimension and strength of H_{ij}. Specifically, we require $k_n = o(n^{1/5})$ to ensure the Yurinskii's coupling of the score with an increasing dimension.[2] This rate is typically assumed in the literature. See, for example, the discussion after Li and Liao (2020, Theorem 1). The second rate requirement is weaker than the first if $\zeta_n = o(k_n)$, which is the case if the support of X_i and W_j are bounded and splines are used to construct the sieve basis. We also note that Assumption 4 allows for k_n to be fixed. Then, Condition 2 in Assumption 3 requires $k_n \geq d_z + 1$.

Assumption 5. *The initial estimator satisfies* $\hat{\theta}_{initial} - \theta_0 = O_P(n^{-1/2})$.

The logistic MLE satisfies this requirement; see Graham (2022). The next theorem shows that our bias-aware CI has the correct asymptotic size control under sparse network asymptotics.

Theorem 1. *Suppose Assumptions 1–5 hold and the bias-aware CI, i.e.,* $\widehat{CI}(\hat{\kappa}^*)$, *is constructed as* (2.15). *Then, for the significance level* $\alpha \in (0,1)$, *we have*

$$\liminf_{n \to \infty} \mathbb{P}(\Psi_{0,n} \in \widehat{CI}_{1-\alpha}(\hat{\kappa}^*)) \geq 1 - \alpha.$$

4 Asymptotic Optimality

In this section, we show that robust one-step estimator $\hat{\Psi}(\hat{\kappa}^*)$ is minimax MSE-optimal compared to any other regular estimator that is asymptotically linear.

---

[2]See, for example, Pollard (2002, Theorem 10) for the statement of the Yurinskii's coupling.

We consider an arbitrary estimator $\breve{\theta}$, which has the following linear expansion:

$$\sqrt{n}(\breve{\theta} - \theta_{0,n}) = \kappa_\theta^\top \left[\frac{n^{3/2}}{NM} \sum_{i=1}^N \sum_{j=1}^M (Y_{ij} - \Lambda(R_{ij}^\top \hat{\theta}_{0,n})) F_{ij} \right] + o_P(1),$$

where $F_{ij} \in \Re^{d_F}$ is d_F dimensional vector of transformations of (X_i, W_j), $\kappa_F \in \Re^{d_F \times d_R}$, and d_F is fixed. We further assume the estimator $\breve{\theta}$ is regular.

Assumption 6. *Suppose* $G_F^\top \kappa_\theta = I_{d_R}$ *with* $G_F = \mathbb{E}\alpha_0 \exp(Z_{ij}^\top \beta_0) F_{ij} R_{ij}^\top$. *Further assume*

$$\lambda_1 \leq \lambda_{\min}(G_F^\top G_F) \leq \lambda_{\max}(G_F^\top G_F) \leq \lambda_2$$

for some constants $0 < \lambda_1 \leq \lambda_2 < \infty$.

The estimator $\breve{\theta}$ can be viewed as a GMM estimator based on the following set of unconditional moments:

$$\mathbb{E}\left[\left(Y_{ij} - \Lambda\left(R_{ij}^\top \theta \right) \right) F_{ij} \right] = 0.$$

A leading example of $\breve{\theta}$ is our initial estimator $\hat{\theta}_{\text{initial}}$, which satisfies the above moment condition with $F_{ij} = R_{ij}$.

Consider the plug-in estimator $\breve{\Psi}$ of $\Psi_{0,n}$ defined as

$$\breve{\Psi} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M \gamma(D_{ij}, 0, \breve{\theta}).$$

Following the arguments that lead to (2.11), we can show that

$$\sqrt{n}(\breve{\Psi} - \Psi_{0,n}) - \omega_{F,n} = o_P(1)$$

and

$$\omega_{F,n} \overset{d}{=} \mathcal{N}\left(\mathbb{E}\eta_{ij}(\kappa_F^\top \alpha_0 \exp(Z_{ij}^\top \beta_0) F_{ij} - \Delta_{v,ij}), \Sigma_\gamma + 2\Sigma_{F,\gamma}^\top \kappa_F + \kappa_F^\top \Sigma_F \kappa_F \right),$$

where Σ_F and $\Sigma_{F,\gamma}$ are similarly defined as Σ_H and $\Sigma_{H,\gamma}$ with H_{ij} replaced by F_{ij}, and

$$\kappa_F = \kappa_\theta \Gamma. \tag{4.1}$$

Specifically, we have

$$\begin{pmatrix} \Sigma_F & \Sigma_{F,\gamma} \\ \Sigma_{F,\gamma}^\top & \Sigma_\gamma \end{pmatrix} = \frac{\Sigma_{F,n}^a}{1 - \phi} + \frac{\Sigma_{F,n}^p}{\phi} + \frac{\Sigma_{2,F,n}}{\phi(1 - \phi)}$$

with

$$\Sigma_{F,n}^a = \mathbb{E} \begin{pmatrix} [\mathbb{E}(s_{Y,ij}F_{ij}|X_i, A_i)][\mathbb{E}(s_{Y,ij}F_{ij}|X_i, A_i)]^\top & \mathbb{E}(s_{Y,ij}F_{ij}|X_i, A_i)\mathbb{E}(s_{\gamma,ij}|X_i, A_i) \\ \mathbb{E}(s_{Y,ij}F_{ij}^\top|X_i, A_i)\mathbb{E}(s_{\gamma,ij}|X_i, A_i) & [\mathbb{E}(s_{\gamma,ij}|X_i, A_i)]^2 \end{pmatrix},$$

$$\Sigma_{F,n}^p = \mathbb{E} \begin{pmatrix} [\mathbb{E}(s_{Y,ij}F_{ij}|W_j, B_j)][\mathbb{E}(s_{Y,ij}H_{ij}|W_j, B_j)]^\top & \mathbb{E}(s_{Y,ij}F_{ij}|W_j, B_j)\mathbb{E}(s_{\gamma,ij}|W_j, B_j) \\ \mathbb{E}(s_{Y,ij}F_{ij}^\top|W_j, B_j)\mathbb{E}(s_{\gamma,ij}|W_j, B_j) & [\mathbb{E}(s_{\gamma,ij}|W_j, B_j)]^2 \end{pmatrix},$$

and

$$\Sigma_{2,F,n} = \mathbb{E} \begin{pmatrix} n^{-1}Var(s_{Y,ij}|X_i, A_i, W_j, B_j)F_{ij}F_{ij}^\top & 0 \\ 0 & 0 \end{pmatrix}.$$

Then, the worst-case mean squared error for the plug-in estimator $\breve{\Psi}$ is

$$MSE_F(\kappa_F) = \bar{\sigma}^2 \left[\kappa_F^\top B_F \kappa_F - 2B_{F,\gamma}^\top \kappa_F + B_\gamma \right] + (\Sigma_\gamma + 2\Sigma_{F,\gamma}^\top \kappa_F + \kappa_F^\top \Sigma_F \kappa_F), \qquad (4.2)$$

where

$$B_F = \alpha_0^2 \mathbb{E} \exp(2Z_{ij}^\top \beta_0)F_{ij}F_{ij}^\top \quad \text{and} \quad B_{F,\gamma} = \alpha_0 \mathbb{E} \exp(Z_{ij}^\top \beta_0)F_{ij}\Delta_{v,ij}.$$

Assumption 7. *Suppose there exists a matrix $\Pi_F \in \Re^{d_F \times k_n}$ and $\delta_{ij} \equiv \delta(X_i, W_j) \in \Re^{d_F}$ for some d_F-dimensional function δ such that*

$$F_{ij} = \Pi_F H_{ij} + \delta_{ij},$$

$||\Pi_F||_{op} = O(1)$, *and* $\mathbb{E}||\delta_{ij}||_2^2 = o(1)$.

Assumption 7 implies each element of F_{ij} can be well approximated by the sieve bases H_{ij}. This condition holds given that F_{ij} is a sufficiently smooth function of (X_i, W_j).

Theorem 2. *Suppose Assumptions 1–7 holds. Then, we have*

$$MSE_H(\kappa^*) \leq MSE_F(\kappa_F) + o(1),$$

where $MSE_H(\cdot)$, κ^*, κ_F, and $MSE_F(\cdot)$ are defined in (2.13), (2.14), (4.1), and (4.2), respectively.

Theorem 2 shows that the one-step robust estimator $\hat{\Psi}(\hat{\kappa}^*)$ attains the smallest worst-case MSE relative to any arbitrary plug-in estimator $\breve{\Psi}$ that estimates regularly the parameter θ based on unconditional moments that are constructed using a general set of covariates F_{ij}. That is, the one-step robust estimator $\hat{\Psi}(\hat{\kappa}^*)$ is minimax-MSE optimal within a clas of regular estimators. This result holds even when the dimension k_n of H_{ij} is fixed, as long as Assumption 7 is valid. To

guarantee Assumption 7 in the fixed dimension case, we can construct H_{ij} by stacking F_{ij} with other functions of (X_i, W_j). Theoretically, the benefit of allowing for a diverging number of H_{ij} is that it expands the range of F_{ij} that can be approximated by linear functions of H_{ij}, thereby broadening the set of estimators to which our one-step estimator is superior. In practice, we do not suggest using a very large k_n when n is moderate because it may (i) violate Assumption 4 and (ii) cause numerical instability.

5 Monte Carlo Simulations

This section presents simulation evidence for the finite sample performance of the one-step robust estimator introduced in Section 2.2. We consider a wide array of Monte Carlo designs that are meant to capture differences in the specification of the local misspecification, sample size and sparsity of the network.

The bipartite network is modelled according to the following data-generating process. For any $i \in [N]$ and $j \in [M]$, the observed attributes X_i and W_j are drawn from independent and identically distributed log-normal distributions with mean $-1/4$ and variance $1/2$. The dyad-specific attributes are computed to account for assortative matching. In particular, we define $Z_{ij} = \log(X_i \cdot W_j)$, which will be distributed normally with mean equal to $-1/2$ and variance 1. The sieve basis is computed using a Hermite polynomial approximation of Z_{ij} of order k_n. As an alternative sieve basis, we consider the tensor product of polynomial expansions on X_i and W_j, which yields similar qualitative results.

The unobserved heterogeneity A_i and B_j are drawn from independent and identically distributed log-normal distributions with mean $-1/12$ and standard deviation given by $1/\sqrt{6}$. Finally, U_{ij} is distributed as a standard exponential distribution. The bipartite graphon is simulated according to the equation

$$Y_{ij} = 1\left[\alpha_0 - \log(n) + Z_{ij}\beta_0 + \log(A_i) + \log(B_i) + n^{-1/2}v_{ij} + U_{ij} \geq 0\right], \tag{5.1}$$

where the term $\log(n)$ will ensure that the average degree of this network is bounded. In particular, under the current setting, the probability of establishing a link decreases at a rate $1/n$.

The distribution of the misspecification component v_{ij} is simulated to capture three different designs: (i) latent homophily, (ii) functional form misspecification, and (iii) semiparametric distribution. In the first design, the misspecification component is modelled as $v_{ij} = 1\left[\sigma > 0\right](1 + \sigma) * \overline{Z}_n * (v_i \cdot v_j)$ where $v_i = 3/4\log(X_i) + 1/4\,\mathrm{N}(1, \sigma)$ and $v_j = 3/4\log(W_j) + 1/4\,\mathrm{N}(1, \sigma)$ represent fixed effects, and $\overline{Z}_n = (NM)^{-1}\sum_{i,j} Z_{ij}$. Here, $\mathrm{N}(1, \sigma)$ stands for a normal random variable with mean 1 and standard deviation σ. In the second design, we simulate $v_{ij} = 1\left[\sigma > 0\right](1 + \sigma) * \overline{Z}_n * \left(Z_{ij}^2 + Z_{ij}^3\right)$.

This design captures the misspecification that stems from truncating the higher-order effects of the observed attributes. Finally, for the last design, the misspecification component is modelled as $v_{ij} = 1[\sigma > 0](1 + \sigma) * N(Z_{ij}, \sigma)$, which represents an heterogeneous latent mixing component in (5.1).

Notice that for all designs, the conditional density of v_{ij} given by π_{ij} depends on (X_i, W_j). Moreover, the magnitude of misspecification is determined by σ^2 with $\sigma^2 \in \{0, 1, 2, 3, 4\}$. A $\sigma^2 = 0$ represents the absence of local misspecification. Across each of the designs, the upper bound on the magnitude of local misspecification $\mathbb{E}\eta_{ij}^2$ is fixed as $\overline{\sigma}^2 = \max_{\sigma^2 \leq 4} \mathbb{E}\eta_{ij}^2$ with $\mathbb{E}\eta_{ij}^2 = var(\mathbb{E}(v_{ij}|X_i, W_j))$. That is, it corresponds to the largest value of $\mathbb{E}\eta_{ij}^2$ that is attained when the simulation design sets $\sigma^2 = 4$. The true DGP design is completed by setting $\alpha_0 = \log(2.56)$, $\beta_0 = \log(4)$, and network size equal to $n = N + M$ with $N, M = 100, 200, 300, 400$.

The one-step robust estimator is computed using the definition in (2.6). As an initial estimator $\hat{\theta}_{\text{initial}}$, we consider a logistic maximum likelihood estimator. We compare the performance of the robust methodology with that of a plug-in estimator with two different initial $\hat{\theta}_{\text{initial}}$: (i) a logistic regression and (ii) a Poisson regression. Given the exponential distribution of U_{ij}, the logistic estimator $\hat{\theta}_n$ will present a bias of order $o(n^{-2})$, as discussed in Graham (2022). This bias is negligible concerning the effect generated by the local misspecification, which induces a nonvanishing bias in the asymptotic distribution of order $n^{-1/2}$. On the contrary, the initial Poisson regressor represents the Composite Maximum Likelihood Estimator when the distribution of the error term is known. We report bias-aware confidence intervals for all the estimators, that is the one-step robust estimator and both plug-in estimators.[3]

As parameter of interest $\Psi_{0,n}$, we consider: (i) the homophily parameter β_0, and (ii) the average out-degree of the network $M\mathbb{E} \int \exp(R_{ij}^\top \theta_{0,n} + n^{-1/2}v)\pi_{ij}dv$. Notice that in case (i), the parameter of interest is not affected directly by the local misspecification; while in (ii), the functional form of $\Psi_{0,n}$ is also misspecified. Importantly, in both cases, the local misspecification plays a central role due to its effect on the conditional probability (2.2).

5.1 Homophily parameters

Table 1 summarizes results from computing the one-step robust estimator across 2,000 Monte Carlo replications when the parameter of interest is $\Psi_{0,n} = \beta_0$, v_{ij} accounts for latent homophily, and the sieves dimension is set to $k_n = 2$.[4] The table includes the true value of β_0, along with the mean coefficient estimates (coeff.), $\sqrt{n}$-bias, standard error (s.e.), standard error-to-standard deviation

---

[3]We also study the performance of the robust methodology when the Poisson regression is set as the initial estimator $\hat{\theta}_{\text{initial}}$. The results are close to being numerically equal to the baseline setting with the logistic estimator as $\hat{\theta}_{\text{initial}}$. This is a consequence of the fact that $\hat{\Psi}(\hat{\kappa})$ is a regular estimator, which ensures that its strong approximation does not depend on the asymptotic distribution of the initial estimator. We report these results in Appendix E.3.

[4]In Appendix E.4, we report the results using larger dimensions for the sieves approximation.

ratio (s.e./sd), root-mean-squared error (rMSE), 95% confidence interval (conf. int.), confidence interval length (length), and 95% coverage probabilities (95% CP). The final column shows the average degree of the network. These values were calculated over all simulations.

The top panel in Table 1 shows the results of computing the robust estimator in a network of size $n = 200$ and across misspecification designs $\sigma^2 \in \{0, 1, 2, 3, 4\}$. The mean estimate and $\sqrt{n}$-Bias show that the robust estimator approximates well the true DGP value. Moreover, the standard error-to-standard deviation ratio (s.e./sd) indicates that the estimator's sampling variability is well approximated. Finally, the bias-aware confidence intervals have the correct asymptotic size control under sparse asymptotic designs as indicated by Theorem 1. Notice that the coverage probabilities are greater than 95%, which is expected as the bias-aware confidence intervals are conservative.

The bottom panels of Table 1 report the results of the robust estimator in larger networks with sizes equal to $n = 400$, 600, and 800. As the sample size grows, the point estimates approximate more precisely the true parameter value. Note that the $\sqrt{n}$-Bias does not vanish completely. This is expected as the average bias is scaled up by a factor of $\sqrt{n}$, and thus, it reflects the first-order effect that the local misspecification has on the limiting distribution of the robust estimator. Notice also that the length of the bias-aware confidence intervals decreases uniformly as n grows throughout all the misspecification designs and even when the average network degree is as low as 0.3%.

Tables 2 and 3 report the results of the plug-in estimators with an initial logistic and Poisson regression, respectively. Table 4 reports the ratios of the plug-in estimators relative to the robust estimator to facilitate the analysis. Relative to the plug-in estimator with an initial logistic estimator, it is clear that the robust estimator exhibits smaller rMSEs. In fact, the rMSE obtained under the logistic estimator is, on average, 1.089 times larger than that of the robust methodology when the network size is $n = 200$. Even as both estimators become more precise in larger networks, the logistic estimator attains larger rMSE relative to the robust estimator as supported by Theorem 2. They are, on average, 1.054 times larger when $n = 400$, 1.036 times larger when $n = 600$, and 1.029 times larger when $n = 800$. Notice that the robust estimator also presents smaller standard errors, significantly less $\sqrt{n}$-Bias, and tighter confidence intervals.

The robust estimator also outperforms the plug-in estimator with an initial Poisson regression in Table 3. When comparing the two, the robust estimator exhibits smaller RMSEs. Albeit, the margins are smaller as the plug-in estimator coincides with the composite MLE. We observe that the Poisson regression has rMSEs that are, on average, 1.019 times larger when $n = 200$, 1.014 times larger when $n = 400$, 1.010 times larger when $n = 600$, and 1.008 times larger when $n = 800$. Notably, both estimators are comparable in terms of standard errors, $\sqrt{n}$-Bias, confidence interval's length, and coverage probabilities. This evidence suggests that under model misspecification, the one-step robust estimator attains shorter rMSEs relative to the composite MLE.

Appendix Tables 16-19, and 20-23 report the simulation results of estimating β_0 for the designs

of functional form misspecification and semiparametric distribution. Qualitatively, we observe similar patterns; that is, the robust one-step estimator dominates the plug-in logistic and Poisson estimators in terms of the rMSE. Moreover, it has comparable performance in terms of $\sqrt{n}$-Bias and coverage probabilities.

5.2 Average out-degree

Table 5 summarizes results from computing the one-step robust estimator across 2,000 Monte Carlo replications when the parameter of interest is the average out-degree of the network, $\Psi_{0,n} = M\mathbb{E}\int\exp\left[R_{ij}^{\top}\theta_{0,n} + n^{-1/2}v\right]\pi_{ij}(v)dv$, v_{ij} accounts for latent homophily, and $k_n = 3$.[5] Notice that, in this case, the local misspecification directly affects the parameter of interest, and thus, the true value of $\Psi_{0,n}$ varies across different specifications of $\sigma^2 \in \{0, 1, 2, 3, 4\}$.

The results in Table 5 suggest that the robust estimator yields reliable inference for the parameter of interest, and its performance improves in larger networks, notwithstanding the large degree of sparsity. The $\sqrt{n}$-Bias is relatively small and shows a decaying pattern as n grows. The bias-aware confidence intervals have the correct asymptotic coverage, and their length decreases uniformly as n grows and throughout all the misspecification designs.

Tables 6 and 7 report the results of the plug-in estimators with an initial logistic and Poisson regression, respectively. Table 8 reports the ratios of the plug-in estimators relative to the robust estimator to facilitate the analysis. When comparing the plug-in estimator with an initial logistic regression and the robust estimator, it is clear that the robust estimator improves significantly on the plug-in estimator by attaining smaller rMSEs. In particular, the Logistic plug-in estimator has rMSE that are, on average, 1.084 times larger when $n = 200$, 1.035 times when $n = 400$, 1.023 times larger when $n = 600$, and 1.016 times larger when $n = 800$. The same pattern is observed when comparing the robust estimator with the plug-in estimator with an initial Poisson regression in Table 7. Although this second plug-in estimator provides better asymptotic guarantees, the results show that using the Poisson estimator yields similar rMSE when $n = 200$, 1.003 times larger rMSE when $n = 400$, 1.002 times larger for $n = 600$, and 1.001 times larger when $n = 800$.

Appendix Tables 24-27 and 28-31 report the simulation results of estimating $\Psi_{0,n}$ for the designs of functional form misspecification and semiparametric distribution. Qualitatively, we observe similar patterns. This evidence suggests that computing the optimal confidence intervals of the robust estimator might lead to a significant improvement when the parameter of interest is affected directly by the local misspecification.

---

[5]In Appendix E.4, we report the results using larger dimensions in the sieves approximation.

6 Empirical Application

In this section, we implement the methodology developed in Section 3 to study a network of scientific collaborations among economists. We utilize articles published in leading American journals of general interest to construct a bipartite network connecting authors and articles. The goal is to estimate the factors influencing an author's decision to participate in a paper, thereby establishing a scientific collaboration with other authors who are also involved in that project.

This study contributes to the growing literature on scientific collaborations in economics. Most of the existing literature has focused on describing stylized features of this network (Newman 2001; Goyal et al. 2006 and Boschini and Sjögren 2007) or measuring the effect that these collaborations have on research productivity (Ductor, Fafchamps, Goyal, and Van der Leij 2014; Ductor 2015 and Colussi 2018). Fewer studies have analyzed the mechanisms that drive the formation of these scientific collaborations (Fafchamps, Van der Leij, and Goyal 2010; Anderson and Richards-Shubik 2022 and Hsieh et al. 2022).[6] We extend these studies by considering a larger number of factors that can influence an author's decision to collaborate on a project while using a computationally tractable method that accounts for two-sided heterogeneity. More importantly, this is the first paper to study the effects that local misspecification has on a bipartite network of scientific collaboration and conduct robust inference on the parameters of interest.

The data include all papers published in top American economic journals from 2000 to 2006: the American Economic Review, Econometrica, Journal of Political Economy, and Quarterly Journal of Economics. It contains author's information, such as gender, university granting PhD degree, graduation year, research fields, institution of employment and position. For articles, the dataset includes citation count since the year of publication, authors' names, publication issue, number of pages, references, three-digit JEL code, and keywords.[7] We used this data to define a single, static bipartite network of collaborations that includes a total of 1776 authors and 1600 articles.[8]

We examine the formation probability of a collaboration link using (2.2) and compare the one-step robust estimator's performance with plug-in estimators based on initial logistic and Poisson regressions. The parameters of interest include assortative matching parameters and the network's average out-degree. We control for author- and project-specific characteristics, along with author-project pair-specific observed attributes. In terms of individual author's attributes, we include binary variables for the categories of female gender and junior economist position and control for the average number of citations received by the author.[9],[10] As project-specific attributes, we

---

[6]Relatively, the literature on the economics of innovation aims to identify individual contributions in teams production (see e.g. Bonhomme 2021)

[7]We thank Tommaso Colussi for sharing this data with us.

[8]Our final sample includes only unique observations of matched authors and articles.

[9]The category of junior economists includes assistant professors and economists in government or research institutions.

[10]The average citation count serves as a proxy measure for the author's productivity.

include the number of authors in the project, indices for the share of female and senior authors in the project, and binary variables indicating whether some authors share the same institution or obtained their PhD from the same university.[11],[12] Table 9 provides summary statistics on the network and author-specific and project-specific observed attributes.

As main mechanisms for assortative matching, we consider: (i) similarity between author i' research fields and project j's classification topic (`jelcode`), (ii) sorting of more productive authors into projects with higher impact (`citations`), (iii) gender sorting across collaborations (`gender`), and (iv) collaboration between junior and senior scientists (`junior_senior`).[13] Table 10 collects the estimation results using the robust methodology with an initial logistic estimator and sieves dimension $k_n = 3$. It reports the coefficient estimates, standard errors, theoretical worst-case rMSEs, confidence intervals, confidence intervals' length and the adjustment term of the one-step robust estimator. The results computed impose an upper bound on $\overline{\sigma}^2 = 4$. In Appendix Tables 13, 14, and 15, we report the results for the robust estimator where the upper bound $\overline{\sigma}^2$ is specified equal to 1, 2, and 3, respectively. Figures 1–6 depict the bias-aware confidence intervals for the parameters of interest across different levels $\overline{\sigma}^2$.

The results in Table 10 indicate that similarity in the research field of expertise is a strong and positive factor influencing the decision to collaborate on a research project. This result aligns with the findings in Ductor 2015 and Hsieh et al. 2022. Additionally, highly productive authors tend to participate in projects with higher impact with a positive probability. This outcome is consistent with the role of "star" economists discussed in Goyal et al. (2006). The evidence also suggests that, during the sample period, sorting of female authors is a strong positive predictor for the establishment of a collaboration network. This pattern is also observed by Boschini and Sjögren (2007) during the 1991-2002 time period. On the contrary, this analysis suggests that junior scientists are less likely to participate in a project with a large share of senior authors. One likely explanation is that successful collaborations tend to perdure, and it is costly to establish new connections (cf. Hsieh et al. 2022). Notice that the estimates predict a negative constant coefficient, which is expected from a sparse bipartite network. The analysis indicates that the average number of authors per project is 1.74.

When comparing these results to those of the plug-in estimator based on initial logistic and Poisson regressions in Tables 11 and 12, we observe that the plug-in estimators fail to capture the effects predicted on citations. Moreover, the one-step robust estimator provides substantial improvements in predicting the homophily coefficients and the average out-degree statistics. In particular, the robust estimator attains smaller theoretical rMSEs for almost all the estimates

---

[11]We use information on the authors collaborating in a given project to construct these project-specific controls.

[12]The category of senior economists includes associate and full professors and senior economists in government or research institutions.

[13]Ideally, the author's productivity should be measured using the citations count prior to year 2000. However, this is not possible because of data limitations. Alternatively, a different measure of productivity could be considered, such as the average number of citations received by the institution granting the PhD to the author.

except for the gender attribute. Notice that the estimates associated with gender present the largest theoretical rMSEs across all the methodologies. This might be related to the fact that the shares of female authors observed in the sample and in a given project are small. The methodologies are comparable in terms of standard errors and confidence intervals' length.

7   Conclusion

We study the effects of local misspecification on a bipartite network. We focus on a class of dyadic network models characterized by conditional moment restrictions that are locally misspecified. The magnitude of misspecification is indexed by the sample size and vanishes at a rate $n^{-1/2}$. We utilize this local asymptotic approach to construct a robust estimator that is MSE-minimax optimal within a prespecified neighborhood of misspecification. Additionally, we introduce bias-aware confidence intervals that account for the effect of the local misspecification. These confidence intervals are asymptotically valid for the structural parameters of interest under sparse network asymptotics, both in the correctly-specified and locally-misspecified case. In an empirical application, we study the formation of a scientific collaboration network among economists. Our analysis documents that homophily in the research field of expertise, sorting of highly productive scientists into higher-impact projects, and participation of female authors in teams with a higher share of female authors are all strong statistical factors that explain the formation of a collaboration network.

