

# DYADIC REGRESSION WITH SAMPLE SELECTION

KENSUKE SAKAMOTO

*University of Wisconsin-Madison*

ABSTRACT. This paper addresses the sample selection problem in panel dyadic regression analysis. Dyadic data often include many zeros in the main outcomes due to the underlying network formation process. This not only contaminates popular estimators used in practice but also complicates the inference due to the dyadic dependence structure. We extend Kyriazidou (1997)'s approach to dyadic data and characterize the asymptotic distribution of our proposed estimator. The convergence rates are  $\sqrt{n}$  or  $\sqrt{n^2 h_n}$ , depending on the degeneracy of the Hájek projection part of the estimator, where  $n$  is the number of nodes and  $h_n$  is a bandwidth. We propose a bias-corrected confidence interval and a variance estimator that adapts to the degeneracy. A Monte Carlo simulation shows the good finite sample performance of our estimator and highlights the importance of bias correction in both asymptotic regimes when the fraction of zeros in outcomes varies. We illustrate our procedure using data from Moretti and Wilson (2017)'s paper on migration.

**Keywords:** Dyadic Data, Sample Selection, Fixed Effects, Network Formation, Bias Correction.

---

*E-mail address:* kenlfo2080@wisc.edu.

*Date:* 07/03/2024.

The author thanks supports and comments from Jack Porter, Bruce Hansen, Xiaoxia Shi, Harold Chiang, Kohei Yata, and participants in New York Camp Econometrics XVII. This paper was supported by the Summer Fellowship from the Department of Economics at University of Wisconsin-Madison.

## 1. INTRODUCTION

Dyadic data describes pairwise outcomes, such as trade volume between countries. Numerous applications have analyzed such data using the regression model, referred to as dyadic regression. Examples include gravity equations in trade, migration, and urban economics (Helpman et al., 2008; Moretti and Wilson, 2017; Monte et al., 2018), and risk-sharing networks in development economics (Fafchamps and Gubert, 2007). One of the prominent features of dyadic data is the non-negligible number of zeros in the outcomes of interest,<sup>1</sup> possibly due to economic mechanisms such as prohibitive fixed costs. This paper deals with panel dyadic data, where zeros are prevalent both cross-sectionally and in the time series.

How should we treat zeros in dyadic regression? In applications, zeros are often discarded due to the log-linear specification (Moretti and Wilson, 2017). The Poisson pseudo-maximum-likelihood (PPML) estimator is also frequently used to avoid discarding zeros and address issues related to log-linearization (Silva and Tenreyro, 2006). These approaches implicitly assume that zeros occur exogenously. Since a zero in a pairwise outcome results from no link between two units, we can associate zeros with the underlying network formation mechanism that determines which pairs appear in a sample. If the network is formed endogenously as a result of an interaction between two agents, the empirical practices mentioned above can be subject to sample selection bias, as in Heckman (1979).

This paper has two primary objectives. First, we aim to jointly model network formation and the outcome generation on such networks. This joint modeling allows identification of the effects of changes in pair-level or individual-level characteristics, separating them from the effects caused by changes in networks. In contrast, the dyadic regression literature has primarily focused on regression with fixed or exogenous networks. Second, we develop a robust inference method that accounts for the dyadic dependence structure. Pairwise outcomes are likely to be dependent on each other through common shocks to individuals. This dyadic dependence can be especially important in the presence of zeros and the network formation because a few individuals can have significantly more links than others,<sup>2</sup> which strengthens the influence of shocks to those individuals on the dyadic dependence. At the same time, it is known that with dyadic data, we can have different asymptotic regimes depending on the nature of those individual-level shocks (Menzel, 2021). To be practitioner-friendly, our inference method needs to consider the dyadic dependence and ensure adaptivity to different resulting asymptotic regimes.

---

<sup>1</sup>Helpman et al. (2008) documents that there was no trade among roughly 50% of country pairs from 1970 to 1997. In 2017, there was no migration among about 60% of country pairs (the author calculated using the data available from the World Bank (<https://www.worldbank.org/en/topic/migrationremittancesdiasporaissues/brief/migration-remittances-data>)).

<sup>2</sup>For example, in Moretti and Wilson (2017)'s migration flow data, star scientists' migration from or to California constituted approximately 14% of the links in the sample on average. This percentage is much higher than the expected 2% when considering all potential links in the sample.

Our setup will be a linear panel dyadic regression model, featuring the network formation process as a sample selection mechanism that generates both zeros and unobservable outcomes. To capture the dyadic dependence structure, we incorporate two types of unobservable individual heterogeneity into the model: time-invariant fixed effects and time-varying random effects, which is a new modeling strategy in the literature. We extend Kyriazidou (1997)’s identification argument, originally designed for individualistic data, to dyadic data, and correspondingly propose a semiparametric, kernel-based estimator that assigns weights to pairs whose selection index remains stable over time. A significant challenge we face when analyzing our estimator is the need to address the dependence structure caused by node-level shocks, which is absent in individualistic data models analyzed in Kyriazidou (1997). To control for this type of dependence, we utilize the U-statistic-like structure of our estimator, which gives us a mutually uncorrelated decomposition into the node-level Hájek projection part and the dyad-level projection error part.

We show that our estimator is asymptotically normal with two different convergence rates depending on the nature of errors. If the Hájek projection is non-degenerate (i.e., each summand has positive variance), our estimator achieves  $\sqrt{n}$ -asymptotic normality, where  $n$  is the number of nodes. In this case, we not only have zero asymptotic bias but also share the same convergence rates as the usual fixed effect estimator and PPML estimator when its leading term is also non-degenerate. The latter point implies that there is no loss in effective sample sizes with our estimator for using a kernel-based local method compared with the usual non-weighted estimator. If the Hájek projection is degenerate, our estimator achieves  $\sqrt{Nh_n}$ -asymptotic normality, where  $N \sim n^2$  is the number of dyads and  $h_n$  is a bandwidth. While the usual fixed effect estimator and the PPML estimator can be non-Gaussian in the limit (Menzel, 2021), our estimator is guaranteed to be asymptotically normal regardless of degeneracy. This result is analogous to Hall (1984)’s central limit theorem for degenerate U-statistics, allowing common statistics of interest, such as confidence intervals, to be constructed in a standard manner. In the degenerate case, our estimator exhibits asymptotic bias, which motivates us to introduce a bias correction.

We propose a variance estimator and bias-corrected confidence intervals that adapt to the degeneracy. Our variance estimator is similar to the one proposed by Graham et al. (2019) for nonparametric dyadic density estimation. We show that our estimator is consistent for the asymptotic variances in both non-degenerate and degenerate cases, after being rescaled by  $\sqrt{n}$  or  $\sqrt{Nh_n}$ , respectively. For the bias correction, we use a consistent estimator for the asymptotic bias in the degenerate case. We show that the correction term is negligible in the non-degenerate case after being rescaled by  $\sqrt{n}$ . Combining both bias-corrected estimator and variance estimator, we can construct bias-corrected confidence intervals for our estimator. These intervals have asymptotically correct sizes regardless of the (non-)degeneracy of the leading term in our estimator.

We conduct a simple simulation exercise to demonstrate the performance of our estimators compared to the usual fixed effect estimator and PPML estimator, as we vary the fraction of

selected dyads from 10% to 90%. Our proposed estimator exhibits better finite sample properties than the other two estimators. Our bias-corrected confidence intervals also outperform the alternatives in coverage probabilities, regardless of degeneracy. This result underscores the importance of bias correction in finite samples, even though the asymptotic bias is zero in the non-degenerate case, which is a new finding in the literature.

As an empirical application, we extend and apply our estimator to the regression specification proposed by Moretti and Wilson (2017), which estimates the effects of state tax differences on the internal migration flows in the U.S. Comparing our proposed estimator with Moretti and Wilson (2017)'s, we find that their conclusion, which suggests that state tax differences have a significant impact on internal migration, may not be robust in the presence of a dyadic dependence structure and sample selection biases.

This paper is closely related to the burgeoning literature on dyadic regression (Cameron and Miller, 2014; Tabord-Meehan, 2019; Bonhomme, 2020; Zeleneev, 2020; Graham, 2020; Graham et al., 2021). Except for Bonhomme (2020) and Zeleneev (2020), these papers do not touch on non-random sample selection but instead focus on the consequence of the dyadic dependence structure. Bonhomme (2020) mainly focuses on the case where the selection is conditionally random with random effects. He also discusses the case where the selection is conditionally non-random without a theoretical analysis. Zeleneev (2020) studies identification and estimation of a dyadic regression model with flexible fixed effects. While his focus is on identification and a rate of convergence analysis, our paper provides the full inference results.

This paper also contributes to the literature on econometric analysis of models with endogenous network formation. Examples include Johnsson and Moon (2021), Auerbach (2022), and Jochmans (2023). While these papers study social interaction/peer effects type models where outcomes of interest are individualistic, our paper studies the direct consequence of network formation on dyadic outcomes.

## 2. MODEL

There are  $n$  nodes in the data (e.g., states, countries), indexed by  $i = 1, \dots, n$ . Let  $\{(X_{it}, Z_{it})_{t=1, \dots, T}\}_{i=1}^n$  be a node-level observation, where  $X_{it} \in \mathbb{R}^{q_x}$  and  $Z_{it} \in \mathbb{R}^{q_z}$ . For each dyad  $ij$  and time  $t$ ,  $Y_{ijt} \in \mathbb{R}$  is a main outcome, and we observe a binary variable  $d_{ijt} \in \{0, 1\}$ , which indicates that  $Y_{ijt}$  is observable<sup>3</sup> only if  $d_{ijt} = 1$ . We can interpret the adjacency matrix  $D_t \equiv [d_{ijt}]_{i,j=1, \dots, n}$  as a network that summarizes the existence of interactions between nodes. In this paper, we restrict our attention to a model with  $T = 2$  and an undirected graph where  $Y_{ijt} = Y_{jit}$ ,  $d_{ijt} = d_{jit}$  for all  $i, j, t$ . We also rule out self-loops by convention:  $Y_{iit} = d_{iit} = 0$  for all  $i, t$ . An extension to  $T > 2$  and a directed graph is discussed in Section 4.1.

<sup>3</sup>Since we focus on a linear model, we can interchange unobservability with zero. Alternatively, we can interpret  $Y_{ijt}$  as the logarithm of  $\tilde{Y}_{ijt} \geq 0$ .

The data is generated according to the following model:

$$W_{ijt} = w(X_{it}, X_{jt}), R_{ijt} = r(Z_{it}, Z_{jt}), \quad (2.1)$$

$$Y_{ijt}^* = W_{ijt}'\beta + A_i + A_j + \epsilon_{ijt}, \quad (2.2)$$

$$d_{ijt} = \mathbf{1}\{R_{ijt}'\gamma + B_i + B_j - \eta_{ijt}\}, \quad (2.3)$$

$$Y_{ijt} = \begin{cases} Y_{ijt}^* & \text{if } d_{ijt} = 1 \\ \text{unobserved} & \text{if } d_{ijt} = 0 \end{cases}. \quad (2.4)$$

The regressors  $W_{ijt} \in \mathbb{R}^{q_w}$  and  $R_{ijt} \in \mathbb{R}^{q_r}$  are constructed from some user-specified symmetric functions  $w : \mathbb{R}^{q_x} \times \mathbb{R}^{q_x} \rightarrow \mathbb{R}^{q_w}$  and  $r : \mathbb{R}^{q_z} \times \mathbb{R}^{q_z} \rightarrow \mathbb{R}^{q_r}$  such that  $w(x, y) = w(y, x)$  and  $r(x', y') = r(y', x')$  for any  $x, y \in \mathbb{R}^{q_x}$  and  $x', y' \in \mathbb{R}^{q_z}$ . For example, we can specify  $w$  to be a pairwise summation  $w(x, y) = x + y$ . The symmetry in these functions is needed as our graphs are undirected; we can relax this requirement with directed graphs, as discussed in Section 4.1. The node-level fixed effects  $A_i, B_i \in \mathbb{R}$  are unobservable, and we allow them to correlate with the regressors, as in the usual fixed effect model.

We specify the structure of errors  $\epsilon_{ijt}, \eta_{ijt}$  as follows: For  $1 \leq i < j \leq n$ ,

$$(\epsilon_{ij1}, \epsilon_{ij2}, \eta_{ij1}, \eta_{ij2}) = \tau(U_{i1}, U_{i2}, U_{j1}, U_{j2}, U_{ij1}, U_{ij2}), \quad (2.5)$$

where  $U_i \equiv (U_{i1}, U_{i2})$  and  $U_{ij} \equiv (U_{ij1}, U_{ij2})$  are node-level and dyad-level random vectors, respectively, and  $\tau$  is an unknown multivariate function.<sup>4</sup>

Let  $\xi_i \equiv (X_{i1}, X_{i2}, Z_{i1}, Z_{i2}, A_i, B_i)$  be a vector that contains observed and unobserved information in the two periods with respect to node  $i$ . We impose the following distributional assumption:

**Assumption 1.**

- (1)  $\xi_i, i = 1, \dots, n$  are independently and identically distributed.
- (2)  $(\epsilon_{ijt}, \eta_{ijt})_{t=1,2}, 1 \leq i < j \leq n$  are generated according to (2.5).
- (3) Conditionally on  $\{\xi_i\}_{i=1}^n, U_i, i = 1, \dots, n$  are independent,  $U_{ij}, 1 \leq i < j \leq n$  are independent, and both of them are mutually independent.
- (4) For  $i < j$ ,  $(U_i, U_j, U_{ij})$  conditional on  $\{\xi_i\}_{i=1}^n$  has the same distribution as  $(U_i, U_j, U_{ij})$  conditional on  $\xi_i, \xi_j$ .
- (5) For  $i < j < k$ , if  $\xi_i = \xi_j = \xi_k$ ,  $(U_i, U_j, U_{ij})$  and  $(U_i, U_k, U_{ik})$  has the same distribution conditional conditionally on  $\xi_i, \xi_j, \xi_k$ .

Part (1) imposes homogeneity on the node-level data-generating process. Parts (2) and (3) are new to the literature on dyadic regression with fixed effects. While the previous literature

<sup>4</sup>Here, we need not specify the dimensions of those vectors and the function since the following results do not depend on them as long as those dimensions are fixed.

assumes conditional independence of dyadic-level errors (Graham, 2017; Zeleneev, 2020; Candalaria, 2020), our error structure (2.5) allows for the conditional dependence between errors with a common node (e.g.,  $\epsilon_{ij1}$  and  $\epsilon_{ik1}$ ) through  $U_i$ , but also includes conditional independence as a special case where node-level random vectors  $U_{it}, U_{jt}$  are degenerate given  $\{\xi_i\}_{i=1}^n$ . Part (4) is the standard assumption in the literature and excludes "externalities," where dyad  $ij$  can be affected by nodes other than  $i$  or  $j$ . Part (5) ensures the conditional exchangeability of  $(\epsilon_{ijt}, \eta_{ijt})_{t=1,2}$  across dyads.

## 2.1. Identification.

The following two assumptions are crucial for the identification of  $\beta$ :

**Assumption 2.**  $(\epsilon_{ij1}, \epsilon_{ij2}, \eta_{ij1}, \eta_{ij2})$  and  $(\epsilon_{ij2}, \epsilon_{ij1}, \eta_{ij2}, \eta_{ij1})$  are identically distributed conditionally on  $\xi_i, \xi_j$ .

**Assumption 3.**  $E[d_{ij1}d_{ij2}\Delta W_{ij}\Delta W'_{ij}|\Delta R'_{ij}\gamma = 0]$  is non-singular where  $\Delta W_{ij} = W_{ij1} - W_{ij2}$  and  $\Delta R_{ij} = R_{ij1} - R_{ij2}$ .

Assumption 2 excludes cases where, for example, the conditional variance of  $\epsilon_{ijt}$  depends only on period  $t$ 's information:  $Var(\epsilon_{ijt}|\xi_i, \xi_j) = \sigma^2 \times W'_{ijt}\beta$ . However, it allows time invariant heteroskedasticity such as  $Var(\epsilon_{ijt}|\xi_i, \xi_j) = \sigma^2(W_{ij1} + W_{ij2})'\beta \times A_i \times A_j$ . From (2.5), this assumption is implied by the conditional exchangeability of  $U_{it}$  and  $U_{ijt}$  with respect to time. Assumption 3 excludes cases where  $W_{ijt}$  is exactly the same as  $R_{ijt}$  and implies that some variables in  $R_{ijt}$  must be excluded from  $W_{ijt}$ . Since

$$\begin{aligned} & E[d_{ij1}d_{ij2}\Delta W_{ij}\Delta W'_{ij}|\Delta R'_{ij}\gamma = 0] \\ &= Pr(d_{ij1}d_{ij2} = 1|\Delta R'_{ij}\gamma = 0) \times E[\Delta W_{ij}\Delta W'_{ij}|d_{ij1}d_{ij2} = 1, \Delta R'_{ij}\gamma = 0], \end{aligned}$$

this assumption also implies that the networks  $D_1, D_2$  are locally dense across time in the sense that  $Pr(d_{ij1}d_{ij2} = 1|\Delta R'_{ij}\gamma = 0) > 0$ .

Our identification argument is summarized in the following two steps, similarly to Kyriazidou (1997). First, take the time-difference on observed outcomes (dyads with  $d_{ij1} = d_{ij2} = 1$ ) to eliminate the fixed effects:

$$\Delta Y_{ij} = \Delta W'_{ij}\beta + \epsilon_{ij1} - \epsilon_{ij2}.$$

If we take expectation of both sides conditionally on  $d_{ij1} = d_{ij2} = 1$  and  $\xi_i, \xi_j$ ,

$$E[\Delta Y_{ij}|d_{ij1}d_{ij2} = 1, \xi_i, \xi_j] = \Delta W'_{ij}\beta + \underbrace{E[\epsilon_{ij1} - \epsilon_{ij2}|d_{ij1}d_{ij2} = 1, \xi_i, \xi_j]}_{\text{Sample selection effect}}.$$

Note that, in general, the sample selection effect is not 0.

Second, we seek to find conditions to eliminate the selection effect. Assumption 2 is equivalent to

$$F(\epsilon_{ij1}, \epsilon_{ij2}, \eta_{ij1}, \eta_{ij2} | \xi_i, \xi_j) = F(\epsilon_{ij2}, \epsilon_{ij1}, \eta_{ij2}, \eta_{ij1} | \xi_i, \xi_j),$$

where  $F$  is the conditional distribution of the errors given  $\xi_i, \xi_j$ . Then, for dyad  $ij$  with  $\Delta R'_{ij}\gamma = R'_{ij1}\gamma - R'_{ij2}\gamma = 0$ ,

$$\begin{aligned} & E[\epsilon_{ij1} | d_{ij1}d_{ij2} = 1, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0] \\ &= E[\epsilon_{ij1} | R'_{ij1}\gamma + B_i + B_j \geq \eta_{ij1}, R'_{ij2}\gamma + B_i + B_j \geq \eta_{ij2}, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0] \\ &= E[\epsilon_{ij2} | R'_{ij2}\gamma + B_i + B_j \geq \eta_{ij2}, R'_{ij1}\gamma + B_i + B_j \geq \eta_{ij1}, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0] \\ &= E[\epsilon_{ij2} | d_{ij1}d_{ij2} = 1, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0]. \end{aligned}$$

Hence, the conditional expectation of  $\Delta Y_{ij}$  given  $d_{ij1}d_{ij2} = 1, \xi_i, \xi_j$ , and  $\Delta R'_{ij}\gamma = 0$  is

$$E[\Delta Y_{ij} | d_{ij1}d_{ij2} = 1, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0] = \Delta W'_{ij}\beta.$$

Multiplying the both sides by  $\Delta W_{ij}$  and aggregating  $\xi_i, \xi_j$ , we get

$$E[\Delta W_{ij}\Delta Y_{ij} | d_{ij1}d_{ij2} = 1, \Delta R'_{ij}\gamma = 0] = E[\Delta W_{ij}\Delta W'_{ij} | d_{ij1}d_{ij2} = 1, \Delta R'_{ij}\gamma = 0]\beta.$$

Then, under Assumption 3,  $\beta$  is uniquely written as

$$\beta = E[d_{ij1}d_{ij2}\Delta W_{ij}\Delta W'_{ij} | \Delta R'_{ij}\gamma = 0]^{-1} E[d_{ij1}d_{ij2}\Delta W_{ij}\Delta Y_{ij} | \Delta R'_{ij}\gamma = 0]. \quad (2.6)$$

## 2.2. Estimation.

Estimation is done in two steps; In the first step, we estimate  $\gamma$  with a consistent estimator  $\hat{\gamma}_n$ , and in the second step we estimate  $\beta$  with  $\hat{\beta}_n$ , a sample analogue of the identified  $\beta$  with  $\gamma$  replaced by  $\hat{\gamma}_n$ .

In the following, we focus on the second step. The sample-analogue of (2.6) is given by

$$\hat{\beta}_n = \left[ \sum_{i < j} d_{ij1}d_{ij2}\Delta W_{ij}\Delta W'_{ij}K_{h_n}(\Delta R'_{ij}\hat{\gamma}_n) \right]^{-1} \left[ \sum_{i < j} d_{ij1}d_{ij2}\Delta W_{ij}\Delta Y_{ij}K_{h_n}(\Delta R'_{ij}\hat{\gamma}_n) \right],$$

where  $\sum_{i < j} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n$ ,  $K_{h_n}(v) = h_n^{-1}K(v/h_n)$  is a kernel, and  $h_n$  is a bandwidth. The weight function is used to smooth the condition  $\Delta R'_{ij}\gamma = 0$  and puts larger weight on observations with small  $\Delta R'_{ij}\hat{\gamma}_n$ .

To evaluate  $\hat{\beta}_n$  in terms of  $\beta$ , rewrite the time-differenced model as

$$\Delta Y_{ij} = \Delta W'_{ij}\beta + \lambda_{ij} + \nu_{ij},$$

where

$$\lambda_{ij} \equiv E[\epsilon_{ij1} - \epsilon_{ij2} | d_{ij1}d_{ij2} = 1, \xi_i, \xi_j]$$

$$\nu_{ij} \equiv \epsilon_{ij1} - \epsilon_{ij2} - \lambda_{ij}.$$

Note that  $E[\nu_{ij} | d_{ij1}d_{ij2} = 1, \xi_i, \xi_j] = 0$  by construction. Define

$$\hat{S}_{WW} \equiv \frac{1}{N} \sum_{i < j} d_{ij1}d_{ij2} \Delta W_{ij} \Delta W'_{ij} K_{h_n}(\Delta R'_{ij} \hat{\gamma}_n),$$

$$\hat{S}_{W\lambda} \equiv \frac{1}{N} \sum_{i < j} d_{ij1}d_{ij2} \Delta W_{ij} \lambda_{ij} K_{h_n}(\Delta R'_{ij} \hat{\gamma}_n),$$

$$\hat{S}_{W\nu} \equiv \frac{1}{N} \sum_{i < j} d_{ij1}d_{ij2} \Delta W_{ij} \nu_{ij} K_{h_n}(\Delta R'_{ij} \hat{\gamma}_n).$$

Substituting  $\Delta Y_{ij}$  into  $\hat{\beta}_n$  yields

$$\hat{\beta}_n = \beta + \hat{S}_{WW}^{-1} \hat{S}_{W\lambda} + \hat{S}_{WW}^{-1} \hat{S}_{W\nu}.$$

The terms  $\hat{S}_{WW}^{-1} \hat{S}_{W\lambda}$  and  $\hat{S}_{WW}^{-1} \hat{S}_{W\nu}$  can be understood as the selection bias term and the stochastic error term in the estimator, respectively.

### 3. ASYMPTOTIC ANALYSIS

#### 3.1. Regularity Conditions.

For ease of notation, we write the following conditions in terms of dyads 12 and 13, which entails no loss of generality under the undirected graph and Assumption 1.

Let  $f_{R\gamma,2}$  be the joint density of  $\Delta R'_{12}\gamma$  and  $\Delta R'_{13}\gamma$  when it exists. Let  $f_{R\gamma}$  be the marginal density and  $f_{R\gamma|\xi_1, U_1}$  be the conditional density given  $\xi_1, U_1$ .

**Assumption 4.** *The joint distribution of  $\Delta R'_{12}\gamma$  and  $\Delta R'_{13}\gamma$  is absolutely continuous, and for some  $\kappa_0 > 0$ , the following hold in the neighborhoods  $(-\kappa_0, \kappa_0)^2$  or  $(-\kappa_0, \kappa_0)$  around  $(0, 0)$  or  $(0)$ , respectively:*

- (1) *The density  $f_{R\gamma,2}(\cdot, \cdot)$  is  $k \geq 2$  times continuously differentiable, and the derivatives  $\frac{\partial^2}{\partial x^p \partial y^q} f_{R\gamma,2}(x, y)$  are uniformly bounded for  $p + q \leq k, p, q \geq 0$  and bounded away from 0.*
- (2) *The marginal density  $f_{R\gamma}(\cdot)$  is bounded away from 0.*
- (3) *The conditional density  $f_{R\gamma|\xi_1, U_1}(\cdot)$  given  $\xi_1, U_1$  is continuous and uniformly bounded almost surely.*

Part (1) is a smoothness assumption on the density as in the nonparametric regression literature. Part (2) ensures that we observe  $\Delta R'_{12}\gamma$  around 0, which is crucial for identification. Part (3) essentially requires well-behaved  $r(\cdot, \cdot)$  in (2.1).



Define  $(w_1, w_2) \mapsto \Lambda(w_1, w_2, \xi_1, \xi_2)$  as

$$\Lambda(w_1, w_2, \xi_1, \xi_2) \equiv E[\epsilon_{12t} | \eta_{12t} \leq w_1, \eta_{12s} \leq w_2, \xi_1, \xi_2]$$

with  $t, s = 1, 2, t \neq s$ . This  $\Lambda$  is the sample selection effect caused by the correlation between errors  $\epsilon_{12t}, \epsilon_{12s}$  and  $\eta_{12t}, \eta_{12s}$ . Note that the function  $\Lambda$  does not depend on time  $t$  or  $s$  because of Assumption 2.

**Assumption 5.** *The function  $(w_1, w_2) \mapsto \Lambda(w_1, w_2, \xi_1, \xi_2)$  is differentiable in the neighborhoods  $(-\kappa_0, \kappa_0)^2$  around  $(0, 0)$  for some  $\kappa_0 > 0$ .*

This assumption is essential for controlling the sample selection effect and characterizing the asymptotic bias in some cases. An implication of this assumption is that for some  $\Lambda_{12} \equiv \tilde{\Lambda}(w_1, w_2, \xi_1, \xi_2)$ ,

$$\Lambda(w_1, w_2, \xi_1, \xi_2) - \Lambda(w_2, w_1, \xi_1, \xi_2) = \Lambda_{12} \times (w_1 - w_2),$$

by the multivariate mean-value theorem. Note that the function  $\Lambda$  does not depend on time  $t$  or  $s$  because of Assumption 2. This assumption is strong because the difference in  $\Lambda$  must be exactly linear in the first and second elements. If we focus on the degenerate case discussed below, since the asymptotic bias is 0 in that case, we can relax the differentiability to Lipschitz-like continuity on  $\Lambda$ :  $|\Lambda(w_1, w_2, \xi_1, \xi_2) - \Lambda(w_2, w_1, \xi_1, \xi_2)| \leq |\Lambda_{12}| \times |w_1 - w_2|$ .

Let  $\|\cdot\|$  denote a Euclidian norm of vectors.

**Assumption 6.** *For some  $\kappa_0 > 0$ , the following hold in the neighborhoods  $(-\kappa_0, \kappa_0)^2$  or  $(-\kappa_0, \kappa_0)$  around  $(0, 0)$  or  $(0)$ , respectively.*

(1) *The following moments are uniformly bounded almost surely:*

$$\begin{aligned} & E[\|\Delta W_{12}\|^8 | \Delta R'_{12} \gamma = \cdot, \xi_1, U_1], E[\|\Delta R_{12}\|^6 | \Delta R'_{12} \gamma = \cdot, \xi_1, U_1], \\ & E[\nu_{12}^8 | \Delta R'_{12} \gamma = \cdot, \xi_1, U_1], E[\Lambda_{12}^6 | \Delta R'_{12} \gamma = \cdot, \xi_1, U_1]. \end{aligned}$$

(2) *The following moments are continuous and bounded, and the first two are positive definite:*

$$\begin{aligned} & E[d_{121} d_{122} \Delta W_{12} \Delta W'_{12} | \Delta R'_{12} \gamma = \cdot], E[d_{121} d_{122} \Delta W_{12} \Delta W'_{12} \nu_{12}^2 | \Delta R'_{12} \gamma = \cdot] \\ & E[d_{121} d_{122} d_{131} d_{132} \Delta W_{12} \Delta W'_{13} \nu_{12} \nu_{13} | \Delta R'_{12} \gamma = \cdot, \Delta R'_{13} \gamma = \cdot]. \end{aligned}$$

(3)  *$g(\cdot) \equiv E[d_{121} d_{122} \Delta W_{12} \Lambda_{12} | \Delta R'_{12} \gamma = \cdot] f_{R\gamma}(\cdot)$  is  $k$ -times continuously differentiable with uniformly bounded derivatives.*

(4)  *$g_{\xi_1, U_1}(\cdot) \equiv E[d_{121} d_{122} \Delta W_{12} \nu_{12} | \Delta R'_{12} \gamma = \cdot, \xi_1, U_1] f_{R\gamma | \xi_1, U_1}(\cdot)$  is  $k$ -times continuously differentiable with uniformly bounded derivatives almost surely.*

Part (1) assumes the existence of conditional moments for the relevant variables. The conditioning on  $\xi_1$  and  $U_1$  is needed for controlling the dyadic dependence structure. Part (2) is crucial

for obtaining the convergence results used below, and the positive definiteness is needed for ensuring the non-degeneracy of our estimator in the limit. Part (3) is used for characterizing the asymptotic bias provided below. Part (4) is essential for the negligibility of the approximation error of our variance estimator.

**Assumption 7.** *The following moments exist:*

$$E[\|\Delta W_{12}\|^8], E[\|\Delta R_{12}\|^8], E[\Lambda_{12}^6], E[\nu_{12}^6]$$

Additionally to Assumption 6, which restricts the moments locally around  $(0, 0)$  or  $(0)$ , we use the existence of these unconditional moments when bounding error terms coming from the usage of  $\hat{\gamma}_n$ .

**Assumption 8.** *A kernel function  $K(\cdot)$  satisfies the following:*

- (1) *For some  $\kappa > 0$ ,  $K$  is 0 outside of  $[-\kappa, \kappa]$ , bounded in  $[-\kappa, \kappa]$ , and three times continuously differentiable with bounded derivatives in  $(-\kappa, \kappa)$ .*
- (2)  $\int K(s)ds = 1$ .
- (3)  $\int s^i K(s)ds = 0$  for  $i = 1, \dots, k$ .

For example, a biweight kernel  $K(x) = 15/16(1 - x^2)^2 \mathbf{1}\{|x| \leq 1\}$  satisfies this assumption with  $\kappa = 1$  and  $k = 2$ .

**Assumption 9.** *The sequence of bandwidths  $\{h_n\}$  satisfies  $h_n \rightarrow 0$  and  $nh_n \rightarrow \infty$  as  $n \rightarrow \infty$ .*

This assumption is standard in the nonparametric regression literature. We impose further conditions on  $\{h_n\}$  in each statement below.

**Assumption 10.** *The first-step estimator  $\hat{\gamma}_n$  satisfies  $\sqrt{Nh_n}(\hat{\gamma}_n - \gamma) = o_p(1)$ .*

This assumption requires the first-step estimator to be consistent and converge faster than our estimator. For example, if  $\eta_{ijt} \sim \text{Logistic}(0, 1)$  independently across  $ij$  and  $t$ , we can show that Chamberlain (1980)'s conditional logit estimator satisfies  $\hat{\gamma}_n - \gamma = O_p(1/\sqrt{N})$  so that  $\sqrt{Nh_n}(\hat{\gamma}_n - \gamma) = O_p(\sqrt{h_n}) = o_p(1)$ . In Section 3.6, we discuss the availability of alternative estimators for  $\gamma$ .

### 3.2. Asymptotic Normality.

Define the following components that will appear in the asymptotic bias and variance expression:

$$\begin{aligned}\Sigma_{WW} &\equiv f_{R\gamma}(0)E[d_{121}d_{122}\Delta W_{12}\Delta W'_{12}|\Delta R'_{12}\gamma = 0] \\ \Sigma_{W\lambda} &\equiv \frac{1}{k!} \frac{\partial^k g(0)}{\partial w^k} \int s^{k+1} K(s) ds, \\ \Sigma_{W\nu,1} &\equiv 4f_{R\gamma,2}(0,0)E[d_{121}d_{122}d_{131}d_{132}\Delta W_{12}\Delta W'_{13}\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0], \\ \Sigma_{W\nu,2} &\equiv f_{R\gamma}(0)E[d_{121}d_{122}\Delta W_{12}\Delta W'_{12}\nu_{12}^2|\Delta R'_{12}\gamma = 0] \int K^2(s) ds.\end{aligned}$$

We have the following result:

**Theorem 1.** *Suppose that Assumptions 1-10 hold. Fix an arbitrary non-zero vector  $c \in \mathbb{R}^{q_w}$  and some constant  $h \in (0, \infty)$ . Let  $c_W = \Sigma_{WW}^{-1}c$ . Then, as  $n \rightarrow \infty$ , we have the following three cases:*

(1) *If  $Nh_n^{2k+3} \rightarrow h$  and  $c'_W \Sigma_{W\nu,1} c_W > 0$ :*

$$\sqrt{n}c'(\hat{\beta}_n - \beta) \rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,1} c_W).$$

(2) *If  $Nh_n^{2k+3} \rightarrow h$  and  $c'_W \Sigma_{W\nu,1} c_W = 0$ :*

$$\sqrt{Nh_n}c'(\hat{\beta}_n - \beta) \rightarrow_d \mathcal{N}(\sqrt{h}c'_W \Sigma_{W\lambda}, c'_W \Sigma_{W\nu,2} c_W).$$

(3) *If  $Nh_n^{2k+3} \rightarrow \infty$  and  $nh_n^{2k+2} \rightarrow \infty$ :*

$$h_n^{-(k+1)}(\hat{\beta}_n - \beta) \rightarrow_p \Sigma_{WW}^{-1} \Sigma_{W\lambda}.$$

Part (1) and (2) of Theorem 1 show that our estimator is asymptotically normal with different convergence rates depending on  $\Sigma_{W\nu,1}$ . Part (1) departs from Kyriazidou (1997)'s result in that we have a parametric convergence rate based on the number of nodes  $n$ , not the number of dyads (or units)  $N$ . Under the condition that  $c'_W \Sigma_{W\nu,1} c_W > 0$  in part (1), the covariance of two summands indexed with a common index (e.g., dyad  $ij$  and  $ik$ ) in the estimator does not vanish in the limit, which results in the reduction in the effective sample size to  $n$ , the number of nodes. At the same time, the leading term is an average of conditional means of the summand given  $\xi_i$  and  $U_i$ , which averages out and drops  $h_n$  in the convergence rate. This  $\sqrt{n}$ -asymptotic normality is aligned with the dyadic non-parametric density estimation literature (Graham et al., 2019). Once we have  $\Sigma_{W\nu,1} = 0$  as in part (2), our result is aligned with Kyriazidou (1997) in that the convergence rates are non-parametric and based on the number of dyads (units). Part (3) of Theorem 1 shows that our estimator converges to the asymptotic bias part with suitable normalization, regardless of the degeneracy. We utilize this result to propose the bias-corrected estimator in the later section.

We can compare our estimator with the usual fixed effect estimator:

$$\hat{\beta}_{FE} = \left[ \sum_{i < j} d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \right]^{-1} \left[ \sum_{i < j} d_{ij1} d_{ij2} \Delta W_{ij} \Delta Y_{ij} \right], \quad (3.1)$$

which is biased because of the selection effect  $\lambda_{ij}$ . First, in the case of non-degeneracy, our estimator and the re-centered (infeasible) fixed effect estimator share the same convergence rates of  $\sqrt{n}$  (Davezies et al., 2021). This implies that there is no reduction in the effective sample size for using our kernel-based local estimator, which amends the need for fairly large samples as discussed in Kyriazidou (1997). Second, in the case of degeneracy, the fixed effect estimator applied to our model can exhibit a non-Gaussian distribution in the limit (Menzel, 2021), while our estimator is asymptotically normal regardless of the degeneracy. This guaranteed asymptotic normality is analogous to Hall (1984)'s central limit theorem for degenerate U-statistics, and thus the common statistics of interest, such as confidence intervals, can be constructed in a standard manner.

If we interpret the structural equation 2.2 as the log-linearized version of the canonical gravity model (Silva and Tenreyro, 2006; Head and Mayer, 2014),

$$\tilde{Y}_{ijt} = \exp(W'_{ijt}\beta + A_i + A_j) \times \underbrace{\eta_{ijt}}_{=d_{ijt}\exp(\epsilon_{ijt})},$$

the Poisson pseudo-maximum-likelihood estimator (PPML) for  $\beta$  can be compared with our estimator. The PPML estimator is given by

$$\sum_{i < j} \sum_{t=1}^2 \left[ \tilde{Y}_{ijt} - \exp(W'_{ijt}\hat{\beta}_{PPML} + a_i \mathbf{1}\{i = 1\} + a_j \mathbf{1}\{j = 1\}) \right] \left( W_{ijt}, \mathbf{1}\{i = 1\}, \mathbf{1}\{j = 1\} \right)' = 0. \quad (3.2)$$

We can make a similar comparison as in the fixed effect estimator based on the results by Davezies et al. (2021) and Menzel (2021):  $\hat{\beta}_{PPML}$  will be biased because of the misspecified errors, and the re-centered  $\hat{\beta}_{PPML}$  is asymptotically normal at the rate of  $\sqrt{n}$  in the non-degenerate case and can be non-Gaussian in the degenerate case.