Table 1: $\hat{\beta}_n$ Robust Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.386	1.385	−0.064	0.147	0.885	2.165	[0.651,	2.112]	1.461	1.000	0.010
1.0	1.386	1.383	−0.075	0.146	0.885	2.166	[0.651,	2.111]	1.460	1.000	0.010
2.0	1.386	1.382	−0.074	0.147	0.884	2.167	[0.651,	2.111]	1.461	1.000	0.010
3.0	1.386	1.381	−0.071	0.147	0.884	2.163	[0.651,	2.111]	1.460	1.000	0.010
4.0	1.386	1.381	−0.066	0.146	0.882	2.155	[0.653,	2.111]	1.458	1.000	0.010
						$n = 400$					
0.0	1.386	1.388	0.049	0.108	0.935	2.228	[0.859,	1.919]	1.060	1.000	0.005
1.0	1.386	1.389	0.054	0.108	0.933	2.227	[0.859,	1.919]	1.060	1.000	0.005
2.0	1.386	1.390	0.051	0.108	0.932	2.227	[0.859,	1.919]	1.059	1.000	0.005
3.0	1.386	1.389	0.054	0.108	0.932	2.225	[0.859,	1.919]	1.059	1.000	0.005
4.0	1.386	1.391	0.053	0.108	0.927	2.224	[0.860,	1.918]	1.059	1.000	0.005
						$n = 600$					
0.0	1.386	1.385	0.022	0.090	0.914	2.254	[0.952,	1.822]	0.870	1.000	0.003
1.0	1.386	1.385	0.018	0.090	0.917	2.254	[0.952,	1.822]	0.870	1.000	0.003
2.0	1.386	1.387	0.019	0.090	0.916	2.255	[0.952,	1.822]	0.870	1.000	0.003
3.0	1.386	1.386	0.024	0.090	0.915	2.255	[0.952,	1.822]	0.870	1.000	0.003
4.0	1.386	1.387	0.015	0.090	0.916	2.255	[0.952,	1.822]	0.870	1.000	0.003
						$n = 800$					
0.0	1.386	1.382	−0.035	0.079	0.930	2.288	[1.005,	1.765]	0.760	1.000	0.003
1.0	1.386	1.383	−0.037	0.079	0.931	2.288	[1.005,	1.765]	0.760	1.000	0.003
2.0	1.386	1.383	−0.036	0.079	0.933	2.288	[1.005,	1.765]	0.760	1.000	0.003
3.0	1.386	1.383	−0.040	0.079	0.931	2.287	[1.005,	1.765]	0.760	1.000	0.003
4.0	1.386	1.382	−0.047	0.079	0.934	2.286	[1.005,	1.764]	0.759	1.000	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] DGP Local misspecification: Latent Homophily. [3] Sieves dimension $k_n = 2$.

Table 2: $\hat{\beta}_n$ Logistic Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf.	int.]	length	95% CP	Degree
						$n = 200$					
0.0	1.386	1.418	0.439	0.160	0.899	2.358	[0.661,	2.174]	1.513	1.000	0.010
1.0	1.386	1.417	0.428	0.159	0.896	2.359	[0.660,	2.173]	1.512	1.000	0.010
2.0	1.386	1.418	0.429	0.159	0.892	2.359	[0.660,	2.173]	1.513	1.000	0.010
3.0	1.386	1.417	0.433	0.159	0.895	2.356	[0.661,	2.173]	1.512	1.000	0.010
4.0	1.386	1.415	0.440	0.158	0.891	2.349	[0.662,	2.173]	1.510	1.000	0.010
						$n = 400$					
0.0	1.386	1.407	0.418	0.114	0.946	2.348	[0.866,	1.948]	1.082	1.000	0.005
1.0	1.386	1.408	0.423	0.114	0.944	2.348	[0.866,	1.949]	1.083	1.000	0.005
2.0	1.386	1.409	0.420	0.114	0.942	2.347	[0.866,	1.948]	1.082	1.000	0.005
3.0	1.386	1.407	0.423	0.114	0.941	2.345	[0.866,	1.948]	1.082	1.000	0.005
4.0	1.386	1.408	0.422	0.114	0.938	2.344	[0.867,	1.948]	1.082	1.000	0.005
						$n = 600$					
0.0	1.386	1.397	0.322	0.093	0.921	2.336	[0.958,	1.841]	0.883	1.000	0.003
1.0	1.386	1.397	0.319	0.093	0.923	2.336	[0.958,	1.841]	0.883	1.000	0.003
2.0	1.386	1.398	0.320	0.093	0.921	2.337	[0.958,	1.841]	0.883	1.000	0.003
3.0	1.386	1.398	0.325	0.093	0.922	2.336	[0.958,	1.841]	0.883	1.000	0.003
4.0	1.386	1.398	0.316	0.093	0.922	2.337	[0.958,	1.841]	0.883	1.000	0.003
						$n = 800$					
0.0	1.386	1.391	0.225	0.081	0.936	2.354	[1.010,	1.779]	0.769	1.000	0.003
1.0	1.386	1.392	0.224	0.081	0.935	2.354	[1.010,	1.779]	0.769	1.000	0.003
2.0	1.386	1.392	0.224	0.081	0.937	2.354	[1.010,	1.779]	0.769	1.000	0.003
3.0	1.386	1.392	0.221	0.081	0.938	2.353	[1.010,	1.778]	0.769	1.000	0.003
4.0	1.386	1.391	0.214	0.081	0.938	2.352	[1.010,	1.778]	0.768	1.000	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] DGP Local misspecification: Latent Homophily.

Table 3: $\hat{\beta}_n$ Poisson Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
					$n = 200$						
0.0	1.386	1.384	-0.082	0.149	0.905	2.206	[0.660,	2.101]	1.440	1.000	0.010
1.0	1.386	1.381	-0.092	0.149	0.907	2.207	[0.660,	2.100]	1.440	1.000	0.010
2.0	1.386	1.380	-0.091	0.150	0.905	2.207	[0.660,	2.100]	1.440	1.000	0.010
3.0	1.386	1.380	-0.089	0.150	0.907	2.204	[0.660,	2.100]	1.439	1.000	0.010
4.0	1.386	1.380	-0.084	0.148	0.900	2.196	[0.662,	2.099]	1.437	1.000	0.010
					$n = 400$						
0.0	1.386	1.388	0.044	0.109	0.948	2.260	[0.862,	1.915]	1.054	1.000	0.005
1.0	1.386	1.389	0.049	0.109	0.946	2.259	[0.862,	1.916]	1.054	1.000	0.005
2.0	1.386	1.390	0.046	0.110	0.945	2.258	[0.862,	1.915]	1.053	1.000	0.005
3.0	1.386	1.389	0.048	0.110	0.946	2.257	[0.862,	1.915]	1.053	1.000	0.005
4.0	1.386	1.391	0.047	0.109	0.941	2.255	[0.862,	1.915]	1.053	1.000	0.005
					$n = 600$						
0.0	1.386	1.385	0.019	0.091	0.923	2.276	[0.954,	1.821]	0.867	1.000	0.003
1.0	1.386	1.385	0.015	0.091	0.925	2.276	[0.953,	1.820]	0.867	1.000	0.003
2.0	1.386	1.386	0.016	0.091	0.924	2.277	[0.953,	1.820]	0.867	1.000	0.003
3.0	1.386	1.386	0.021	0.091	0.924	2.277	[0.954,	1.821]	0.867	1.000	0.003
4.0	1.386	1.387	0.012	0.091	0.924	2.277	[0.953,	1.820]	0.867	1.000	0.003
					$n = 800$						
0.0	1.386	1.382	-0.037	0.080	0.938	2.306	[1.006,	1.764]	0.758	1.000	0.003
1.0	1.386	1.383	-0.039	0.080	0.937	2.306	[1.006,	1.764]	0.758	1.000	0.003
2.0	1.386	1.383	-0.038	0.080	0.939	2.306	[1.006,	1.764]	0.758	1.000	0.003
3.0	1.386	1.383	-0.042	0.080	0.939	2.305	[1.006,	1.764]	0.758	1.000	0.003
4.0	1.386	1.382	-0.049	0.080	0.941	2.304	[1.006,	1.763]	0.758	1.000	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] DGP Local misspecification: Latent Homophily.

Table 4: $\hat{\beta}_n$ Ratios

$\bar{\sigma}^2$	Logit/Robust				Poisson/Robust			
	s.e	bias	rMSE	length	s.e	bias	rMSE	length
				$n = 200$				
0.0	1.090	6.859	1.089	1.036	1.018	1.275	1.019	0.986
1.0	1.087	5.712	1.089	1.036	1.020	1.235	1.019	0.986
2.0	1.083	5.834	1.089	1.036	1.018	1.240	1.019	0.986
3.0	1.086	6.103	1.089	1.036	1.021	1.250	1.019	0.986
4.0	1.084	6.642	1.090	1.036	1.015	1.269	1.019	0.986
				$n = 400$				
0.0	1.052	8.472	1.054	1.022	1.013	0.883	1.014	0.994
1.0	1.051	7.784	1.054	1.022	1.012	0.893	1.014	0.994
2.0	1.050	8.173	1.054	1.022	1.013	0.888	1.014	0.994
3.0	1.050	7.886	1.054	1.022	1.014	0.892	1.014	0.994
4.0	1.052	7.999	1.054	1.022	1.014	0.890	1.014	0.994
				$n = 600$				
0.0	1.035	14.924	1.036	1.015	1.009	0.863	1.010	0.997
1.0	1.032	17.598	1.036	1.015	1.008	0.837	1.010	0.997
2.0	1.032	16.727	1.036	1.015	1.008	0.845	1.010	0.997
3.0	1.034	13.386	1.036	1.015	1.009	0.878	1.010	0.997
4.0	1.033	20.570	1.036	1.015	1.008	0.807	1.010	0.997
				$n = 800$				
0.0	1.027	6.402	1.029	1.012	1.008	1.054	1.008	0.998
1.0	1.025	6.104	1.029	1.012	1.006	1.051	1.008	0.998
2.0	1.025	6.240	1.029	1.012	1.007	1.052	1.008	0.998
3.0	1.028	5.517	1.029	1.012	1.009	1.047	1.008	0.998
4.0	1.024	4.526	1.029	1.012	1.006	1.040	1.008	0.998

[1] Number of Monte Carlo simulations is 2,000.

[2] DGP Local misspecification: Latent Homophily.

[3] Sieves dimension $k_n = 2$.

Table 5: $\hat{\Psi}_n$ Robust Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.041	1.021	-0.118	0.162	0.950	2.406	[0.697,	1.368]	0.672	0.931	0.010
1.0	1.041	1.021	-0.134	0.162	0.950	2.405	[0.696,	1.368]	0.672	0.930	0.010
2.0	1.041	1.020	-0.126	0.162	0.950	2.407	[0.696,	1.368]	0.672	0.928	0.010
3.0	1.043	1.020	-0.135	0.162	0.955	2.409	[0.697,	1.370]	0.672	0.927	0.010
4.0	1.048	1.026	-0.163	0.163	0.957	2.414	[0.700,	1.374]	0.674	0.926	0.010
						$n = 400$					
0.0	1.029	1.029	0.129	0.117	0.955	2.417	[0.796,	1.276]	0.479	0.945	0.005
1.0	1.030	1.029	0.118	0.118	0.957	2.416	[0.796,	1.275]	0.479	0.945	0.005
2.0	1.030	1.029	0.114	0.118	0.956	2.416	[0.796,	1.275]	0.479	0.942	0.005
3.0	1.031	1.029	0.110	0.118	0.955	2.418	[0.796,	1.276]	0.480	0.942	0.005
4.0	1.033	1.032	0.112	0.118	0.953	2.422	[0.798,	1.279]	0.480	0.944	0.005
						$n = 600$					
0.0	1.040	1.029	-0.189	0.097	0.984	2.417	[0.837,	1.229]	0.392	0.942	0.003
1.0	1.040	1.028	-0.194	0.097	0.986	2.416	[0.836,	1.228]	0.392	0.942	0.003
2.0	1.040	1.030	-0.194	0.097	0.987	2.417	[0.836,	1.228]	0.392	0.942	0.003
3.0	1.040	1.030	-0.184	0.097	0.989	2.418	[0.837,	1.229]	0.392	0.943	0.003
4.0	1.042	1.030	-0.186	0.097	0.987	2.421	[0.838,	1.231]	0.393	0.943	0.003
						$n = 800$					
0.0	1.033	1.030	-0.013	0.084	0.965	2.427	[0.862,	1.203]	0.340	0.946	0.003
1.0	1.033	1.030	0.003	0.084	0.965	2.426	[0.862,	1.203]	0.340	0.947	0.003
2.0	1.032	1.030	0.005	0.084	0.964	2.426	[0.862,	1.203]	0.340	0.946	0.003
3.0	1.033	1.030	0.011	0.084	0.963	2.427	[0.863,	1.203]	0.340	0.947	0.003
4.0	1.034	1.032	0.008	0.084	0.964	2.429	[0.864,	1.205]	0.341	0.947	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] DGP Local misspecification: Latent Homophily. [3] Sieves dimension $k_n = 3$.

Table 6: $\hat{\Psi}_n$ Logistic Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.041	1.053	0.359	0.172	0.936	2.606	[0.715,	1.417]	0.701	0.938	0.010
1.0	1.041	1.052	0.343	0.172	0.936	2.605	[0.715,	1.416]	0.701	0.936	0.010
2.0	1.041	1.053	0.350	0.172	0.937	2.606	[0.715,	1.417]	0.701	0.935	0.010
3.0	1.043	1.053	0.343	0.172	0.940	2.610	[0.716,	1.418]	0.702	0.939	0.010
4.0	1.048	1.057	0.317	0.173	0.943	2.615	[0.719,	1.423]	0.704	0.936	0.010
						$n = 400$					
0.0	1.029	1.046	0.452	0.121	0.952	2.502	[0.810,	1.294]	0.485	0.945	0.005
1.0	1.030	1.045	0.441	0.121	0.953	2.501	[0.809,	1.294]	0.485	0.947	0.005
2.0	1.030	1.046	0.437	0.121	0.953	2.502	[0.809,	1.294]	0.485	0.945	0.005
3.0	1.031	1.045	0.433	0.121	0.951	2.503	[0.810,	1.295]	0.485	0.949	0.005
4.0	1.033	1.047	0.437	0.121	0.950	2.507	[0.812,	1.297]	0.486	0.946	0.005
						$n = 600$					
0.0	1.040	1.039	0.072	0.099	0.985	2.473	[0.847,	1.240]	0.393	0.946	0.003
1.0	1.040	1.038	0.068	0.099	0.987	2.472	[0.847,	1.239]	0.393	0.945	0.003
2.0	1.040	1.039	0.068	0.099	0.988	2.472	[0.847,	1.240]	0.393	0.943	0.003
3.0	1.040	1.039	0.078	0.099	0.989	2.473	[0.847,	1.240]	0.393	0.946	0.003
4.0	1.042	1.040	0.077	0.099	0.986	2.476	[0.849,	1.242]	0.393	0.946	0.003
						$n = 800$					
0.0	1.033	1.038	0.212	0.086	0.963	2.465	[0.871,	1.210]	0.340	0.944	0.003
1.0	1.033	1.038	0.227	0.086	0.964	2.465	[0.871,	1.210]	0.340	0.942	0.003
2.0	1.032	1.037	0.229	0.086	0.962	2.465	[0.871,	1.210]	0.340	0.942	0.003
3.0	1.033	1.037	0.236	0.086	0.964	2.465	[0.871,	1.211]	0.340	0.945	0.003
4.0	1.034	1.039	0.233	0.086	0.964	2.468	[0.872,	1.212]	0.340	0.945	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] DGP Local misspecification: Latent Homophily.

Table 7: $\hat{\Psi}_n$ Poisson Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.041	1.020	-0.085	0.163	0.960	2.406	[0.708,	1.362]	0.654	0.926	0.010
1.0	1.041	1.020	-0.100	0.163	0.961	2.406	[0.707,	1.361]	0.654	0.924	0.010
2.0	1.041	1.020	-0.092	0.163	0.960	2.406	[0.707,	1.362]	0.654	0.921	0.010
3.0	1.043	1.020	-0.101	0.163	0.964	2.408	[0.708,	1.363]	0.655	0.924	0.010
4.0	1.048	1.030	-0.129	0.164	0.964	2.414	[0.711,	1.368]	0.657	0.922	0.010
						$n = 400$					
0.0	1.029	1.030	0.158	0.118	0.960	2.424	[0.802,	1.272]	0.470	0.944	0.005
1.0	1.030	1.030	0.147	0.118	0.961	2.423	[0.802,	1.272]	0.470	0.944	0.005
2.0	1.030	1.030	0.143	0.118	0.962	2.423	[0.802,	1.272]	0.470	0.940	0.005
3.0	1.031	1.030	0.139	0.118	0.961	2.425	[0.802,	1.273]	0.470	0.940	0.005
4.0	1.033	1.035	0.141	0.118	0.958	2.428	[0.804,	1.275]	0.471	0.942	0.005
						$n = 600$					
0.0	1.040	1.030	-0.163	0.097	0.991	2.423	[0.841,	1.226]	0.385	0.941	0.003
1.0	1.040	1.030	-0.167	0.097	0.992	2.422	[0.841,	1.226]	0.385	0.940	0.003
2.0	1.040	1.030	-0.167	0.098	0.993	2.423	[0.841,	1.226]	0.385	0.940	0.003
3.0	1.040	1.030	-0.157	0.098	0.994	2.424	[0.841,	1.227]	0.385	0.941	0.003
4.0	1.042	1.030	-0.159	0.097	0.991	2.427	[0.843,	1.229]	0.386	0.942	0.003
						$n = 800$					
0.0	1.033	1.030	0.009	0.084	0.967	2.429	[0.866,	1.201]	0.335	0.944	0.003
1.0	1.033	1.030	0.025	0.085	0.970	2.428	[0.866,	1.201]	0.335	0.943	0.003
2.0	1.032	1.030	0.027	0.085	0.968	2.428	[0.866,	1.201]	0.335	0.944	0.003
3.0	1.033	1.030	0.033	0.085	0.969	2.429	[0.866,	1.201]	0.335	0.945	0.003
4.0	1.034	1.033	0.030	0.085	0.969	2.431	[0.867,	1.203]	0.335	0.944	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] DGP Local misspecification: Latent Homophily.

Table 8: $\hat{\Psi}_n$ Ratios

$\bar{\sigma}^2$	Logit/Robust				Poisson/Robust			
	s.e	bias	rMSE	length	s.e	bias	rMSE	length
				$n = 200$				
0.0	1.062	3.031	1.083	1.044	1.008	0.715	1.000	0.974
1.0	1.062	2.559	1.083	1.044	1.010	0.745	1.000	0.974
2.0	1.063	2.771	1.083	1.044	1.008	0.731	1.000	0.974
3.0	1.062	2.550	1.084	1.044	1.007	0.750	1.000	0.974
4.0	1.063	1.948	1.084	1.045	1.005	0.794	1.000	0.974
				$n = 400$				
0.0	1.033	3.506	1.035	1.011	1.006	1.226	1.003	0.981
1.0	1.032	3.735	1.035	1.011	1.005	1.247	1.003	0.981
2.0	1.032	3.842	1.035	1.011	1.006	1.256	1.003	0.981
3.0	1.031	3.951	1.035	1.011	1.006	1.265	1.003	0.981
4.0	1.033	3.898	1.035	1.011	1.005	1.260	1.002	0.980
				$n = 600$				
0.0	1.024	0.383	1.023	1.002	1.006	0.859	1.002	0.983
1.0	1.024	0.349	1.023	1.002	1.006	0.862	1.002	0.983
2.0	1.024	0.349	1.023	1.002	1.006	0.861	1.002	0.983
3.0	1.023	0.424	1.023	1.002	1.004	0.852	1.002	0.983
4.0	1.022	0.414	1.023	1.002	1.004	0.855	1.002	0.982
				$n = 800$				
0.0	1.016	16.855	1.016	0.998	1.002	0.739	1.001	0.983
1.0	1.017	89.595	1.016	0.998	1.005	9.672	1.001	0.983
2.0	1.017	48.532	1.016	0.998	1.004	5.640	1.001	0.983
3.0	1.019	21.420	1.016	0.998	1.006	3.003	1.001	0.984
4.0	1.019	29.352	1.016	0.998	1.005	3.778	1.001	0.984

[1] Number of Monte Carlo simulations is $2,000$.

[2] DGP Local misspecification: Latent Homophily.

[3] Sieves dimension $k_n = 3$.

Table 9: Descriptive Statistics

	mean	median	std	min	max
network_links	1.584	1.000	1.161	0.000	14.000
Author's attributes					
gender	0.106	0.000	0.308	0.000	1.000
junior dummy	0.330	0.000	0.470	0.000	1.000
citations	45.701	35.000	36.065	0.000	163.000
Project's attributes					
share_gender	0.095	0.000	0.233	0.000	1.000
number_authors	1.958	2.000	0.775	1.000	5.000
share_seniors	0.486	0.500	0.395	0.000	1.000
same_institution	0.181	0.000	0.385	0.000	1.000
same_PhD	0.179	0.000	0.384	0.000	1.000

[1] Total sample includes $N = 1776$ authors and $M = 1600$ articles.

Table 10: Results Robust Estimator

	coeff.	s.e.	max rMSE	[conf. int.]	length	adjm
Homophily parameters						
jelcode	1.633	0.045	10.244	[1.375, 1.891]	0.516	0.127
citations	0.391	0.015	10.253	[0.187, 0.596]	0.409	0.088
gender	5.938	0.180	69.120	[4.408, 7.467]	3.059	0.481
junior_senior	−3.671	0.130	23.324	[−4.306, −3.036]	1.270	−0.623
constant	−0.834	0.080	15.207	[−1.239, −0.428]	0.811	−0.069
Average out-degree						
Psi	1.739	0.031	2.063	[1.661, 1.817]	0.156	−0.032

[1] Total sample includes $N = 1776$ authors and $M = 1600$ articles.

Table 11: Results Logit Estimator

	coeff.	s.e.	max rMSE	[conf. int.]	length	adjm
Homophily parameters						
jelcode	1.507	0.043	11.773	[1.225, 1.788]	0.563	0.000
citations	0.304	0.009	21.633	[−0.087, 0.694]	0.782	0.000
gender	5.457	0.175	71.557	[3.895, 7.018]	3.123	0.000
junior_senior	−3.048	0.132	28.520	[−3.780, −2.316]	1.465	0.000
constant	−0.765	0.080	16.870	[−1.201, −0.329]	0.872	0.000
Average out-degree						
Psi	1.771	0.036	2.111	[1.700, 1.842]	0.142	0.000

[1] Total sample includes $N = 1776$ authors and $M = 1600$ articles.

Table 12: Results Poisson Estimator

	coeff.	s.e.	max rMSE	[conf. int.]	length	adjm
Homophily parameters						
jelcode	1.499	0.042	10.271	[1.244, 1.754]	0.510	0.000
citations	0.287	0.009	20.408	[−0.082, 0.656]	0.739	0.000
gender	5.415	0.173	66.372	[3.947, 6.883]	2.936	0.000
junior_senior	−3.036	0.132	26.363	[−3.729, −2.343]	1.386	0.000
constant	−0.756	0.080	15.915	[−1.175, −0.338]	0.836	0.000
Average out-degree						
Psi	1.759	0.036	2.070	[1.689, 1.829]	0.140	0.000

[1] Total sample includes $N = 1776$ authors and $M = 1600$ articles.

Figure 1: Confidence Intervals: `jelcode`
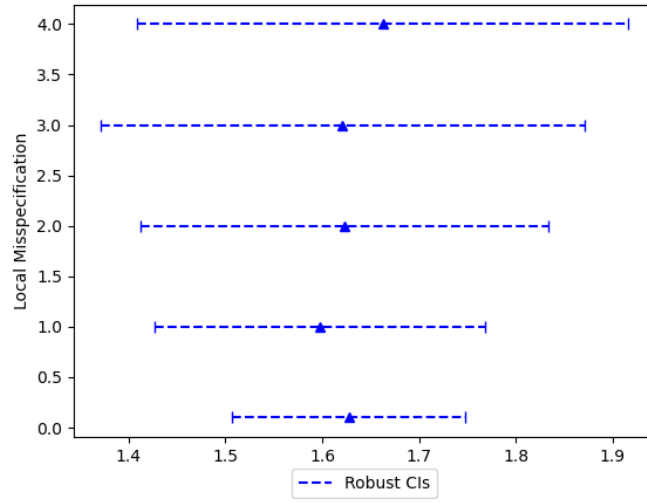
Figure 2: Confidence Intervals: `citations`
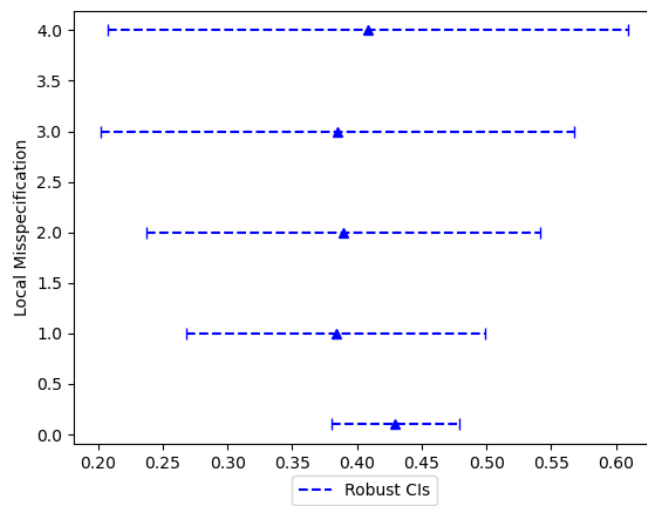
Figure 3: Confidence Intervals: `gender`
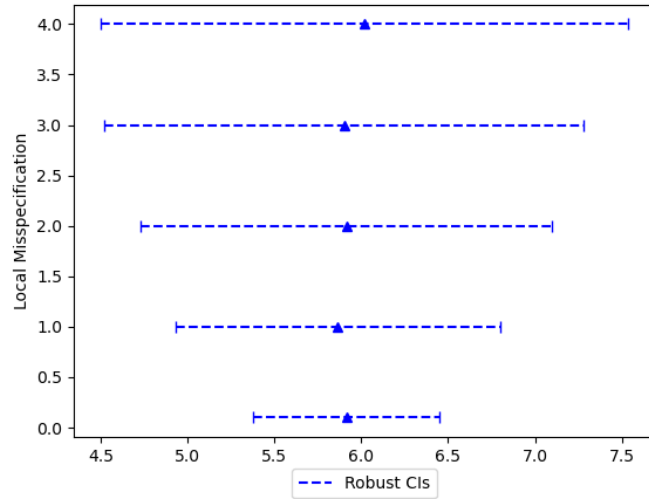
Figure 4: Confidence Intervals: `junior-senior`
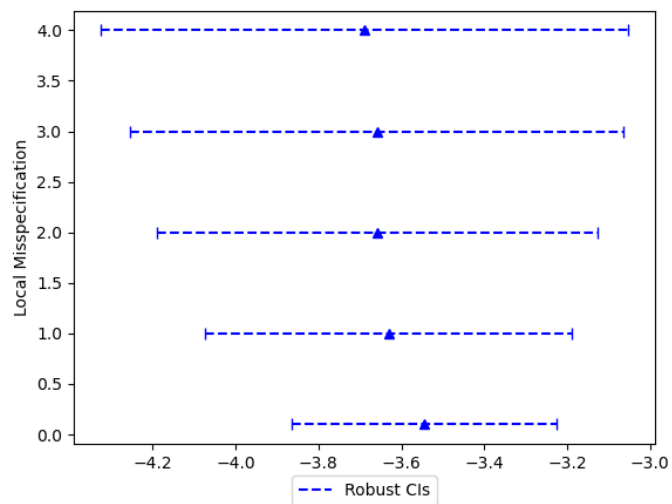
Figure 5: Confidence Intervals: `constant`
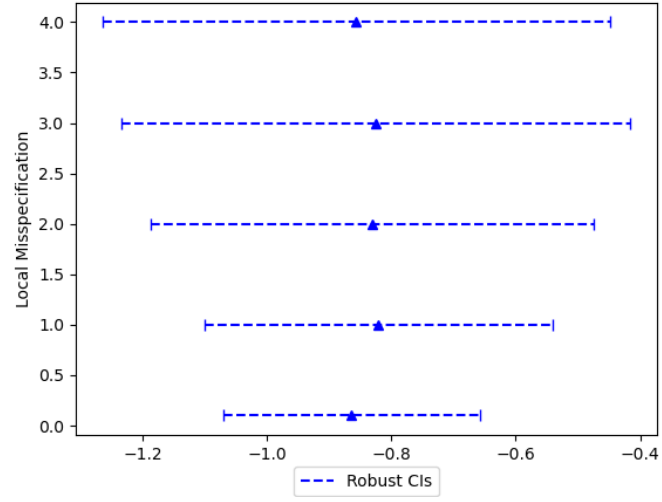
Figure 6: Confidence Intervals: $\Psi_{0,n}$
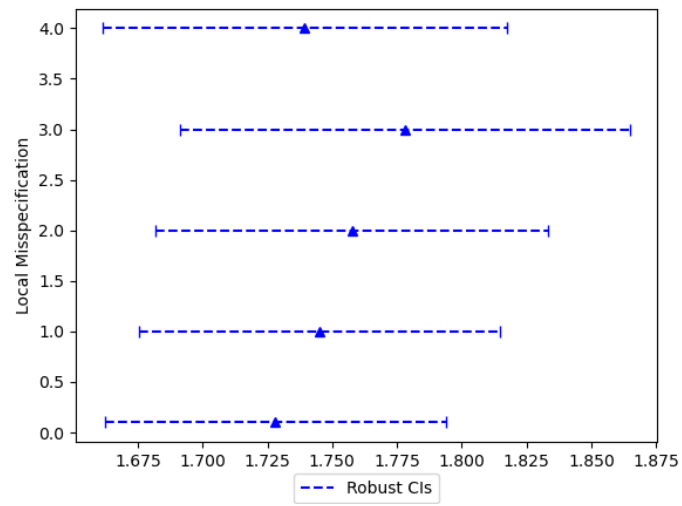
Mathematical Appendix

Throughout the appendix, we define $(X_1, \cdots, X_N)$, $(A_1, \cdots, A_N)$, $(W_1, \cdots, W_M)$, and $(B_1, \cdots, B_M)$ as $\mathcal{X}_N$, $\mathcal{A}_N$, $\mathcal{W}_M$, and $\mathcal{B}_M$, respectively.

A   Proof of Theorem 1

We divide the proof into two steps. In the first step, we establish the strong approximation

$$\sqrt{n}(\hat{\Psi}(\hat{\kappa}^*) - \Psi_{0,n}) - \omega_n(\kappa^*) = o_P(1), \tag{A.1}$$

where

$$\omega_n(\kappa^*) \stackrel{d}{=} \mathcal{N}\left(\mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top\beta_0)H_{ij} - \Delta_{v,ij}), \Sigma_\gamma + 2\Sigma_{H,\gamma}^\top\kappa^* + \kappa^{*,\top}\Sigma_H\kappa^*\right).$$

In the second step, we prove the desired result. The proof relies on the following lemmas, which are proved in Section C.

Lemma A.1. *Suppose the Assumptions in Theorem 1 hold. Define* $\hat{G}(\theta) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} n\exp(R_{ij}^\top\theta)H_{ij}R_{ij}^\top$ *so that* $\hat{G} = \hat{G}(\theta_{initial})$. *Then, we have that*

$$\sup_{\theta\in\mathbb{B}(\theta_{0,n}, C/\sqrt{n})} \left\|\hat{G}(\theta) - G_0\right\|_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2\log^2(n)}{n}}\right) \quad \text{for any constant } C > 0,$$

$$\left\|\hat{B}_H - B_H\right\|_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2\log^2(n)}{n}}\right), \quad \left\|\hat{B}_{H,\gamma} - B_{H,\gamma}\right\|_2 = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2\log^2(n)}{n}}\right),$$

$$\sup_{\theta\in\mathbb{B}(\theta_{0,n}, C/\sqrt{n})} \left\|\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\partial_\theta\gamma_n(D_{ij}, 0, \theta) - \Gamma\right\|_{op} = O_P\left(\frac{1}{\sqrt{n}}\right), \quad \|\hat{B}_\gamma - B_\gamma\|_2 = O_P\left(\frac{1}{\sqrt{n}}\right),$$

$$\left\|\hat{\Sigma}_H - \Sigma_H\right\|_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right), \quad \left\|\hat{\Sigma}_{H,\gamma} - \Sigma_{H,\gamma}\right\|_2 = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right),$$

$$|\hat{\Sigma}_\gamma - \Sigma_\gamma| = O_P(\zeta_n n^{-1/2}), \quad \|\hat{\kappa}^* - \kappa^*\|_2 = o_P(k_n^{-1/2}), \quad \text{and}$$

$$\|\kappa^*\|_2 \leq C < \infty \quad \text{for some constant } C > 0,$$

where $\mathbb{B}(\theta_{0,n}, C/\sqrt{n})$ *is a ball centered at* $\theta_{0,n}$ *with radius* $C/\sqrt{n}$.

Lemma A.2. *Suppose the Assumptions in Theorem 1 hold. Then there exists a* $(k_n + 1) \times 1$

Gaussian vector g_n such that $g_n \overset{d}{=} \mathcal{N}(0, \Sigma_n)$ and

$$\left\| \sqrt{n} \begin{pmatrix} \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n \left(Y_{ij} - \int \Lambda(R_{ij}^\top \theta_{0,n} + n^{-1/2}v)\pi_{ij}(v)dv \right) H_{ij} \\ \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} [\gamma_n(D_{ij}, 0, \theta_{0,n}) - \mathbb{E}\gamma_n(D_{ij}, 0, \theta_{0,n})] \end{pmatrix} - g_n \right\|_2 = O_P(\sqrt{k_n^5/n}),$$

where Σ_n is defined in (2.16).

A.1   Step 1: Proof of (A.1)

We have

$$\sqrt{n}(\hat{\Psi}(\hat{\kappa}^*) - \Psi_{0,n}) = \frac{\sqrt{n}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (\gamma(D_{ij}, 0, \hat{\theta}_{\text{initial}}) - \mathbb{E}\gamma(D_{ij}, 0, \theta_{0,n})) - \mathbb{E}\eta_{ij}\Delta_{v,ij}$$

$$+ \hat{\kappa}^{*,\top} \left[\frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (Y_{ij} - \Lambda(R_{ij}^\top \hat{\theta}_{\text{initial}})) H_{ij} \right] + o(1)$$

$$= \frac{\sqrt{n}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (\gamma(D_{ij}, 0, \theta_{0,n}) - \mathbb{E}\gamma(D_{ij}, 0, \theta_{0,n})) - \mathbb{E}\eta_{ij}\Delta_{v,ij}$$

$$+ \left[\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \partial_\theta\gamma_n(D_{ij}, 0, \tilde{\theta})^\top - \hat{\kappa}^{*,\top}\hat{G}(\tilde{\theta}) \right] \sqrt{n}(\hat{\theta}_{\text{initial}} - \theta_{0,n})$$

$$+ \hat{\kappa}^{*,\top} \left[\frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (Y_{ij} - \Lambda(R_{ij}^\top \theta_{0,n})) H_{ij} \right] + o_P(1)$$

$$= \frac{\sqrt{n}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (\gamma(D_{ij}, 0, \theta_{0,n}) - \mathbb{E}\gamma(D_{ij}, 0, \theta_{0,n})) - \mathbb{E}\eta_{ij}\Delta_{v,ij}$$

$$+ \hat{\kappa}^{*,\top} \left[\frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (Y_{ij} - \int \Lambda(R_{ij}^\top \theta_{0,n} + n^{-1/2}v)d\pi_{ij}H_{ij} \right]$$

$$+ \hat{\kappa}^{*,\top} \left[\frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \int (\Lambda(R_{ij}^\top \theta_{0,n} + n^{-1/2}v) - \Lambda(R_{ij}^\top \theta_{0,n}))d\pi_{ij}H_{ij} \right] + o_P(1)$$

$$= \frac{\sqrt{n}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (\gamma(D_{ij}, 0, \theta_{0,n}) - \mathbb{E}\gamma(D_{ij}, 0, \theta_{0,n})) + \mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top\beta_0)H_{ij} - \Delta_{v,ij})$$

$$+ \kappa^{*,\top} \left[\frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (Y_{ij} - \int \Lambda(R_{ij}^\top \theta_{0,n} + vn^{-1/2})d\pi_{ij})H_{ij} \right] + o_P(1), \qquad (A.2)$$

where the first equality is by (3.1), the second equality is by the mean-value theorem where $\tilde{\theta}$ is between $\hat{\theta}_{\text{initial}}$ and $\theta_{0,n}$ and the facts that

$$\hat{\kappa}^{*,\top} \left[\frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (\Lambda(R_{ij}^{\top} \hat{\theta}_{\text{initial}})) H_{ij} - \Lambda(R_{ij}^{\top} \theta_{0,n})) H_{ij} \right]$$

$$= \hat{\kappa}^{*,\top} \left[\frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (\Lambda(R_{ij}^{\top} \tilde{\theta}))(1 - \Lambda(R_{ij}^{\top} \tilde{\theta})) H_{ij} R_{ij}^{\top} \right] (\hat{\theta}_{\text{initial}} - \theta_{0,n})$$

$$= \kappa^{*,\top} \left[\hat{G}(\tilde{\theta}) - \frac{n}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\exp(2R_{ij}^{\top} \hat{\theta}_{\text{initial}})(2 + \exp(R_{ij}^{\top} \hat{\theta}_{\text{initial}}))}{(1 + \exp(R_{ij}^{\top} \hat{\theta}_{\text{initial}}))^2} \right] H_{ij} R_{ij} \sqrt{n} (\hat{\theta}_{\text{initial}} - \theta_{0,n})$$

and

$$\left| \kappa^{*,\top} \left[\frac{n}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\exp(2R_{ij}^{\top} \hat{\theta}_{\text{initial}})(2 + \exp(R_{ij}^{\top} \hat{\theta}_{\text{initial}}))}{(1 + \exp(R_{ij}^{\top} \hat{\theta}_{\text{initial}}))^2} \right] \sqrt{n} (\hat{\theta}_{\text{initial}} - \theta_{0,n}) \right|$$

$$\leq 2 \| \kappa^{*,\top} \|_2 \left\| \frac{n}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \exp(2R_{ij}^{\top} \hat{\theta}_{\text{initial}}) H_{ij} R_{ij} \right\|_{op} \left\| 2\sqrt{n} (\hat{\theta}_{\text{initial}} - \theta_{0,n}) \right\|_2$$

$$\leq O_P(n^{-1}) \times \left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \exp\left(2(\hat{\alpha}_n - \log n) + 2Z_{ij}^{\top} \hat{\beta}_n \right) H_{ij} R_{ij} \right\|_{op} = O_P(n^{-1}),$$

and the third equality is by Lemma A.1,

$$\left\| \hat{G}^{\top}(\tilde{\theta}) \hat{\kappa}^* - \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \partial_\theta \gamma_n(D_{ij}, 0, \tilde{\theta}) \right\|_{op}$$

$$\leq \left\| (\hat{G}(\tilde{\theta}) - G_0) \right\|_{op} \| \hat{\kappa}^* \|_2 + \| G_0 \|_{op} \| \hat{\kappa}^* - \kappa^* \|_2 + \left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \partial_\theta \gamma_n(D_{ij}, 0, \tilde{\theta}) - \Gamma \right\|_{op}$$

$$= O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}} \right) + o_P(1) = o_P(1)$$

and $\sqrt{n}(\hat{\theta}_{\text{initial}} - \theta_{0,n}) = O_P(1)$. To see the last equality of (A.2), we note that

$$\left| (\kappa^* - \hat{\kappa}^*)^{\top} \left[\frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (Y_{ij} - \int \Lambda(R_{ij}^{\top} \theta_{0,n} + vn^{-1/2}) d\pi_{ij}) H_{ij} \right] \right|$$

$$\leq \| \kappa^* - \hat{\kappa}^* \|_2 \left\| \frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (Y_{ij} - \int \Lambda(R_{ij}^{\top} \theta_{0,n} + vn^{-1/2}) d\pi_{ij}) H_{ij} \right\|_2$$

$$\leq \| \kappa^* - \hat{\kappa}^* \|_2 (\| g_n \|_2 + o_P(1)) = o_P(1),$$

where the second inequality is by Lemma A.2 and the facts that $||\kappa^* - \hat{\kappa}^*||_2 = o_P(k_n^{-1/2})$ as shown in Lemma A.1 and $||g_n||_2 = O_P(k_n^{1/2})$.[14]

In addition, we have

$$
\left\| \hat{\kappa}^{*,\top} \left[\frac{n^{3/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \int (\Lambda(R_{ij}^\top \theta_{0,n} + n^{-1/2}v) - \Lambda(R_{ij}^\top \theta_{0,n})) d\pi_{ij} H_{ij} \right] - \kappa^{*,\top} \mathbb{E}\alpha_0 \exp(Z_{ij}^\top \beta_0) H_{ij} \eta_{ij} \right\|_2
$$

$$
\leq \left\| (\hat{\kappa}^{*,\top} - \kappa^{*,\top}) \mathbb{E}\alpha_0 \exp(Z_{ij}^\top \beta_0) H_{ij} \eta_{ij} \right\|_2
$$

$$
+ \left\| \hat{\kappa}^{*,\top} \left[\frac{n}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \Lambda(R_{ij}^\top \theta_{0,n})(1 - \Lambda(R_{ij}^\top \theta_{0,n})) H_{ij} \eta_{ij} - \mathbb{E}\alpha_0 \exp(Z_{ij}^\top \beta_0) H_{ij} \eta_{ij} \right] \right\|_2
$$

$$
+ \left\| \hat{\kappa}^{*,\top} \frac{n^{1/2}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \int \left(\int_0^v \Lambda(R_{ij}^\top \theta_{0,n} + \tilde{v})(1 - \Lambda(R_{ij}^\top \theta_{0,n} + \tilde{v}))(1 - 2\Lambda(R_{ij}^\top \theta_{0,n} + \tilde{v}))\tilde{v} d\tilde{v} \right) d\pi_{ij} \right\|_2
$$

$$
= o_P(1),
$$

where the inequality is by Taylor's expansion with the integral remainder term and the equality is by the fact that

$$
\left| n \left(\int \Lambda(R_{ij}^\top \theta_{0,n} + \tilde{v})(1 - \Lambda(R_{ij}^\top \theta_{0,n} + \tilde{v}))(1 - 2\Lambda(R_{ij}^\top \theta_{0,n} + \tilde{v}))\tilde{v} d\tilde{v} \right) d\pi_{ij} \right|
$$

$$
\leq C \int \int_0^{|v|} \exp(\tilde{v})\tilde{v} d\tilde{v} \pi_{ij}(v) dv \leq C
$$

by Assumption 1. Combining the above two bounds, we can establish the last equality in (A.2).

Then, by Lemma A.2, we have

$$
\sqrt{n}(\hat{\Psi}(\hat{\kappa}^*) - \Psi_{0,n})
$$

$$
= (1, \kappa^*)^\top g_n + \mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0) H_{ij} - \Delta_{v,ij})
$$

$$
+ (1, \kappa^*)^\top \left(\sqrt{n} \left(\begin{array}{c} \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n \left(Y_{ij} - \int \Lambda(R_{ij}^\top \theta_{0,n} + n^{-1/2}v)\pi_{ij}(v) dv \right) H_{ij} \\ \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} [\gamma_n(D_{ij}, 0, \theta_{0,n}) - \mathbb{E}\gamma_n(D_{ij}, 0, \theta_{0,n})] \end{array} \right) - g_n \right) + o_P(1)
$$

$$
= (1, \kappa^*)^\top g_n + \mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0) H_{ij} - \Delta_{v,ij}) + (1 + ||\kappa^*||_2)O_P\left(\sqrt{k_n^5/n} \right) + o_P(1)
$$

$$
= (1, \kappa^*)^\top g_n + \mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0) H_{ij} - \Delta_{v,ij}) + o_P(1),
$$

where the last equality holds because $||\kappa^*||_2 = O(1)$ and $k_n^5/n = o(1)$.

---

[14]When k_n is fixed, we interpret $o_P(k_n^{-1/2})$ and $O_P(k_n^{1/2})$ as $o_P(1)$ and $O_P(1)$, respectively.

In addition, we note that

$$|\mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij})| \le (\mathbb{E}\eta_{ij}^2)^{1/2}(\mathbb{E}((\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij}))))^{1/2}$$
$$\le \overline{\sigma}(\mathbb{E}((\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij})))^{1/2} = \mathcal{B}(\kappa^*).$$

A.2 Step 2: Proof of the Desired Result.

We note that

$$\mathbb{P}(\Psi_{0,n} \in \widehat{CI}_{1-\alpha}(\hat{\kappa}^*))$$
$$= \mathbb{P}\left(\frac{|(1,\kappa^*)^\top g_n + \mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij}) + o_P(1)|}{(\hat{\Sigma}_\gamma + 2\hat{\Sigma}_{H,\gamma}^\top \hat{\kappa}^* + \hat{\kappa}^{*,\top}\hat{\Sigma}_H \hat{\kappa}^*)^{1/2}} \le CV_\alpha\left(\frac{\hat{\mathcal{B}}(\hat{\kappa}^*)}{(\hat{\Sigma}_\gamma + 2\hat{\Sigma}_{H,\gamma}^\top \hat{\kappa}^* + \hat{\kappa}^{*,\top}\hat{\Sigma}_H \hat{\kappa}^*)^{1/2}}\right)\right)$$
$$\ge \mathbb{P}\left(\begin{array}{c}\frac{|(1,\kappa^*)^\top g_n + \mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij})|}{(\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^*)^{1/2}} \\ \le CV_\alpha\left(\frac{\mathcal{B}(\kappa^*)}{(\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^*)^{1/2}}\right) - o_P(1)\end{array}\right)$$
$$= \mathbb{P}\left(\begin{array}{c}\frac{|(1,\kappa^*)^\top g_n + \mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij})|}{(\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^*)^{1/2}} \\ \le CV_\alpha\left(\frac{\mathcal{B}(\kappa^*)}{(\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^*)^{1/2}}\right)\end{array}\right) - \sup_{t\in\Re}\mathbb{P}\left(\frac{|(1,\kappa^*)^\top g_n + t|}{(\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^*)^{1/2}} \le o_P(1)\right)$$
$$= \mathbb{P}\left(\begin{array}{c}\frac{|(1,\kappa^*)^\top g_n + \mathbb{E}\eta_{ij}(\kappa^{*,\top}\alpha_0 \exp(Z_{ij}^\top \beta_0)H_{ij} - \Delta_{v,ij})|}{(\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^*)^{1/2}} \\ \le CV_\alpha\left(\frac{\mathcal{B}(\kappa^*)}{(\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^*)^{1/2}}\right)\end{array}\right) - \sup_{t\in\Re}\mathbb{P}(|\mathcal{N}(0,1) + t| \le o_P(1))$$
$$\ge 1 - \alpha - o(1),$$

where the first inequality holds because by Lemma A.1, we have $|\hat{\mathcal{B}}(\hat{\kappa}^*) - \mathcal{B}(\hat{\kappa}^*)| = o_P(1)$,

$$|(\hat{\Sigma}_\gamma + 2\hat{\Sigma}_{H,\gamma}^\top \hat{\kappa}^* + \hat{\kappa}^{*,\top}\hat{\Sigma}_H \hat{\kappa}^*)^{1/2} - (\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^*)^{1/2}| = o_P(1),$$

and

$$\left|CV_\alpha\left(\frac{\hat{\mathcal{B}}(\hat{\kappa}^*)}{(\hat{\Sigma}_\gamma + 2\hat{\Sigma}_{H,\gamma}^\top \hat{\kappa}^* + \hat{\kappa}^{*,\top}\hat{\Sigma}_H \hat{\kappa}^*)^{1/2}}\right) - CV_\alpha\left(\frac{\mathcal{B}(\kappa^*)}{(\Sigma_\gamma + 2\Sigma_{H,\gamma}^\top \kappa^* + \kappa^{*,\top}\Sigma_H \kappa^*)^{1/2}}\right)\right| = o_P(1).$$

The last inequality follows from the fact that the univariate standard normal density is bounded and the construction of the critical value function $CV_\alpha(\cdot)$. By taking $\liminf_{n\to\infty}$ on both sides of the above display, we obtain the desired result.

44

B   Proof of Theorem 2

Because κ_θ satisfies Assumption 6, we can rewrite κ_θ as

$$\kappa_\theta = G_F(G_F^\top G_F)^{-1}(I_{d_R} - G_F^\top \Xi) + \Xi,$$

where Ξ is some deterministic $d_F \times d_R$ matrix that does not depend on n. Then, we have

$$\kappa_F = \left[G_F(G_F^\top G_F)^{-1}(I_{d_R} - G_F^\top \Xi) + \Xi \right] \Gamma.$$

Further define

$$\tilde{\kappa}_F = \left[\Pi_F G_0(G_0^\top \Pi_F^\top \Pi_F G_0)^{-1}(I_{d_R} - G_0^\top \Pi_F^\top \Xi) + \Xi \right] \Gamma$$

and

$$\kappa_H = \Pi_F^\top \tilde{\kappa}_F.$$

We rely the following three claims to establish the desired result: (1) $\|\kappa_F - \tilde{\kappa}_F\|_2 = o(1)$, (2) $MSE_F(\kappa_F) = MSE_F(\tilde{\kappa}_F) + o(1)$, and (3) $MSE_F(\tilde{\kappa}_F) = MSE_H(\kappa_H) + o(1)$. Claims (2) and (3) imply that

$$MSE_F(\kappa_F) = MSE_H(\kappa_H) + o(1).$$

In addition, we note that $G_0^\top \kappa_H = \Gamma$, implying that κ_H satisfies (2.10). Therefore, by the definition of κ^*, we have

$$MSE_H(\kappa_H) \geq MSE_H(\kappa^*),$$

which is the desired result.

Next, we prove the three claims.

Proof of Claim (1). We have

$$G_F = \mathbb{E}\alpha_0 \exp(Z_{ij}^\top \beta_0)(\Pi_F H_{ij} + \delta_{ij})R_{ij}^\top = \Pi_F G_0 + G_\delta,$$

where $\|G_\delta\|_{op} = o(1)$. Then, by Assumption 6, we have

$$\left\| (G_F^\top G_F)^{-1} - (G_0^\top \Pi_F^\top \Pi_F G_0)^{-1} \right\|_{op} = o(1),$$

which implies

$$\|\kappa_F - \tilde{\kappa}_F\|_2 \leq C \left\| G_F(G_F^\top G_F)^{-1}(I_{d_R} - G_F^\top \Xi) - \Pi_F G_0(G_0^\top \Pi_F^\top \Pi_F G_0)^{-1}(I_{d_R} - G_0^\top \Pi_F^\top \Xi) \right\|_2$$
$$\leq C \left\| G_F(G_F^\top G_F)^{-1} - \Pi_F G_0(G_0^\top \Pi_F^\top \Pi_F G_0)^{-1} \right\|_2$$
$$+ C \left\| G_F(G_F^\top G_F)^{-1} G_F^\top - \Pi_F G_0(G_0^\top \Pi_F^\top \Pi_F G_0)^{-1} G_0^\top \Pi_F^\top \right\|_2 = o(1).$$

Proof of Claim (2). Given Claim (1), Claim (2) holds by the continuity of the quadratic function $MSE_F(\kappa_F)(\cdot)$.

Proof of Claim (3). We note that

$$\left| (\tilde{\kappa}_F^\top B_F \tilde{\kappa}_F)^{1/2} - (\kappa_H^\top B_H \kappa_H)^{1/2} \right|$$
$$\leq C \left| \left(\tilde{\kappa}_F^\top \mathbb{E} \exp(2Z_{ij}^\top \beta_0)(\Pi_F H_{ij} + \delta_{ij})(\Pi_F H_{ij} + \delta_{ij})^\top \tilde{\kappa}_F \right)^{1/2} - \left(\kappa_F^\top \mathbb{E} \exp(2Z_{ij}^\top \beta_0)(\Pi_F H_{ij})(\Pi_F H_{ij})^\top \kappa_F \right)^{1/2} \right|$$
$$\leq C \left[\mathbb{E} \exp(2Z_{ij}^\top \beta_0)(\tilde{\kappa}_F^\top \delta_{ij})^2 \right]^{1/2} \leq C \left(\mathbb{E} \|\delta_{ij}\|_2^2 \right)^{1/2} = o(1).$$

This implies

$$\left| \tilde{\kappa}_F^\top B_F \tilde{\kappa}_F - \kappa_H^\top B_H \kappa_H \right| = o(1),$$

given that $\tilde{\kappa}_F^\top B_F \tilde{\kappa}_F$ is fixed and

$$\kappa_H^\top B_H \kappa_H = ((\tilde{\kappa}_F^\top B_F \tilde{\kappa}_F)^{1/2} + o(1))^2 = O(1).$$

In the same manner, we can show that

$$B_{F,\gamma}^\top \tilde{\kappa}_F = B_{H,\gamma}^\top \kappa_H + o(1), \quad \tilde{\kappa}_F^\top \Sigma_F \tilde{\kappa}_F = \kappa_H^\top \Sigma_H \kappa_H + o(1), \quad \text{and} \quad \Sigma_{F,\gamma}^\top \kappa_F = \Sigma_{H,\gamma}^\top \kappa_H + o(1).$$

This concludes the proof.

C   Proof of Lemmas A.1 and A.2

We use the following notation throughout this section. For a scalar function $f(U_i, V_j)$ for dyadic data (U_i, V_j), we denote it as f_{ij} and write $\mathbb{P}_{N,M} f_{ij} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} f_{ij}$, $\mathbb{P}_{i,\cdot} f_{ij} = \mathbb{E}(f_{ij}|U_i)$, $\mathbb{P}_{\cdot,j} f_{ij} = \mathbb{E}(f_{ij}|V_j)$, $\mathbb{P} f_{ij} = \mathbb{E} f_{ij}$, $\mathbb{P}_N f_{ij} = \frac{1}{N} \sum_{i=1}^{N} f_{ij}$, $\mathbb{P}_M f_{ij} = \frac{1}{M} \sum_{j=1}^{M} f_{ij}$, and $\mathbb{U}_{N,M} f_{ij} = \mathbb{P}_{N,M} f_{ij} - \mathbb{P}_N \mathbb{P}_{i,\cdot} f_{ij} - \mathbb{P}_M \mathbb{P}_{\cdot,j} f_{ij} + \mathbb{P} f_{ij}$. Based on these definitions, we have $\mathbb{P}_{N,M} f_{ij} = \mathbb{P}_N \mathbb{P}_M f_{ij} = \mathbb{P}_M \mathbb{P}_N f_{ij}$.