### 3.3. Variance Estimation.

Since our estimator exhibits different asymptotic distributions depending on  $\Sigma_{W\nu,1}$ , it is desirable to have a variance estimator that adapts to the degeneracy.

First, we estimate  $\Sigma_{W\nu,1}$ . Define

$$\hat{S}_{ij} \equiv 2d_{ij1}d_{ij2}K_{h_n}(\Delta R'_{ij}\hat{\gamma}_n)\Delta W_{ij}\Delta\hat{\epsilon}_{ij},$$

where  $\Delta\hat{\epsilon}_{ij}$  is a residual  $\Delta Y_{ij} - \Delta W'_{ij}\hat{\beta}_n$ . Then, we propose an estimator for  $\Sigma_{W\nu,1}$  as

$$\hat{\Sigma}_{W\nu,1} = \binom{n}{3}^{-1} \sum_{i<j<k} \frac{1}{3} (\hat{S}_{ij}\hat{S}'_{ik} + \hat{S}_{ij}\hat{S}'_{jk} + \hat{S}_{ik}\hat{S}'_{jk}).$$

Next, we estimate  $\Sigma_{W\nu,2}$  by

$$\hat{\Sigma}_{W\nu,2} = \frac{h_n}{N} \sum_{i<j} d_{ij1}d_{ij2}K_{h_n}(\Delta R'_{ij}\hat{\gamma}_n)^2 \Delta W_{ij}\Delta W'_{ij}\Delta\hat{\epsilon}_{ij}.$$

The following result shows consistency of these estimators and their usefulness in adaptive variance estimation.

**Proposition 1.** *Suppose that Assumptions 1-10 hold. Set  $h_n = hN^{-1/(2k+3)}$  for some  $h \in (0, \infty)$ . We have*

$$\begin{aligned} \hat{\Sigma}_{W\nu,1} &\rightarrow_p \Sigma_{W\nu,1}, \\ \hat{\Sigma}_{W\nu,2} &\rightarrow_p \Sigma_{W\nu,2}, \end{aligned}$$

as  $n \rightarrow \infty$ . If  $c'_W \Sigma_{W\nu,1} c_W = 0$  with  $c_W = \Sigma_{WW}^{-1}c$  for some  $c \in \mathbb{R}^{q_w}$ , we have

$$nh_n c' \hat{S}_{WW}^{-1} \hat{\Sigma}_{W\nu,1} \hat{S}_{WW}^{-1} c \rightarrow_p 0,$$

as  $n \rightarrow \infty$ .

We now propose our variance estimator as follows:

$$\hat{\Sigma} \equiv \hat{S}_{WW}^{-1} \left[ \frac{n-2}{n(n-1)} \hat{\Sigma}_{W\nu,1} + \frac{1}{Nh_n} \hat{\Sigma}_{W\nu,2} \right] \hat{S}_{WW}^{-1}.$$

We can see that this estimator is adaptive to the degeneracy: When  $\Sigma_{W\nu,1}$  is positive definite, since  $n/(Nh_n) = o(1)$ ,

$$nc' \hat{\Sigma} c = c' \hat{S}_{WW}^{-1} \left[ \frac{n-2}{n-1} \hat{\Sigma}_{W\nu,1} + \frac{n}{Nh_n} \hat{\Sigma}_{W\nu,2} \right] \hat{S}_{WW}^{-1} c \rightarrow_p c'_W \Sigma_{W\nu,1} c_W,$$

as  $n \rightarrow \infty$  by Proposition 1 and Lemma 1 in Appendix A. When  $c'_W \Sigma_{W\nu,1} c_W = 0$ , since  $nh_n c' \hat{S}_{WW}^{-1} \hat{\Sigma}_{W\nu,1} \hat{S}_{WW}^{-1} c = o_p(1)$  by Proposition 1,

$$Nh_n c' \hat{\Sigma} c = c' \hat{S}_{WW}^{-1} \left[ 2(n-2)h_n \hat{\Sigma}_{W\nu,1} + \hat{\Sigma}_{W\nu,2} \right] \hat{S}_{WW}^{-1} c \rightarrow_p c'_W \Sigma_{W\nu,2} c_W,$$

as  $n \rightarrow \infty$ .

Our variance estimator is adapted from the one provided in Graham et al. (2019) for a dyadic non-parametric density estimator. They show that this type of estimator can be adaptive to the "knife edge" case, where  $nh_n$  is bounded from above and below asymptotically so that  $Nh_n \sim n$ . Here, we additionally show that the estimator is adaptive to the degeneracy by showing that the term involving  $\hat{\Sigma}_{W\nu,1}$  decays fast enough to be negligible when the convergence rate is  $\sqrt{Nh_n}$ .

### 3.4. Bandwidth Selection.

From the asymptotic distributional approximation result in Theorem 1, we can write down

the mean squared error of our estimator (without negligible parts)

$$MSE(c'\hat{\beta}_n) = h_n^{2(k+1)}(c'_W \Sigma_{W\lambda})^2 + \frac{1}{n} c'_W \Sigma_{W\nu,1} c_W + \frac{1}{Nh_n} c'_W \Sigma_{W\nu,1} c_W.$$

The optimal solution for minimizing this mean squared error with respect to  $h_n$  is given by

$$\begin{aligned} h_n^* &= \left( \frac{c'_W \Sigma_{W\nu,2} c_W}{2(k+1)N(c'_W \Sigma_{W\nu})^2} \right)^{\frac{1}{2k+3}} \\ &= h^* N^{-\frac{1}{2k+3}}. \end{aligned}$$

We can estimate  $h^*$  by the plug-in method. By Proposition 1, we have a consistent estimator for the variance part. For the bias part, we use a pilot bandwidth given by

$$h_{n,\delta} = hN^{-\delta/(2k+3)},$$

for some  $\delta \in (0, \frac{2k+3}{4k+4})$  and  $h > 0$ . Let  $\hat{\beta}_{n,\delta}$  be our estimator calculated with  $h_{n,\delta}$ . We can check that this bandwidth satisfies  $Nh_{n,\delta}^{2k+3} \rightarrow \infty$  and  $nh_{n,\delta}^{2k+2} \rightarrow \infty$ . Thus, by Theorem 1,

$$h_{n,\delta}^{-(k+1)}(\hat{\beta}_{n,\delta} - \beta) \rightarrow_p \Sigma_{WW}^{-1} \Sigma_{W\nu},$$

as  $n \rightarrow \infty$ . By replacing  $\beta$  by  $\hat{\beta}_n$ , calculated with  $h_n = hN^{-\frac{1}{2k+3}}$ , we have the following result:

**Proposition 2.** *Suppose that Assumptions 1-10 hold. Let  $\hat{\beta}_n$  and  $\hat{\beta}_{n,\delta}$  be the proposed estimators with bandwidths  $h_n = hN^{-1/(2k+3)}$  and  $h_{n,\delta} = hN^{-\delta/(2k+3)}$ , respectively, for some  $h > 0$  and  $\delta \in (0, \frac{2k+3}{4k+4})$ . Then,*

$$h_{n,\delta}^{-(k+1)}(\hat{\beta}_{n,\delta} - \hat{\beta}_n) \rightarrow_p \Sigma_{WW}^{-1} \Sigma_{W\lambda},$$

as  $n \rightarrow \infty$ .

Thus,

$$\hat{h}^* = \left( \frac{c' \hat{\Sigma}_{WW}^{-1} \hat{\Sigma}_{W\nu,2} \hat{\Sigma}_{WW}^{-1} c}{2(k+1)\{h_{n,\delta}^{-(k+1)} c'(\hat{\beta}_{n,\delta} - \hat{\beta}_n)\}^2} \right)^{\frac{1}{2k+3}}$$

is a consistent estimator for  $h^*$  by Propositions 1 and 2.

### 3.5. Bias Correction.

Notice that our estimator has the asymptotic bias of  $\sqrt{h} \Sigma_{WW}^{-1} \Sigma_{W\lambda}$  in the case of degeneracy,  $\Sigma_{W\nu,1} = 0$  from Theorem 1. If the bias is non-negligible, it distorts the coverage probability of the confidence interval. Correcting the bias part is desirable as it is generally unknown whether the degeneracy occurs. Fortunately, given the similar asymptotic distributional result as Kyriazidou (1997) in the degenerate case, we can use her bias correction strategy as follows.

Note that  $h_{n,\delta}^{-(k+1)}(\hat{\beta}_{n,\delta} - \beta)$  directly estimates the asymptotic bias from Theorem 1. We can construct a bias-corrected estimator  $\hat{\beta}_{n,bc}(\beta)$  by subtracting this bias estimator from the original estimator with suitable normalization: Let  $r_{n,\delta} = N^{(1-\delta)/(2k+3)}$ . The bias-corrected estimator is

given by

$$\hat{\beta}_{n,bc}(\beta) = \hat{\beta}_n - r_{n,\delta}^{-(k+1)}(\hat{\beta}_{n,\delta} - \beta).$$

We can check that this estimator is asymptotically unbiased regardless of the degeneracy: When  $c'_W \Sigma_{W\nu,1} c_W > 0$ ,

$$\begin{aligned} \sqrt{n}c'(\hat{\beta}_{n,bc}(\beta) - \beta) &= \sqrt{n}c'(\hat{\beta}_n - \beta) - \sqrt{nr}r_{n,\delta}^{-(k+1)}c'(\hat{\beta}_{n,\delta} - \beta) \\ &= \underbrace{\sqrt{n}(\hat{\beta}_n - \beta)}_{\rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,1} c_W)} - \underbrace{\sqrt{n}h^{k+1}N^{-(k+1)/(2k+3)}}_{\rightarrow 0} \underbrace{h_{n,\delta}^{-(k+1)}c'(\hat{\beta}_{n,\delta} - \beta)}_{\rightarrow_p c'_W \Sigma_{W\lambda}} \\ &\rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,1} c_W), \end{aligned}$$

as  $n \rightarrow \infty$ . When  $c'_W \Sigma_{W\nu,1} c_W = 0$ ,

$$\begin{aligned} \sqrt{Nh_n}c'(\hat{\beta}_{n,bc}(\beta) - \beta) &= \sqrt{Nh_n}c'(\hat{\beta}_n - \beta) - \sqrt{Nh_n}h_{n,1-\delta}^{-(k+1)}c'(\hat{\beta}_{n,\delta} - \beta) \\ &= \underbrace{\sqrt{Nh_n}c'(\hat{\beta}_n - \beta)}_{\rightarrow_d \mathcal{N}(\sqrt{h^{2k+3}}c'_W \Sigma_{W\lambda}, c'_W \Sigma_{W\nu,2} c_W)} - \underbrace{\sqrt{h}h_{n,\delta}^{-(k+1)}c'(\hat{\beta}_{n,\delta} - \beta)}_{\rightarrow_p \sqrt{h^{2k+3}}c'_W \Sigma_{W\lambda}} \\ &\rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,2} c_W), \end{aligned}$$

as  $n \rightarrow \infty$ . Thus, given the adaptivity of  $\hat{\Sigma}$  to the degeneracy, we have

$$(c'\hat{\Sigma}c)^{-1/2}c'(\hat{\beta}_{n,bc}(\beta) - \beta) \rightarrow_d \mathcal{N}(0, 1),$$

as  $n \rightarrow \infty$  for an arbitrary non-zero vector  $c \in \mathbb{R}^{qw}$ .

Then, we can construct the bias-corrected confidence interval as follows: Letting  $\Phi_{1-\alpha/2}^{-1}$  be  $1 - \alpha/2$  quantile of the standard normal distribution, we have

$$\begin{aligned} -\Phi_{1-\alpha/2}^{-1} &\leq (c'\hat{\Sigma}c)^{-1/2}c'(\hat{\beta}_{n,bc}(\beta) - \beta) \leq \Phi_{1-\alpha/2}^{-1} \\ \iff -(c'\hat{\Sigma}c)^{-1/2}\Phi_{1-\alpha/2}^{-1} &\leq c'\hat{\beta}_n - h_{n,1-\delta}^{-(k+1)}c'\hat{\beta}_{n,\delta} - (1 - h_{n,1-\delta}^{-(k+1)})c'\beta \leq (c'\hat{\Sigma}c)^{-1/2}\Phi_{1-\alpha/2}^{-1} \\ \iff CI_{L,\alpha,c} &\leq c'\beta \leq CI_{U,\alpha,c}, \end{aligned}$$

where

$$\begin{aligned} CI_{L,\alpha,c} &\equiv (1 - h_{n,1-\delta}^{-(k+1)})^{-1} \left[ c'\hat{\beta}_n - h_{n,1-\delta}^{-(k+1)}c'\hat{\beta}_{n,\delta} - (c'\hat{\Sigma}c)^{-1/2}\Phi_{1-\alpha/2}^{-1} \right], \\ CI_{U,\alpha,c} &\equiv (1 - h_{n,1-\delta}^{-(k+1)})^{-1} \left[ c'\hat{\beta}_n - h_{n,1-\delta}^{-(k+1)}c'\hat{\beta}_{n,\delta} + (c'\hat{\Sigma}c)^{-1/2}\Phi_{1-\alpha/2}^{-1} \right]. \end{aligned}$$

The full inference procedure is summarized as follows:

- (1) Compute the first step estimator  $\hat{\gamma}_n$ .
- (2) Choose  $k \geq 2, \delta \in (0, \frac{2k+3}{4k+4})$ , and  $h > 0$  to compute  $\hat{\beta}_n$  and  $\hat{\beta}_{n,\delta}$  with bandwidths  $h_n = hN^{-1/(2k+3)}$  and  $h_{n,\delta} = hN^{-\delta/(2k+3)}$ , respectively.
- (3) Compute  $\hat{\Sigma}$  and  $h_{n,\delta}^{-(k+1)}(\hat{\beta}_{n,\delta} - \hat{\beta}_n)$  to estimate the asymptotic variance and bias and obtain  $\hat{h}^*$ .

- (4) Update  $\hat{\beta}_n$  and  $\hat{\beta}_{n,\delta}$  with bandwidths  $h_n = \hat{h}^* N^{-1/(2k+3)}$  and  $h_{n,\delta} = \hat{h}^* N^{-\delta/(2k+3)}$ , respectively.
- (5) Construct the confidence interval by computing  $CI_{L,\alpha,c}$  and  $CI_{U,\alpha,c}$  from  $\hat{\beta}_n, \hat{\beta}_{n,\delta}$ , and  $c'\hat{\Sigma}c$ .

### 3.6. First-step Estimator.

Remember that we want to estimate  $\gamma$  from the selection equation or network formation process (2.3):

$$d_{ijt} = \mathbf{1}\{R'_{ijt}\gamma + B_i + B_j - \eta_{ijt} \geq 0\}.$$

This DGP can be interpreted as a panel discrete choice model as well as a network formation model. Estimators for discrete choice models such as Chamberlain (1980), Manski (1987), or Horowitz (1992) can be candidates for estimating  $\gamma$ . Also, estimators for network formation models such as Graham (2017) or Candelaria (2020) can be applicable under additional conditions.

Whether those estimators can be used as our first-step estimator  $\hat{\gamma}_n$  boils down to their convergence rates: Recall that Assumption 10 requires that  $\sqrt{N}h_n(\hat{\gamma}_n - \gamma) = o_p(1)$ , which implies that the first-step estimator needs to converge faster than  $\hat{\beta}_n$ . We can conjecture that, without additional conditions on  $\eta_{ijt}$ , the convergence rates of those estimators are  $\sqrt{n}$  in worst cases due to the conditional dependence across dyads. Obviously,  $\sqrt{n}$ -rate is incompatible with Assumption 10. In the following, we discuss what kind of additional conditions are needed to ensure Assumption 10.

We may assume additive separability for  $\eta_{ijt}$ :  $\eta_{ijt} = V_{it} + V_{jt} + V_{ijt}$ , where conditionally on  $\{\xi_i\}_{i=1}^n, (V_{i1}, V_{i2}), i = 1, \dots, n$  is independent,  $(V_{ij1}, V_{ij2}), 1 \leq i < j \leq n$  is independent, and both are mutually independent. This assumption is weaker than assuming  $(\eta_{ij1}, \eta_{ij2}), 1 \leq i < j \leq n$  is conditionally independent given  $\{\xi_i\}_{i=1}^n$ , where  $V_{it}$  is treated as degenerate. With additional conditions, we can directly apply Graham (2017)'s joint maximum likelihood estimator or Candelaria (2020)'s semiparametric estimator, both of which leverage the cross-sectional variation in  $d_{ijt}$  and  $R_{ijt}$ . We can show that in our setting (especially Assumptions 1 and 6), the limiting networks are dense, which implies that both Graham (2017) and Candelaria (2020)'s estimators satisfy  $\sqrt{N}(\hat{\gamma}_n - \gamma) = O_p(1)$  and Assumption 10.

Alternatively, we may assume that  $(\eta_{ij1}, \eta_{ij2}), 1 \leq i < j \leq n$  is conditionally independent given  $\{\xi_i\}_{i=1}^n$ . Graham (2017) and Candelaria (2020)'s estimators still satisfy Assumption 10, but we can also show that Chamberlain (1980)'s conditional logit estimator and Horowitz (1992)'s smoothed maximum score estimator can satisfy Assumption 10. Under the conditional independence assumption, the latter two estimators can be written in an asymptotically locally linear form where the corresponding influence function is indexed by  $ij$  with 0 covariances. Thus, the



convergence rates are based on  $N$  and Assumption 10 can be satisfied depending on the tuning parameters.

## 4. EXTENSION

### 4.1. Directed Graph with Multiple Periods.

In the above analysis, we restricted our attention to an undirected graph; the variables are all symmetric with respect to nodes (e.g.,  $Y_{ijt} = Y_{jit}$ ). Also, there were only two time periods,  $t = 1, 2$ . The extension to a directed graph case with  $t = 1, \dots, T$  ( $T \geq 2$ ) is straightforward; Letting  $\Delta_{st}A \equiv A_s - A_t$  denote the time difference between  $s$  and  $t$ , we propose the following estimator:

$$\hat{\beta}_n = \left[ \sum_{s<t} \sum_{i=1}^n \sum_{j \neq i} d_{ijs} d_{ijt} \Delta_{st} W_{ij} \Delta_{st} W'_{ij} K_{h_n}(\Delta_{st} R'_{ij} \hat{\gamma}_n) \right]^{-1} \\ \times \left[ \sum_{s<t} \sum_{i=1}^n \sum_{j \neq i} d_{ijs} d_{ijt} \Delta_{st} W_{ij} \Delta_{st} Y_{ij} K_{h_n}(\Delta_{st} R'_{ij} \hat{\gamma}_n) \right].$$

All the results and their proofs are valid with some modification because we can always rewrite the double sum  $\sum_{i=1}^n \sum_{j \neq i} A_{ij}$  as  $\sum_{i<j} (A_{ij} + A_{ji})$  for any variables  $\{A_{ij}\}$ . We will use this version of the estimator in our empirical application.