C.1   Proof of Lemma A.1

Proof of $\sup_{\theta \in \mathbb{B}(\theta_{0,n}, C/\sqrt{n})} \left\| \hat{G}(\theta) - G_0 \right\|_{op} = O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}} \right)$.

Note that

$$\sup_{\theta \in \mathbb{B}(\theta_{0,n}, C/\sqrt{n})} \left\| \hat{G}(\theta) - G_0 \right\|_{op}$$

$$\leq \sup_{\theta \in \mathbb{B}(\theta_{0,n}, C/\sqrt{n})} \left\| \hat{G}(\theta) - G(\theta) \right\|_{op} + \| G(\theta_{0,n}) - G_0 \|$$

$$= \sup_{v_1 \in \Re^{kn}, ||v_1||_2=1, v_2 \in \Re^{d_R}, ||v_2||_2=1, (\alpha,\beta) \in \mathbb{B}((\alpha_0,\beta_0),C/\sqrt{n})} |(\mathbb{P}_{N,M} - \mathbb{P}) \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2| + O_P(n^{-1/2})$$

$$\leq \sup_{v_1 \in \Re^{kn}, ||v_1||_2=1, v_2 \in \Re^{d_R}, ||v_2||_2=1, (\alpha,\beta) \in \mathbb{B}((\alpha_0,\beta_0),C/\sqrt{n})} |\mathbb{P}_N (\mathbb{P}_M - \mathbb{P}_{i,\cdot}) \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2|$$

$$+ \sup_{v_1 \in \Re^{kn}, ||v_1||_2=1, v_2 \in \Re^{d_R}, ||v_2||_2=1, (\alpha,\beta) \in \mathbb{B}((\alpha_0,\beta_0),C/\sqrt{n})} |(\mathbb{P}_N - \mathbb{P}) \left[\mathbb{P}_{i,\cdot} \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2 \right] | + O_P(n^{-1/2})$$

$$\leq \sup_{i=1,\cdots,N, v_1 \in \Re^{kn}, ||v_1||_2=1, v_2 \in \Re^{d_R}, ||v_2||_2=1, (\alpha,\beta) \in \mathbb{B}((\alpha_0,\beta_0),C/\sqrt{n})} |(\mathbb{P}_M - \mathbb{P}_{i,\cdot}) \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2|$$

$$+ \sup_{v_1 \in \Re^{kn}, ||v_1||_2=1, v_2 \in \Re^{d_R}, ||v_2||_2=1, (\alpha,\beta) \in \mathbb{B}((\alpha_0,\beta_0),C/\sqrt{n})} |(\mathbb{P}_N - \mathbb{P}) \left[\mathbb{P}_{i,\cdot} \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2 \right] | + O_P(n^{-1/2}).$$

We note that $\exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2$ is independent across $j = 1, \cdots, M$ and

$$\sup_{i=1,\cdots,N, v_1 \in \Re^{kn}, ||v_1||_2=1, v_2 \in \Re^{d_R}, ||v_2||_2=1, (\alpha,\beta) \in \mathbb{B}((\alpha_0,\beta_0),C/\sqrt{n})} \left| \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2 \right|$$

$$\leq \max_{i \in [N]} \exp(\alpha_0 + Z_{ij}^\top \beta_0 + C||Z_{ij}||_2/\sqrt{n}) ||H_{ij}||_2 ||R_{ij}||_2 \leq C k_n^{1/2} \zeta_n \equiv F.$$

Then, the class of functions

$$\mathcal{F} = \cup_{i=1}^{N} \mathcal{F}_i$$

$$\mathcal{F}_i = \left\{ \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2 : v_1 \in \Re^{kn}, ||v_1||_2 = 1, v_2 \in \Re^{d_R}, ||v_2||_2 = 1, (\alpha, \beta) \in \mathbb{B}((\alpha_0, \beta_0), C/\sqrt{n}) \right\}$$

is of the VC type in the sense that

$$\sup_Q N(\mathcal{F}, ||\cdot||_{Q,2}, \epsilon||F||_{Q,2}) \leq \sum_{i=1}^{N} \sup_Q N(\mathcal{F}_j, ||\cdot||_{Q,2}, \epsilon||F||_{Q,2}) \leq N\left(\frac{a}{\epsilon}\right)^{ck_n} \leq \left(\frac{a}{\epsilon}\right)^{c(k_n \vee \log n)}, \; 0 < \forall \epsilon \leq 1,$$

where $\sup_Q$ means the supremum over all finitely discrete distributions. Therefore, by Chernozhukov, Chetverikov, and Kato (2014, Corollary 5.1) we have

$$\mathbb{E} \sup_{i=1,\cdots,N,v_1 \in \Re^{k_n}, ||v_1||_2=1, v_2 \in \Re^{d_R}, ||v_2||_2=1, (\alpha,\beta) \in \mathbb{B}((\alpha_0,\beta_0),C/\sqrt{n})} |(\mathbb{P}_M - \mathbb{P}_{i,\cdot}) \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2|$$

$$\leq C\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}} + \frac{k_n^{3/2} \zeta_n \log^2(n)}{n}\right) \leq C\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}},$$

which implies

$$\sup_{i=1,\cdots,N,v_1 \in \Re^{k_n}, ||v_1||_2=1, v_2 \in \Re^{d_R}, ||v_2||_2=1, (\alpha,\beta) \in \mathbb{B}((\alpha_0,\beta_0),C/\sqrt{n})} \left|(\mathbb{P}_M - \mathbb{P}_{i,\cdot}) \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2\right|$$

$$= O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right).$$

Similarly, we can show that

$$\sup_{v_1 \in \Re^{k_n}, ||v_1||_2=1, v_2 \in \Re^{d_R}, ||v_2||_2=1, (\alpha,\beta) \in \mathbb{B}((\alpha_0,\beta_0),C/\sqrt{n})} \left|(\mathbb{P}_N - \mathbb{P})\left[\mathbb{P}_{i,\cdot} \exp(\alpha + Z_{ij}^\top \beta) v_1^\top H_{ij} R_{ij}^\top v_2\right]\right|$$

$$= O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right),$$

which implies the desired result.

We can establish $\left\|\hat{B}_H - B_H\right\|_{op} = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right)$ and $\left\|\hat{B}_{H,\gamma} - B_{H,\gamma}\right\|_2 = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right)$ in the same manner.

Proof of $\sup_{\theta \in \mathbb{B}(\theta_{0,n},C/\sqrt{n})} \left\|\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \partial_\theta \gamma_n(D_{ij},0,\theta) - \Gamma\right\|_2 = O_P(n^{-1/2})$. We have

$$\left\|\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \partial_\theta \gamma_n(D_{ij},0,\theta) - \Gamma\right\|_2$$

$$\leq \left\|\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \partial_{\theta,\theta^\top} \gamma_n(D_{ij},0,\tilde{\theta})(\theta - \theta_{0,n})\right\|_2 + \left\|\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \partial_\theta \gamma_n(D_{ij},0,\theta_{0,n}) - \Gamma\right\|_2$$

$$\leq \left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \partial_{\theta,\theta^\top} \gamma_n(D_{ij}, 0, \tilde{\theta}) \right\|_{op} \|(\theta - \theta_{0,n})\|_2 + O_P(n^{-1/2}),$$

where $\tilde{\theta}$ is between θ and $\theta_{0,n}$ and thus belongs to $\mathbb{B}(\theta_{0,n}, C/\sqrt{n})$. Then, by Assumption 2.2 and taking $\sup_{\theta \in \mathbb{B}(\theta_{0,n}, C/\sqrt{n})}$ on both sides, we obtain the desired result.

Proof of $\|\hat{\Sigma}_H - \Sigma_H\|_{op} = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}}\right)$ **and** $\left\|\hat{\Sigma}_{H,\gamma} - \Sigma_{H,\gamma}\right\|_2 = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}}\right)$. We note that

$$\hat{\Sigma}_H = \hat{\Sigma}^a_{YY,n} + \hat{\Sigma}^p_{YY,n} + \frac{1}{n}\left[\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \hat{s}^2_{Y,ij} H_{ij} H_{ij}^\top - \hat{\Sigma}^a_{YY,n} - \hat{\Sigma}^p_{YY,n}\right].$$

By Lemma D.1, we have

$$||\hat{\Sigma}^a_{YY,n} - \Sigma^a_{YY,n}||_{op} = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}}\right) \tag{C.1}$$

and

$$||\hat{\Sigma}^p_{YY,n} - \Sigma^p_{YY,n}||_{op} = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}}\right). \tag{C.2}$$

Next, we focus on $\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \hat{s}^2_{Y,ij} H_{ij} H_{ij}^\top$. Recall $s_{Y,ij} = n\left(Y_{ij} - \int \Lambda(R_{ij}^\top \theta_{0,n} + n^{-1/2}v)\pi_{ij}dv\right) H_{ij}$ and $p_{ij} = \mathbb{E}(Y_{ij}|X_i, A_i, W_j, B_j)$, and define

$$\bar{s}_{Y,ij} = \mathbb{E}(s_{Y,ij}|X_i, A_i, W_j, B_j) = n(p_{ij} - \mathbb{E}(p_{ij}|X_i, W_j)).$$

Then, we have

$$\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \hat{s}^2_{Y,ij} H_{ij} H_{ij}^\top = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} s^2_{Y,ij} H_{ij} H_{ij}^\top + \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} (\hat{s}_{Y,ij} + s_{Y,ij})(\hat{s}_{Y,ij} - s_{Y,ij}) H_{ij} H_{ij}^\top. \tag{C.3}$$

For the first term on the RHS of (C.3), we have

$$\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} s^2_{Y,ij} H_{ij} H_{ij}^\top = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \bar{s}^2_{Y,ij} H_{ij} H_{ij}^\top + \frac{2}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \bar{s}_{Y,ij}(s_{Y,ij} - \bar{s}_{Y,ij}) H_{ij} H_{ij}^\top$$

$$+ \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} (s_{Y,ij} - \bar{s}_{Y,ij})^2 H_{ij} H_{ij}^\top$$

$$= \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \bar{s}_{Y,ij}^2 H_{ij} H_{ij}^\top + \frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n\bar{s}_{Y,ij}(Y_{ij} - p_{ij}) H_{ij} H_{ij}^\top$$

$$+ \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n^2 (Y_{ij} - p_{ij}) H_{ij} H_{ij}^\top - \frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n^2 (Y_{ij} - p_{ij}) p_{ij} H_{ij} H_{ij}^\top$$

$$+ \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n^2 p_{ij}(1 - p_{ij}) H_{ij} H_{ij}^\top$$

$$\equiv I + 2II + III - 2IV + V. \tag{C.4}$$

For term I in the RHS of (C.4), we have

$$\bar{s}_{Y,ij}^2 = n^2 (p_{ij} - \mathbb{E}(p_{ij}|X_i, W_j))^2 \leq C,$$

for some constant $C < \infty$, and thus,

$$\|I\|_{op} \leq C \left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} H_{ij} H_{ij}^\top \right\|_{op} = O_P(1). \tag{C.5}$$

For term II in the RHS of (C.4), we have

$$\mathbb{E} \left(\|II\|_F^2 \mid \mathcal{X}_N, \mathcal{A}_N, \mathcal{W}_M, \mathcal{B}_M \right)$$

$$= \sum_{(\ell,\ell') \in [k_n] \times [k_n]} \mathbb{E} \left\{ \left[\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n\bar{s}_{Y,ij}(Y_{ij} - p_{ij}) H_{ij,\ell} H_{ij,\ell'} \right]^2 \mid \mathcal{X}_N, \mathcal{A}_N, \mathcal{W}_M, \mathcal{B}_M \right\}$$

$$= \sum_{(\ell,\ell') \in [k_n] \times [k_n]} \frac{1}{N^2 M^2} \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbb{E} \left\{ \left[n\bar{s}_{Y,ij}(Y_{ij} - p_{ij}) H_{ij,\ell} H_{ij,\ell'} \right]^2 \mid \mathcal{X}_N, \mathcal{A}_N, \mathcal{W}_M, \mathcal{B}_M \right\}$$

$$\leq C \frac{n}{NM} \mathbb{P}_{N,M} \|H_{ij}\|_2^4 = O_P \left(\frac{k_n^2 \zeta_n^2}{n} \right),$$

which implies

$$\|II\|_{op} \leq \|II\|_F = O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}} \right). \tag{C.6}$$

In the same manner, we can show that

$$\left\| n^{-1} III \right\|_{op} = O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}} \right) \tag{C.7}$$

and

$$\|IV\|_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right). \tag{C.8}$$

Last, for term V in the RHS of (C.4), by Hoeffding decomposition, we have

$$n^{-1}V - \mathbb{E}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)$$
$$= (\mathbb{P}_{N,M} - \mathbb{P})\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)$$
$$= (\mathbb{P}_N - \mathbb{P})\left[\mathbb{P}_{i,}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)\right] + (\mathbb{P}_M - \mathbb{P})\left[\mathbb{P}_{,j}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)\right]$$
$$+ \mathbb{U}_{N,M}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right). \tag{C.9}$$

We have

$$\left\|(\mathbb{P}_N - \mathbb{P})\left[\mathbb{P}_{i,}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)\right]\right\|_{op}$$
$$\leq \left\|(\mathbb{P}_N - \mathbb{P})\left[\mathbb{P}_{i,}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)\right]\right\|_F = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right)$$

and

$$\left\|(\mathbb{P}_M - \mathbb{P})\left[\mathbb{P}_{,j}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)\right]\right\|_{op}$$
$$\leq \left\|(\mathbb{P}_M - \mathbb{P})\left[\mathbb{P}_{,j}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)\right]\right\|_F = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right).$$

To analyze the last term in the RHS of (C.9), we note that

$$\mathbb{E}\left\|\mathbb{U}_{N,M}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)\right\|_F^2$$
$$= \sum_{(\ell,\ell')\in[k_n]\times[k_n]} \mathbb{E}\left[\mathbb{U}_{N,M}\left(np_{ij}(1-p_{ij})H_{ij,\ell}H_{ij,\ell'}\right)\right]^2$$
$$\leq \frac{C}{(NM)^2} \sum_{(\ell,\ell')\in[k_n]\times[k_n]} \sum_{i=1}^N \sum_{j=1}^M \mathbb{E}\left(np_{ij}(1-p_{ij})H_{ij,\ell}H_{ij,\ell'}\right)^2 \leq \frac{Ck_n^2\zeta_n^2}{NM},$$

which implies

$$\left\|\mathbb{U}_{N,M}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)\right\|_{op} \leq \left\|\mathbb{U}_{N,M}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top\right)\right\|_F = O_P\left(\frac{k_n\zeta_n}{n}\right).$$

Therefore, following (C.9), we have

$$\left\| n^{-1}V - \mathbb{E}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top \right) \right\|_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}} \right). \tag{C.10}$$

Combining (C.4), (C.5), (C.6), (C.7), (C.8), and (C.10), we have

$$\left\| \frac{1}{nNM} \sum_{i=1}^{N} \sum_{j=1}^{M} s_{Y,ij}^2 H_{ij}H_{ij}^\top - \mathbb{E}\left(np_{ij}(1-p_{ij})H_{ij}H_{ij}^\top \right) \right\|_{op}$$

$$= \left\| \frac{1}{nNM} \sum_{i=1}^{N} \sum_{j=1}^{M} s_{Y,ij}^2 H_{ij}H_{ij}^\top - n^{-1}\mathbb{E}\left[Var(s_{Y,ij} \mid X_i, A_i, W_j, B_j)H_{ij}H_{ij}^\top \right] \right\|_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}} \right). \tag{C.11}$$

For the second term on the RHS of (C.3), we have

$$\left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (\hat{s}_{Y,ij} + s_{Y,ij})(\hat{s}_{Y,ij} - s_{Y,ij})H_{ij}H_{ij}^\top \right\|_{op}$$

$$\leq \left\| \frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n^2 Y_{ij} \exp(R_{ij}^\top\tilde{\theta})\|R_{ij}\|_2 H_{ij}H_{ij}^\top \right\|_{op} \|\hat{\theta}_{\text{initial}} - \theta_{0,n}\|_2$$

$$\leq \left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} n Y_{ij} H_{ij}H_{ij}^\top \right\|_{op} \times O_P(n^{-1/2})$$

$$\leq \left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \left[n(Y_{ij} - p_{ij}) + np_{ij} \right] H_{ij}H_{ij}^\top \right\|_{op} \times O_P(n^{-1/2})$$

$$\leq \left[O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}} \right) + \left\| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} H_{ij}H_{ij}^\top \right\|_{op} \right] \times O_P(n^{1/2}) = O_P(n^{1/2}). \tag{C.12}$$

where $\tilde{\theta} = (\tilde{\alpha}_n, \tilde{\beta}_n)$ is between $\hat{\theta}_{\text{initial}}$ and 0 so that $\log n + \tilde{\alpha}_n = O_P(1)$, $\tilde{\beta}_n = O_P(1)$, and

$$n \exp(R_{ij}^\top\tilde{\theta}) \leq \exp(|\log n + \tilde{\alpha}_n| + \|Z_{ij}\|_2\|\tilde{\beta}\|_2) \leq \exp(|\log n + \tilde{\alpha}_n| + C\|\tilde{\beta}_n\|_2),$$

and the last inequality is due to (C.7). Further note that

$$\left\| \hat{\Sigma}_{YY,n}^a \right\|_{op} \leq \left\| \Sigma_{YY,n}^a \right\|_{op} + \left\| \hat{\Sigma}_{YY,n}^a - \Sigma_{YY,n}^a \right\|_{op} = O_P(1)$$

and, similarly, $\left\| \hat{\Sigma}_{YY,n}^p \right\|_{op} = O_P(1)$.

Combining (C.3), (C.11), (C.12), and above fact, we obtain the desired result that

$$\left\| \frac{1}{n} \left[\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \hat{s}_{Y,ij}^2 H_{ij} H_{ij}^\top - \hat{\Sigma}_{YY,n}^a - \hat{\Sigma}_{YY,n}^p \right] - n^{-1} \mathbb{E} \left[Var(s_{Y,ij} \mid X_i, A_i, W_j, B_j) H_{ij} H_{ij}^\top \right] \right\|_{op}$$
$$= O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}} \right).$$

We can show $\left\| \hat{\Sigma}_{H,\gamma} - \Sigma_{H,\gamma} \right\|_2 = O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}} \right)$ in the same manner.

Proof of $|\hat{\Sigma}_\gamma - \Sigma_\gamma| = O_P(\zeta_n n^{-1/2})$. We note that

$$\hat{\Sigma}_\gamma = \frac{\hat{\Sigma}_{\gamma\gamma,n}^a}{1-\phi} + \frac{\hat{\Sigma}_{\gamma\gamma,n}^p}{\phi}, \tag{C.13}$$

where $\hat{\Sigma}_{\gamma\gamma,n}^a$ and $\hat{\Sigma}_{\gamma\gamma,n}^p$ are the bottom-right elements of $\hat{\Sigma}_n^a$ and $\hat{\Sigma}_n^p$, respectively. Specifically, we have

$$\hat{\Sigma}_{\gamma\gamma,n}^a = \frac{1}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} \hat{s}_{\gamma,ij} \hat{s}_{\gamma,ik}$$

and

$$\hat{\Sigma}_{\gamma\gamma,n}^p = \frac{1}{M} \sum_{j=1}^{M} \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{k=i+1}^{N} \hat{s}_{\gamma,ij} \hat{s}_{\gamma,ik}.$$

Then, similar to the argument in the proof of Lemma D.1 (with the dimension $k_n = 1$), we can show that

$$\left\| \hat{\Sigma}_{\gamma\gamma,n}^a - \Sigma_{\gamma\gamma,n}^a \right\|_{op} = O_P \left(\sqrt{\frac{\zeta_n^2}{n}} \right) \quad \text{and} \quad \left\| \hat{\Sigma}_{\gamma\gamma,n}^p - \Sigma_{\gamma\gamma,n}^p \right\|_{op} = O_P \left(\sqrt{\frac{\zeta_n^2}{n}} \right),$$

where

$$\Sigma_{\gamma\gamma,n}^a = \mathbb{E}[\mathbb{E}(s_{\gamma,ij}|X_i, A_i)]^2 \quad \text{and} \quad \Sigma_{\gamma\gamma,n}^p = \mathbb{E}[\mathbb{E}(s_{\gamma,ij}|W_j, B_j)]^2.$$

This is the desired result.

Proof of $\|\hat{\kappa}^* - \kappa^*\|_2 = o_P(k_n^{-1/2})$. Based on the previous results, we have

$$\left\| \hat{\Phi}_1 - \Phi_1 \right\|_{op} \leq \bar{\sigma}^2 \left\| \hat{B}_H - B_H \right\|_{op} + \left\| \hat{\Sigma}_H - \Sigma_H \right\|_{op} = O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}} \right),$$

$$\left\|\hat{\Phi}_2 - \Phi_2\right\|_2 \leq \bar{\sigma}^2 \left\|\hat{B}_{H,\gamma} - B_{H,\gamma}\right\|_2 + \left\|\hat{\Sigma}_{H,\gamma} - \Sigma_{H,\gamma}\right\|_2 = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right),$$

$$\left\|\hat{G} - G_0\right\|_{op} = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right),$$

and $\|\hat{\Gamma} - \Gamma\|_2 = O_P(n^{-1/2})$. By Assumption 3.2, we have

$$C \geq \lambda_{\max}(\Phi_1) \geq \lambda_{\min}(\Phi_1) \geq c \quad \text{and} \quad C \geq \lambda_{\max}(G_0^\top \Phi_1^{-1} G_0) \geq \lambda_{\min}(G_0^\top \Phi_1^{-1} G_0) \geq c$$

for some constants (C, c) such that $\infty > C \geq c > 0$. In addition, we have $\|G_0\|_{op} = O(1)$, $\|\Phi_2\|_2 = O(1)$, and $\|\Gamma\|_2 = O(1)$.

Now we are ready to bound $\|\hat{\kappa}^* - \kappa^*\|_2$. Note that

$\|\hat{\kappa}^* - \kappa^*\|_2$

$$= \left\| \hat{\Phi}_1^{-1} \left\{ \hat{\Phi}_2 - \hat{G} \left[\hat{G}^\top \hat{\Phi}_1^{-1} \hat{G}\right]^{-1} \left[\hat{G}^\top \hat{\Phi}_1^{-1} \hat{\Phi}_2 + \hat{\Gamma}\right] \right\} - \Phi_1^{-1} \left\{ \Phi_2 - G_0 \left[G_0^\top \Phi_1^{-1} G_0\right]^{-1} \left[G_0^\top \Phi_1^{-1} \Phi_2 + \Gamma\right] \right\} \right\|_2$$

$$\leq \left\| \hat{\Phi}_1^{-1} \hat{\Phi}_2 - \Phi_1^{-1} \Phi_2 \right\|_2 + \left\| \hat{\Phi}_1^{-1} \hat{G} \left[\hat{G}^\top \hat{\Phi}_1^{-1} \hat{G}\right]^{-1} \hat{G}^\top \hat{\Phi}_1^{-1} \hat{\Phi}_2 - \Phi_1^{-1} G_0 \left[G_0^\top \Phi_1^{-1} G_0\right]^{-1} G_0^\top \Phi_1^{-1} \Phi_2 \right\|_2$$

$$+ \left\| \hat{\Phi}_1^{-1} \hat{G} \left[\hat{G}^\top \hat{\Phi}_1^{-1} \hat{G}\right]^{-1} \hat{\Gamma} - \Phi_1^{-1} G_0 \left[G_0^\top \Phi_1^{-1} G_0\right]^{-1} \Gamma \right\|_2$$

$$\equiv I + II + III. \tag{C.14}$$

For term I on the RHS of (C.14), we have

$$\left\| \hat{\Phi}_1^{-1} \hat{\Phi}_2 - \Phi_1^{-1} \Phi_2 \right\|_2 \leq \left\| \hat{\Phi}_1^{-1} (\hat{\Phi}_2 - \Phi_2) \right\|_2 + \left\| (\hat{\Phi}_1^{-1} - \Phi_1^{-1}) \Phi_2 \right\|_2$$

$$\leq \left\| \hat{\Phi}_1^{-1} \right\|_{op} \left\| \hat{\Phi}_2 - \Phi_2 \right\|_2 + \left\| \hat{\Phi}_1^{-1} - \Phi_1^{-1} \right\|_{op} \|\Phi_2\|_2 = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right).$$

For term II on the RHS of (C.14), we have

$$\left\| \hat{\Phi}_1^{-1} \hat{G} \left[\hat{G}^\top \hat{\Phi}_1^{-1} \hat{G}\right]^{-1} \hat{G}^\top \hat{\Phi}_1^{-1} \hat{\Phi}_2 - \Phi_1^{-1} G_0 \left[G_0^\top \Phi_1^{-1} G_0\right]^{-1} G_0^\top \Phi_1^{-1} \Phi_2 \right\|_2$$

$$\leq \left\| \hat{\Phi}_1^{-1} \hat{G} - \Phi_1^{-1} G_0 \right\|_{op} \left\| \left[\hat{G}^\top \hat{\Phi}_1^{-1} \hat{G}\right]^{-1} \hat{G}^\top \hat{\Phi}_1^{-1} \hat{\Phi}_2 \right\|_2 + \left\| \Phi_1^{-1} G_0 \right\|_{op} \left\| \left[\hat{G}^\top \hat{\Phi}_1^{-1} \hat{G}\right]^{-1} - \left[G_0^\top \Phi_1^{-1} G_0\right]^{-1} \right\| \|\hat{\Phi}_2\|_2$$

$$+ \left\| \Phi_1^{-1} G_0 \left[G_0^\top \Phi_1^{-1} G_0\right]^{-1} G_0^\top \Phi_1^{-1} \right\|_{op} \left\| \hat{\Phi}_2 - \Phi_2 \right\|_2$$

$$= O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right).$$

Similarly, we have $||III||_2 = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right)$, which implies

$$\|\hat{\kappa}^* - \kappa^*\|_2 = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2 \log^2(n)}{n}}\right) = o_P(k_n^{-1/2}),$$

where the last equality holds by Assumption 4.

Proof of $\|\kappa^*\|_2 = O(1)$. Note that

$$
\begin{aligned}
\|\kappa^*\|_2 &= \left\|\Phi_1^{-1}\left\{\Phi_2 - G_0\left[G_0^\top \Phi_1^{-1} G_0\right]^{-1}\left[G_0^\top \Phi_1^{-1}\Phi_2 + \Gamma\right]\right\}\right\|_2 \\
&\leq \left\|\Phi_1^{-1}\Phi_2\right\|_2 + \left\|\Phi_1^{-1} G_0\left[G_0^\top \Phi_1^{-1} G_0\right]^{-1} G_0^\top \Phi_1^{-1}\Phi_2\right\|_2 \\
&\quad + \left\|\Phi_1^{-1} G_0\left[G_0^\top \Phi_1^{-1} G_0\right]^{-1}\Gamma\right\|_2 \\
&\leq \left\|\Phi_1^{-1}\right\|_{op}\|\Phi_2\|_2 + \left\|\Phi_1^{-1/2}\right\|_{op}^2\|\Phi_2\|_2 + \left\|\Phi_1^{-1}\right\|_{op}\|G_0\|_{op}\left\|\left[G_0^\top \Phi_1^{-1} G_0\right]^{-1}\right\|_{op}\|\Gamma\|_2 \\
&= O(1).
\end{aligned}
$$

This concludes the proof.