### 4.2. Pairwise Fixed Effects.

In the model (2.2) and (2.3), all the fixed effects are node-wise. Since we are interested in coefficients on time-varying dyadic variables, it is possible to include pairwise fixed effects  $A_{ij}$  and  $B_{ij}$  in each equation, additionally to  $A_i, A_j$  and  $B_i, B_j$ . Clearly, with pairwise fixed effects, the identification and estimator will be the same as with node-wise fixed effects since we are leveraging the time variation. Thus, a similar asymptotic analysis will also hold as long as  $(A_{ij}, B_{ij}), 1 \leq i < j \leq n$  are independently distributed conditionally on  $\{\xi_i\}_{i=1}^n$ .

Alternatively, we can also do away with the additive separability by incorporating node-wise fixed effects into pairwise ones:

$$A_{ij} = \tilde{\tau} \left( \tilde{A}_i, \tilde{A}_j, \tilde{A}_{ij} \right),$$

where  $\tilde{\tau}$  is some unknown function,  $\tilde{A}_i$  is a node-wise fixed effect, and  $\tilde{A}_{ij}$  is a pairwise fixed effect. We can impose a similar structure for  $B_{ij}$ . Again, the asymptotic analysis will hold as long as  $(\tilde{A}_{ij}, \tilde{B}_{ij}), 1 \leq i < j \leq n$  are conditionally independent. With a more general dependence structure, we could show a similar asymptotic result using Kojevnikov et al. (2021)'s central limit theorem for  $\psi$ -dependent data.

### 4.3. Sparsity.

Above, we argue that our model and assumptions imply that the limiting networks  $D_1$  and  $D_2$  are locally dense around  $\Delta R'_{ij} \gamma \sim 0$ . Thus, we limit our attention to cases where the

number of dyads in the sample must be proportional to  $N$ . Our modeling is appropriate in some applications, such as trade or migration, where the number of dyads is rather dense. However, ours can be inappropriate for some applications where the networks are sparse such as employee-employer, bank-firm matched data (e.g., Abowd et al. (1999), Jiménez et al. (2014)).

We can accommodate sparse networks by the following modification; let us modify Assumption 1 so that  $\xi_i, i = 1, \dots, n$  are drawn from some distribution that is allowed to depend on  $n$ . For example, as argued in Graham (2017), we can consider a distribution where the fixed effects are such that  $\liminf_{1 \leq i \leq n} B_i = -\infty$ . Then, we can discuss identification and estimation with fixed  $n$ , and the moments of interest are all dependent on  $n$ . Especially, we can consider the sequence of networks such that  $Pr(d_{121}d_{122} = 1 | \Delta R'_{12}\gamma = 0) \rightarrow 0$  and  $r_n Pr(d_{121}d_{122} = 1 | \Delta R'_{12}\gamma = 0) = \Omega(1)$  for some  $r_n \rightarrow \infty$  to incorporate sparsity. We do not pursue sparsity in this paper and leave it for future projects.

## 5. SIMULATION

To see the performance of the estimator, we conduct some simulation exercises. Consider the following data-generating process:

$$\begin{aligned} W_{ijt} &= X_{it} + X_{jt}, R_{it} = (W_{ijt}, Z_{it} + Z_{jt})', \\ A_i &= \frac{X_{i1} + X_{i2}}{2}, B_i = \frac{Z_{i1} + Z_{i2}}{2}, \\ d_{ijt} &= \mathbf{1}\{R'_{ijt}(1, 1)' + \theta \times (B_i + B_j) - \eta_{ijt} \geq 0\}, \\ Y_{ijt} &= d_{ijt}(W_{ijt} + A_i + A_j + \epsilon_{ijt}) \end{aligned}$$

where

$$\begin{aligned} X_{it}, Z_{it} &\sim \mathcal{N}(2, 1), i.i.d. \text{ across } i, t, \\ \eta_{ijt} &\sim \text{Logistic}(0, 1), i.i.d. \text{ across } ij, t, \\ \epsilon_{ijt} &= U_{it} + U_{jt} + \eta_{ijt}, \text{ where } U_{it} \sim \mathcal{N}(0, \sigma), i.i.d. \text{ across } i, t. \end{aligned}$$

Note that  $\beta = 1$  and  $\gamma = (1, 1)'$ . We have  $\theta \in \{-0.3, -2.0, -3.0\}$  inside of  $d_{ijt}$  to control for the fraction of zeros in the simulated data set:

$$Pr(d_{121} \times d_{122} = 0) \sim \begin{cases} 20\% & \text{if } \theta = -0.3 \\ 75\% & \text{if } \theta = -2.0 \\ 90\% & \text{if } \theta = -3.0 \end{cases} .$$

We also change  $\sigma \in \{0.0, 1.0\}$  for  $U_{it}$  so that  $\sigma = 0.0$  ( $\sigma = 1.0$ ) corresponds to the degenerate (non-degenerate) case.

As described above, we can interpret this data-generating process as a log-linearized version of the canonical gravity model (Head and Mayer (2014)); by writing  $\tilde{Y}_{ijt}$  as an observable outcome,

we redefine the main equation as

$$\tilde{Y}_{ijt} = \exp(W_{ijt} + A_i + A_j) \times \underbrace{\eta_{ijt}}_{=d_{ijt}\exp(\epsilon_{ijt})}.$$

We can take a log and recover the original model for a unit with  $d_{ijt} = 1$ . This modeling allows a mass at  $\tilde{Y}_{ijt} = 0$ , one important feature of dyadic data.

We conduct experiments for  $n \in \{50, 100, 150, 200\}$ ,  $\theta \in \{-0.3, -2.0, -3.0\}$ , and  $\sigma \in \{0.0, 1.0\}$ , and iterate 2000 times for each one. We calculate  $\hat{\gamma}_n$  by Chamberlain (1980)'s conditional logit estimator:

$$\hat{\gamma}_n = \underset{g \in \mathcal{G}}{\operatorname{argmax}} \sum_{i < j: d_{ij1} + d_{ij2} = 1} M_{ij}(g)$$

where  $\mathcal{G}$  is a compact subset of  $\mathbb{R}^{q_r}$  and

$$M_{ij}(g) = \mathbf{1}\{d_{ij1} = 1\} \ln \left( \frac{\exp(\Delta R'_{ij} g)}{1 + \exp(\Delta R'_{ij} g)} \right) + \mathbf{1}\{d_{ij2} = 1\} \ln \left( \frac{1}{1 + \exp(\Delta R'_{ij} g)} \right).$$

For  $\hat{\beta}_n$ , we use a biweight kernel for  $K(\cdot)$ , given by  $K(x) = 15/16(1 - x^2)^2 \mathbf{1}\{|x| \leq 1\}$ . This choice implies that we assume that the smoothness of the model is given by  $k = 2$ . We set  $\delta = 0.4$  and  $h = 3.0$  and calculate each estimator and confidence interval according to the inference procedure discussed above.

For comparison, we calculate the fixed effect estimator  $\hat{\beta}_{FE}$  given by (3.1). The standard error is calculated by  $\hat{\Sigma}$ , with  $K_{h_n}(\cdot)$  replaced by 1. We also calculate the Poisson pseudo-maximum-likelihood (PPML) estimator  $\hat{\beta}_{PPML}$  given by (3.2). We compute  $\hat{\beta}_{PPML}$  and its standard error by the *penppml* package in *R* (Ferrerias Garrucho and Zylkin, 2023). The standard error is clustered at the node level, which is close to  $\hat{\Sigma}_{WW}^{-2} \hat{\Sigma}_{W\nu,1}$  in our setting (Graham, 2020).

The result is summarized in the following TABLE 1 and 2. In TABLE 1, we evaluate the three estimators by mean and median biases (MeanBias), root mean square error (RMSE) for  $\sigma = 0, 1$ . In TABLE 2, we compute 95% coverage probabilities (Coverage) of four different confidence intervals:  $CI_{conv}$  (conventional CI from  $\hat{\beta}_n$  and  $\hat{\Sigma}$ ),  $CI_{bc}$  (bias-corrected CI given by  $CI_{L,0.05}$  and  $CI_{U,0.05}$ ),  $CI_{FE}$  (conventional CI from  $\hat{\beta}_{FE}$  and  $\hat{\Sigma}$  with a flat kernel.), and  $CI_{PPML}$  (conventional CI from  $\hat{\beta}_{PPML}$  and its node-level clustered standard error).

From TABLE 1a and 1b, we can see that our estimator performs better than the fixed effect estimator and the PPML estimator in terms of bias, which shows that the weights given by the first step estimator work well in eliminating the bias. Our estimator also outperforms the competitors regarding RMSE, which implies that the loss in precision is not severe. Our estimator also performs well even when there is a large fraction of zeros in  $Y$  ( $Pr(D_{ij1} \times D_{ij2}) \sim 90\%$  when  $\theta = -3.0$ ). There is little difference between  $\sigma = 0$  and  $\sigma = 1$  other than added variances in the estimators.

From TABLE 2a and 2b, we can see that  $CI_{bc}$  is close to 95% regardless of the degeneracy ( $\Sigma_{W\nu,1} = 0$  or  $> 0$ ) while the others are off from the targeted nominal coverage. This result confirms the effectiveness of the bias correction strategy as well as the adaptivity of our variance estimator, as claimed in Section 3.3. Also, it is notable to see that the bias correction is important for obtaining correct coverage probabilities even though the asymptotic bias is 0 in the case of  $\sigma = 1.0$  so that  $\Sigma_{W\nu,1} > 0$  (Theorem 1) and  $CI_{conv}$  would return an asymptotically correct coverage.

TABLE 1. Finite sample properties of  $\hat{\beta}_n$ ,  $\hat{\beta}_{FE}$ , and  $\hat{\beta}_{PPML}$

(A)  $\sigma = 1.0$

$\theta$	$n$	MeanBias			RMSE		
		$\hat{\beta}_n$	$\hat{\beta}_{FE}$	$\hat{\beta}_{PPML}$	$\hat{\beta}_n$	$\hat{\beta}_{FE}$	$\hat{\beta}_{PPML}$
-0.3	50	0.045	0.133	0.185	0.122	0.160	0.430
-2.0	50	0.141	0.352	0.467	0.210	0.377	0.617
-3.0	50	0.162	0.369	0.582	0.273	0.415	0.752
-0.3	100	0.038	0.136	0.195	0.087	0.148	0.376
-2.0	100	0.099	0.349	0.438	0.142	0.359	0.536
-3.0	100	0.117	0.359	0.542	0.184	0.378	0.657
-0.3	150	0.028	0.135	0.193	0.070	0.143	0.327
-2.0	150	0.075	0.346	0.427	0.112	0.353	0.496
-3.0	150	0.095	0.356	0.527	0.145	0.367	0.607
-0.3	200	0.024	0.134	0.193	0.060	0.140	0.305
-2.0	200	0.061	0.344	0.417	0.091	0.348	0.471
-3.0	200	0.076	0.352	0.510	0.118	0.360	0.572

(B)  $\sigma = 0.0$

$\theta$	$n$	MeanBias			RMSE		
		$\hat{\beta}_n$	$\hat{\beta}_{FE}$	$\hat{\beta}_{PPML}$	$\hat{\beta}_n$	$\hat{\beta}_{FE}$	$\hat{\beta}_{PPML}$
-0.3	50	0.048	0.134	0.194	0.082	0.142	0.399
-2.0	50	0.140	0.352	0.468	0.176	0.365	0.586
-3.0	50	0.161	0.369	0.581	0.229	0.397	0.714
-0.3	100	0.037	0.135	0.193	0.053	0.138	0.332
-2.0	100	0.093	0.348	0.438	0.110	0.352	0.508
-3.0	100	0.113	0.359	0.546	0.145	0.368	0.630
-0.3	150	0.028	0.135	0.191	0.039	0.136	0.301
-2.0	150	0.071	0.345	0.427	0.082	0.348	0.477
-3.0	150	0.089	0.354	0.529	0.108	0.359	0.592
-0.3	200	0.024	0.135	0.191	0.031	0.136	0.278
-2.0	200	0.058	0.345	0.415	0.067	0.347	0.451
-3.0	200	0.074	0.355	0.508	0.089	0.358	0.553

TABLE 2. 95% coverage probabilities of  $CI_{conv}$ ,  $CI_{bc}$ ,  $CI_{FE}$ , and  $CI_{PPML}$

(A)  $\sigma = 1.0$

$\theta$	n	Coverage			
		$CI_{conv}$	$CI_{bc}$	$CI_{FE}$	$CI_{PPML}$
-0.3	50	0.790	0.961	0.498	0.537
-2.0	50	0.646	0.963	0.150	0.236
-3.0	50	0.640	0.901	0.311	0.211
-0.3	100	0.785	0.978	0.233	0.498
-2.0	100	0.668	0.970	0.011	0.173
-3.0	100	0.674	0.953	0.072	0.143
-0.3	150	0.790	0.971	0.103	0.472
-2.0	150	0.689	0.949	0.001	0.117
-3.0	150	0.688	0.944	0.016	0.09
-0.3	200	0.817	0.964	0.040	0.426
-2.0	200	0.730	0.947	0.000	0.08
-3.0	200	0.720	0.946	0.004	0.08

(B)  $\sigma = 0.0$

$\theta$	n	Coverage			
		$CI_{conv}$	$CI_{bc}$	$CI_{FE}$	$CI_{PPML}$
-0.3	50	0.698	0.918	0.141	0.547
-2.0	50	0.535	0.935	0.026	0.204
-3.0	50	0.592	0.869	0.168	0.182
-0.3	100	0.655	0.960	0.001	0.515
-2.0	100	0.482	0.958	0.000	0.12
-3.0	100	0.571	0.944	0.004	0.106
-0.3	150	0.673	0.977	0.000	0.45
-2.0	150	0.471	0.945	0.000	0.073
-3.0	150	0.532	0.949	0.001	0.065
-0.3	200	0.660	0.970	0.000	0.407
-2.0	200	0.444	0.939	0.000	0.052
-3.0	200	0.520	0.933	0.000	0.047

## 6. EMPIRICAL EXAMPLE

### 6.1. Background.

As a leading application of our model, consider Moretti and Wilson (2017). They study how state-level tax differences affect migration by top scientists in the U.S. Specifically, they estimate

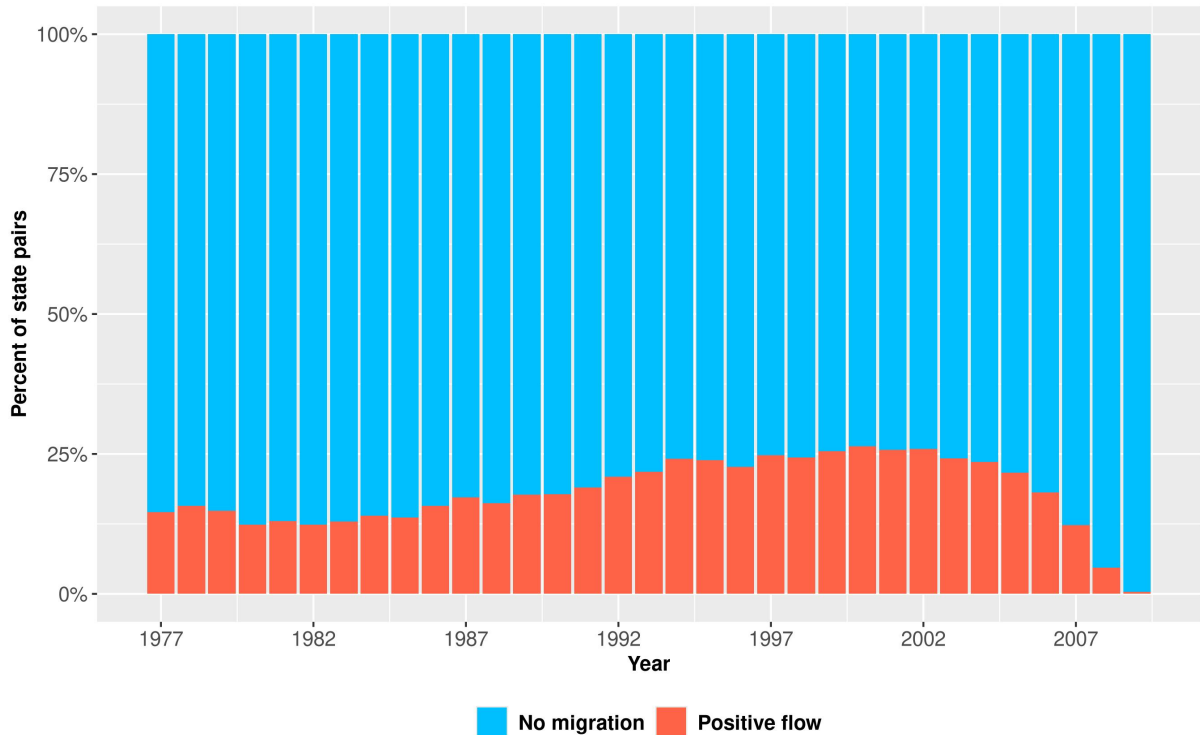
the following model implied by their economic theory:

$$\begin{aligned} \log(P_{ijt}/P_{iit}) = & \eta [\log(1 - \tau_{jt}) - \log(1 - \tau_{it})] \\ & + \eta' [\log(1 - \tau'_{jt}) - \log(1 - \tau'_{it})] + \gamma_j + \gamma_i + u_{ijt}, \end{aligned}$$

where  $P_{ijt}$  is the number of scientists migrating to state  $j$  from state  $i$  at year  $t$ ,  $\tau_{it}$  and  $\tau'_{it}$  are personal and corporate taxes imposed in state  $i$  at year  $t$ ,  $\gamma_i$  is a state fixed effect, and  $u_{ijt}$  is an error term.