C.2 Proof of Lemma A.2

Recall

$$s_{ij} = \begin{pmatrix} n\left(Y_{ij} - \int \Lambda(R_{ij}^\top\theta_{0,n} + n^{-1/2}v)\pi_{ij}(v)dv\right)H_{ij} \\ \gamma_n(D_{ij}, 0, \theta_{0,n}) - \mathbb{E}\gamma_n(D_{ij}, 0, \theta_{0,n}) \end{pmatrix} \equiv (s_{Y,ij}H_{ij}^\top, s_{\gamma,ij})^\top,$$

$\bar{s}_{ij} = \mathbb{E}(s_{ij}|X_i, A_i, W_j, B_j)$, $\bar{s}_{1i}^a = \mathbb{E}(\bar{s}_{ij}|X_i, A_i)$, $\bar{s}_{1j}^p = \mathbb{E}(\bar{s}_{ij}|W_j, B_j)$, $\Sigma_n^a = \mathbb{E}\bar{s}_{1i}^a(\bar{s}_{1i}^a)^\top$, $\Sigma_n^p = \mathbb{E}\bar{s}_{1j}^p(\bar{s}_{1j}^p)^\top$, and $\Sigma_{2n} = \mathbb{E}(s_{ij} - \bar{s}_{ij})(s_{ij} - \bar{s}_{ij})^\top$.

Following Graham (2022), we have

$$
\begin{aligned}
\frac{\sqrt{n}}{NM}\sum_{i=1}^N\sum_{j=1}^M s_{ij} &= \frac{\sqrt{n}}{N}\sum_{i=1}^N \bar{s}_{1i}^a + \frac{\sqrt{n}}{M}\sum_{j=1}^M \bar{s}_{1j}^p + \frac{\sqrt{n}}{NM}\sum_{i=1}^N\sum_{j=1}^M(s_{ij} - \bar{s}_{ij}) + \frac{\sqrt{n}}{NM}\sum_{i=1}^N\sum_{j=1}^M(\bar{s}_{ij} - \bar{s}_{1i}^a - \bar{s}_{1j}^p) \\
&\equiv \sum_{l=1}^4 U_{l,n},
\end{aligned}
$$

where

$$U_{1,n} = \sum_{i \in [N]} \frac{\sqrt{n}}{N} \bar{s}_{1i}^a,$$

$$U_{2,n} = \sum_{j \in [M]} \frac{\sqrt{n}}{M} \bar{s}_{1j}^p,$$

$$U_{3,n} = \frac{\sqrt{n}}{NM} \sum_{i=1}^N \sum_{j=1}^M (s_{ij} - \bar{s}_{ij}),$$

$$U_{4,n} = \frac{\sqrt{n}}{NM} \sum_{i=1}^N \sum_{j=1}^M (\bar{s}_{ij} - \bar{s}_{1i}^a - \bar{s}_{1j}^p).$$

Let $\mathcal{F}_n$ be the sigma field generated by $\{X_i, A_i\}_{i \in [N]}$ and $\{W_j, B_j\}_{j \in [M]}$. Then, we note that $(U_{1,n}, U_{2,n})$ belong to $\mathcal{F}_n$, and conditionally on $\mathcal{F}_n$, $U_{3,n} = (U_{3,n}^{*,\top}, 0)^\top$,

$$U_{3,n}^* = \frac{\sqrt{n}}{NM} \sum_{i=1}^N \sum_{j=1}^M n\left(Y_{ij} - \mathbb{E}(Y_{i,j}|\mathcal{F}_n)\right) H_{ij} = \frac{\sqrt{n}}{NM} \sum_{i=1}^N \sum_{j=1}^M n\left(Y_{ij} - p_{ij}\right) H_{ij}$$

is a sequence of independent random variables. Therefore, by Yurinskii's Coupling (Pollard (2002, Theorem 10)), there exists $g_{3,n} = (g_{3,n}^{*,\top}, 0)^\top$ such that, for $g_{3,n}^* \stackrel{d}{=} \mathcal{N}(0, I_{k_n})$, $g_{3,n}^*$ is independent of $\mathcal{F}_n$ and for any $\mathcal{F}_n$-measurable random variable Υ_n with $\Upsilon_n > 0$, we have

$$\mathbb{P}\left\{ ||U_{3,n}^* - \hat{\Omega}_{3,n}^{1/2} g_{3,n}^*||_2 > 3\Upsilon_n \mid \mathcal{F}_n \right\} \leq C_0 \hat{B}_{3,n} \left(1 + \frac{|\log(1/\hat{B}_{3,n})|}{k_n} \right),$$

where C_0 is a universal constant,

$$\hat{\Omega}_{3,n} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M \frac{n^3 p_{ij}(1 - p_{ij}) H_{ij} H_{ij}^\top}{NM},$$

$$\hat{B}_{3,n} = \hat{D}_{3,n} k_n \Upsilon_n^{-3}, \quad \text{and} \quad \hat{D}_{3,n} = \frac{n^{3/2}}{(NM)^3} \sum_{i \in [N]} \sum_{j \in [M]} \mathbb{E}\left(||s_{ij} - \bar{s}_{ij}||_2^3 \mid \mathcal{F}_n \right).$$

Consequently, by choosing $\Upsilon_n = \hat{D}_{3,n}^{1/3} \delta_n^{-1/3}$ for some deterministic sequence $\delta_n > 0$ which will be specified later, we have

$$\hat{B}_{3,n} = k_n \delta_n$$

and

$$\mathbb{P}\left\{ ||U_{3,n}^* - \hat{\Omega}_{3,n}^{1/2} g_{3,n}^*||_2 > 3 D_{3,n}^{1/3} \delta_n^{-1/3} \mid \mathcal{F}_n \right\} \leq C_0 k_n \delta_n \left(1 + \frac{|\log(1/(k_n \delta_n))|}{k_n} \right).$$

We further note that $\hat{D}_{3,n} = O_P\left(\sqrt{\frac{k_n^3}{n}}\right)$, and $\hat{\Omega}_{3,n}$ and $\hat{B}_{3,n}$ belong to $\mathcal{F}_n$. In addition, let $\Omega_{3,n} = \mathbb{E}\hat{\Omega}_{3,n}$. Then, we have $\left\|\Omega_{3,n} - \hat{\Omega}_{3,n}\right\|_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right)$ by (C.10), and thus,

$$\left\|\left(\hat{\Omega}_{3,n}^{1/2} - \Omega_{3,n}^{1/2}\right)g_{3,n}^*\right\|_2 \leq \left\|\hat{\Omega}_{3,n}^{1/2} - \Omega_{3,n}^{1/2}\right\|_{op}\|g_{3,n}^*\|_2 = O_P\left(\sqrt{\frac{k_n^3\zeta_n^2}{n}}\right),$$

where we use the fact that $\lambda_{\min}\left(\Omega_{3,n}\right) \geq c > 0$ for some constant c and $\|g_{3,n}^*\| = O_P(\sqrt{k_n})$.

Similarly, we can strongly approximate $U_{1,n}$ and $U_{2,n}$ by $\Omega_{1,n}^{1/2}g_{1,n}$ and $\Omega_{2,n}^{1/2}g_{2,n}$, respectively, where $g_{1,n}$ and $g_{2,n}$ are (k_n+1)-dimensional standard normal random variables,

$$\Omega_{1,n} = \frac{n}{N^2}\sum_{i=1}^N \mathbb{E}(\bar{s}_{1i}^a)^2, \quad \Omega_{2,n} = \frac{n}{M^2}\sum_{j=1}^M \mathbb{E}(\bar{s}_{1j}^p)^2.$$

More specifically, for $l = 1, 2$, we have

$$\mathbb{P}\left(\left\|U_{l,n} - \Omega_{l,n}^{1/2}g_{l,n}\right\|_2 > 3D_{l,n}^{1/3}\delta_n\right) \leq C_0 k_n\delta_n\left(1 + \frac{|\log(1/(k_n\delta_n))|}{k_n}\right),$$

where

$$D_{1,n} = \frac{n^{3/2}}{N^3}\sum_{i\in[N]}\mathbb{E}\left(\|\bar{s}_{1i}^a\|_2^3\right) = O\left(\sqrt{\frac{k_n^3}{n}}\right) \quad \text{and} \quad D_{2,n} = \frac{n^{3/2}}{M^3}\sum_{j\in[M]}\mathbb{E}\left(\|\bar{s}_{1j}^p\|_2^3\right) = O\left(\sqrt{\frac{k_n^3}{n}}\right).$$

In addition, because $(X_i, A_i)_{i\in[N]}$ and $(W_j, B_j)_{j\in[M]}$ are independent, we have $g_{1,n} \perp g_{2,n}$. Last, because $g_{3,n}^*$ is independent of $\mathcal{F}_n$, we have $(g_{1,n}, g_{2,n}, g_{3,n}^*)$ are independent.

Next, we turn to $U_{4,n}$. We note that

$$\|U_{4,n}\|_2 \leq \sqrt{k_n}\|U_{4,n}\|_\infty.$$

In addition, by Chiang, Kato, and Sasaki (2023, Corollary 3) with their $k = 2$, $q = 2$, and $e = \{1, 1\}$, we have

$$\begin{aligned}
\frac{NM}{\sqrt{n}}\left(\mathbb{E}\left[\|U_{4,n}\|_\infty^2\right]\right)^{1/2} &\leq C(\log(k_n))\sqrt{NM}\left(\mathbb{E}\|\bar{s}_{ij} - \bar{s}_{1i}^a - \bar{s}_{1j}^p\|_\infty^2\right)^{1/2} \\
&\leq C(\log(k_n))\sqrt{NM}\left(\mathbb{E}\|\bar{s}_{ij} - \bar{s}_{1i}^a - \bar{s}_{1j}^p\|_2^2\right)^{1/2} \\
&\leq C\log(k_n)\sqrt{k_n}\sqrt{NM}.
\end{aligned}$$

This implies

$$||U_{4,n}||_2 = O_P\left(\sqrt{\frac{\log(n)k_n^2}{n}}\right).$$

Therefore, by letting $\delta_n = (Ck_n)^{-1}$ for a large constant $C > 0$, we have

$$\left\|\frac{\sqrt{n}}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}s_{ij} - \left(\Omega_{1,n}^{1/2}g_{1,n} + \Omega_{2,n}^{1/2}g_{2,n} + \begin{pmatrix}\Omega_{3,n}^{*,1/2}g_{3,n}^* \\ 0\end{pmatrix}\right)\right\| = O_P\left(\sqrt{\frac{\log(n)k_n^2}{n}} + \sqrt{\frac{k_n^3\zeta_n^2}{n}} + \left(\frac{k_n^3}{n}\right)^{1/6}k_n^{1/3}\right)$$

$$= o_P(1).$$

In addition, we have

$$g_n \equiv \Omega_{1,n}^{1/2}g_{1,n} + \Omega_{2,n}^{1/2}g_{2,n} + \begin{pmatrix}\Omega_{3,n}^{*,1/2}g_{3,n}^* \\ 0\end{pmatrix} \sim \mathcal{N}\left(0, \Omega_{1,n} + \Omega_{2,n} + \begin{pmatrix}\Omega_{3,n}^* & 0 \\ 0 & 0\end{pmatrix}\right) \stackrel{d}{=} \mathcal{N}(0, \Sigma_n).$$

This concludes the proof.

D   Additional Lemmas

Lemma D.1. *Suppose conditions in Theorem 1 holds. Then, we have*

$$||\hat{\Sigma}_{YY,n}^a - \Sigma_{YY,n}^a||_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right)$$

and

$$||\hat{\Sigma}_{YY,n}^p - \Sigma_{YY,n}^p||_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right).$$

Proof. We have

$$\left|\hat{\Sigma}_{YY,n}^a - \mathbb{E}\left\{[\mathbb{E}(s_{Y,ij}H_{ij}|X_i, A_i)][\mathbb{E}(s_{Y,ij}H_{ij}^\top|X_i, A_i)]\right\}\right|$$

$$\leq \left\|\frac{1}{N}\sum_{i=1}^{N}\frac{2}{M(M-1)}\sum_{j=1}^{M-1}\sum_{k=j+1}^{M}(\hat{s}_{Y,ij}\hat{s}_{Y,ik} - s_{Y,ij}s_{Y,ik})H_{ik}H_{ij}^\top\right\|_{op}$$

$$+ \left\|\frac{1}{N}\sum_{i=1}^{N}\frac{2}{M(M-1)}\sum_{j=1}^{M-1}\sum_{k=j+1}^{M}s_{Y,ij}s_{Y,ik}H_{ij}H_{ij}^\top - \mathbb{E}\left\{[\mathbb{E}(s_{Y,ij}H_{ij}|X_i, A_i)][\mathbb{E}(s_{Y,ij}H_{ij}^\top|X_i, A_i)]\right\}\right\|_{op}$$

$$\leq \left\| \frac{1}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} (\hat{s}_{Y,ij} \hat{s}_{Y,ik} - s_{Y,ij} s_{Y,ik}) H_{ik} H_{ij}^\top \right\|_{op}$$

$$+ \left\| \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} \frac{1}{N} \sum_{i=1}^{N} \left\{ s_{Y,ij} s_{Y,ik} H_{ij} H_{ik}^\top - \left\{ [\mathbb{E}(s_{Y,ij} H_{ij}|X_i, A_i)][\mathbb{E}(s_{Y,ij} H_{ij}^\top|X_i, A_i)] \right\} \right\} \right\|_{op}$$

$$+ \left\| \frac{1}{N} \sum_{i=1}^{N} \left\{ [\mathbb{E}(s_{Y,ij} H_{ij}|X_i, A_i)][\mathbb{E}(s_{Y,ij} H_{ij}^\top|X_i, A_i)] \right\} - \mathbb{E} \left\{ [\mathbb{E}(s_{Y,ij} H_{ij}|X_i, A_i)][\mathbb{E}(s_{Y,ij} H_{ij}^\top|X_i, A_i)] \right\} \right\|_{op}$$

$$\equiv ||I||_{op} + ||II||_{op} + ||III||_{op}. \tag{D.1}$$

Bound for $||I||_{op}$. We will establish that

$$||I||_{op} \leq C \left\| \frac{1}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} H_{ik} H_{ij} \right\|_{op} ||\hat{\theta}_{\text{initial}} - \theta_{0,n}||_2 = O_P(n^{-1/2}). \tag{D.2}$$

First, we note that

$$||I||_{op} = \sup_{u,v \in \Re^{k_n}, ||u||_2 = ||v||_2 = 1} \left| \frac{1}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} (\hat{s}_{Y,ij} - s_{Y,ij}) \hat{s}_{Y,ik} u^\top H_{ik} H_{ij}^\top v \right|$$

$$+ \sup_{u,v \in \Re^{k_n}, ||u||_2 = ||v||_2 = 1} \left| \frac{1}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} (\hat{s}_{Y,ij} - s_{Y,ij}) s_{Y,ik} u^\top H_{ik} H_{ij}^\top v \right|$$

$$\leq \frac{C}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} n(Y_{ik} + \Lambda(R_{ik}^\top \hat{\theta}_{\text{initial}}))||H_{ik}||_2 ||H_{ij}||_2 ||\hat{\theta}_{\text{initial}} - \theta_{0,n}||_2$$

$$+ \frac{C}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} n(Y_{ik} + \Lambda(R_{i,k}^\top \theta_{0,n}))||H_{ik}||_2 ||H_{ij}||_2 ||\hat{\theta}_{\text{initial}} - \theta_{0,n}||_2$$

$$\leq \frac{C}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} n(Y_{ik} - p_{ik})||H_{ik}||_2 ||H_{ij}||_2 \times O_P(n^{-1/2})$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} ||H_{ik}||_2 ||H_{ij}||_2 \times O_P(n^{-1/2}), \tag{D.3}$$

where we use the fact that np_{ij} is bounded, $\max_{i,j} n\Lambda(R_{ij}^\top \hat{\theta}_{\text{initial}}) = O_P(1)$, and

$$|\hat{s}_{Y,ij} - S_{Y,ij}| = n|\Lambda(R_{ij}^\top \hat{\theta}_{\text{initial}}) - \Lambda(R_{ij}^\top \theta_{0,n})| \leq C||\hat{\theta}_{\text{initial}} - \theta_{0,n}||_2.$$

Then, we have

$$
\mathbb{E} \left| \frac{1}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} n(Y_{ik} - p_{ik}) ||H_{ik}||_2 ||H_{ij}||_2 \right|^2
$$

$$
\leq \mathbb{E} \left| \frac{1}{NM(M-1)} \sum_{i=1}^{N} \sum_{j=1}^{M-1} \left[n(Y_{ij} - p_{ij}) ||H_{ij}||_2 \right] \left[\sum_{k=j+1}^{M} ||H_{ik}||_2 \right] \right|^2
$$

$$
= \frac{1}{N^2 M^2 (M-1)^2} \sum_{i=1}^{N} \sum_{j=1}^{M-1} \mathbb{E} \left[n(Y_{ij} - p_{ij}) ||H_{ij,l}||_2 \right]^2 \left[\sum_{k=j+1}^{M} ||H_{ik}||_2 \right]^2
$$

$$
\leq \frac{CnM^2 k_n \zeta_n^2}{N^2 M^2 (M-1)^2} \sum_{i=1}^{N} \sum_{j=1}^{M-1} ||H_{ij}||_2^2 = O_P \left(\frac{k_n^2 \zeta_n^2}{n} \right).
$$

This implies

$$
\frac{1}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} n(Y_{ik} - p_{ik}) ||H_{ik}||_2 ||H_{ij}||_2 = O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}} \right).
$$

In addition, we have

$$
\frac{1}{N} \sum_{i=1}^{N} \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} ||H_{ik}||_2 ||H_{ij}||_2 \leq \frac{\sqrt{k_n} \zeta_n}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M-1} ||H_{ij}||_2 \leq O_P(k_n \zeta_n).
$$

Therefore, given that $||\hat{\theta}_{\text{initial}} - \theta_{0,n}||_2 = O_P(n^{-1/2})$, we have $||I||_{op} = O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}} \right)$.

Bound for $||II||_{op}$**.** We first define $\mathcal{D}_j = \{Y_{1j}, \cdots, Y_{Nj}, W_j, B_j\}$ and make our argument conditional on $\mathcal{X}_N$ and $\mathcal{A}_N$. In this view,

$$
\frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} \frac{1}{N} \sum_{i=1}^{N} \left\{ s_{Y,ij} s_{Y,ik} H_{ij} H_{ik}^\top - \left\{ [\mathbb{E}(s_{Y,ij} H_{ij} | X_i, A_i)][\mathbb{E}(s_{Y,ij} H_{ij}^\top | X_i, A_i)] \right\} \right\}
$$

$$
\equiv \frac{1}{M(M-1)} \sum_{j,k \in [M], j \neq k} f(\mathcal{D}_j, \mathcal{D}_k),
$$

where

$$
f(\mathcal{D}_j, \mathcal{D}_k) = \frac{1}{N} \sum_{i=1}^{N} \left\{ s_{Y,ij} s_{Y,ik} H_{ij} H_{ik}^\top - [\mathbb{E}(s_{Y,ij} H_{ij} | X_i, A_i)][\mathbb{E}(s_{Y,ij} H_{ij}^\top | X_i, A_i)] \right\}.
$$

By Hoeffding decomposition, we have

$$
II = \frac{1}{M(M-1)} \sum_{j,k \in [M], j \neq k} f(\mathcal{D}_j, \mathcal{D}_k)
$$

$$
= \frac{2}{M} \sum_{j=1}^{M} \mathbb{E}\left[f(\mathcal{D}_j, \mathcal{D}_k) | \mathcal{D}_j, \mathcal{X}_N, \mathcal{A}_N\right]
$$

$$
+ \frac{1}{M(M-1)} \sum_{j,k \in [M], j \neq k} \{f(\mathcal{D}_j, \mathcal{D}_k) - \mathbb{E}\left[f(\mathcal{D}_j, \mathcal{D}_k; u) | \mathcal{D}_k, \mathcal{X}_N, \mathcal{A}_N\right] - \mathbb{E}\left[f(\mathcal{D}_j, \mathcal{D}_k; u) | \mathcal{D}_j, \mathcal{X}_N, \mathcal{A}_N\right]\}
$$

$$
\equiv 2II_1 + II_2. \tag{D.4}
$$

Bound for $\|II_1\|_{op}$**.** For the linear term II_1 on the RHS of (D.4), we have

$$
II_1 = \frac{1}{M} \sum_{j=1}^{M} \mathbb{E}\left[f(\mathcal{D}_j, \mathcal{D}_k) | \mathcal{D}_j, \mathcal{X}_N, \mathcal{A}_N\right]
$$

$$
= \frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} \left[s_{Y,ij} H_{ij} - \mathbb{E}(s_{Y,ij} H_{ij} | X_i, A_i)\right] \mathbb{E}(s_{Y,ij} H_{ij}^{\top} | X_i, A_i)
$$

$$
= \frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} (s_{Y,ij} - \bar{s}_{Y,ij}) H_{ij} \mathbb{E}(s_{Y,ij} H_{ij}^{\top} | X_i, A_i)
$$

$$
+ \frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} \left[\bar{s}_{Y,ij} H_{ij} - \mathbb{E}(s_{Y,ij} H_{ij} | X_i, A_i)\right] \mathbb{E}(s_{Y,ij} H_{ij}^{\top} | X_i, A_i)
$$

$$
\equiv II_{1,1} + II_{1,2}. \tag{D.5}
$$

Conditionally on $(\mathcal{X}_N, \mathcal{A}_N, \mathcal{W}_M, \mathcal{B}_M)$, $\{(s_{Y,ij} - \bar{s}_{Y,ij}) H_{ij} \mathbb{E}(s_{Y,ij} H_{ij}^{\top} | X_i, A_i)\}_{i \in [N], j \in [M]}$ is a sequence of independent and mean-zero random variables. Therefore, we have

$$
\mathbb{E}\left(\|II_{1,1}\|_F^2 \mid \mathcal{X}_N, \mathcal{A}_N, \mathcal{W}_M, \mathcal{B}_M\right)
$$

$$
\leq \frac{1}{(MN)^2} \sum_{(\ell,\ell') \in [k_n] \times [k_n]} \sum_{i=1}^{N} \sum_{j=1}^{M} n^2 p_{ij} (1 - p_{ij}) H_{ij,\ell}^2 \left[\mathbb{E}(s_{Y,ij} H_{ij,\ell'} | X_i, A_i)\right]^2
$$

$$
\leq C \frac{k_n n \zeta_n^2}{NM} \left[\sum_{\ell=1}^{k_n} \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{1}{NM} H_{ij,\ell}^2\right] = O_P\left(\frac{k_n^2 \zeta_n^2}{n}\right),
$$

which implies

$$
\|II_{1,1}\|_{op} = O_P\left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}}\right).
$$

For $II_{1,2}$, we have

$$\mathbb{E}\left(||II_{1,2}||_F^2 \mid \mathcal{X}_N, \mathcal{A}_N\right)$$

$$= \frac{1}{M^2} \sum_{(\ell,\ell')\in[k_n]\times[k_n]} \sum_{j=1}^{M} \mathbb{E}\left(\left\{\frac{1}{N}\sum_{i=1}^{N}[\bar{s}_{Y,ij}H_{ij,\ell} - \mathbb{E}(s_{Y,ij}H_{ij,\ell}|X_i,A_i)]\left[\mathbb{E}(s_{Y,ij}H_{ij,\ell'}|X_i,A_i)\right]\right\}^2 \mid \mathcal{X}_N, \mathcal{A}_N\right)$$

$$\leq \frac{1}{M^2} \sum_{(\ell,\ell')\in[k_n]\times[k_n]} \sum_{j=1}^{M} \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left\{[\bar{s}_{Y,ij}H_{ij,\ell} - \mathbb{E}(s_{Y,ij}H_{ij,\ell}|X_i,A_i)]\left[\mathbb{E}(s_{Y,ij}H_{ij,\ell'}|X_i,A_i)\right]\right\}^2 \mid \mathcal{X}_N, \mathcal{A}_N\right),$$

and thus,

$$\mathbb{E}\left(||II_{1,2}||_F^2\right)$$

$$\leq \frac{1}{NM^2} \sum_{(\ell,\ell')\in[k_n]\times[k_n]} \sum_{i=1}^{N}\sum_{j=1}^{M} \mathbb{E}\left\{[\bar{s}_{Y,ij}H_{ij,\ell} - \mathbb{E}(s_{Y,ij}H_{ij,\ell}|X_i,A_i)]\left[\mathbb{E}(s_{Y,ij}H_{ij,\ell'}|X_i,A_i)\right]\right\}^2$$

$$= O_P\left(\frac{k_n^2\zeta_n^2}{n}\right),$$

where use the fact that $|\bar{s}_{Y,ij}| \leq C < \infty$.