Note that if there is no migration from  $i$  to  $j$  at year  $t$ ,  $P_{ijt} = 0$  and  $\log(P_{ijt}/P_{iit})$  is undefined. In Moretti and Wilson (2017)'s dataset, more than 70% of state-pairs exhibit no migration flow:

FIGURE 6.1. Fraction of positive migration flows in Moretti and Wilson (2017)'s dataset. The migration flow is positive in a given year if there is at least one scientist moving from state  $i$  to  $j$  (scientists are "star"; They are at or above 95% quantile in number of patents over the past ten years)



When running a regression, they are concerned with a potential sample selection bias stemming from these undefined outcomes. They argue that if the main regressors are not systemically associated with whether there is positive migration flow or not, the selection bias should be minimal. Running OLS on the linear probability model, they find little correlation between the main regressors and no flow. Recalculating their regression with our estimator helps check the validity of their argument and the appropriateness of using the linear probability model.

When applying our model to their context, we must consider what  $R_{ijt}$  should be. Since Moretti and Wilson (2017)'s underlying theory is based on scientists' and firms' discrete choice, one consistent way to generate zero migration flows between some states is to consider endogenous choice sets as in Dubé et al. (2021). Formally, we can write the choice set of representative scientists in state  $i$  as  $C_{it} = \{j \in \{1, \dots, 51\} : d_{ijt} = 1\}$ . Here,  $\{d_{ijt}\}$  represents the job-market network; if  $d_{ijt} = 1$ , it is possible to move from  $i$  to  $j$ , and vice versa. We can attribute the determinants of the network to the utilities and profits of scientists and firms, as well as the matching costs between the two parties. Such costs are not present in the structural equation if those costs are not compensated through wages; The structural equation consists of the determinants of log wage differences between two states. Thus in the selection equation (2.3), in addition to  $W_{ijt}$ , we can include variables in  $R_{ijt}$  that capture non-monetary matching costs between two states  $i$  and  $j$ , which does not violate Assumption 3 as  $R_{ijt}$  satisfies the exclusion restriction.

## 6.2. Implementation.

For  $W_{ijt}$ , as in Moretti and Wilson (2017), we include the state-to-state differences in (i) an individual income average income tax rate (ATR) faced by a hypothetical taxpayer at 99% quantile of the national income distribution, (ii) the corporate tax rate (CIT), (iii) the investment tax credit (ITC), and (iv) the R&D tax credit (R&D credit). This is the same set of regressors as Moretti and Wilson (2017)'s baseline regression. For  $R_{ijt}$ , we use  $W_{ijt}$  plus state-to-state difference in the logarithm of population (POP) and a dummy variable that indicates whether  $i$  and  $j$  share their governors' political parties (GOV). The additional variables in  $R_{ijt}$  arguably measure non-monetary costs of connecting firms and workers in two states.

We implement the first step estimation as follows. We use the conditional logit estimator extended to a directed graph with multi-periods case, which is given by

$$\hat{\gamma}_n = \underset{g \in \mathcal{G}}{\operatorname{argmax}} \sum_{s < t} \sum_{i, j} M_{ij, st}(g),$$

where  $\mathcal{G}$  is a compact subset of  $\mathbb{R}^{q_r}$  and

$$M_{ij, st}(g) = \mathbf{1}\{d_{ijs} + d_{ijt} = 1\} \times [\mathbf{1}\{d_{ijs} = 1\} \ln(e_{ij, st}) + \mathbf{1}\{d_{ijt} = 1\} \ln(1 - e_{ij, st})],$$

$$e_{ij, st} = \frac{\exp(\Delta_{st} R'_{ij} g)}{1 + \exp(\Delta_{st} R'_{ij} g)}.$$

TABLE 3 reports the first step estimation result. We can see that the coefficients on the newly added variables  $GOV$  and  $POP$  deviate from zero, which implies that a part of the identification assumptions (Assumption 3) is satisfied. Also, for these two variables, the estimated coefficients imply that the job market network exhibits homophily; similar states are more likely to be connected.

TABLE 3. First Step Estimation Result

Variable	$\hat{\gamma}_n$
GOV	0.162
POP	-13.144
ATR	8.561
CIT	5.850
ITC	5.350
R&D credit	-0.230

For the second step estimator, we use our  $\hat{\beta}_n$  defined above and extend it to the directed graph with multiple period cases, as discussed in Section 4. We use a biweight kernel for  $K(\cdot)$  (so  $k = 2$ ), choose  $h = 3.0$  as an initial constant for pilot bandwidths, and use  $\delta = 0.4$  for calculating  $h_{n,\delta}$ . We extend and use  $\hat{\Sigma}$  to calculate the standard error while taking into account the correlation across time (case 1). We also calculate the bias-corrected 95%-confidence interval by computing  $CI_{L,0.05}$  and  $CI_{U,0.05}$  as defined above. Also, we list  $\hat{\beta}_{MW}$ , an estimate from Moretti and Wilson (2017) (page 1883, TABLE 2A, specification (3)) and calculate the conventional 95%-confidence interval based on their standard errors. Note that  $\hat{\beta}_{MW}$  is the fixed estimator, where its standard error is calculated by clustering across time and the origin, destination, and origin-destination pairs.

We summarize the result in TABLE 4. We can see that  $\hat{\beta}_n$  returns similar values as  $\hat{\beta}_{WM}$ , which claims the robust positive effect of income and corporate-related tax differences on migration. Thus, Moretti and Wilson (2017)'s estimates are not likely to be qualitatively affected by the sample selection effects. However, while Moretti and Wilson (2017)'s estimates are statistically significant at 5% level, our confidence intervals show that all of the estimates are no longer statistically significant at that level except for ITC. Our insignificance result is driven by both the increase in standard errors <sup>5</sup> and the asymptotic bias correction. Thus, our exercise shows that some of the results in Moretti and Wilson (2017) may not be robust to the presence of sample selection due to the endogeneity of the job market network.

<sup>5</sup>Our standard error is from  $\hat{\Sigma}$ , which takes fully into account the dependence among pairs that share origin and destination, such as California→Wisconsin and New York→California. Moretti and Wilson (2017)'s standard error calculation ignores such dependence structure.



TABLE 4. Comparison of our estimator and Moretti and Wilson (2017)

Variable	(A) This paper		(B) Moretti and Wilson (2017)	
	Estimator (s.e)	CI	Estimator (s.e)	CI
ATR	1.634 (1.886)	[−2.656, 5.872]	1.926 (0.514)	[0.918, 2.933]
CIT	1.666 (2.040)	[−2.948, 6.276]	1.840 (0.588)	[0.687, 2.992]
ITC	1.980 (0.584)	[0.651, 3.290]	1.793 (0.411)	[0.987, 2.598]
R&D credit	0.429 (0.780)	[−1.352, 2.117]	0.368 (0.182)	[0.011, 0.724]

## 7. CONCLUSION

This paper studies identification and inference of a panel dyadic data sample selection model. We show that Kyriazidou (1997)’s identification strategy can be extended to our dyadic data setting, and we prove asymptotic normality of the proposed estimator.

Our estimator has some appealing properties. The distributional result implies that our estimator has the same convergence rates as the usual estimators used in practice in the non-degenerate case, and there is no loss of effective sample size for using our nonparametric type estimator. Also, our estimator is guaranteed to be asymptotically normal, while others can be non-Gaussian in the limit.

We also provide consistent estimators for asymptotic bias and variance that adapts to the degeneracy. Specifically, the bias corrected confidence interval has an asymptotically correct size. Our simple simulation exercise confirms the validity of these estimators and highlights the importance of bias correction in both degenerate and non-degenerate cases.

## REFERENCES

- ABOWD, J. M., F. KRAMARZ, D. N. MARGOLIS, B. Y. J. M. ABOWD, F. KRAMARZ, AND D. N. MARGOLIS (1999): “High Wage Workers and High Wage Firms,” *Econometrica*, 67, 251–333.
- AHN, H. AND J. L. POWELL (1993): “Semiparametric estimation of censored selection models with a nonparametric selection mechanism,” *Journal of Econometrics*, 58, 3–29.
- AUERBACH, E. (2022): “Identification and Estimation of a Partially Linear Regression Model Using Network Data,” *Econometrica*, 90, 347–365.
- BONHOMME, S. (2020): “Econometric analysis of bipartite networks,” *The Econometric Analysis of Network Data*.
- CAMERON, A. C. AND D. MILLER (2014): “Robust Inference for Dyadic Data,” *Unpublished manuscript*.
- CANDELARIA, L. E. (2020): “A Semiparametric Network Formation Model with Unobserved Linear Heterogeneity,” .
- CHAMBERLAIN, G. (1980): “Analysis of Covariance with Qualitative Data,” *The Review of Economic Studies*, 47, 225–238.
- DAVEZIES, L., X. D’HAULTFŒUILLE, AND Y. GUYONVARCH (2021): “Empirical process results for exchangeable arrays,” *The Annals of Statistics*, 49, 845 – 862.
- DUBÉ, J. P., A. HORTAÇSU, AND J. JOO (2021): “Random-coefficients logit demand estimation with zero-valued market shares,” *Marketing Science*, 40.
- FAFCHAMPS, M. AND F. GUBERT (2007): “The formation of risk sharing networks,” *Journal of Development Economics*, 83.
- FERRERAS GARRUCHO, D. AND T. ZYLKIN (2023): *penppml: Penalized Poisson Pseudo Maximum Likelihood Regression*, r package version 0.2.3.
- GRAHAM, B. S. (2017): “An Econometric Model of Network Formation With Degree Heterogeneity,” *Econometrica*, 85, 1033–1063.
- (2020): “Dyadic regression,” *The Econometric Analysis of Network Data*.
- GRAHAM, B. S., F. NIU, AND J. L. POWELL (2019): “Kernel Density Estimation for Undirected Dyadic Data,” .
- (2021): “Minimax Risk and Uniform Convergence Rates for Nonparametric Dyadic Regression,” .
- HALL, P. (1984): “Central limit theorem for integrated square error of multivariate nonparametric density estimators,” *Journal of Multivariate Analysis*, 14, 1–16.
- HEAD, K. AND T. MAYER (2014): “Gravity Equations: Workhorse, Toolkit, and Cookbook,” *Handbook of International Economics*, 4, 131–195.
- HECKMAN, J. (1979): “Sample Selection Bias as a Specification Error,” *Econometrica*, 47, 153–161.

- HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): “Estimating trade flows: Trading partners and trading volumes,” *Quarterly Journal of Economics*, 123, 441–487.
- HOEFFDING, W. (1961): “The strong law of large numbers for U-statistics,” .
- HOROWITZ, J. L. (1992): “A Smoothed Maximum Score Estimator for the Binary Response Model,” *Econometrica*, 60.
- JIMÉNEZ, G., S. ONGENA, J.-L. PEYDRÓ, AND J. SAURINA (2014): “Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk-Taking?” *Econometrica*, 82, 463–505.
- JOCHMANS, K. (2023): “Peer effects and endogenous social interactions,” *Journal of Econometrics*, 235, 1203–1214.
- JOHANSSON, I. AND H. R. MOON (2021): “Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach,” *The Review of Economics and Statistics*, 103, 328–345.
- KOJEVNIKOV, D., V. MARMER, AND K. SONG (2021): “Limit theorems for network dependent random variables,” *Journal of Econometrics*, 222, 882–908.
- KYRIAZIDOU, E. (1997): “Estimation of a Panel Data Sample Selection Model,” *Econometrica*, 65.
- MANSKI, C. F. (1987): “Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data,” *Econometrica*, 55, 357–362.
- MENZEL, K. (2021): “Bootstrap With Cluster-Dependence in Two or More Dimensions,” *Econometrica*, 89, 2143–2188.
- MONTE, F., S. J. REDDING, AND E. ROSSI-HANSBERG (2018): “Commuting, migration, and local employment elasticities,” *American Economic Review*, 108.
- MORETTI, E. AND D. J. WILSON (2017): “The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists,” *American Economic Review*, 107, 1858–1903.
- SILVA, J. M. S. AND S. TENREYRO (2006): “The log of gravity,” *Review of Economics and Statistics*, 88.
- TABORD-MEEHAN, M. (2019): “Inference With Dyadic Data: Asymptotic Behavior of the Dyadic-Robust t-Statistic,” *Journal of Business and Economic Statistics*, 37.
- WHITE, H. (2001): *Asymptotic Theory for Econometricians*, Academic Press.
- ZELENEEV, A. (2020): “Identification and Estimation of Network Models with Nonparametric Unobserved Heterogeneity,” [https://www.princeton.edu/~zeleneev/azeleneev\\_jmp.pdf](https://www.princeton.edu/~zeleneev/azeleneev_jmp.pdf).

APPENDIX A. PROOFS

**Proof of Theorem 1.**

*Proof.* First, consider the infeasible version of  $\beta_n$ , where  $\hat{\gamma}_n$  is replaced by true  $\gamma$ :

$$\tilde{\beta}_n = \beta + S_{WW}^{-1}S_{W\lambda} + S_{WW}^{-1}S_{W\nu},$$

where  $S_{WW}, S_{W\lambda}$ , and  $S_{W\nu}$  are the same as  $\hat{S}_{WW}, \hat{S}_{W\lambda}$ , and  $\hat{S}_{W\nu}$  except  $\hat{\gamma}_n$  replaced by  $\gamma$ . We use the following lemmas

**Lemma 1.** *Suppose Assumptions 1-9 hold. Then,*

$$S_{WW} \rightarrow_p \Sigma_{WW},$$

as  $n \rightarrow \infty$ .

**Lemma 2.** *Suppose Assumptions 1-9 hold. Fix some  $h \in (0, \infty)$ . If  $Nh_n^{2k+3} \rightarrow h$ , then*

$$\sqrt{Nh_n}S_{W\lambda} \rightarrow_p \sqrt{h}\Sigma_{W\lambda},$$

as  $n \rightarrow \infty$ . If  $Nh_n^{2k+3} \rightarrow \infty$  and  $nh_n^{2k+2} \rightarrow \infty$ , then

$$h_n^{-(k+1)}S_{W\lambda} \rightarrow_p \Sigma_{W\nu},$$

as  $n \rightarrow \infty$ .

**Lemma 3.** *Suppose Assumptions 1-9 hold. Fix an arbitrary nonzero vector  $c \in \mathbb{R}^{q_w}$  and some constant  $h \in (0, \infty]$ . Let  $c_W = \Sigma_{WW}^{-1}c$ . If  $c'_W \Sigma_{W\nu,1} c_W > 0$  and  $Nh_n^{2k+3} \rightarrow h$ , then*

$$\sqrt{n}c' S_{WW}^{-1} S_{W\nu} \rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,1} c_W),$$

as  $n \rightarrow \infty$ . If  $c'_W \Sigma_{W\nu,1} c_W = 0$  and  $Nh_n^{2k+3} \rightarrow h$ , then

$$\sqrt{Nh_n}c' S_{WW}^{-1} S_{W\nu} \rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,2} c_W),$$

as  $n \rightarrow \infty$ .

By combining Lemmas 1-3, the statement of Theorem 1 follows for  $\tilde{\beta}_n$ . The following lemmas are used to show the negligibility of  $\hat{\beta}_n - \tilde{\beta}_n$ :

**Lemma 4.** *Suppose Assumptions 1-10 hold. Fix some constant  $h \in (0, \infty]$ . If  $Nh_n^{2k+3} \rightarrow h$ , then,*

$$\hat{S}_{WW} = S_{WW} + o_p(1).$$

**Lemma 5.** *Suppose Assumptions 1-10 hold. Fix some constant  $h \in (0, \infty]$ . If  $Nh_n^{2k+3} \rightarrow h$ , then*

$$\hat{S}_{W\lambda} = S_{W\lambda} + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

**Lemma 6.** *Suppose Assumptions 1-10 hold. Fix some constant  $h \in (0, \infty]$ . If  $Nh_n^{2k+3} \rightarrow h$ , then*

$$\hat{S}_{W\nu} = S_{W\nu} + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

By combining Lemmas 4-6, we have

$$\hat{\beta}_n - \beta = \tilde{\beta}_n - \beta + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

Thus, the normalization  $r_n \in \{\sqrt{n}, \sqrt{Nh_n}, h_n^{-(k+1)}\}$  corresponding to each case results in

$$r_n(\hat{\beta}_n - \beta) = r_n(\tilde{\beta}_n - \beta) + o_p(1).$$

Since  $\tilde{\beta}_n$  satisfies the statement of Theorem 1, this completes the proof.  $\square$

**Proof of Proposition 1.**

*Proof.* We show the claim by the following steps.

*Step 1:*  $\hat{\Sigma}_{W\nu,2} \rightarrow_p \Sigma_{W\nu,2}$ .

By expanding  $K^2(\Delta R'_{ij}\hat{\gamma}_n/h_n)$  around  $\Delta R'_{ij}\gamma$ , we get

$$K^2(\Delta R'_{ij}\hat{\gamma}_n/h_n) = K^2(\Delta R'_{ij}\gamma/h_n) + 2\Delta R'_{ij}(\hat{\gamma}_n - \gamma)/h_n K'(c_{ij,n}^*/h_n)K(c_{12n}^*/h_n),$$

where  $c_{ij,n}^*$  is between  $\Delta R'_{ij}\gamma$  and  $\Delta R'_{ij}\hat{\gamma}_n$ . Then,

$$\begin{aligned} \hat{\Sigma}_{W\nu} &= \underbrace{\frac{1}{Nh_n} \sum_{i < j} K^2(\Delta R'_{12}\gamma/h_n) d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \Delta \hat{\varepsilon}_{ij}^2}_{D_{p1,1}} \\ &\quad + \underbrace{\frac{2}{Nh_n^2} \sum_{i < j} \Delta R'_{ij}(\hat{\gamma}_n - \gamma) K'(c_{ij,n}^*/h_n) K(c_{ij,n}^*/h_n) \Delta W_{ij} \Delta W'_{ij} \Delta \hat{\varepsilon}_{ij}^2}_{D_{p1,2}}. \end{aligned}$$

*Sub-Step 1:*  $D_{p1,1} \rightarrow_p \Sigma_{W\nu,2}$ .

Observe that

$$\begin{aligned} \Delta \hat{\varepsilon}_{ij}^2 - \nu_{ij}^2 &= \left(\Delta W'_{ij}(\beta - \hat{\beta}_n)\right)^2 + \lambda_{ij}^2 + 2\Delta W'_{ij}(\beta - \hat{\beta}_n)\lambda_{ij} \\ &\quad + 2\Delta W'_{ij}(\beta - \hat{\beta}_n)\nu_{ij} + 2\lambda_{ij}\nu_{ij}. \end{aligned}$$

Thus,

$$\begin{aligned}
D_{p1,1} &= \frac{1}{Nh_n} \sum_{i < j} K^2(\Delta R'_{ij}\gamma/h_n) d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \nu_{ij}^2 \\
&+ \frac{1}{Nh_n} \sum_{i < j} K^2(\Delta R'_{ij}\gamma/h_n) d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \left( \Delta W'_{ij} (\beta - \hat{\beta}_n) \right)^2 \\
&+ \frac{1}{Nh_n} \sum_{i < j} K^2(\Delta R'_{ij}\gamma/h_n) d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \lambda_{ij}^2 \\
&+ \frac{2}{Nh_n} \sum_{i < j} K^2(\Delta R'_{ij}\gamma/h_n) d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \Delta W'_{ij} (\beta - \hat{\beta}_n) \lambda_{ij} \\
&+ \frac{2}{Nh_n} \sum_{i < j} K^2(\Delta R'_{ij}\gamma/h_n) d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \Delta W'_{ij} (\beta - \hat{\beta}_n) \nu_{ij} \\
&+ \frac{2}{Nh_n} \sum_{i < j} K^2(\Delta R'_{ij}\gamma/h_n) d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \lambda_{ij} \nu_{ij}.
\end{aligned}$$

We call each term by  $D_{p1,1}^i$  for  $i = 1, \dots, 6$  that is corresponding to each row.

The first term  $D_{p1,1}^1$  converges to  $\Sigma_{W\nu}$ . Its expectation coincides with  $\Sigma_{W\nu}$  in the limit as

$$\begin{aligned}
E[D_{p1,1}^1] &= \frac{1}{h_n} \int E[d_{121} d_{122} \Delta W_{12} \Delta W'_{12} \nu_{12}^2 | \Delta R'_{12}\gamma = r] K^2(r/h_n) f_{R\gamma}(r) dr \\
&= \int E[d_{121} d_{122} \Delta W_{12} \Delta W'_{12} \nu_{12}^2 | \Delta R'_{12}\gamma = rh_n] K^2(r) f_{R\gamma}(rh_n) dr \\
&= \Sigma_{W\nu,2} + o(1),
\end{aligned}$$

where the last line holds by the dominated convergence theorem under Assumptions 4, 6, and 8.

For the variance, denoting each summand by  $D_{p1,1,ij}^1$  and for any vector  $a$  with  $\|a\| = 1$ , we have

$$\begin{aligned}
Var[\|D_{p1,1}^1\|] &\leq \frac{1}{Nh_n^2} E[\|D_{p1,1,12}^1\|^2] \\
&+ \frac{2(n-2)}{Nh_n^2} E[\|D_{p1,1,12}^1\| \times \|D_{p1,1,13}^1\|].
\end{aligned}$$

The first term in the right hand side is  $O(1/(Nh_n))$  because,

$$\begin{aligned}
E[\|D_{p1,1,12}^1\|^2] &\leq \int E[\|\Delta W_{12}\|^4 \nu_{12}^4 | \Delta R'_{12}\gamma = r] K^2(r/h_n) f_{R\gamma}(r) dr \\
&= h_n \int E[\|\Delta W_{12}\|^4 \nu_{12}^4 | \Delta R'_{12}\gamma = rh_n] K^2(r) f_{R\gamma}(rh_n) dr \\
&= O(h_n),
\end{aligned}$$

where the last line holds from Assumptions 4, 6, and 8. The second term on the right-hand side is  $O(1/n)$  because

$$\begin{aligned}
E[\|D_{p1,1,12}^1\| \times \|D_{p1,1,13}^1\|] &\leq E\left[\int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12} \gamma = r_1, \xi_1, U_1] \right. \\
&\quad \times E[\|\Delta W_{13}\|^2 \nu_{13}^2 | \Delta R'_{13} \gamma = r_1, \xi_1, U_1] \\
&\quad \times K^2(r_1/h_n) K^2(r_2/h_n) f_{R\gamma|\xi_1, U_1}(r_1) f_{R\gamma|\xi_1, U_1}(r_2) dr_1 dr_2 \left. \right] \\
&= h_n^2 E\left[\int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12} \gamma = r_1 h_n, \xi_1, U_1] \right. \\
&\quad \times E[\|\Delta W_{13}\|^2 \nu_{13}^2 | \Delta R'_{13} \gamma = r_1 h_n, \xi_1, U_1] \\
&\quad \times K^2(r_1) K^2(r_2) f_{R\gamma|\xi_1, U_1}(r_1 h_n) f_{R\gamma|\xi_1, U_1}(r_2 h_n) dr_1 dr_2 \left. \right] \\
&= O(h_n^2),
\end{aligned}$$

where the first line follows from Assumptions 4, 6, and 8. Thus,

$$Var[\|D_{p1,1}^1\|] = O\left(\frac{1}{Nh_n}\right) + O\left(\frac{1}{n}\right) = o(1).$$

This implies that  $D_{p1,1}^1 \rightarrow_p \Sigma_{W\nu,2}$  as  $n \rightarrow \infty$ .

The second term  $D_{p1,1}^2$  converges to 0. Observe that, as  $K$  is bounded by Assumption 8, for some absolute constant  $C > 0$ ,

$$\|D_{p1,1}^2\| \leq \frac{\|\beta - \beta_n\|^2}{h_n} \times \frac{C}{N} \sum_{i < j} \|\Delta W_{ij}\|^4.$$

Since  $E[\|\Delta W_{ij}\|^4] < \infty$  by Assumption 7, we can apply the law of large numbers for U-statistics (Hoeffding (1961)) to  $C/N \sum_{i < j} \|\Delta W_{ij}\|^4$ , which is  $O_p(1)$ . Also, since  $\|\beta - \beta_n\| = O_p(1/\sqrt{n})$  (which is the worst-case rate for the specified  $h_n$  by Theorem 1), we have  $\|\beta - \beta_n\|^2/h_n = O_p(1/(nh_n^2)) = o_p(1)$  as  $nh_n^2 \sim n \times n^{-2/(2k+3)} = n^{(2k+1)/(2k+3)}$  diverges. Thus,

$$\|D_{p1,1}^2\| = o_p(1) \times O_p(1) = o_p(1),$$

and  $D_{p1,1}^2 \rightarrow_p 0$  as  $n \rightarrow \infty$ .

The third term  $D_{p1,1}^3$  converges to 0. Observe that,

$$\begin{aligned}
E[\|D_{p1,1}^3\|] &\leq \int E[\|\Delta W_{12}^2 \Lambda_{12}^2 | \Delta R'_{12} \gamma = r\| r^2 K^2(r/h_n) f_{R\gamma}(r) dr \\
&= h_n^3 E[\|\Delta W_{12}^2 \Lambda_{12}^2 | \Delta R'_{12} \gamma = rh_n\| r^2 K^2(r) f_{R\gamma}(rh_n) dr \\
&= O(h_n^3)
\end{aligned}$$

where the last line follows from Assumptions 4, 6, and 8. Thus,  $E[\|D_{p1,1}^3\|] = o(1)$ . Observe that, by writing each summand of  $D_{p1,1}^3$  as  $D_{p1,1,ij}^3$ ,

$$Var[\|D_{p1,1}^3\|] \leq \frac{1}{Nh_n^2} E[\|D_{p1,1,12}^3\|^2] + \frac{2(n-2)}{Nh_n^2} E[\|D_{p1,1,12}^3\| \times \|D_{p1,1,13}^3\|].$$

The first term on the right hand is  $O(h_n^3/N)$  because

$$\begin{aligned} E[\|D_{p,1,12}^3\|^2] &\leq \int E[\|\Delta W_{12}\|^4 \Lambda_{12}^4 | \Delta R'_{12}\gamma = r] r^2 K^2(r/h_n) f_{R\gamma}(r) dr \\ &= h_n^5 E[\|\Delta W_{12}\|^4 \Lambda_{12}^4 | \Delta R'_{12}\gamma = rh_n] r^4 K^2(r) f_{R\gamma}(rh_n) dr \\ &= O(h_n^5), \end{aligned}$$

where the last line holds from Assumptions 4, 6, and 8. The second term on the right hand side is  $O(h_n^4/n)$  because

$$\begin{aligned} E[\|D_{p,1,12}^3\| \times \|D_{p,1,13}^3\|] &\leq E\left[ \int E[\|\Delta W_{12}\|^2 \Lambda_{12}^2 | \Delta R'_{12}\gamma = r_1, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{13}\|^2 \Lambda_{13}^2 | \Delta R'_{13}\gamma = r_1, \xi_1, U_1] \\ &\quad \times r_1^2 r_2^2 K^2(r_1/h_n) K^2(r_2/h_n) f_{R\gamma|\xi_1, U_1}(r_1) f_{R\gamma|\xi_1, U_1}(r_2) dr_1 dr_2 \Big] \\ &= h_n^6 E\left[ \int E[\|\Delta W_{12}\|^2 \Lambda_{12}^2 | \Delta R'_{12}\gamma = r_1 h_n, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{13}\|^2 \Lambda_{13}^2 | \Delta R'_{13}\gamma = r_2 h_n, \xi_1, U_1] \\ &\quad \times r_1^2 r_2^2 K^2(r_1) K^2(r_2) f_{R\gamma|\xi_1, U_1}(r_1 h_n) f_{R\gamma|\xi_1, U_1}(r_2 h_n) dr_1 dr_2 \Big] \\ &= O(h_n^6), \end{aligned}$$

where the last line follows from Assumptions 4, 6, and 8. Hence, we have

$$\text{Var}[\|D_{p,1,1}^3\|] = O\left(\frac{h_n^3}{N}\right) + O\left(\frac{h_n^4}{n}\right) = o(1).$$

This implies that  $D_{p,1,1}^3 \rightarrow_p 0$  as  $n \rightarrow \infty$ .

The fourth term  $D_{p,1,1}^4$  converges to 0. Observe that, since  $K$  is bounded by Assumption 8 and  $\|\gamma\| < \infty$ , for some constant  $C > 0$ ,

$$\|D_{p,1,1}^4\| \leq \frac{C\|\beta - \hat{\beta}_n\|}{h_n} \times \frac{1}{N} \sum_{i < j} \|\Delta W_{ij}\|^3 \|\Delta R_{ij}\| \|\Lambda_{12}\|$$

The sum part converges to the expectation of summand by the law of large numbers for U-statistics (Hoeffding (1961)) as  $E[\|\Delta W_{12}\|^3 \|\Delta R_{12}\| \|\Lambda_{12}\|] < \infty$  is bounded by Cauchy-Schwartz and Assumption 7. Thus, this part is  $O_p(1)$ . Also note that

$$\frac{\|\beta - \hat{\beta}_n\|}{h_n} = O_p\left(\frac{1}{nh_n}\right) = o_p(1),$$

by Assumption 9. Hence,

$$\|D_{p,1,1}^4\| = o_p(1).$$

This shows that  $D_{p,1,1}^4 \rightarrow_p 0$  as  $n \rightarrow \infty$ .



The fifth term  $D_{p1,1}^5$  converges to 0. Observe that, since  $K$  is bounded by Assumption 8, for some constant  $C > 0$

$$\|D_{p1,1}^5\| \leq \frac{C}{N} \sum_{i < j} \|\Delta W_{ij}\|^3 |\nu_{ij}| \times \frac{\|\beta - \hat{\beta}_n\|}{h_n}.$$

The sum part is  $O_p(1)$  because

$$E[\|\Delta W_{12}\|^3 |\nu_{12}|] < \infty,$$

by Assumption 7 and

$$\text{Var} \left[ \frac{1}{N} \sum \|\Delta W_{ij}\| |\nu_{ij}| \right] \leq \frac{E[\|\Delta W_{12}\|^6 \nu_{12}^2]}{N} + \frac{2(n-2)}{N} E[\|\Delta W_{12}\|^3 \|\Delta W_{13}\|^3 |\nu_{12}| |\nu_{13}|] = o(1),$$

as these two moments are bounded by Assumption 7. Thus,

$$\|D_{p1,1}^5\| = o_p(1),$$

by the previous calculation for the term involving  $\hat{\beta}_n - \beta$ . This shows that  $D_{p1,1}^5 \rightarrow_p 0$  as  $n \rightarrow \infty$ .

The sixth term  $D_{p1,1}^6$  converges to 0. Its expectation is exactly 0 by the conditional mean independence of  $\nu_{ij}$ . Also, by repeating the similar calculation as  $\text{Var}[\|D_{p1,1}^2\|]$  (by replacing  $\nu_{ij}^2$  by  $\lambda_{ij}\nu_{ij}$ ), we have

$$\text{Var}[\|D_{p1,1}^6\|] = O\left(\frac{1}{N}\right) + O\left(\frac{h_n^2}{n}\right) = o(1).$$

This shows that  $\tilde{D}_{1,6} \rightarrow_p 0$  as  $n \rightarrow \infty$ .

*Sub-Step 2:*  $D_{p1,2} \rightarrow_p 0$ .

As before, we can decompose  $D_{p1,2}$  into  $D_{p1,2}^i$  for  $i = 1, \dots, 6$ . Unlike in  $D_{p1,1}$ , we can no longer have the moments scaled by  $h_n^\alpha$  because the middle values  $c_{ij,n}^*$  are in the kernels. Thus, by the previous calculation for  $D_{p1,1}$ , the  $D_{p1,2}^i$  that involves  $\nu_{ij}^2, \lambda_{ij}^2$ , or  $\lambda_{ij}\nu_{ij}$  will have the slowest convergence rate. So, it suffices to show that those terms converge to 0 in probability.

Pick up such  $D_{p1,2}^i$  with  $\nu_{ij}^2$ , which is  $D_{p1,2}^1$  and given by

$$D_{p1,2}^1 = \frac{2}{Nh_n^2} \sum_{i < j} d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \Delta R'_{ij} \nu_{ij}^2 K'(c_{ij,n}^*/h_n) K(c_{ij,n}^*) (\hat{\gamma}_n - \gamma)$$

Observe that, for some constant  $C > 0$

$$\|D_{p1,2}^1\| \leq \frac{C}{N} \sum_{i < j} \|\Delta W_{ij}\|^2 \|\Delta R_{ij}\| \nu_{ij}^2 \times \frac{\|\hat{\gamma}_n - \gamma\|}{h_n^2}$$

The sum part is  $O_p(1)$  because

$$E[\|\Delta W_{12}\|^2 \|\Delta R_{12}\| \nu_{12}^2] < \infty,$$

by Assumption 7, and

$$\begin{aligned} & \text{Var} \left[ \frac{1}{N} \sum_{i < j} \|\Delta W_{ij}\|^2 \|\Delta R_{ij}\| \nu_{ij}^2 \right] \\ & \leq \frac{E[\|\Delta W_{12}\|^4 \|\Delta R_{12}\|^2 \nu_{12}^4]}{N} + \frac{2(n-2)}{N} E[\|\Delta W_{12}\|^2 \|\Delta W_{13}\|^2 \|\Delta R_{12}\| \|\Delta R_{13}\| \nu_{12}^2 \nu_{13}^2] \\ & = o(1), \end{aligned}$$

as these moments are bounded by Assumption 7. The term involving  $\hat{\gamma}_n$  is  $o_p(1)$  because

$$\frac{\|\hat{\gamma}_n - \gamma\|}{h_n^2} = \frac{\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|}{\sqrt{Nh_n^5}} = o_p(1),$$

by Assumption 10 and  $Nh_n^5 = Nh_n^{2k+3} \times h_n^{-2k+2}$  diverges for  $k \geq 2$ . Hence,

$$\|D_{p1,2}^1\| = O_p(1) \times o_p(1) = o_p(1).$$

This shows that  $D_{p1,2}^1 \rightarrow_p 0$  as  $n \rightarrow \infty$ . Thus, by the above argument, it follows that  $D_{p1,2} \rightarrow_p 0$  as  $n \rightarrow \infty$ .

These two sub-steps conclude that

$$\hat{\Sigma}_{W\nu,2} \rightarrow_p \Sigma_{W\nu,2},$$

as  $n \rightarrow \infty$ . This finishes Step 1.

*Step 2:*  $\hat{\Sigma}_{W\nu,1} \rightarrow_p \Sigma_{W\nu,1}$ .

Define

$$S_{ij} \equiv 2d_{ij1}d_{ij2}K_{h_n}(\Delta R_{ij}'\gamma)\Delta W_{ij}\Delta\hat{\epsilon}_{ij},$$

and let  $\tilde{\Sigma}_{W\nu,1}$  be  $\hat{\Sigma}_{W\nu,1}$  with  $\hat{S}_{ij}$  replaced by  $S_{ij}$ . First, we use the following result:

**Lemma 7.** *Suppose that Assumptions 1-10 hold. If  $h_n = hN^{-1/(2k+3)}$  for some  $h > 0$ , we have*

$$\tilde{\Sigma}_{W\nu,1} \rightarrow_p \Sigma_{W\nu,1},$$

as  $n \rightarrow \infty$ .

Then, it is enough to show that  $\hat{\Sigma}_{W\nu,1}$  is well approximated by  $\tilde{\Sigma}_{W\nu,1}$ :

**Lemma 8.** *Suppose that Assumptions 1-10 hold. If  $h_n = hN^{-1/(2k+3)}$  for some  $h > 0$ , we have*

$$\|\hat{\Sigma}_{W\nu,1} - \tilde{\Sigma}_{W\nu,1}\| = o_p(1).$$

Lemmas 7 and 8 imply that

$$\|\hat{\Sigma}_{W\nu,1} - \Sigma_{W\nu,1}\| \leq \|\hat{\Sigma}_{W\nu,1} - \tilde{\Sigma}_{W\nu,1}\| + \|\tilde{\Sigma}_{W\nu,1} - \Sigma_{W\nu,1}\| = o_p(1),$$

which shows the consistency of  $\hat{\Sigma}_{W\nu,1}$  for  $\Sigma_{W\nu,1}$ . This finishes Step 2.

*Step 3:  $c'_W \Sigma_{W\nu,1} c_W = 0$  case.*

Observe that, by some algebra,

$$\begin{aligned} & nh_n c' \hat{S}_{WW}^{-1} \hat{\Sigma}_{W\nu,1} \hat{S}_{WW}^{-1} c \\ &= nh_n c' (\hat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) \hat{\Sigma}_{W\nu,1} (\hat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) c \\ &+ nh_n c' \Sigma_{WW}^{-1} \hat{\Sigma}_{W\nu,1} (\hat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) c + nh_n c' (\hat{S}_{WW}^{-1}) \hat{\Sigma}_{W\nu,1} \Sigma_{WW}^{-1} c \\ &+ 3nh_n c' \Sigma_{WW}^{-1} \hat{\Sigma}_{W\nu,1} \Sigma_{WW}^{-1} c. \end{aligned}$$

We show the negligibility of the first line in the right hand side of this decomposition. By the proof of Lemma 1, we have that

$$\hat{S}_{WW}^{-1} - \Sigma_{WW}^{-1} = o_p(n^{-\alpha/2})$$

for any  $\alpha \in (0, 1)$ . Thus,

$$nh_n c' (\hat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) \hat{\Sigma}_{W\nu,1} (\hat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) c = n^{1-\alpha} h_n o_p(1) \hat{\Sigma}_{W\nu,1} o_p(1) = o_p(1),$$

for  $\alpha \in [(2k+1)/(2k+3), 1)$  as  $n^{1-\alpha} h_n = n^{(2k+1-\alpha(2k+3))/(2k+3)} = o(1)$  and  $\hat{\Sigma}_{W\nu,1} = O_p(1)$  by the above Step 2.