This implies $||II_{1,2}||_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right)||$, and thus,

$$||II_1||_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right). \tag{D.6}$$

Bound for $||II_2||_{op}$. Next, we turn to $II_2(u)$ in (D.4). We note that, conditionally on $(\mathcal{X}_N, \mathcal{A}_N)$, II_2 is a second-order degenerate U-statistic (based on observations $(\mathcal{D}_j, \mathcal{D}_k)$) which has an increasing dimension of $k_n \times k_n$. Let

$$f_{\ell,\ell'}(\mathcal{D}_j, \mathcal{D}_k) = \frac{1}{N}\sum_{i=1}^{N}\left\{s_{Y,ij}s_{Y,ik}H_{ij,\ell}H_{ik,\ell'} - [\mathbb{E}(s_{Y,ij}H_{ij,\ell}|X_i,A_i)][\mathbb{E}(s_{Y,ij}H_{ij,\ell'}|X_i,A_i)]\right\}.$$

We have

$$\mathbb{E}\left(||II_2||_F^2 \mid \mathcal{X}_N, \mathcal{A}_N\right) \leq \frac{1}{M^4} \sum_{(\ell,\ell')\in[k_n]\times[k_n]} \sum_{(j,k)\in[M]\times[M],j\neq k} \mathbb{E}\left(f_{\ell,\ell'}^2(\mathcal{D}_j, \mathcal{D}_k) \mid \mathcal{X}_N, \mathcal{A}_N\right).$$

Further note that $\{s_{Y,ij}s_{Y,ik}H_{ij,\ell}H_{ik,\ell'}\}_{i\in[N]}$ are independent conditionally on $(\mathcal{X}_N, \mathcal{A}_N, \mathcal{W}_M, \mathcal{B}_M)$ and

$$\mathbb{E}\left[s_{Y,ij}s_{Y,ik}H_{ij,\ell}H_{ik,\ell'} \mid \mathcal{X}_N, \mathcal{A}_N, \mathcal{W}_M, \mathcal{B}_M\right] = n^2 p_{ij}p_{ik}H_{ij,\ell}H_{ik,\ell'}.$$

Then, we have

$$f_{\ell,\ell'}^2(\mathcal{D}_j, \mathcal{D}_k) \leq 2\left[\frac{1}{N}\sum_{i=1}^N (s_{Y,ij}s_{Y,ik} - n^2 p_{ij}p_{ik})H_{ij,\ell}H_{ik,\ell'}\right]^2$$

$$+ 2\left[\frac{1}{N}\sum_{i=1}^N n^2 p_{ij}p_{ik}H_{ij,\ell}H_{ik,\ell'} - [\mathbb{E}(s_{Y,ij}H_{ij,\ell}|X_i, A_i)][\mathbb{E}(s_{Y,ij}H_{ij,\ell'}|X_i, A_i)]\right]^2$$

$$\leq C\left[\frac{1}{N}\sum_{i=1}^N (s_{Y,ij}s_{Y,ik} - n^2 p_{ij}p_{ik})H_{ij,\ell}H_{ik,\ell'}\right]^2 + C\zeta_n^4.$$

and thus,

$$\mathbb{E}\left(\|II_2\|_F^2\right)$$

$$\leq \frac{C}{M^4}\sum_{(\ell,\ell')\in[k_n]\times[k_n]}\sum_{(j,k)\in[M]\times[M], j\neq k}\mathbb{E}\left(f_{\ell,\ell'}^2(\mathcal{D}_j, \mathcal{D}_k)\right)$$

$$\leq \frac{C}{M^4}\sum_{(\ell,\ell')\in[k_n]\times[k_n]}\sum_{(j,k)\in[M]\times[M], j\neq k}\left\{\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N (s_{Y,ij}s_{Y,ik} - n^2 p_{ij}p_{ik})H_{ij,\ell}H_{ik,\ell'}\right]^2 + \zeta_n^4\right\}$$

$$= \frac{1}{M^4}\sum_{(\ell,\ell')\in[k_n]\times[k_n]}\sum_{(j,k)\in[M]\times[M], j\neq k}\mathbb{E}\left\{\mathbb{E}\left(\left[\frac{1}{N}\sum_{i=1}^N (s_{Y,ij}s_{Y,ik} - n^2 p_{ij}p_{ik})H_{ij,\ell}H_{ik,\ell'}\right]^2 \mid \mathcal{X}_N, \mathcal{A}_N, \mathcal{W}_M, \mathcal{B}_M\right)\right\}$$

$$+ \frac{Ck_n^2\zeta_n^4}{n^2}$$

$$= \frac{C}{M^4 N^2}\sum_{(\ell,\ell')\in[k_n]\times[k_n]}\sum_{(j,k)\in[M]\times[M], j\neq k}\mathbb{E}\left[(s_{Y,ij}s_{Y,ik} - n^2 p_{ij}p_{ik})H_{ij,\ell}H_{ik,\ell'}\right]^2 + \frac{Ck_n^2\zeta_n^4}{n^2}$$

$$\leq \frac{n^2 k_n\zeta_n^2}{N^2 M^4}\sum_{(\ell,\ell')\in[k_n]\times[k_n]}\sum_{(j,k)\in[M]\times[M], j\neq k}\sum_{i=1}^N \mathbb{E}H_{ij,\ell}^2 + \frac{Ck_n^2\zeta_n^4}{n^2}$$

$$= O\left(\frac{k_n^2\varsigma_n^2}{n}\right).$$

This implies

$$\|II_2\|_{op} \leq \|II_2\|_F = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right). \tag{D.7}$$

Combining (D.7) with (D.6), we have

$$\|II\|_{op} = O_P\left(\sqrt{\frac{k_n^2\zeta_n^2}{n}}\right).$$

Bound for $||III||_{op}$**.** By a similar argument to the analysis for II_1, we have

$$||III||_{op} = O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}} \right). \tag{D.8}$$

Combining (D.2), (D.7), and (D.8), we have the desired result that

$$||\hat{\Sigma}_{YY,n}^a - \Sigma_{YY,n}^a||_{op} = O_P \left(\sqrt{\frac{k_n^2 \zeta_n^2}{n}} \right).$$

The other result can be established in the same manner. $\qquad\qquad\square$

E Additional Tables

E.1 Empirical Application

Table 13: Results Robust Estimator with $\overline{\sigma}^2 = 1$

	coeff.	s.e.	max rMSE	[conf. int.]		length	adjm
Homophily parameters							
jelcode	1.611	0.044	5.403	[1.443,	1.779]	0.335	0.104
citations	0.390	0.015	5.076	[0.275,	0.505]	0.230	0.087
gender	5.895	0.181	35.256	[4.961,	6.829]	1.868	0.438
junior_senior	−3.639	0.129	13.235	[−4.080,	−3.197]	0.882	−0.591
constant	−0.832	0.077	8.158	[−1.101,	−0.563]	0.537	−0.067
Average out-degree							
Psi	1.745	0.032	1.897	[1.675,	1.815]	0.139	−0.026

[1] Total sample includes $N = 1776$ authors and $M = 1600$ articles.

Table 14: Results Robust Estimator with $\overline{\sigma}^2 = 2$

	coeff.	s.e.	max rMSE	[conf. int.]		length	adjm
Homophily parameters							
jelcode	1.644	0.047	8.016	[1.422,	1.867]	0.445	0.138
citations	0.401	0.014	7.406	[0.248,	0.554]	0.306	0.098
gender	5.969	0.178	50.036	[4.777,	7.161]	2.384	0.512
junior_senior	−3.670	0.133	17.742	[−4.205,	−3.134]	1.071	−0.622
constant	−0.847	0.087	12.232	[−1.209,	−0.485]	0.724	−0.082
Average out-degree							
Psi	1.758	0.031	2.004	[1.682,	1.833]	0.152	−0.014

[1] Total sample includes $N = 1776$ authors and $M = 1600$ articles.

Table 15: Results Robust Estimator with $\overline{\sigma}^2 = 3$

	coeff.	s.e.	max rMSE	[conf. int.]		length	adjm
Homophily parameters							
jelcode	1.650	0.045	9.087	[1.412,	1.888]	0.477	0.143
citations	0.402	0.014	8.874	[0.223,	0.580]	0.357	0.098
gender	5.979	0.179	60.135	[4.610,	7.349]	2.740	0.523
junior_senior	−3.678	0.132	20.771	[−4.268,	−3.087]	1.181	−0.630
constant	−0.848	0.083	13.888	[−1.235,	−0.461]	0.774	−0.083
Average out-degree							
Psi	1.778	0.031	2.349	[1.691,	1.865]	0.174	0.007

[1] Total sample includes $N = 1776$ authors and $M = 1600$ articles.

E.2 Additional DGPs for Local Misspecification

Table 16: $\hat{\beta}_n$ Robust Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf.	int.]	length	95% CP	Degree
					$n = 200$						
0.0	1.386	1.385	−0.025	0.146	0.878	2.146	[0.657,	2.112]	1.455	0.999	0.010
1.0	1.386	1.687	4.216	0.145	0.875	2.158	[0.811,	2.558]	1.747	1.000	0.011
2.0	1.386	1.774	5.428	0.145	0.887	2.172	[0.842,	2.699]	1.857	1.000	0.012
3.0	1.386	1.841	6.395	0.145	0.885	2.200	[0.859,	2.818]	1.960	1.000	0.013
4.0	1.386	1.888	7.038	0.147	0.900	2.252	[0.861,	2.907]	2.045	1.000	0.014
					$n = 400$						
0.0	1.386	1.388	0.097	0.109	0.917	2.226	[0.862,	1.921]	1.059	1.000	0.005
1.0	1.386	1.599	4.274	0.108	0.903	2.227	[0.998,	2.202]	1.203	1.000	0.005
2.0	1.386	1.663	5.508	0.108	0.894	2.241	[1.033,	2.290]	1.257	1.000	0.006
3.0	1.386	1.724	6.667	0.108	0.880	2.261	[1.063,	2.376]	1.313	1.000	0.006
4.0	1.386	1.783	7.822	0.108	0.874	2.267	[1.092,	2.463]	1.371	0.999	0.006
					$n = 600$						
0.0	1.386	1.386	−0.001	0.090	0.928	2.246	[0.951,	1.821]	0.870	1.000	0.003
1.0	1.386	1.557	4.186	0.090	0.919	2.265	[1.073,	2.042]	0.969	1.000	0.004
2.0	1.386	1.604	5.342	0.090	0.913	2.285	[1.103,	2.106]	1.003	1.000	0.004
3.0	1.386	1.649	6.477	0.091	0.904	2.310	[1.131,	2.171]	1.040	1.000	0.004
4.0	1.386	1.698	7.652	0.092	0.896	2.339	[1.158,	2.239]	1.081	0.997	0.004
					$n = 800$						
0.0	1.386	1.387	−0.039	0.079	0.934	2.274	[1.006,	1.763]	0.757	1.000	0.003
1.0	1.386	1.534	4.113	0.079	0.918	2.297	[1.117,	1.947]	0.830	1.000	0.003
2.0	1.386	1.573	5.242	0.079	0.908	2.316	[1.144,	1.999]	0.855	1.000	0.003
3.0	1.386	1.613	6.330	0.080	0.911	2.351	[1.169,	2.051]	0.882	0.999	0.003
4.0	1.386	1.651	7.433	0.082	0.917	2.393	[1.193,	2.105]	0.912	0.999	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Functional misspecification. [3] Sieves dimension $k_n = 2$.

Table 17: $\hat{\beta}_n$ Logistic Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
					$n = 200$						
0.0	1.386	1.419	0.475	0.159	0.894	2.343	[0.666,	2.174]	1.508	0.999	0.010
1.0	1.386	1.751	5.159	0.167	0.895	2.498	[0.833,	2.669]	1.836	1.000	0.011
2.0	1.386	1.855	6.584	0.171	0.908	2.566	[0.872,	2.832]	1.960	1.000	0.012
3.0	1.386	1.940	7.771	0.174	0.907	2.632	[0.900,	2.972]	2.072	1.000	0.013
4.0	1.386	2.000	8.619	0.177	0.905	2.703	[0.914,	3.077]	2.163	1.000	0.014
					$n = 400$						
0.0	1.386	1.405	0.465	0.115	0.928	2.345	[0.869,	1.951]	1.082	1.000	0.005
1.0	1.386	1.628	4.865	0.116	0.912	2.422	[1.009,	2.250]	1.240	1.000	0.005
2.0	1.386	1.695	6.200	0.118	0.904	2.481	[1.045,	2.348]	1.303	1.000	0.006
3.0	1.386	1.763	7.476	0.122	0.901	2.555	[1.075,	2.445]	1.370	1.000	0.006
4.0	1.386	1.830	8.772	0.124	0.901	2.616	[1.106,	2.544]	1.438	0.998	0.006
					$n = 600$						
0.0	1.386	1.398	0.301	0.093	0.934	2.330	[0.957,	1.840]	0.883	1.000	0.003
1.0	1.386	1.574	4.633	0.094	0.925	2.397	[1.081,	2.070]	0.989	1.000	0.004
2.0	1.386	1.624	5.848	0.096	0.919	2.444	[1.111,	2.139]	1.028	1.000	0.004
3.0	1.386	1.672	7.052	0.098	0.915	2.505	[1.139,	2.209]	1.070	0.998	0.004
4.0	1.386	1.724	8.310	0.100	0.911	2.578	[1.167,	2.284]	1.117	0.996	0.004
					$n = 800$						
0.0	1.386	1.396	0.221	0.081	0.940	2.339	[1.011,	1.777]	0.766	1.000	0.003
1.0	1.386	1.547	4.478	0.082	0.923	2.393	[1.123,	1.966]	0.843	1.000	0.003
2.0	1.386	1.587	5.646	0.083	0.913	2.429	[1.151,	2.021]	0.870	1.000	0.003
3.0	1.386	1.628	6.780	0.084	0.918	2.487	[1.176,	2.076]	0.900	0.999	0.003
4.0	1.386	1.668	7.935	0.086	0.924	2.559	[1.200,	2.134]	0.934	0.998	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Functional misspecification.

Table 18: $\hat{\beta}_n$ Poisson Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.386	1.384	-0.041	0.149	0.898	2.191	[0.665,	2.102]	1.437	0.999	0.010
1.0	1.386	1.683	4.158	0.147	0.901	2.191	[0.844,	2.517]	1.673	1.000	0.011
2.0	1.386	1.768	5.340	0.145	0.904	2.189	[0.890,	2.638]	1.749	1.000	0.012
3.0	1.386	1.832	6.271	0.144	0.897	2.188	[0.924,	2.735]	1.811	1.000	0.013
4.0	1.386	1.875	6.873	0.142	0.895	2.202	[0.944,	2.800]	1.856	1.000	0.014
						$n = 400$					
0.0	1.386	1.388	0.091	0.110	0.929	2.257	[0.864,	1.918]	1.053	1.000	0.005
1.0	1.386	1.599	4.260	0.110	0.922	2.270	[1.006,	2.192]	1.186	1.000	0.005
2.0	1.386	1.663	5.488	0.110	0.916	2.294	[1.043,	2.278]	1.236	1.000	0.006
3.0	1.386	1.723	6.639	0.112	0.912	2.326	[1.075,	2.361]	1.286	1.000	0.006
4.0	1.386	1.781	7.784	0.112	0.912	2.341	[1.108,	2.443]	1.335	0.999	0.006
						$n = 600$					
0.0	1.386	1.386	-0.004	0.091	0.936	2.269	[0.952,	1.820]	0.867	1.000	0.003
1.0	1.386	1.556	4.179	0.091	0.931	2.298	[1.076,	2.038]	0.962	1.000	0.004
2.0	1.386	1.604	5.334	0.092	0.929	2.325	[1.107,	2.101]	0.995	1.000	0.004
3.0	1.386	1.649	6.467	0.093	0.923	2.362	[1.136,	2.165]	1.030	0.999	0.004
4.0	1.386	1.697	7.638	0.095	0.923	2.404	[1.164,	2.232]	1.068	0.997	0.004
						$n = 800$					
0.0	1.386	1.387	-0.041	0.079	0.941	2.292	[1.007,	1.762]	0.755	1.000	0.003
1.0	1.386	1.534	4.110	0.080	0.927	2.322	[1.118,	1.945]	0.826	1.000	0.003
2.0	1.386	1.572	5.237	0.080	0.919	2.347	[1.146,	1.997]	0.850	1.000	0.003
3.0	1.386	1.612	6.325	0.082	0.925	2.389	[1.171,	2.048]	0.877	0.999	0.003
4.0	1.386	1.651	7.427	0.083	0.932	2.441	[1.196,	2.102]	0.906	0.999	0.003

[1] Number of Monte Carlo simulations is 2,000. [2] Local misspecification: Functional misspecification.

Table 19: $\hat{\beta}_n$ Ratios

$\bar{\sigma}^2$	Logit/Robust				Poisson/Robust			
	s.e	bias	rMSE	length	s.e	bias	rMSE	length
				$n = 200$				
0.0	1.087	19.273	1.092	1.037	1.019	1.677	1.021	0.987
1.0	1.060	0.768	1.055	1.024	1.010	1.006	1.009	0.990
2.0	1.053	0.811	1.051	1.023	1.006	1.004	1.008	0.990
3.0	1.051	0.836	1.048	1.021	1.005	1.004	1.007	0.991
4.0	1.048	0.852	1.046	1.021	1.006	1.003	1.007	0.991
				$n = 400$				
0.0	1.051	4.803	1.054	1.022	1.012	0.941	1.014	0.995
1.0	1.034	0.816	1.034	1.015	1.007	1.002	1.007	0.995
2.0	1.029	0.853	1.032	1.014	1.004	1.002	1.006	0.995
3.0	1.028	0.872	1.030	1.013	1.005	1.002	1.006	0.995
4.0	1.030	0.885	1.029	1.013	1.005	1.001	1.005	0.995
				$n = 600$				
0.0	1.033	413.730	1.037	1.015	1.009	5.141	1.010	0.996
1.0	1.025	0.855	1.025	1.010	1.006	1.001	1.005	0.996
2.0	1.025	0.882	1.023	1.010	1.005	1.001	1.005	0.996
3.0	1.020	0.897	1.022	1.009	1.003	1.001	1.004	0.996
4.0	1.021	0.907	1.021	1.009	1.004	1.001	1.004	0.996
				$n = 800$				
0.0	1.026	5.601	1.028	1.011	1.007	1.048	1.008	0.998
1.0	1.019	0.877	1.020	1.008	1.004	1.001	1.005	0.997
2.0	1.018	0.900	1.018	1.008	1.004	1.001	1.004	0.997
3.0	1.016	0.913	1.018	1.008	1.003	1.001	1.004	0.997
4.0	1.016	0.921	1.017	1.007	1.003	1.000	1.004	0.997

[1] Number of Monte Carlo simulations is $2,000$.

[2] DGP Local misspecification: Functional misspecification.

[3] Sieves dimension $k_n = 2$.

Table 20: $\hat{\beta}_n$ Robust Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
					$n = 200$						
0.0	1.386	1.385	−0.025	0.146	0.878	2.146	[0.657,	2.112]	1.455	0.999	0.010
1.0	1.386	1.529	1.990	0.146	0.881	2.141	[0.738,	2.316]	1.577	1.000	0.011
2.0	1.386	1.556	2.387	0.146	0.884	2.145	[0.752,	2.358]	1.605	1.000	0.011
3.0	1.386	1.580	2.716	0.146	0.879	2.145	[0.764,	2.393]	1.629	1.000	0.011
4.0	1.386	1.595	2.974	0.146	0.878	2.144	[0.773,	2.420]	1.647	1.000	0.011
					$n = 400$						
0.0	1.386	1.388	0.097	0.109	0.917	2.226	[0.862,	1.921]	1.059	1.000	0.005
1.0	1.386	1.487	2.097	0.108	0.908	2.219	[0.930,	2.052]	1.122	1.000	0.005
2.0	1.386	1.509	2.508	0.108	0.908	2.215	[0.944,	2.079]	1.135	1.000	0.005
3.0	1.386	1.523	2.832	0.108	0.906	2.214	[0.955,	2.101]	1.146	1.000	0.005
4.0	1.386	1.538	3.103	0.108	0.906	2.211	[0.964,	2.119]	1.155	1.000	0.005
					$n = 600$						
0.0	1.386	1.386	−0.001	0.090	0.928	2.246	[0.951,	1.821]	0.870	1.000	0.003
1.0	1.386	1.467	2.005	0.090	0.930	2.245	[1.011,	1.925]	0.914	1.000	0.004
2.0	1.386	1.485	2.434	0.090	0.925	2.246	[1.024,	1.947]	0.923	1.000	0.004
3.0	1.386	1.499	2.743	0.090	0.923	2.246	[1.033,	1.964]	0.931	1.000	0.004
4.0	1.386	1.509	3.014	0.089	0.920	2.246	[1.041,	1.978]	0.937	1.000	0.004
					$n = 800$						
0.0	1.386	1.387	−0.039	0.079	0.934	2.274	[1.006,	1.763]	0.757	1.000	0.003
1.0	1.386	1.458	1.924	0.078	0.928	2.280	[1.060,	1.849]	0.789	1.000	0.003
2.0	1.386	1.472	2.337	0.078	0.924	2.282	[1.071,	1.867]	0.796	1.000	0.003
3.0	1.386	1.483	2.659	0.078	0.924	2.283	[1.079,	1.881]	0.802	1.000	0.003
4.0	1.386	1.493	2.925	0.078	0.920	2.285	[1.086,	1.893]	0.807	1.000	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Semiparametric distribution. [3] Sieves dimension $k_n = 2$.

Table 21: $\hat{\beta}_n$ Logistic Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf.	int.]	length	95% CP	Degree
					$n = 200$						
0.0	1.386	1.419	0.475	0.159	0.894	2.343	[0.666,	2.174]	1.508	0.999	0.010
1.0	1.386	1.577	2.661	0.162	0.898	2.391	[0.752,	2.396]	1.644	1.000	0.011
2.0	1.386	1.603	3.100	0.162	0.896	2.407	[0.768,	2.443]	1.675	1.000	0.011
3.0	1.386	1.631	3.465	0.163	0.892	2.420	[0.781,	2.482]	1.701	1.000	0.011
4.0	1.386	1.648	3.754	0.163	0.886	2.426	[0.791,	2.512]	1.721	1.000	0.011
					$n = 400$						
0.0	1.386	1.405	0.465	0.115	0.928	2.345	[0.869,	1.951]	1.082	1.000	0.005
1.0	1.386	1.510	2.555	0.114	0.917	2.366	[0.939,	2.089]	1.150	1.000	0.005
2.0	1.386	1.531	2.988	0.115	0.918	2.368	[0.953,	2.118]	1.165	1.000	0.005
3.0	1.386	1.547	3.331	0.115	0.913	2.372	[0.965,	2.141]	1.177	1.000	0.005
4.0	1.386	1.563	3.618	0.115	0.914	2.373	[0.974,	2.160]	1.186	1.000	0.005
					$n = 600$						
0.0	1.386	1.398	0.301	0.093	0.934	2.330	[0.957,	1.840]	0.883	1.000	0.003
1.0	1.386	1.482	2.367	0.093	0.937	2.346	[1.018,	1.948]	0.929	1.000	0.004
2.0	1.386	1.501	2.812	0.093	0.932	2.352	[1.031,	1.971]	0.940	1.000	0.004
3.0	1.386	1.515	3.131	0.093	0.932	2.355	[1.040,	1.988]	0.947	1.000	0.004
4.0	1.386	1.525	3.412	0.093	0.928	2.359	[1.048,	2.003]	0.954	1.000	0.004
					$n = 800$						
0.0	1.386	1.396	0.221	0.081	0.940	2.339	[1.011,	1.777]	0.766	1.000	0.003
1.0	1.386	1.468	2.228	0.081	0.932	2.356	[1.065,	1.865]	0.799	1.000	0.003
2.0	1.386	1.483	2.651	0.081	0.929	2.360	[1.077,	1.883]	0.807	1.000	0.003
3.0	1.386	1.494	2.982	0.081	0.929	2.364	[1.085,	1.898]	0.813	1.000	0.003
4.0	1.386	1.504	3.255	0.081	0.924	2.368	[1.092,	1.910]	0.818	1.000	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Semiparametric distribution.

Table 22: $\hat{\beta}_n$ Poisson Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.386	1.384	−0.041	0.149	0.898	2.191	[0.665,	2.102]	1.437	0.999	0.010
1.0	1.386	1.527	1.961	0.148	0.901	2.183	[0.755,	2.295]	1.540	1.000	0.011
2.0	1.386	1.554	2.354	0.148	0.903	2.185	[0.771,	2.334]	1.563	1.000	0.011
3.0	1.386	1.577	2.679	0.149	0.903	2.184	[0.785,	2.366]	1.581	1.000	0.011
4.0	1.386	1.592	2.934	0.148	0.898	2.181	[0.796,	2.391]	1.595	1.000	0.011
						$n = 400$					
0.0	1.386	1.388	0.091	0.110	0.929	2.257	[0.864,	1.918]	1.053	1.000	0.005
1.0	1.386	1.487	2.088	0.110	0.924	2.254	[0.935,	2.047]	1.112	1.000	0.005
2.0	1.386	1.508	2.498	0.110	0.925	2.251	[0.949,	2.073]	1.124	1.000	0.005
3.0	1.386	1.522	2.822	0.109	0.918	2.250	[0.960,	2.095]	1.134	1.000	0.005
4.0	1.386	1.538	3.092	0.109	0.919	2.247	[0.970,	2.112]	1.142	1.000	0.005
						$n = 600$					
0.0	1.386	1.386	−0.004	0.091	0.936	2.269	[0.952,	1.820]	0.867	1.000	0.003
1.0	1.386	1.467	2.001	0.091	0.941	2.271	[1.014,	1.922]	0.909	1.000	0.004
2.0	1.386	1.485	2.430	0.091	0.936	2.273	[1.026,	1.945]	0.918	1.000	0.004
3.0	1.386	1.499	2.738	0.091	0.935	2.273	[1.036,	1.961]	0.925	1.000	0.004
4.0	1.386	1.509	3.008	0.090	0.931	2.274	[1.044,	1.975]	0.931	1.000	0.004
						$n = 800$					
0.0	1.386	1.387	−0.041	0.079	0.941	2.292	[1.007,	1.762]	0.755	1.000	0.003
1.0	1.386	1.457	1.921	0.079	0.934	2.300	[1.061,	1.847]	0.786	1.000	0.003
2.0	1.386	1.471	2.334	0.079	0.932	2.302	[1.072,	1.866]	0.793	1.000	0.003
3.0	1.386	1.483	2.656	0.079	0.931	2.304	[1.081,	1.880]	0.799	1.000	0.003
4.0	1.386	1.492	2.922	0.079	0.928	2.307	[1.088,	1.892]	0.804	1.000	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Semiparametric distribution.

Table 23: $\hat{\beta}_n$ Ratios

$\bar{\sigma}^2$	Logit/Robust				Poisson/Robust			
	s.e	bias	rMSE	length	s.e	bias	rMSE	length
				$n = 200$				
0.0	1.087	19.273	1.092	1.037	1.019	1.677	1.021	0.987
1.0	1.110	1.337	1.117	1.042	1.016	0.985	1.020	0.976
2.0	1.111	1.298	1.123	1.043	1.014	0.986	1.019	0.973
3.0	1.116	1.276	1.128	1.044	1.019	0.986	1.018	0.971
4.0	1.115	1.262	1.132	1.045	1.014	0.987	1.017	0.969
				$n = 400$				
0.0	1.051	4.803	1.054	1.022	1.012	0.941	1.014	0.995
1.0	1.059	1.219	1.066	1.025	1.015	0.996	1.016	0.991
2.0	1.064	1.192	1.069	1.026	1.017	0.996	1.016	0.990
3.0	1.063	1.176	1.071	1.026	1.012	0.996	1.016	0.989
4.0	1.066	1.166	1.073	1.027	1.012	0.996	1.016	0.989
				$n = 600$				
0.0	1.033	413.730	1.037	1.015	1.009	5.141	1.010	0.996
1.0	1.040	1.181	1.045	1.017	1.011	0.998	1.012	0.995
2.0	1.041	1.155	1.047	1.018	1.011	0.998	1.012	0.994
3.0	1.044	1.142	1.049	1.018	1.012	0.998	1.012	0.994
4.0	1.044	1.132	1.050	1.018	1.011	0.998	1.012	0.994
				$n = 800$				
0.0	1.026	5.601	1.028	1.011	1.007	1.048	1.008	0.998
1.0	1.029	1.158	1.033	1.013	1.007	0.999	1.009	0.997
2.0	1.030	1.135	1.034	1.013	1.008	0.999	1.009	0.996
3.0	1.031	1.121	1.035	1.014	1.008	0.999	1.009	0.996
4.0	1.031	1.113	1.036	1.014	1.009	0.999	1.009	0.996

[1] Number of Monte Carlo simulations is $2,000$.

[2] DGP Local misspecification: Semiparametric distribution.

[3] Sieves dimension $k_n = 2$.