The remaining terms are shown to be negligible by applying the following lemmas:

**Lemma 9.** *Suppose that Assumptions 1-10 hold. If  $c'_W \Sigma_{W\nu,1} c_W = 0$  and  $h_n = hN^{-1/(2k+3)}$  for some  $h \in (0, \infty)$ , we have*

$$\begin{aligned} n^{1-\alpha/2} h_n \hat{\Sigma}_{W\nu,1} c_W &\rightarrow_p 0, \\ n^{1-\alpha/2} h_n c'_W \hat{\Sigma}_{W\nu,1} &\rightarrow_p 0, \end{aligned}$$

as  $n \rightarrow \infty$  for  $\alpha \in [6/(2k+3), 1)$ .

**Lemma 10.** *Suppose that Assumptions 1-10 hold. If  $c'_W \Sigma_{W\nu,1} c_W = 0$  and  $h_n = hN^{-1/(2k+3)}$  for some  $h \in (0, \infty)$ , we have*

$$nh_n c'_W \hat{\Sigma}_{W\nu,1} c_W \rightarrow_p 0,$$

as  $n \rightarrow \infty$ .

Then, by Lemmas 9 and 10, the last two lines are shown to be

$$\begin{aligned} & nh_n c' \Sigma_{WW}^{-1} \hat{\Sigma}_{W\nu,1} (\hat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) c + nh_n c' (\hat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) \hat{\Sigma}_{W\nu,1} \Sigma_{WW}^{-1} c \\ &+ 3nh_n c' \Sigma_{WW}^{-1} \hat{\Sigma}_{W\nu,1} \Sigma_{WW}^{-1} c \\ &= n^{1-\alpha/2} h_n c' \Sigma_{WW}^{-1} \hat{\Sigma}_{W\nu,1} o_p(1) + c' o_p(1) n^{1-\alpha/2} h_n \hat{\Sigma}_{W\nu,1} \Sigma_{WW}^{-1} c + o_p(1) \\ &= o_p(1). \end{aligned}$$

Hence,

$$nh_n c' \hat{S}_{WW}^{-1} \hat{\Sigma}_{W\nu,1} \hat{S}_{WW} c = o_p(1).$$

Steps 1-3 finish the proof of Proposition 1.  $\square$

### Proof of Proposition 2.

*Proof.* Since

$$h_{n,\delta}^{-(k+1)} (\hat{\beta}_{n,\delta} - \hat{\beta}_n) = h_{n,\delta}^{-(k+1)} (\hat{\beta}_{n,\delta} - \beta) - h_{n,\delta}^{-(k+1)} (\hat{\beta}_n - \beta),$$

where the first term on the right hand side converges to  $\Sigma_{WW}^{-1} \Sigma_{W\lambda}$  by Theorem 1 as  $Nh_{n,\delta}^{2k+3} \rightarrow \infty$ , it suffices to show that

$$h_{n,\delta}^{-(k+1)} (\hat{\beta}_n - \beta) = o_p(1).$$

Take an arbitrary non-zero vector  $c \in \mathbb{R}^{q_w}$ . Since  $\hat{\beta}_n$  is calculated based on  $h_n = hN^{-1/(2k+3)}$  such that  $Nh_n^{2k+3} \rightarrow h$ , by Theorem 1,

$$h_{n,\delta}^{-(k+1)} c' (\hat{\beta}_n - \beta) = \frac{1}{\sqrt{nh_{n,\delta}^{2(k+1)}}} \times \underbrace{\sqrt{n} c' (\hat{\beta}_n - \beta)}_{=O_p(1)} = o_p(1)$$

since

$$nh_{n,\delta}^{2(k+1)} \sim n \times n^{-4\delta(k+1)/(2k+3)} = n^{\frac{2k+3-4\delta(k+1)}{2k+3}}$$

diverges for  $\delta \in (0, \frac{2k+3}{4k+4})$ , which is assumed by the hypothesis. Since  $c$  is arbitrary,  $h_{n,\delta}^{-(k+1)} (\hat{\beta}_n - \beta) = o_p(1)$ , which completes the proof.  $\square$

### Proofs of Lemmas.

*Proof of Lemma 1.*

*Proof.* Write each summand of  $S_{WW}$  as  $S_{WW,ij}$ . Since it suffices to show the element-wise convergence of  $S_{WW}$  to  $\Sigma_{WW}$ , we use a unit vector  $e \in \mathbb{R}^{q_w}$  with the arbitrary element being 1 and 0 elsewhere. Observe that

$$\begin{aligned} E[e' S_{WW} e] &= E[e' S_{WW,ij} e] = \frac{1}{h_n} \int E[d_{121} d_{122} e' \Delta W_{12} \Delta W'_{12} e | \Delta R'_{12} \gamma = r] K(r/h_n) f_{R\gamma}(r) dr \\ &= \int E[d_{121} d_{122} e' \Delta W_{12} \Delta W'_{12} e | \Delta R'_{12} \gamma = rh_n] K(r) f_{R\gamma}(rh_n) dr \\ &= e' \Sigma_{WW} e + o_p(1), \end{aligned}$$

where the last line holds from the dominated convergence theorem under Assumptions 4, 6 and 8. Since  $S_{WW,ij}$  and  $S_{WW,kl}$  are independent if  $i \neq k, l$  and  $j \neq k, l$  by Assumption 1, observe

that

$$\text{Var}(e'S_{WW}e) = \frac{\text{Var}(S_{WW,12})}{N} + \frac{2(n-2)}{N} \text{Cov}(e'S_{WW,12}e, e'S_{WW,13}e).$$

For the variance, we have

$$\begin{aligned} \text{Var}(S_{WW,12}) &\leq E[(e'S_{WW,12}e)^2] \\ &\leq \frac{1}{h_n^2} \int E[\|\Delta W_{12}\|^4 | \Delta R'_{12}\gamma = r] K^2(r/h_n) f_{R\gamma}(r) dr \\ &= \frac{1}{h_n} \int E[\|\Delta W_{12}\|^4 | \Delta R'_{12}\gamma = rh_n] K^2(r) f_{R\gamma}(rh_n) dr \\ &= O\left(\frac{1}{h_n}\right), \end{aligned}$$

where the last line holds from Assumptions 4, 6 and 8. For the covariance, by the conditional independence of  $\Delta W_{12}$  and  $\Delta W_{13}$ , we have

$$\begin{aligned} &\text{Cov}(e'S_{WW,12}e, e'S_{WW,13}e) \\ &\leq E[e'S_{WW,12}e \times e'S_{WW,13}e] \\ &\leq \frac{1}{h_n} \int E[\|\Delta W_{12}\|^2 | \Delta R'_{12}\gamma = r_1, \xi_1, U_1] \times E[\|\Delta W_{13}\|^2 | \Delta R'_{13}\gamma = r_2, \xi_1, U_1] \\ &\quad \times |K(r_1/h_n)| |K(r_2/h_n)| f_{R\gamma|\xi_1, U_1}(r_1) f_{R\gamma|\xi_1, U_1}(r_2) dr_1 dr_2 \\ &= h_n \int E[\|\Delta W_{12}\|^2 | \Delta R'_{12}\gamma = r_1 h_n, \xi_1, U_1] \times E[\|\Delta W_{13}\|^2 | \Delta R'_{13}\gamma = r_2 h_n, \xi_1, U_1] \\ &\quad \times |K(r_1)| |K(r_2)| f_{R\gamma|\xi_1, U_1}(r_1 h_n) f_{R\gamma|\xi_1, U_1}(r_2 h_n) dr_1 dr_2 \\ &= O(h_n), \end{aligned}$$

where the last line holds from Assumptions 4, 6 and 8. Thus,

$$\text{Var}(e'S_{WW}e) = O\left(\frac{1}{Nh_n}\right) + O\left(\frac{h_n}{n}\right) = o(1).$$

By Chebychev's inequality,  $e'S_{WW}e \rightarrow_p e'\Sigma_{WW}e$  as  $n \rightarrow \infty$ . Since  $e$  is arbitrary, this completes the proof. □

*Proof of Lemma 2.*

*Proof.* Write each summand of  $S_{W\lambda}$  as  $S_{W\lambda,ij}$ . We use a unit vector  $e \in \mathbb{R}^{qw}$  with an arbitrary element being 1 and 0 elsewhere. Observe that, for large enough  $n$ ,

$$\begin{aligned} E[e'S_{W\lambda}] &= E[e'S_{W\lambda,ij}] = \frac{1}{h_n} \int E[d_{121}d_{122}e'\Delta W_{12}\lambda_{12}|\Delta R'_{12}\gamma = r]K(r/h_n)f_{R\gamma}(r)dr \\ &= h_n \int e'g(rh_n)rK(r)dr \\ &= \frac{h_n^{k+1}}{k!} \int \left( e' \frac{\partial^k g(rh_n)}{\partial r^k} + o(1) \right) r^{k+1}K(r)dr \\ &= h_n^{k+1}e'\Sigma_{W\lambda} + o(h_n^{k+1}), \end{aligned}$$

where the second line holds from  $\lambda_{12} = \Lambda_{12} \times \Delta R'_{12}\gamma$ , the third line holds from Assumption 8 eliminating  $\int s^i K(s)$  for  $i = 1, \dots, k$ , and the last line holds from the dominated convergence theorem under Assumptions 6 and 8. Observe that

$$\text{Var}[e'S_{W\lambda}] = \frac{\text{Var}[e'S_{W\lambda,12}]}{N} + \frac{2(n-2)}{N} \text{Cov}[e'S_{W\lambda,12}, e'S_{W\lambda,13}].$$

For the variance, we have

$$\begin{aligned} \text{Var}[e'S_{W\lambda,12}] &\leq E[(e'S_{W\lambda,12})^2] \\ &\leq \frac{1}{h_n^2} \int E[\|\Delta W_{12}\|^2 \lambda_{12}^2 |\Delta R'_{12}\gamma = r] K^2(r/h_n) f_{R\gamma}(r) dr \\ &= h_n \int E[\|\Delta W_{12}\|^2 \Lambda_{12}^2 |\Delta R'_{12}\gamma = rh_n] r^2 K^2(r) f_{R\gamma}(rh_n) dr \\ &= O(h_n), \end{aligned}$$

where the last line holds from Cauchy-Schwartz and Assumptions 4, 6, and 8. For the covariance, we have

$$\begin{aligned} &\text{Cov}[e'S_{W\lambda,12}, e'S_{W\lambda,13}] \\ &\leq E[e'S_{W\lambda,12} \times e'S_{W\lambda,13}] \\ &= \frac{1}{h_n^2} E \left[ \int E[d_{121}d_{122}e'\Delta W_{12}\Lambda_{12}|\Delta R'_{12}\gamma = r_1, \xi_1, U_1] E[d_{131}d_{132}e'\Delta W_{13}\Lambda_{13}|\Delta R'_{13}\gamma = r_1, \xi_1, U_1] \right. \\ &\quad \left. \times r_1 r_2 K(r_1/h_n) K(r_2/h_n) f_{R\gamma, \xi_1, U_1}(r_1) f_{R\gamma, \xi_1, U_1}(r_2) dr_1 dr_2 \right] \\ &\leq h_n^2 E \left[ \int E[\|\Delta W_{12}\| \|\Lambda_{12}\| |\Delta R'_{12}\gamma = r_1 h_n, \xi_1, U_1] E[\|\Delta W_{13}\| \|\Lambda_{13}\| |\Delta R'_{13}\gamma = r_2 h_n, \xi_1, U_1] \right. \\ &\quad \left. \times r_1 r_2 K(r_1) K(r_2) f_{R\gamma, \xi_1, U_1}(r_1 h_n) f_{R\gamma, \xi_1, U_1}(r_2 h_n) dr_1 dr_2 \right] \\ &= O(h_n^2), \end{aligned}$$

where the last line holds from Cauchy-Schwartz and Assumptions 4, 6, and 8. Thus,

$$\text{Var}[e'S_{W\lambda}] = O\left(\frac{h_n}{N}\right) + O\left(\frac{h_n^2}{n}\right) = O\left(\frac{h_n^2}{n}\right),$$

since  $O(h_n/N) = O(h_n^2/n) \times O(1/(nh_n)) = o(h_n^2/n)$  under Assumption 9.

If  $Nh_n^{2k+3} \rightarrow h$  for some  $0 < h < \infty$ , note that

$$\sqrt{Nh_n}E[e'S_{W\lambda}] = \sqrt{Nh_n^{2k+3}}e'\Sigma_{W\lambda} + o(\sqrt{Nh_n^{2k+3}}) \rightarrow \sqrt{h}\Sigma_{W\lambda},$$

as  $n \rightarrow \infty$ . Also,

$$\text{Var}[\sqrt{Nh_n}e'S_{W\lambda}] = O\left(\frac{Nh_n^3}{n}\right) = O(nh_n) \times O(h_n^2) = o(1),$$

by Assumption 9. Thus, by Chebyshev's inequality, we have

$$\sqrt{Nh_n}e'S_{W\lambda} \rightarrow_p \sqrt{h}e'\Sigma_{W\lambda},$$

as  $n \rightarrow \infty$ .

If  $Nh_n^{2k+3} \rightarrow \infty$ , note that

$$h_n^{-(k+1)}E[e'S_{W\lambda}] = e'\Sigma_{W\lambda} + o(1) \rightarrow e'\Sigma_{W\lambda},$$

as  $n \rightarrow \infty$ . Also,

$$\text{Var}[h_n^{-(k+1)}e'S_{W\lambda}] = O\left(\frac{h_n^2}{nh_n^{2k+2}}\right) = o(1),$$

as  $nh_n^{2k+2} \rightarrow \infty$  by the hypothesis. Thus, by Chebyshev's inequality, we have

$$h_n^{-(k+1)}e'S_{W\lambda} \rightarrow_p e'\Sigma_{W\lambda},$$

as  $n \rightarrow \infty$ . Since  $e$  is arbitrary, this completes the proof.  $\square$

*Proof of Lemma 3.*

*Proof.* The proof is done in the following steps:

*Step 0: Decomposition.*

Observe that

$$c'S_{WW}^{-1}S_{W\nu} = c'(S_{WW}^{-1} - \Sigma_{WW}^{-1})S_{W\nu} + c'\Sigma_{WW}^{-1}S_{W\nu}$$

In Steps 1-2, we verify the asymptotic normality of  $S_{W\nu}$ , with the worst-case convergence rate being  $\sqrt{n}$ . Given that result, the first term on the right-hand side is shown to be negligible even when normalized by  $\sqrt{Nh_n}$ :

$$\begin{aligned} \sqrt{Nh_n}c'(S_{WW}^{-1} - \Sigma_{WW}^{-1})S_{W\nu} &= \sqrt{Nh_n}o_p(n^{-\alpha/2})O_p(1/\sqrt{n}) \\ &= \sqrt{n^{1-\alpha}h_n}o_p(1) = o_p(1) \end{aligned}$$

because by Lemma 1,  $S_{WW}^{-1} - \Sigma_{WW}^{-1} = o_p(n^{-\alpha/2})$  for any  $\alpha \in (0, 1)$  and  $n^{1-\alpha}h_n = o(1)$  for sufficiently large  $\alpha$  under the hypothesis. Thus,

$$c'S_{WW}^{-1}S_{W\nu} = c'\Sigma_{WW}^{-1}S_{W\nu} + o_p(1/\sqrt{Nh_n}),$$

and it suffices to establish the asymptotic normality of  $S_{W\nu}$ . Write  $c = c_W$  for short. Observe that, since  $E[S_{W\nu}] = 0$  by the definition of  $\nu_{ij}$ ,  $c'S_{W\nu}$  can be decomposed as

$$c'S_{W\nu} = \underbrace{\frac{1}{n} \sum_{i=1}^n L_{i,W\nu}}_{L_{W\nu}} + \underbrace{\frac{1}{N} \sum_{i<j} Q_{ij,W\nu}}_{Q_{W\nu}},$$

where

$$L_{i,W\nu} \equiv 2E[d_{ij1}d_{ij2}c'\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma)|\xi_i, U_i]$$

$$Q_{ij,W\nu} = d_{ij1}d_{ij2}c'\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma) - E[d_{ij1}d_{ij2}c'\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma)|\xi_i, U_i] \\ - E[d_{ij1}d_{ij2}c'\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma)|\xi_j, U_j]$$

By design, we have that  $Cov[L_{i,W\nu}, L_{j,W\nu}] = Cov[L_{i,W\nu}, Q_{kl,W\nu}] = Cov[Q_{ij,W\nu}, Q_{kl,W\nu}] = 0$  for any  $i \neq j$ ,  $k \neq l$ , and  $ij \neq kl$ . We show the asymptotic normality of  $c'S_{W\nu}$  in the following.

*Step 1: Asymptotic Normality of  $L_{W\nu}$ .*

Define  $V_L$  by

$$V_L = \sqrt{n}L_{W\nu} = \sum_{i=1}^n \underbrace{\frac{L_{i,W\nu}}{\sqrt{n}}}_{V_{i,L}}$$

Note that  $E[V_{i,W\nu}] = 0$  by the mean independence of  $\nu_{ij}$ . Observe that

$$\begin{aligned} Var[V_L] &= Var[L_{i,W\nu}] \\ &= 4E[E[d_{121}d_{122}c'\Delta W_{12}\nu_{12}K_{h_n}(\Delta R'_{12}\gamma)|\xi_1, U_1]^2] \\ &= 4E[d_{121}d_{122}d_{131}d_{132}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13}K_{h_n}(\Delta R'_{12}\gamma)K_{h_n}(\Delta R'_{13}\gamma)] \\ &= \frac{4}{h_n^2} \int E[d_{121}d_{122}d_{131}d_{132}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = r_1, \Delta R'_{13}\gamma = r_2] \\ &\quad \times K(r_1/h_n)K(r_2/h_n)f_{R\gamma,2}(r_1, r_2)dr_1dr_2 \\ &= 4 \int E[d_{121}d_{122}d_{131}d_{132}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = r_1h_n, \Delta R'_{13}\gamma = r_2h_n] \\ &\quad \times K(r_1)K(r_2)f_{R\gamma,2}(r_1h_n, r_2h_n)dr_1dr_2 \\ &= c'\Sigma_{W\nu,1}c + o_p(1), \end{aligned}$$



where the last line holds from the dominated convergence theorem under Assumptions 4, 6, and 8. Furthermore, note that

$$\begin{aligned}
& E[d_{121}d_{122}c'\Delta W_{12}\nu_{12}K_{h_n}(\Delta R'_{12}\gamma)|\xi_1, U_1] \\
&= \frac{1}{h_n} \int E[d_{121}d_{122}c'\Delta W_{12}\nu_{12}|\Delta R'_{12}\gamma = r, \xi_1, U_1]K(r/h_n)f_{R\gamma|\xi_1, U_1}(r)dr \\
&= \int E[d_{121}d_{122}c'\Delta W_{12}\nu_{12}|\Delta R'_{12}\gamma = r, \xi_1, U_1]K(r/h_n)f_{R\gamma|\xi_1, U_1}(r)dr \\
&= O(1),
\end{aligned}$$

almost surely for sufficiently large  $n$  by Assumptions 4, 6, and 8. Thus, we have

$$\sum_{i=1}^n E[|V_{i,L}|^3] = n \times O\left(\frac{1}{n\sqrt{n}}\right) = o(1).$$

If  $\Sigma_{W\nu,1}$  is positive definite, we have  $c'\Sigma_{W\nu,1}c > 0$ . Thus, by Lyapunov CLT, we have

$$V_L/\sqrt{\text{Var}[V_L]} \rightarrow_d \mathcal{N}(0, 1),$$

as  $n \rightarrow \infty$ . Thus,  $V_L = \sqrt{n}L_{W\nu} \rightarrow_d \mathcal{N}(0, c'\Sigma_{W\nu,1}c)$  as  $n \rightarrow \infty$ .