Table 24: $\hat{\Psi}_n$ Robust Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
					$n = 200$						
0.0	1.026	1.024	0.132	0.163	0.949	2.407	[0.699,	1.371]	0.672	0.942	0.010
1.0	0.977	0.978	0.105	0.149	0.937	2.162	[0.678,	1.290]	0.612	0.939	0.010
2.0	0.969	0.969	0.093	0.146	0.935	2.125	[0.674,	1.277]	0.603	0.942	0.010
3.0	0.963	0.963	0.086	0.145	0.935	2.098	[0.671,	1.267]	0.596	0.942	0.010
4.0	0.958	0.956	0.078	0.143	0.929	2.079	[0.668,	1.259]	0.591	0.942	0.010
					$n = 400$						
0.0	1.045	1.030	−0.215	0.118	0.960	2.410	[0.795,	1.273]	0.478	0.940	0.005
1.0	1.006	0.993	−0.191	0.109	0.947	2.229	[0.774,	1.220]	0.446	0.939	0.005
2.0	0.999	0.986	−0.191	0.108	0.945	2.199	[0.770,	1.210]	0.440	0.937	0.005
3.0	0.995	0.980	−0.197	0.107	0.944	2.177	[0.766,	1.203]	0.436	0.936	0.005
4.0	0.990	0.978	−0.193	0.106	0.944	2.160	[0.764,	1.197]	0.433	0.936	0.005
					$n = 600$						
0.0	1.035	1.029	−0.059	0.097	0.972	2.423	[0.836,	1.229]	0.393	0.943	0.003
1.0	1.004	0.996	−0.088	0.091	0.965	2.264	[0.817,	1.185]	0.369	0.944	0.003
2.0	0.999	0.991	−0.087	0.090	0.963	2.237	[0.813,	1.177]	0.365	0.943	0.003
3.0	0.994	0.986	−0.093	0.089	0.960	2.217	[0.810,	1.172]	0.362	0.942	0.003
4.0	0.991	0.982	−0.096	0.089	0.958	2.201	[0.807,	1.167]	0.360	0.942	0.003
					$n = 800$						
0.0	1.033	1.025	−0.085	0.085	0.999	2.425	[0.860,	1.201]	0.341	0.954	0.003
1.0	1.006	0.997	−0.108	0.080	0.989	2.284	[0.842,	1.163]	0.321	0.954	0.003
2.0	1.001	0.992	−0.118	0.079	0.985	2.260	[0.838,	1.156]	0.318	0.955	0.002
3.0	0.998	0.988	−0.132	0.078	0.982	2.242	[0.835,	1.151]	0.316	0.952	0.002
4.0	0.995	0.985	−0.129	0.078	0.980	2.227	[0.833,	1.147]	0.314	0.952	0.002

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Functional misspecification. [3] Sieves dimension $k_n = 3$.

Table 25: $\hat{\Psi}_n$ Logistic Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf.	int.]	length	95% CP	Degree
					$n = 200$						
0.0	1.026	1.055	0.607	0.174	0.938	2.619	[0.716,	1.421]	0.705	0.947	0.010
1.0	0.977	1.004	0.486	0.159	0.952	2.330	[0.694,	1.328]	0.634	0.945	0.010
2.0	0.969	0.995	0.465	0.156	0.952	2.289	[0.690,	1.313]	0.623	0.945	0.010
3.0	0.963	0.989	0.451	0.155	0.952	2.259	[0.687,	1.303]	0.616	0.946	0.010
4.0	0.958	0.980	0.437	0.153	0.947	2.238	[0.684,	1.294]	0.610	0.949	0.010
					$n = 400$						
0.0	1.045	1.045	0.106	0.122	0.961	2.499	[0.808,	1.292]	0.484	0.946	0.005
1.0	1.006	1.005	0.090	0.114	0.965	2.334	[0.784,	1.237]	0.453	0.944	0.005
2.0	0.999	1.001	0.086	0.113	0.963	2.306	[0.780,	1.228]	0.448	0.942	0.005
3.0	0.995	0.995	0.078	0.112	0.964	2.284	[0.777,	1.220]	0.444	0.941	0.005
4.0	0.990	0.990	0.079	0.111	0.962	2.268	[0.774,	1.215]	0.441	0.942	0.005
					$n = 600$						
0.0	1.035	1.039	0.203	0.099	0.971	2.475	[0.847,	1.240]	0.393	0.946	0.003
1.0	1.004	1.005	0.146	0.094	0.979	2.339	[0.824,	1.196]	0.372	0.946	0.003
2.0	0.999	1.000	0.144	0.093	0.978	2.315	[0.820,	1.189]	0.368	0.948	0.003
3.0	0.994	0.995	0.137	0.093	0.979	2.298	[0.817,	1.183]	0.366	0.949	0.003
4.0	0.991	0.992	0.133	0.092	0.978	2.284	[0.815,	1.178]	0.363	0.949	0.003
					$n = 800$						
0.0	1.033	1.032	0.139	0.086	0.997	2.463	[0.868,	1.208]	0.340	0.951	0.003
1.0	1.006	1.004	0.094	0.082	0.999	2.346	[0.848,	1.171]	0.324	0.951	0.003
2.0	1.001	0.999	0.082	0.081	0.997	2.326	[0.844,	1.165]	0.321	0.951	0.002
3.0	0.998	0.995	0.069	0.081	0.997	2.310	[0.841,	1.160]	0.319	0.950	0.002
4.0	0.995	0.991	0.071	0.080	0.996	2.297	[0.839,	1.155]	0.317	0.950	0.002

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Functional misspecification.

Table 26: $\hat{\Psi}_n$ Poisson Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.026	1.030	0.164	0.165	0.960	2.415	[0.709,	1.366]	0.657	0.940	0.010
1.0	0.977	0.980	0.153	0.153	0.964	2.218	[0.685,	1.290]	0.606	0.940	0.010
2.0	0.969	0.970	0.146	0.150	0.961	2.187	[0.680,	1.278]	0.597	0.943	0.010
3.0	0.963	0.970	0.143	0.149	0.963	2.163	[0.677,	1.268]	0.591	0.942	0.010
4.0	0.958	0.960	0.137	0.148	0.960	2.146	[0.675,	1.261]	0.586	0.941	0.010
						$n = 400$					
0.0	1.045	1.030	−0.188	0.119	0.970	2.421	[0.801,	1.271]	0.470	0.939	0.005
1.0	1.006	0.995	−0.151	0.112	0.970	2.279	[0.777,	1.220]	0.443	0.939	0.005
2.0	0.999	0.990	−0.147	0.111	0.970	2.254	[0.773,	1.211]	0.438	0.938	0.005
3.0	0.995	0.985	−0.149	0.110	0.970	2.235	[0.770,	1.204]	0.434	0.938	0.005
4.0	0.990	0.980	−0.143	0.109	0.968	2.220	[0.767,	1.199]	0.432	0.939	0.005
						$n = 600$					
0.0	1.035	1.030	−0.034	0.097	0.977	2.425	[0.841,	1.226]	0.385	0.943	0.003
1.0	1.004	0.997	−0.054	0.093	0.982	2.302	[0.819,	1.185]	0.366	0.945	0.003
2.0	0.999	0.993	−0.049	0.092	0.982	2.280	[0.815,	1.178]	0.363	0.945	0.003
3.0	0.994	0.987	−0.053	0.091	0.982	2.264	[0.812,	1.172]	0.360	0.945	0.003
4.0	0.991	0.983	−0.053	0.091	0.981	2.251	[0.810,	1.168]	0.358	0.945	0.003
						$n = 800$					
0.0	1.033	1.025	−0.063	0.085	1.001	2.427	[0.864,	1.198]	0.335	0.950	0.003
1.0	1.006	0.998	−0.081	0.081	1.002	2.318	[0.843,	1.163]	0.320	0.955	0.003
2.0	1.001	0.993	−0.088	0.080	1.000	2.299	[0.840,	1.157]	0.317	0.953	0.002
3.0	0.998	0.990	−0.098	0.080	0.998	2.284	[0.837,	1.152]	0.315	0.953	0.002
4.0	0.995	0.985	−0.092	0.079	0.999	2.272	[0.834,	1.148]	0.314	0.951	0.002

[1] Number of Monte Carlo simulations is 2,000. [2] Local misspecification: Functional misspecification.

Table 27: $\hat{\Psi}_n$ Ratios

$\bar{\sigma}^2$	Logit/Robust				Poisson/Robust			
	s.e	bias	rMSE	length	s.e	bias	rMSE	length
				$n = 200$				
0.0	1.064	4.593	1.088	1.048	1.009	1.239	1.004	0.978
1.0	1.070	4.651	1.078	1.036	1.028	1.461	1.026	0.990
2.0	1.070	4.971	1.077	1.034	1.028	1.563	1.029	0.991
3.0	1.069	5.227	1.077	1.033	1.030	1.654	1.031	0.992
4.0	1.070	5.606	1.076	1.031	1.033	1.757	1.032	0.992
				$n = 400$				
0.0	1.036	0.491	1.037	1.012	1.009	0.872	1.005	0.982
1.0	1.048	0.468	1.047	1.017	1.025	0.791	1.022	0.994
2.0	1.048	0.447	1.048	1.017	1.028	0.769	1.025	0.995
3.0	1.049	0.395	1.049	1.017	1.028	0.758	1.026	0.995
4.0	1.048	0.411	1.050	1.017	1.028	0.741	1.028	0.996
				$n = 600$				
0.0	1.023	3.412	1.022	1.001	1.005	0.564	1.001	0.982
1.0	1.034	1.659	1.033	1.009	1.018	0.614	1.017	0.993
2.0	1.034	1.665	1.035	1.010	1.020	0.570	1.019	0.995
3.0	1.038	1.467	1.037	1.010	1.023	0.566	1.021	0.995
4.0	1.039	1.380	1.038	1.010	1.025	0.551	1.023	0.996
				$n = 800$				
0.0	1.014	1.642	1.016	0.997	1.001	0.740	1.001	0.983
1.0	1.025	0.867	1.027	1.007	1.014	0.747	1.015	0.995
2.0	1.027	0.695	1.029	1.007	1.016	0.742	1.017	0.996
3.0	1.029	0.521	1.030	1.008	1.017	0.745	1.019	0.997
4.0	1.030	0.555	1.031	1.008	1.020	0.718	1.020	0.997

[1] Number of Monte Carlo simulations is $2,000$.

[2] DGP Local misspecification: Functional misspecification.

[3] Sieves dimension $k_n = 3$.

Table 28: $\hat{\Psi}_n$ Robust Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
					$n = 200$						
0.0	1.026	1.024	0.132	0.163	0.949	2.407	[0.699,	1.371]	0.672	0.942	0.010
1.0	1.061	1.061	0.154	0.177	0.933	2.654	[0.704,	1.440]	0.736	0.936	0.011
2.0	1.071	1.071	0.158	0.181	0.934	2.716	[0.706,	1.458]	0.752	0.936	0.011
3.0	1.081	1.080	0.148	0.184	0.933	2.768	[0.709,	1.475]	0.766	0.935	0.011
4.0	1.089	1.090	0.170	0.188	0.935	2.813	[0.712,	1.489]	0.777	0.933	0.011
					$n = 400$						
0.0	1.045	1.030	−0.215	0.118	0.960	2.410	[0.795,	1.273]	0.478	0.940	0.005
1.0	1.070	1.050	−0.221	0.125	0.960	2.578	[0.804,	1.314]	0.510	0.938	0.005
2.0	1.076	1.056	−0.214	0.127	0.959	2.622	[0.806,	1.325]	0.519	0.940	0.005
3.0	1.081	1.064	−0.195	0.128	0.959	2.656	[0.809,	1.334]	0.525	0.939	0.005
4.0	1.087	1.069	−0.202	0.130	0.961	2.684	[0.812,	1.342]	0.531	0.938	0.005
					$n = 600$						
0.0	1.035	1.029	−0.059	0.097	0.972	2.423	[0.836,	1.229]	0.393	0.943	0.003
1.0	1.055	1.047	−0.073	0.102	0.969	2.564	[0.845,	1.260]	0.415	0.942	0.004
2.0	1.061	1.053	−0.080	0.104	0.965	2.595	[0.848,	1.267]	0.419	0.943	0.004
3.0	1.065	1.057	−0.079	0.104	0.963	2.620	[0.850,	1.273]	0.423	0.943	0.004
4.0	1.069	1.061	−0.090	0.105	0.963	2.642	[0.852,	1.279]	0.427	0.943	0.004
					$n = 800$						
0.0	1.033	1.025	−0.085	0.085	0.999	2.425	[0.860,	1.201]	0.341	0.954	0.003
1.0	1.050	1.040	−0.126	0.088	1.005	2.541	[0.868,	1.224]	0.357	0.955	0.003
2.0	1.054	1.045	−0.124	0.089	1.003	2.569	[0.870,	1.230]	0.360	0.956	0.003
3.0	1.058	1.048	−0.129	0.090	1.002	2.590	[0.872,	1.235]	0.363	0.954	0.003
4.0	1.062	1.052	−0.149	0.091	1.005	2.609	[0.874,	1.239]	0.366	0.955	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Semiparametric distribution. [3] Sieves dimension $k_n = 3$.

Table 29: $\hat{\Psi}_n$ Logistic Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.026	1.055	0.607	0.174	0.938	2.619	[0.716,	1.421]	0.705	0.947	0.010
1.0	1.061	1.101	0.775	0.193	0.913	3.030	[0.717,	1.515]	0.798	0.948	0.011
2.0	1.071	1.114	0.820	0.198	0.908	3.149	[0.717,	1.541]	0.824	0.950	0.011
3.0	1.081	1.124	0.846	0.201	0.900	3.252	[0.718,	1.564]	0.845	0.949	0.011
4.0	1.089	1.133	0.900	0.206	0.901	3.348	[0.720,	1.585]	0.865	0.949	0.011
						$n = 400$					
0.0	1.045	1.045	0.106	0.122	0.961	2.499	[0.808,	1.292]	0.484	0.946	0.005
1.0	1.070	1.068	0.161	0.131	0.962	2.701	[0.817,	1.339]	0.522	0.948	0.005
2.0	1.076	1.075	0.185	0.133	0.960	2.753	[0.820,	1.351]	0.532	0.946	0.005
3.0	1.081	1.083	0.217	0.135	0.963	2.794	[0.823,	1.362]	0.539	0.946	0.005
4.0	1.087	1.087	0.221	0.136	0.964	2.829	[0.825,	1.371]	0.546	0.945	0.005
						$n = 600$					
0.0	1.035	1.039	0.203	0.099	0.971	2.475	[0.847,	1.240]	0.393	0.946	0.003
1.0	1.055	1.058	0.228	0.105	0.967	2.631	[0.856,	1.273]	0.417	0.948	0.004
2.0	1.061	1.065	0.230	0.106	0.964	2.666	[0.859,	1.281]	0.423	0.947	0.004
3.0	1.065	1.069	0.238	0.107	0.961	2.695	[0.861,	1.288]	0.427	0.944	0.004
4.0	1.069	1.074	0.235	0.108	0.961	2.720	[0.863,	1.294]	0.431	0.946	0.004
						$n = 800$					
0.0	1.033	1.032	0.139	0.086	0.997	2.463	[0.868,	1.208]	0.340	0.951	0.003
1.0	1.050	1.048	0.126	0.090	1.006	2.588	[0.876,	1.233]	0.357	0.957	0.003
2.0	1.054	1.053	0.135	0.091	1.005	2.618	[0.879,	1.239]	0.361	0.954	0.003
3.0	1.058	1.058	0.136	0.092	1.004	2.641	[0.881,	1.245]	0.364	0.953	0.003
4.0	1.062	1.061	0.120	0.093	1.007	2.661	[0.883,	1.249]	0.367	0.950	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Semiparametric distribution.

Table 30: $\hat{\Psi}_n$ Poisson Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.026	1.030	0.164	0.165	0.960	2.415	[0.709,	1.366]	0.657	0.940	0.010
1.0	1.061	1.060	0.182	0.179	0.944	2.658	[0.714,	1.434]	0.720	0.934	0.011
2.0	1.071	1.070	0.185	0.183	0.944	2.719	[0.716,	1.452]	0.736	0.933	0.011
3.0	1.081	1.080	0.174	0.186	0.943	2.771	[0.719,	1.468]	0.750	0.932	0.011
4.0	1.089	1.090	0.195	0.189	0.945	2.817	[0.721,	1.483]	0.762	0.931	0.011
						$n = 400$					
0.0	1.045	1.030	-0.188	0.119	0.970	2.421	[0.801,	1.271]	0.470	0.939	0.005
1.0	1.070	1.050	-0.195	0.127	0.972	2.589	[0.809,	1.311]	0.502	0.935	0.005
2.0	1.076	1.060	-0.188	0.128	0.970	2.631	[0.812,	1.322]	0.510	0.938	0.005
3.0	1.081	1.065	-0.170	0.130	0.973	2.664	[0.815,	1.331]	0.516	0.935	0.005
4.0	1.087	1.070	-0.177	0.131	0.974	2.693	[0.817,	1.339]	0.521	0.937	0.005
						$n = 600$					
0.0	1.035	1.030	-0.034	0.097	0.977	2.425	[0.841,	1.226]	0.385	0.943	0.003
1.0	1.055	1.047	-0.049	0.103	0.973	2.564	[0.850,	1.257]	0.407	0.942	0.004
2.0	1.061	1.053	-0.057	0.104	0.969	2.594	[0.852,	1.264]	0.412	0.942	0.004
3.0	1.065	1.057	-0.057	0.105	0.967	2.619	[0.855,	1.270]	0.416	0.941	0.004
4.0	1.069	1.063	-0.067	0.106	0.966	2.641	[0.856,	1.276]	0.419	0.942	0.004
						$n = 800$					
0.0	1.033	1.025	-0.063	0.085	1.001	2.427	[0.864,	1.198]	0.335	0.950	0.003
1.0	1.050	1.040	-0.105	0.089	1.010	2.542	[0.871,	1.222]	0.350	0.951	0.003
2.0	1.054	1.045	-0.103	0.090	1.008	2.569	[0.874,	1.228]	0.354	0.949	0.003
3.0	1.058	1.050	-0.108	0.090	1.007	2.590	[0.876,	1.233]	0.357	0.949	0.003
4.0	1.062	1.053	-0.128	0.091	1.010	2.609	[0.877,	1.237]	0.360	0.952	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Semiparametric distribution.

Table 31: $\hat{\Psi}_n$ Ratios

$\bar{\sigma}^2$	Logit/Robust				Poisson/Robust			
	s.e	bias	rMSE	length	s.e	bias	rMSE	length
				$n = 200$				
0.0	1.064	4.593	1.088	1.048	1.009	1.239	1.004	0.978
1.0	1.088	5.033	1.142	1.085	1.010	1.181	1.002	0.979
2.0	1.092	5.189	1.159	1.095	1.009	1.168	1.001	0.979
3.0	1.092	5.703	1.175	1.104	1.009	1.175	1.001	0.979
4.0	1.098	5.295	1.190	1.112	1.008	1.146	1.002	0.980
				$n = 400$				
0.0	1.036	0.491	1.037	1.012	1.009	0.872	1.005	0.982
1.0	1.046	0.727	1.048	1.023	1.012	0.882	1.004	0.983
2.0	1.047	0.863	1.050	1.025	1.011	0.880	1.003	0.983
3.0	1.052	1.112	1.052	1.027	1.014	0.871	1.003	0.983
4.0	1.052	1.094	1.054	1.029	1.013	0.878	1.003	0.983
				$n = 600$				
0.0	1.023	3.412	1.022	1.001	1.005	0.564	1.001	0.982
1.0	1.027	3.131	1.026	1.006	1.005	0.674	1.000	0.982
2.0	1.028	2.879	1.027	1.008	1.004	0.709	1.000	0.982
3.0	1.029	2.994	1.028	1.009	1.004	0.711	1.000	0.982
4.0	1.029	2.621	1.029	1.010	1.003	0.746	1.000	0.982
				$n = 800$				
0.0	1.014	1.642	1.016	0.997	1.001	0.740	1.001	0.983
1.0	1.021	1.000	1.018	1.000	1.005	0.830	1.000	0.983
2.0	1.022	1.091	1.019	1.001	1.005	0.829	1.000	0.983
3.0	1.023	1.055	1.020	1.001	1.005	0.837	1.000	0.983
4.0	1.023	0.808	1.020	1.002	1.005	0.859	1.000	0.983

[1] Number of Monte Carlo simulations is 2,000.

[2] DGP Local misspecification: Semiparametric distribution.

[3] Sieves dimension $k_n = 3$.

E.3 Additional Simulations with Initial Poisson Estimator

Table 32: $\hat{\beta}_n$ Robust Estimator with $\hat{\theta}_{\text{initial}}$ Poisson

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.386	1.385	−0.064	0.147	0.885	2.165	[0.651,	2.112]	1.461	1.000	0.010
1.0	1.386	1.383	−0.075	0.146	0.885	2.166	[0.651,	2.111]	1.460	1.000	0.010
2.0	1.386	1.382	−0.074	0.147	0.884	2.167	[0.651,	2.111]	1.461	1.000	0.010
3.0	1.386	1.381	−0.071	0.147	0.884	2.163	[0.651,	2.111]	1.460	1.000	0.010
4.0	1.386	1.381	−0.066	0.146	0.882	2.155	[0.653,	2.111]	1.458	1.000	0.010
						$n = 400$					
0.0	1.386	1.388	0.049	0.108	0.935	2.228	[0.859,	1.919]	1.060	1.000	0.005
1.0	1.386	1.389	0.054	0.108	0.933	2.227	[0.859,	1.919]	1.060	1.000	0.005
2.0	1.386	1.390	0.051	0.108	0.932	2.227	[0.859,	1.919]	1.059	1.000	0.005
3.0	1.386	1.389	0.054	0.108	0.932	2.225	[0.859,	1.919]	1.059	1.000	0.005
4.0	1.386	1.391	0.053	0.108	0.927	2.224	[0.860,	1.918]	1.059	1.000	0.005
						$n = 600$					
0.0	1.386	1.385	0.022	0.090	0.914	2.254	[0.952,	1.822]	0.870	1.000	0.003
1.0	1.386	1.385	0.018	0.090	0.917	2.254	[0.952,	1.822]	0.870	1.000	0.003
2.0	1.386	1.387	0.019	0.090	0.916	2.255	[0.952,	1.822]	0.870	1.000	0.003
3.0	1.386	1.386	0.024	0.090	0.915	2.255	[0.952,	1.822]	0.870	1.000	0.003
4.0	1.386	1.387	0.015	0.090	0.916	2.255	[0.952,	1.822]	0.870	1.000	0.003
						$n = 800$					
0.0	1.386	1.382	−0.035	0.079	0.930	2.288	[1.005,	1.765]	0.760	1.000	0.003
1.0	1.386	1.383	−0.037	0.079	0.931	2.288	[1.005,	1.765]	0.760	1.000	0.003
2.0	1.386	1.383	−0.036	0.079	0.933	2.288	[1.005,	1.765]	0.760	1.000	0.003
3.0	1.386	1.383	−0.040	0.079	0.931	2.287	[1.005,	1.765]	0.760	1.000	0.003
4.0	1.386	1.382	−0.047	0.079	0.934	2.286	[1.005,	1.764]	0.759	1.000	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Latent Homophily. [3] Sieves dimension $k_n = 2$.

Table 33: $\hat{\beta}_n$ Robust Estimator with $\hat{\theta}_{\text{initial}}$ Poisson

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
					$n = 200$						
0.0	1.386	1.385	-0.025	0.146	0.878	2.146	[0.657,	2.112]	1.455	0.999	0.010
1.0	1.386	1.687	4.216	0.145	0.875	2.158	[0.811,	2.558]	1.747	1.000	0.011
2.0	1.386	1.774	5.428	0.145	0.887	2.172	[0.842,	2.699]	1.857	1.000	0.012
3.0	1.386	1.841	6.395	0.145	0.885	2.200	[0.859,	2.818]	1.960	1.000	0.013
4.0	1.386	1.888	7.038	0.147	0.900	2.252	[0.861,	2.907]	2.045	1.000	0.014
					$n = 400$						
0.0	1.386	1.388	0.097	0.109	0.917	2.226	[0.862,	1.921]	1.059	1.000	0.005
1.0	1.386	1.599	4.274	0.108	0.903	2.227	[0.998,	2.202]	1.203	1.000	0.005
2.0	1.386	1.663	5.508	0.108	0.894	2.241	[1.033,	2.290]	1.257	1.000	0.006
3.0	1.386	1.724	6.667	0.108	0.880	2.261	[1.063,	2.376]	1.313	1.000	0.006
4.0	1.386	1.783	7.822	0.108	0.874	2.267	[1.092,	2.463]	1.371	0.999	0.006
					$n = 600$						
0.0	1.386	1.386	-0.001	0.090	0.928	2.246	[0.951,	1.821]	0.870	1.000	0.003
1.0	1.386	1.557	4.186	0.090	0.919	2.265	[1.073,	2.042]	0.969	1.000	0.004
2.0	1.386	1.604	5.342	0.090	0.913	2.285	[1.103,	2.106]	1.003	1.000	0.004
3.0	1.386	1.649	6.477	0.091	0.904	2.310	[1.131,	2.171]	1.040	1.000	0.004
4.0	1.386	1.698	7.652	0.092	0.896	2.339	[1.158,	2.239]	1.081	0.997	0.004
					$n = 800$						
0.0	1.386	1.387	-0.039	0.079	0.934	2.274	[1.006,	1.763]	0.757	1.000	0.003
1.0	1.386	1.534	4.113	0.079	0.918	2.297	[1.117,	1.947]	0.830	1.000	0.003
2.0	1.386	1.573	5.242	0.079	0.908	2.316	[1.144,	1.999]	0.855	1.000	0.003
3.0	1.386	1.613	6.330	0.080	0.911	2.351	[1.169,	2.051]	0.882	0.999	0.003
4.0	1.386	1.651	7.433	0.082	0.917	2.393	[1.193,	2.105]	0.912	0.999	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Functional misspecification. [3] Sieves dimension $k_n = 2$.

Table 34: $\hat{\beta}_n$ Robust Estimator with $\hat{\theta}_{\text{initial}}$ Poisson

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
					$n = 200$						
0.0	1.386	1.385	−0.025	0.146	0.878	2.146	[0.657,	2.112]	1.455	0.999	0.010
1.0	1.386	1.529	1.990	0.146	0.881	2.141	[0.738,	2.316]	1.577	1.000	0.011
2.0	1.386	1.556	2.387	0.146	0.884	2.145	[0.752,	2.358]	1.605	1.000	0.011
3.0	1.386	1.580	2.716	0.146	0.879	2.145	[0.764,	2.393]	1.629	1.000	0.011
4.0	1.386	1.595	2.974	0.146	0.878	2.144	[0.773,	2.420]	1.647	1.000	0.011
					$n = 400$						
0.0	1.386	1.388	0.097	0.109	0.917	2.226	[0.862,	1.921]	1.059	1.000	0.005
1.0	1.386	1.487	2.097	0.108	0.908	2.219	[0.930,	2.052]	1.122	1.000	0.005
2.0	1.386	1.509	2.508	0.108	0.908	2.215	[0.944,	2.079]	1.135	1.000	0.005
3.0	1.386	1.523	2.832	0.108	0.906	2.214	[0.955,	2.101]	1.146	1.000	0.005
4.0	1.386	1.538	3.103	0.108	0.906	2.211	[0.964,	2.119]	1.155	1.000	0.005
					$n = 600$						
0.0	1.386	1.386	−0.001	0.090	0.928	2.246	[0.951,	1.821]	0.870	1.000	0.003
1.0	1.386	1.467	2.005	0.090	0.930	2.245	[1.011,	1.925]	0.914	1.000	0.004
2.0	1.386	1.485	2.434	0.090	0.925	2.246	[1.024,	1.947]	0.923	1.000	0.004
3.0	1.386	1.499	2.743	0.090	0.923	2.246	[1.033,	1.964]	0.931	1.000	0.004
4.0	1.386	1.509	3.014	0.089	0.920	2.246	[1.041,	1.978]	0.937	1.000	0.004
					$n = 800$						
0.0	1.386	1.387	−0.039	0.079	0.934	2.274	[1.006,	1.763]	0.757	1.000	0.003
1.0	1.386	1.458	1.924	0.078	0.928	2.280	[1.060,	1.849]	0.789	1.000	0.003
2.0	1.386	1.472	2.337	0.078	0.924	2.282	[1.071,	1.867]	0.796	1.000	0.003
3.0	1.386	1.483	2.659	0.078	0.924	2.283	[1.079,	1.881]	0.802	1.000	0.003
4.0	1.386	1.493	2.925	0.078	0.920	2.285	[1.086,	1.893]	0.807	1.000	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Semiparametric distribution. [3] Sieves dimension $k_n = 2$.

Table 35: $\hat{\Psi}_n$ Robust Estimator with $\hat{\theta}_{\text{initial}}$ Poisson

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
					$n = 200$						
0.0	1.041	1.021	−0.118	0.162	0.950	2.406	[0.697,	1.368]	0.672	0.931	0.010
1.0	1.041	1.021	−0.134	0.162	0.950	2.405	[0.696,	1.368]	0.672	0.930	0.010
2.0	1.041	1.020	−0.126	0.162	0.950	2.407	[0.696,	1.368]	0.672	0.928	0.010
3.0	1.043	1.020	−0.135	0.162	0.955	2.409	[0.697,	1.370]	0.672	0.927	0.010
4.0	1.048	1.026	−0.163	0.163	0.957	2.414	[0.700,	1.374]	0.674	0.926	0.010
					$n = 400$						
0.0	1.029	1.029	0.129	0.117	0.955	2.417	[0.796,	1.276]	0.479	0.945	0.005
1.0	1.030	1.029	0.118	0.118	0.957	2.416	[0.796,	1.275]	0.479	0.945	0.005
2.0	1.030	1.029	0.114	0.118	0.956	2.416	[0.796,	1.275]	0.479	0.942	0.005
3.0	1.031	1.029	0.110	0.118	0.955	2.418	[0.796,	1.276]	0.480	0.942	0.005
4.0	1.033	1.032	0.112	0.118	0.953	2.422	[0.798,	1.279]	0.480	0.944	0.005
					$n = 600$						
0.0	1.040	1.029	−0.189	0.097	0.984	2.417	[0.837,	1.229]	0.392	0.942	0.003
1.0	1.040	1.028	−0.194	0.097	0.986	2.416	[0.836,	1.228]	0.392	0.942	0.003
2.0	1.040	1.030	−0.194	0.097	0.987	2.417	[0.836,	1.228]	0.392	0.942	0.003
3.0	1.040	1.030	−0.184	0.097	0.989	2.418	[0.837,	1.229]	0.392	0.943	0.003
4.0	1.042	1.030	−0.186	0.097	0.987	2.421	[0.838,	1.231]	0.393	0.943	0.003
					$n = 800$						
0.0	1.033	1.030	−0.013	0.084	0.965	2.427	[0.862,	1.203]	0.340	0.946	0.003
1.0	1.033	1.030	0.003	0.084	0.965	2.426	[0.862,	1.203]	0.340	0.947	0.003
2.0	1.032	1.030	0.005	0.084	0.964	2.426	[0.862,	1.203]	0.340	0.946	0.003
3.0	1.033	1.030	0.011	0.084	0.963	2.427	[0.863,	1.203]	0.340	0.947	0.003
4.0	1.034	1.032	0.008	0.084	0.964	2.429	[0.864,	1.205]	0.341	0.947	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Latent Homophily. [3] Sieves dimension $k_n = 3$.

Table 36: $\hat{\Psi}_n$ Robust Estimator with $\hat{\theta}_{\text{initial}}$ Poisson

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.026	1.024	0.132	0.163	0.949	2.407	[0.699,	1.371]	0.672	0.942	0.010
1.0	0.977	0.978	0.105	0.149	0.937	2.162	[0.678,	1.290]	0.612	0.939	0.010
2.0	0.969	0.969	0.093	0.146	0.935	2.125	[0.674,	1.277]	0.603	0.942	0.010
3.0	0.963	0.963	0.086	0.145	0.935	2.098	[0.671,	1.267]	0.596	0.942	0.010
4.0	0.958	0.956	0.078	0.143	0.929	2.079	[0.668,	1.259]	0.591	0.942	0.010
						$n = 400$					
0.0	1.045	1.030	−0.215	0.118	0.960	2.410	[0.795,	1.273]	0.478	0.940	0.005
1.0	1.006	0.993	−0.191	0.109	0.947	2.229	[0.774,	1.220]	0.446	0.939	0.005
2.0	0.999	0.986	−0.191	0.108	0.945	2.199	[0.770,	1.210]	0.440	0.937	0.005
3.0	0.995	0.980	−0.197	0.107	0.944	2.177	[0.766,	1.203]	0.436	0.936	0.005
4.0	0.990	0.978	−0.193	0.106	0.944	2.160	[0.764,	1.197]	0.433	0.936	0.005
						$n = 600$					
0.0	1.035	1.029	−0.059	0.097	0.972	2.423	[0.836,	1.229]	0.393	0.943	0.003
1.0	1.004	0.996	−0.088	0.091	0.965	2.264	[0.817,	1.185]	0.369	0.944	0.003
2.0	0.999	0.991	−0.087	0.090	0.963	2.237	[0.813,	1.177]	0.365	0.943	0.003
3.0	0.994	0.986	−0.093	0.089	0.960	2.217	[0.810,	1.172]	0.362	0.942	0.003
4.0	0.991	0.982	−0.096	0.089	0.958	2.201	[0.807,	1.167]	0.360	0.942	0.003
						$n = 800$					
0.0	1.033	1.025	−0.085	0.085	0.999	2.425	[0.860,	1.201]	0.341	0.954	0.003
1.0	1.006	0.997	−0.108	0.080	0.989	2.284	[0.842,	1.163]	0.321	0.954	0.003
2.0	1.001	0.992	−0.118	0.079	0.985	2.260	[0.838,	1.156]	0.318	0.955	0.002
3.0	0.998	0.988	−0.132	0.078	0.982	2.242	[0.835,	1.151]	0.316	0.952	0.002
4.0	0.995	0.985	−0.129	0.078	0.980	2.227	[0.833,	1.147]	0.314	0.952	0.002

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Functional misspecification. [3] Sieves dimension $k_n = 3$.

Table 37: $\hat{\Psi}_n$ Robust Estimator with $\hat{\theta}_{\text{initial}}$ Poisson

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf.	int.]	length	95% CP	Degree
					$n = 200$						
0.0	1.026	1.024	0.132	0.163	0.949	2.407	[0.699,	1.371]	0.672	0.942	0.010
1.0	1.061	1.061	0.154	0.177	0.933	2.654	[0.704,	1.440]	0.736	0.936	0.011
2.0	1.071	1.071	0.158	0.181	0.934	2.716	[0.706,	1.458]	0.752	0.936	0.011
3.0	1.081	1.080	0.148	0.184	0.933	2.768	[0.709,	1.475]	0.766	0.935	0.011
4.0	1.089	1.090	0.170	0.188	0.935	2.813	[0.712,	1.489]	0.777	0.933	0.011
					$n = 400$						
0.0	1.045	1.030	-0.215	0.118	0.960	2.410	[0.795,	1.273]	0.478	0.940	0.005
1.0	1.070	1.050	-0.221	0.125	0.960	2.578	[0.804,	1.314]	0.510	0.938	0.005
2.0	1.076	1.056	-0.214	0.127	0.959	2.622	[0.806,	1.325]	0.519	0.940	0.005
3.0	1.081	1.064	-0.195	0.128	0.959	2.656	[0.809,	1.334]	0.525	0.939	0.005
4.0	1.087	1.069	-0.202	0.130	0.961	2.684	[0.812,	1.342]	0.531	0.938	0.005
					$n = 600$						
0.0	1.035	1.029	-0.059	0.097	0.972	2.423	[0.836,	1.229]	0.393	0.943	0.003
1.0	1.055	1.047	-0.073	0.102	0.969	2.564	[0.845,	1.260]	0.415	0.942	0.004
2.0	1.061	1.053	-0.080	0.104	0.965	2.595	[0.848,	1.267]	0.419	0.943	0.004
3.0	1.065	1.057	-0.079	0.104	0.963	2.620	[0.850,	1.273]	0.423	0.943	0.004
4.0	1.069	1.061	-0.090	0.105	0.963	2.642	[0.852,	1.279]	0.427	0.943	0.004
					$n = 800$						
0.0	1.033	1.025	-0.085	0.085	0.999	2.425	[0.860,	1.201]	0.341	0.954	0.003
1.0	1.050	1.040	-0.126	0.088	1.005	2.541	[0.868,	1.224]	0.357	0.955	0.003
2.0	1.054	1.045	-0.124	0.089	1.003	2.569	[0.870,	1.230]	0.360	0.956	0.003
3.0	1.058	1.048	-0.129	0.090	1.002	2.590	[0.872,	1.235]	0.363	0.954	0.003
4.0	1.062	1.052	-0.149	0.091	1.005	2.609	[0.874,	1.239]	0.366	0.955	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Semiparametric distribution. [3] Sieves dimension $k_n = 3$.

E.4 Additional Simulations with Larger Sieves Dimension

Table 38: $\hat{\beta}_n$ Robust Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.386	1.415	0.418	0.136	0.764	1.991	[0.706,	2.125]	1.419	1.000	0.010
1.0	1.386	1.414	0.398	0.136	0.764	1.992	[0.705,	2.123]	1.418	1.000	0.010
2.0	1.386	1.415	0.406	0.137	0.761	1.993	[0.708,	2.122]	1.414	1.000	0.010
3.0	1.386	1.417	0.417	0.137	0.765	1.994	[0.712,	2.120]	1.408	0.999	0.010
4.0	1.386	1.415	0.408	0.136	0.768	1.993	[0.715,	2.115]	1.400	1.000	0.010
						$n = 400$					
0.0	1.386	1.408	0.351	0.103	0.854	2.104	[0.902,	1.906]	1.003	0.999	0.005
1.0	1.386	1.408	0.355	0.103	0.854	2.104	[0.903,	1.905]	1.003	0.999	0.005
2.0	1.386	1.406	0.348	0.103	0.854	2.103	[0.903,	1.905]	1.002	0.999	0.005
3.0	1.386	1.407	0.348	0.103	0.852	2.101	[0.903,	1.904]	1.001	0.999	0.005
4.0	1.386	1.408	0.345	0.103	0.854	2.103	[0.905,	1.902]	0.997	0.999	0.005
						$n = 600$					
0.0	1.386	1.393	0.212	0.087	0.878	2.154	[0.986,	1.804]	0.817	1.000	0.003
1.0	1.386	1.393	0.208	0.087	0.879	2.155	[0.986,	1.803]	0.817	1.000	0.003
2.0	1.386	1.393	0.205	0.087	0.877	2.155	[0.987,	1.802]	0.816	1.000	0.003
3.0	1.386	1.392	0.208	0.087	0.877	2.155	[0.987,	1.802]	0.815	1.000	0.003
4.0	1.386	1.392	0.205	0.087	0.877	2.156	[0.988,	1.801]	0.813	1.000	0.003
						$n = 800$					
0.0	1.386	1.390	0.113	0.077	0.903	2.189	[1.039,	1.742]	0.702	0.999	0.003
1.0	1.386	1.389	0.108	0.077	0.902	2.189	[1.039,	1.741]	0.702	0.999	0.003
2.0	1.386	1.390	0.117	0.077	0.903	2.189	[1.039,	1.741]	0.702	0.999	0.003
3.0	1.386	1.389	0.116	0.077	0.903	2.189	[1.040,	1.741]	0.702	0.999	0.003
4.0	1.386	1.388	0.099	0.077	0.910	2.189	[1.039,	1.740]	0.701	0.999	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Latent Homophily. [3] Sieves dimension $k_n = 4$.

Table 39: $\hat{\beta}_n$ Robust Estimator

$\overline{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.386	1.441	0.810	0.128	0.672	1.894	[0.740,	2.147]	1.407	0.997	0.010
1.0	1.386	1.438	0.790	0.128	0.673	1.894	[0.739,	2.145]	1.406	0.997	0.010
2.0	1.386	1.437	0.790	0.129	0.670	1.896	[0.741,	2.143]	1.401	0.997	0.010
3.0	1.386	1.438	0.801	0.129	0.676	1.896	[0.745,	2.140]	1.395	0.997	0.010
4.0	1.386	1.438	0.788	0.128	0.678	1.894	[0.748,	2.136]	1.388	0.997	0.010
						$n = 400$					
0.0	1.386	1.422	0.701	0.099	0.790	2.037	[0.920,	1.923]	1.003	0.998	0.005
1.0	1.386	1.420	0.704	0.099	0.789	2.037	[0.920,	1.923]	1.003	0.998	0.005
2.0	1.386	1.420	0.697	0.099	0.788	2.037	[0.920,	1.922]	1.002	0.998	0.005
3.0	1.386	1.421	0.697	0.100	0.788	2.036	[0.921,	1.921]	1.000	0.998	0.005
4.0	1.386	1.424	0.689	0.100	0.791	2.039	[0.922,	1.919]	0.997	0.999	0.005
						$n = 600$					
0.0	1.386	1.403	0.456	0.085	0.824	2.108	[0.994,	1.815]	0.821	0.999	0.003
1.0	1.386	1.403	0.455	0.085	0.824	2.108	[0.995,	1.815]	0.820	0.999	0.003
2.0	1.386	1.402	0.450	0.085	0.823	2.108	[0.995,	1.814]	0.819	0.999	0.003
3.0	1.386	1.402	0.452	0.085	0.823	2.108	[0.996,	1.814]	0.818	0.999	0.003
4.0	1.386	1.402	0.450	0.085	0.822	2.108	[0.997,	1.813]	0.816	0.998	0.003
						$n = 800$					
0.0	1.386	1.395	0.323	0.075	0.863	2.147	[1.043,	1.752]	0.709	0.999	0.003
1.0	1.386	1.395	0.319	0.075	0.865	2.147	[1.043,	1.752]	0.709	0.999	0.003
2.0	1.386	1.395	0.328	0.075	0.865	2.147	[1.044,	1.752]	0.709	0.999	0.003
3.0	1.386	1.395	0.327	0.076	0.866	2.147	[1.044,	1.752]	0.708	0.999	0.003
4.0	1.386	1.394	0.307	0.075	0.871	2.146	[1.043,	1.751]	0.707	0.999	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Latent Homophily. [3] Sieves dimension $k_n = 5$.

Table 40: $\hat{\Psi}_n$ Robust Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
						$n = 200$					
0.0	1.041	1.018	−0.202	0.163	0.956	2.399	[0.687,	1.366]	0.680	0.935	0.010
1.0	1.041	1.017	−0.220	0.162	0.952	2.398	[0.686,	1.365]	0.679	0.933	0.010
2.0	1.041	1.018	−0.216	0.162	0.953	2.400	[0.686,	1.366]	0.680	0.928	0.010
3.0	1.043	1.015	−0.223	0.163	0.955	2.403	[0.687,	1.367]	0.680	0.926	0.010
4.0	1.048	1.019	−0.249	0.163	0.958	2.410	[0.690,	1.372]	0.682	0.924	0.010
						$n = 400$					
0.0	1.029	1.025	0.061	0.117	0.954	2.411	[0.791,	1.274]	0.483	0.946	0.005
1.0	1.030	1.025	0.051	0.117	0.955	2.409	[0.791,	1.274]	0.483	0.947	0.005
2.0	1.030	1.025	0.046	0.117	0.955	2.408	[0.791,	1.274]	0.483	0.944	0.005
3.0	1.031	1.025	0.038	0.117	0.954	2.410	[0.791,	1.274]	0.483	0.942	0.005
4.0	1.033	1.028	0.042	0.117	0.953	2.418	[0.793,	1.277]	0.485	0.945	0.005
						$n = 600$					
0.0	1.040	1.026	−0.223	0.097	0.984	2.412	[0.834,	1.228]	0.394	0.941	0.003
1.0	1.040	1.025	−0.228	0.097	0.985	2.411	[0.834,	1.228]	0.394	0.941	0.003
2.0	1.040	1.027	−0.229	0.097	0.987	2.411	[0.834,	1.228]	0.394	0.940	0.003
3.0	1.040	1.027	−0.220	0.097	0.987	2.411	[0.834,	1.229]	0.394	0.943	0.003
4.0	1.042	1.027	−0.220	0.097	0.987	2.416	[0.836,	1.231]	0.395	0.944	0.003
						$n = 800$					
0.0	1.033	1.032	0.025	0.084	0.970	2.428	[0.862,	1.206]	0.344	0.949	0.003
1.0	1.033	1.033	0.041	0.084	0.973	2.428	[0.862,	1.206]	0.343	0.950	0.003
2.0	1.032	1.032	0.043	0.084	0.971	2.428	[0.862,	1.206]	0.343	0.949	0.003
3.0	1.033	1.032	0.047	0.084	0.971	2.428	[0.862,	1.206]	0.344	0.949	0.003
4.0	1.034	1.034	0.043	0.084	0.971	2.430	[0.864,	1.207]	0.344	0.946	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Latent Homophily. [3] Sieves dimension $k_n = 4$.

Table 41: $\hat{\Psi}_n$ Robust Estimator

$\bar{\sigma}^2$	True	coeff.	$\sqrt{n}$-Bias	s.e.	s.e./sd	rMSE	[conf. int.]		length	95% CP	Degree
					$n = 200$						
0.0	1.041	1.017	-0.219	0.163	0.951	2.423	[0.679,	1.371]	0.692	0.934	0.010
1.0	1.041	1.016	-0.236	0.163	0.953	2.427	[0.678,	1.371]	0.693	0.927	0.010
2.0	1.041	1.015	-0.233	0.163	0.948	2.433	[0.678,	1.371]	0.694	0.927	0.010
3.0	1.043	1.015	-0.241	0.163	0.950	2.432	[0.679,	1.372]	0.693	0.925	0.010
4.0	1.048	1.019	-0.269	0.163	0.949	2.442	[0.681,	1.378]	0.697	0.924	0.010
					$n = 400$						
0.0	1.029	1.024	0.042	0.117	0.952	2.436	[0.786,	1.277]	0.491	0.946	0.005
1.0	1.030	1.023	0.032	0.117	0.953	2.434	[0.786,	1.277]	0.491	0.947	0.005
2.0	1.030	1.023	0.028	0.117	0.953	2.434	[0.786,	1.277]	0.491	0.947	0.005
3.0	1.031	1.024	0.020	0.117	0.951	2.434	[0.786,	1.277]	0.491	0.943	0.005
4.0	1.033	1.026	0.023	0.117	0.948	2.436	[0.788,	1.280]	0.492	0.947	0.005
					$n = 600$						
0.0	1.040	1.026	-0.229	0.097	0.980	2.418	[0.832,	1.230]	0.398	0.945	0.003
1.0	1.040	1.025	-0.235	0.097	0.980	2.416	[0.832,	1.230]	0.398	0.941	0.003
2.0	1.040	1.027	-0.236	0.097	0.983	2.417	[0.832,	1.230]	0.398	0.943	0.003
3.0	1.040	1.026	-0.227	0.097	0.983	2.417	[0.832,	1.230]	0.398	0.943	0.003
4.0	1.042	1.027	-0.227	0.097	0.985	2.420	[0.834,	1.232]	0.398	0.945	0.003
					$n = 800$						
0.0	1.033	1.032	0.021	0.084	0.967	2.433	[0.861,	1.207]	0.346	0.949	0.003
1.0	1.033	1.032	0.036	0.084	0.970	2.435	[0.861,	1.207]	0.346	0.950	0.003
2.0	1.032	1.032	0.038	0.084	0.968	2.436	[0.861,	1.207]	0.346	0.949	0.003
3.0	1.033	1.032	0.042	0.084	0.968	2.437	[0.861,	1.207]	0.347	0.949	0.003
4.0	1.034	1.034	0.037	0.084	0.970	2.439	[0.862,	1.209]	0.347	0.946	0.003

[1] Number of Monte Carlo simulations is $2,000$. [2] Local misspecification: Latent Homophily. [3] Sieves dimension $k_n = 5$.

References

ANDERSON, K. A. AND S. RICHARDS-SHUBIK (2022): "Collaborative production in science: An empirical analysis of coauthorships in economics," *Review of Economics and Statistics*, 104, 1241–1255.

ANDREWS, I., M. GENTZKOW, AND J. M. SHAPIRO (2017): "Measuring the sensitivity of parameter estimates to estimation moments," *The Quarterly Journal of Economics*, 132, 1553–1592.

——— (2020): "On the informativeness of descriptive statistics for structural estimates," *Econometrica*, 88, 2231–2258.

ARMSTRONG, T. B., P. KLINE, AND L. SUN (2023): "Adapting to Misspecification," *arXiv preprint arXiv:2305.14265*.

ARMSTRONG, T. B. AND M. KOLESÁR (2021): "Sensitivity analysis using approximate moment condition models," *Quantitative Economics*, 12, 77–108.

ARMSTRONG, T. B., M. WEIDNER, AND A. ZELENEEV (2022): "Robust Estimation and Inference in Panels with Interactive Fixed Effects," *arXiv preprint arXiv:2210.06639*.

BONHOMME, S. (2021): "Teams: Heterogeneity, Sorting, and Complementarity," *arXiv preprint arXiv:2102.01802*.

BONHOMME, S. AND M. WEIDNER (2022): "Minimizing sensitivity to model misspecification," *Quantitative Economics*, 13, 907–954.

BOSCHINI, A. AND A. SJÖGREN (2007): "Is team formation gender neutral? Evidence from coauthorship patterns," *Journal of Labor Economics*, 25, 325–365.

BUGNI, F. A., I. A. CANAY, AND P. GUGGENBERGER (2012): "Distortions of asymptotic confidence size in locally misspecified moment inequality models," *Econometrica*, 80, 1741–1768.

BUGNI, F. A. AND T. URA (2019): "Inference in dynamic discrete choice problems under local misspecification," *Quantitative Economics*, 10, 67–103.

CHEN, X., E. T. TAMER, AND A. TORGOVITSKY (2011): "Sensitivity analysis in semiparametric likelihood models," *Cowles Foundation discussion paper*.

CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2014): "Gaussian approximation of suprema of empirical processes," *Annals of Statistics*, 42, 1564–1597.

CHIANG, H. D., K. KATO, AND Y. SASAKI (2023): "Inference for high-dimensional exchangeable arrays," *Journal of the American Statistical Association*, 118, 1595–1605.

CHRISTENSEN, T. AND B. CONNAULT (2023): "Counterfactual sensitivity and robustness," *Econometrica*, 91, 263–298.

COLUSSI, T. (2018): "Social ties in academia: A friend is a treasure," *Review of Economics and Statistics*, 100, 45–50.

CONLEY, T. G., C. B. HANSEN, AND P. E. ROSSI (2012): "Plausibly exogenous," *Review of Economics and Statistics*, 94, 260–272.

Davezies, L., X. D'Haultfœuille, and Y. Guyonvarch (2021): "Empirical Process Results for Exchangeable Arrays," *The Annals of Statistics*, 49, 845–862.

de Paula, Á. (2020): "Econometric models of network formation," *Annual Review of Economics*, 12, 775–799.

Donoho, D. L. (1994): "Statistical estimation and optimal recovery," *The Annals of Statistics*, 22, 238–270.

Ductor, L. (2015): "Does co-authorship lead to higher academic productivity?" *Oxford Bulletin of Economics and Statistics*, 77, 385–407.

Ductor, L., M. Fafchamps, S. Goyal, and M. J. Van der Leij (2014): "Social networks and research output," *Review of Economics and Statistics*, 96, 936–948.

Fafchamps, M., M. J. Van der Leij, and S. Goyal (2010): "Matching and network effects," *Journal of the European Economic Association*, 8, 203–231.

Goyal, S., M. J. Van Der Leij, and J. L. Moraga-González (2006): "Economics: An emerging small world," *Journal of Political Economy*, 114, 403–412.

Graham, B. (2020): "Handbook of Econometrics, volume 7A, chapter Network data," .

Graham, B. S. (2017): "An econometric model of network formation with degree heterogeneity," *Econometrica*, 85, 1033–1063.

——— (2022): "Sparse network asymptotics for logistic regression," *arXiv preprint arXiv:2010.04703*.

Graham, B. S., F. Niu, and J. L. Powell (2019): "Kernel density estimation for undirected dyadic data," *arXiv preprint arXiv:1907.13630*.

——— (2021): "Minimax risk and uniform convergence rates for nonparametric dyadic regression," Tech. rep., National Bureau of Economic Research.

Hsieh, C.-S., M. Konig, X. Liu, and C. Zimmermann (2022): "Collaboration in bipartite networks," Tech. rep., Working Paper.

Huber, P. (1964): "Robust estimation of a location parameter," *Ann. Math. Statist.*, 35, 73–101.

Huber, P. J. and E. M. Ronchetti (2009): *Robust statistics*, John Wiley & Sons Hoboken, NJ, USA.

Imbens, G. W. (2003): "Sensitivity to exogeneity assumptions in program evaluation," *American Economic Review*, 93, 126–132.

Kitamura, Y., T. Otsu, and K. Evdokimov (2013): "Robustness, infinitesimal neighborhoods, and moment restrictions," *Econometrica*, 81, 1185–1201.

Leamer, E. E. (1985): "Sensitivity analyses would help," *The American Economic Review*, 75, 308–313.

Leung, M. (2015): "Two-step estimation of network-formation models with incomplete information," *Journal of Econometrics*, 188, 182–195.

LI, J. AND Z. LIAO (2020): "Uniform nonparametric inference for time series," *Journal of Econometrics*, 219, 38–51.

MASTEN, M. A. AND A. POIRIER (2020): "Inference on breakdown frontiers," *Quantitative Economics*, 11, 41–111.

——— (2021): "Salvaging falsified instrumental variable models," *Econometrica*, 89, 1449–1469.

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

NEVO, A. AND A. M. ROSEN (2012): "Identification with imperfect instruments," *Review of Economics and Statistics*, 94, 659–671.

NEWEY, W. K. (1985): "Generalized method of moments specification testing," *Journal of Econometrics*, 29, 229–256.

——— (1990): "Semiparametric efficiency bounds," *Journal of Applied Econometrics*, 5, 99–135.

NEWMAN, M. E. (2001): "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences*, 98, 404–409.

NOACK, C. AND C. ROTHE (2019): "Bias-aware inference in fuzzy regression discontinuity designs," *arXiv preprint arXiv:1906.04631*.

PELICAN, A. AND B. S. GRAHAM (2020): "An optimal test for strategic interaction in social and economic network formation between heterogeneous agents," Tech. rep., National Bureau of Economic Research.

POLLARD, D. (2002): *A user's guide to measure theoretic probability*, 8, Cambridge University Press.

RIDDER, G. AND S. SHENG (2020): "Estimation of large network formation games," Working paper.

RIEDER, H. (1994): *Robust Asymptotic Statistics*, Springer Verlag, New York, NY.

TABORD-MEEHAN, M. (2019): "Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic," *Journal of Business & Economic Statistics*, 37, 671–680.