If  $c'\Sigma_{W\nu,1}c = 0$ , observe that, for some constant  $C > 0$

$$\begin{aligned}
& \text{Var}[L_{i,W\nu}] \\
&= 4E \left[ \left( \int h_n^{-1} E[d_{121}d_{122}c'\Delta W_{12}\nu_{12}|\Delta R'_{12}\gamma = r, \xi_1, U_1]K(r/h_n)f_{R\gamma|\xi_1, U_1}(r)dr \right)^2 \right] \\
&= 4E \left[ \left( \int g_{\xi_1, U_1}(rh_n)K(r)dr \right)^2 \right] \\
&\leq 4E \left[ \left( g_{\xi_1, U_1}(0) + Ch_n^k \right)^2 \right] \\
&= O(h_n^{2k}),
\end{aligned}$$

where the first inequality holds from the Taylor expansion of  $g_{\xi_1, U_1}(rh_n)$  and  $K$  eliminating  $\int r^i K(r)dr = 0$  for  $i = 1, \dots, k$ , and the last line holds since

$$\begin{aligned}
& E \left[ (g_{\xi_1, U_1}(0))^2 \right] \\
&= E \left[ \left( E[d_{121}d_{122}c'\Delta W_{12}\nu_{12}f_{R\gamma|\xi_1, U_1}(0)|\Delta R'_{12}\gamma = 0, \xi_1, U_1] \right)^2 \right] \\
&= E \left[ E[d_{121}d_{122}d_{131}d_{132}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13}f_{R\gamma,2}(0,0)|\Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0, \xi_1, U_1] \right] \\
&= f_{R\gamma,2}(0,0)E[d_{121}d_{122}d_{131}d_{132}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0] \\
&= c'\Sigma_{W\nu,1}c/4 = 0,
\end{aligned}$$

and  $E[g_{\xi_1, U_1}] = 0$  by the conditional mean independence of  $\nu_{12}$ . Thus,

$$\text{Var}[\sqrt{Nh_n}L_{W\nu}] = nh_n \times O(h_n^{2k}) = h_n^{k-1/2} \times O(\sqrt{Nh_n^{2k+3}}) = o(1),$$

since  $\sqrt{Nh_n^{2k+3}} = O(1)$  by the hypothesis and  $k \geq 2$  so that  $h_n^{k-1/2} = o(1)$ . Hence, by Chebyshev's inequality,

$$\sqrt{Nh_n}L_{W\nu} \xrightarrow{p} 0,$$

as  $n \rightarrow \infty$  when  $\Sigma_{W\nu,1} = 0$ .

*Step 2: Asymptotic Normality of  $Q_{W\nu}$ .*

We use the CLT for martingale differences (Theorem 5.24 and Corollary 5.26 in White (2001)). Define  $V_{n,t} (1 \leq t \leq N)$ , a triangular array, as

$$\begin{aligned} V_{n,1} &= \frac{1}{N}Q_{12,W\nu}, \\ V_{n,2} &= \frac{1}{N}Q_{13,W\nu}, \\ &\vdots \\ V_{n,n+n-1} &= \frac{1}{N}Q_{1n,W\nu}, \\ &\vdots \\ V_{n,N} &= \frac{1}{N}Q_{n-1n,W\nu}. \end{aligned}$$

Notice that  $Q_{ij,W\nu}$  is independent of  $Q_{km,W\nu}$  if  $i \neq k, m$  and  $j \neq k, m$ . Also,  $Q_{ij,W\nu}$  is conditionally independent of  $Q_{km,W\nu}$  even if  $i = k$  or  $m$ , or  $j = k$  or  $m$ ; Note that  $(\epsilon_{ijt}, \eta_{ijt})_{t=1,2}$  and  $(\epsilon_{imt}, \eta_{imt})_{t=1,2}$  are conditionally independent given  $\xi_i, U_i$  by Assumption 1. Since  $(\xi_i, U_i), i = 1, \dots, n$  is i.i.d., this implies that, for  $1 < t \leq N$ ,

$$\begin{aligned} E[V_{n,t}|\{V_{n,s}; s < t\}] &= E[E[V_{n,t}|\{V_{n,s}; s < t\}, \{\xi_i, U_i\}_{i=1}^n]|\{V_{n,s}; s < t\}] \\ &= E[E[V_{n,t}|\xi_t, U_t]|\{V_{n,s}; s < t\}] \\ &= 0, \end{aligned}$$

where

$$\begin{aligned} &E[V_{n,t}|\xi_t, U_t] \\ &= \frac{1}{N}E[d_{121}d_{122}c'\Delta W_{12\nu_{12}}K_{h_n}(\Delta R'_{12}\gamma)|\xi_1, U_1] - \frac{1}{N}E[d_{121}d_{122}c'\Delta W_{12\nu_{12}}K_{h_n}(\Delta R'_{12}\gamma)|\xi_1, U_1] \\ &\quad - \frac{1}{N}E[E[d_{121}d_{122}c'\Delta W_{12\nu_{12}}K_{h_n}(\Delta R'_{12}\gamma)|\xi_2, U_2]|\xi_1, U_1] \\ &= -\frac{1}{N}E[d_{121}d_{122}c'\Delta W_{12\nu_{12}}K_{h_n}(\Delta R'_{12}\gamma)] \\ &= 0 \end{aligned}$$

because  $\xi_1, U_1$  is independent of  $\xi_2, U_2$ . Thus, letting  $\mathcal{F}_t \equiv \sigma(V_s | 1 \leq s \leq t)$  be a sigma algebra generated by  $V_1, \dots, V_{t-1}$  ( $\mathcal{F}_1$  is set to be a trivial  $\sigma$ -algebra) and  $\mathbb{F} \equiv (\mathcal{F}_t)_{1 \leq t \leq N}$  be a filtration,

we have

$$E[V_{n,t}|\mathcal{F}_{t-1}] = 0$$

for  $1 \leq t \leq N$ . Also, for each  $t$ , for some constant  $C > 0$ ,

$$E[|V_{n,t}|] \leq \frac{3}{N} E[|\Delta W_{12}| |\nu_{12}| |K_{h_n}(\Delta R'_{12}\gamma)|] \leq \frac{C}{Nh_n} E[|\Delta W_{12}|^2]^{1/2} E[\nu_{12}^2]^{1/2} < \infty,$$

by Assumptions 7 and 8. This shows that  $\{V_{n,t}\}$  is a martingale difference sequence.

Let  $V_n = \sum_{t=1}^N V_{n,t}$ . Define the variance of this sequence by

$$v_n^2 = \text{Var} \left[ \sum_{t=1}^N V_{n,t} \right] = N \text{Var}[V_{n,1}] = \frac{1}{N} E[Q_{12,W\nu}^2].$$

We can calculate that

$$\begin{aligned} E[Q_{12,W\nu}^2] &= E \left[ d_{121} d_{122} (c' \Delta W_{12})^2 \nu_{12}^2 K^2(\Delta R'_{12}\gamma) \right] \\ &\quad - 2E \left[ E[d_{121} d_{122} c' \Delta W_{12} \nu_{12} K_{h_n}(\Delta R'_{12}\gamma) | \xi_1, U_1]^2 \right]. \end{aligned}$$

Observe that

$$\begin{aligned} &E \left[ d_{121} d_{122} (c' \Delta W_{12})^2 \nu_{12}^2 K^2(\Delta R'_{12}\gamma) \right] \\ &= \frac{1}{h_n^2} \int E[d_{121} d_{122} (c' \Delta W_{12})^2 \nu_{12}^2 | \Delta R'_{12}\gamma = r] K(r/h_n) f_{R\gamma}(r) dr \\ &= \frac{1}{h_n} \int E[d_{121} d_{122} (c' \Delta W_{12})^2 \nu_{12}^2 | \Delta R'_{12}\gamma = rh_n] K(r) f_{R\gamma}(rh_n) dr \\ &= \frac{1}{h_n} c' \Sigma_{W\nu, 2} c + o\left(\frac{1}{h_n}\right), \end{aligned}$$

where the last line holds from the dominated convergence theorem under Assumptions 4, 6, and 8, and

$$\begin{aligned} &E[d_{121} d_{122} c' \Delta W_{12} \nu_{12} K_{h_n}(\Delta R'_{12}\gamma) | \xi_1, U_1] \\ &= \frac{1}{h_n} \int E[d_{121} d_{122} c' \Delta W_{12} \nu_{12} | \Delta R'_{12}\gamma = r, \xi_1, U_1] K(r/h_n) f_{R\gamma|\xi_1, U_1}(r) dr \\ &= \int E[d_{121} d_{122} c' \Delta W_{12} \nu_{12} | \Delta R'_{12}\gamma = rh_n, \xi_1, U_1] K(r) f_{R\gamma|\xi_1, U_1}(rh_n) dr \\ &= O(1), \end{aligned}$$

almost surely for sufficiently large  $n$  by Assumptions 4, 6, and 8. Since  $1/h_n$  diverges as  $n \rightarrow \infty$ ,

$$E[Q_{12,W\nu}^2] = \frac{1}{h_n} c' \Sigma_{W\nu, 2} c + o\left(\frac{1}{h_n}\right).$$

Hence, we have

$$v_n^2 = \frac{1}{Nh_n} c' \Sigma_{W\nu, 2} c + o\left(\frac{1}{Nh_n}\right).$$

The CLT for martingale differences holds if we can show the following two conditions:

$$\sum_{t=1}^N E \left[ \left( \frac{V_{n,t}}{v_n} \right)^{2+\delta} \right] \rightarrow 0 \text{ (Lyapunov),}$$

for some  $\delta > 0$  as  $n \rightarrow \infty$  and

$$\sum_{t=1}^N \left( \frac{V_{n,t}}{v_n} \right)^2 \rightarrow_p 1 \text{ (Stability),}$$

as  $n \rightarrow \infty$ . If these conditions are met, we can apply Theorem 5.24 and Corollary 5.26 in White (2001) to show that

$$\frac{V_n}{v_n} \rightarrow_d \mathcal{N}(0, 1),$$

as  $n \rightarrow \infty$ . Since  $\sqrt{Nh_n v_n} \rightarrow \sqrt{c' \Sigma_{W\nu, 2} c}$ , by Slutsky's lemma,

$$\sqrt{Nh_n} V_n \rightarrow_d \mathcal{N}(0, c' \Sigma_{W\nu, 2} c),$$

which is equivalent to

$$\sqrt{Nh_n} Q_{W\nu} \rightarrow_d \mathcal{N}(0, c' \Sigma_{W\nu, 2} c),$$

as  $n \rightarrow \infty$ .

For Lyapunov's condition, observe that for some constant  $C > 0$ ,

$$\begin{aligned} E[|V_{n,1}|^3] &\leq \frac{C}{(Nh_n)^3} \int E[|\Delta W_{12}|^3 |\nu_{12}|^3 |\Delta R'_{12} \gamma = r| |K(r/h_n)|^3 f_{R\gamma}(r) dr] \\ &= \frac{Ch_n}{(Nh_n)^3} \int E[|\Delta W_{12}|^3 |\nu_{12}|^3 |\Delta R'_{12} \gamma = rh_n| |K(r)|^3 f_{R\gamma}(rh_n) dr] \\ &= O\left(\frac{1}{N^3 h_n^2}\right), \end{aligned}$$

where the first inequality follows from Jensen's inequality, the last line follows from Cauchy-Schwartz and Assumptions 4, 6, and 8. Since  $v_n = O(1/\sqrt{Nh_n})$ , we have

$$\sum_{t=1}^N E \left[ \left| \frac{V_{n,t}}{v_n} \right|^3 \right] = NO \left( \frac{\sqrt{Nh_n}}{N^3 h_n^2} \right) = O \left( \frac{1}{(Nh_n)^{3/2}} \right) = o(1)$$

by Assumption 9. Thus Lyapunov's condition holds.

For the stability condition, we can alternatively show that

$$\frac{1}{v_n^2} \sum_{t=1}^N (V_{n,t}^2 - E[V_{n,t}^2]) \rightarrow_p 0,$$

as  $n \rightarrow \infty$ . Note that

$$\frac{1}{v_n^2} \sum_{t=1}^N (V_{n,t}^2 - E[V_{n,t}^2]) = \frac{1}{Nv_n^2} \left( \frac{1}{N} \sum_{i < j} Q_{ij, W\nu}^2 - E[Q_{12, W\nu}^2] \right).$$

Since  $Nv_n^2 = O(1/h_n)$ , we need to show that the remaining term is  $o_p(1/h_n)$ . Since  $Q_{ij,W\nu}$  is independent from  $Q_{km,W\nu}$  if there is no common node,

$$\begin{aligned} & E \left[ \left( \frac{1}{N} \sum_{i < j} Q_{ij,W\nu}^2 - E[Q_{12,W\nu}^2] \right)^2 \right] \\ &= \frac{\text{Var}[Q_{12,W\nu}^2]}{N} + \frac{2(n-2)}{N} \text{Cov}[Q_{12,W\nu}^2, Q_{13,W\nu}^2] \\ &\leq \frac{E[Q_{12,W\nu}^4]}{N} + \frac{2(n-2)}{N} E[Q_{12,W\nu}^2 \times Q_{13,W\nu}^2]. \end{aligned}$$

The first term in the far right-hand side is bounded as follows: For some constant  $C > 0$ ,

$$\begin{aligned} \frac{E[Q_{12,W\nu}^4]}{N} &\leq \frac{C}{Nh_n^4} \int E[\|\Delta W_{12}\|^4 \nu_{12}^4 | \Delta R'_{12}\gamma = r] K^4(r) f_{R\gamma}(r) dr \\ &= \frac{C}{Nh_n^3} \int E[\|\Delta W_{12}\|^4 \nu_{12}^4 | \Delta R'_{12}\gamma = rh_n] K^4(r) f_{R\gamma}(rh_n) dr \\ &= O\left(\frac{1}{Nh_n^3}\right), \end{aligned}$$

where the first inequality follows from Jensen's inequality, and the last line follows from Cauchy-Schwartz and Assumptions 4, 6, and 8. The second term on the far right-hand side is bounded as follows: For some constant  $C > 0$ ,

$$\begin{aligned} & \frac{2(n-2)}{N} E[Q_{12,W\nu}^2 \times Q_{13,W\nu}^2] \\ &\leq \frac{C(n-2)}{Nh_n^4} \int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12}\gamma = r_1, \xi_1, U_1] \times E[\|\Delta W_{13}\|^2 \nu_{13}^2 | \Delta R'_{13}\gamma = r_2, \xi_1, U_1] \\ &\quad \times K^2(r_1/h_n) K^2(r_2/h_n) f_{R\gamma|\xi_1, U_1}(r_1) f_{R\gamma|\xi_1, U_1}(r_2) dr_1 dr_2 \\ &= \frac{C(n-2)}{Nh_n^2} \int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12}\gamma = r_1 h_n, \xi_1, U_1] \times E[\|\Delta W_{13}\|^2 \nu_{13}^2 | \Delta R'_{13}\gamma = r_2 h_n, \xi_1, U_1] \\ &\quad \times K^2(r_1) K^2(r_2) f_{R\gamma|\xi_1, U_1}(r_1 h_n) f_{R\gamma|\xi_1, U_1}(r_2 h_n) dr_1 dr_2 \\ &= O\left(\frac{1}{nh_n^2}\right) \end{aligned}$$

Thus,

$$h_n E \left[ \left( \frac{1}{N} \sum_{i < j} Q_{ij,W\nu}^2 - E[Q_{12,W\nu}^2] \right)^2 \right] = O\left(\frac{1}{Nh_n^2}\right) + O\left(\frac{1}{nh_n}\right) = o(1),$$

and by Markov's inequality,

$$\sqrt{h_n} \left( \frac{1}{N} \sum_{i < j} Q_{ij,W\nu}^2 - E[Q_{12,W\nu}^2] \right) = o_p(1).$$

Then,

$$\frac{1}{v_n^2} \sum_{t=1}^N (V_{n,t}^2 - E[V_{n,t}^2]) = O(h_n) \times o_p\left(\frac{1}{\sqrt{h_n}}\right) = o_p(\sqrt{h_n}) = o_p(1),$$

which shows the stability condition.

*Step 3: Conclusion.*

By Steps 0-2, we have established that if  $c'_W \Sigma_{W\nu,1} c_W > 0$ ,

$$\begin{aligned} & \sqrt{n} c' S_{WW}^{-1} S_{W\nu} \\ &= \underbrace{\sqrt{n} L_{W\nu}}_{\rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,1} c_W)} + \frac{\sqrt{n}}{\sqrt{N h_n}} \times \underbrace{\sqrt{N h_n} Q_{W\nu}}_{\rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,2} c_W)} + o_p(\sqrt{n}/\sqrt{N h_n}) \rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,1} c_W), \end{aligned}$$

as  $n \rightarrow \infty$  by Assumption 9, and if  $c'_W \Sigma_{W\nu,1} c_W = 0$ ,

$$\sqrt{N h_n} c' S_{WW}^{-1} S_{W\nu} = \underbrace{\sqrt{N h_n} L_{W\nu}}_{\rightarrow_p 0} + \underbrace{\sqrt{N h_n} Q_{W\nu}}_{\rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,2} c_W)} + o_p(1) \rightarrow_d \mathcal{N}(0, c'_W \Sigma_{W\nu,2} c_W),$$

as  $n \rightarrow \infty$ . This completes the proof.  $\square$

*Proof of Lemma 4.*

*Proof.* By expanding  $K(\Delta R'_{ij} \hat{\gamma}_n / h_n)$  around  $\Delta R'_{ij} \gamma$ , we have

$$\hat{S}_{WW} = S_{WW} + \frac{1}{N h_n^2} \sum_{i < j} d_{ij1} d_{ij2} \Delta W_{ij} \Delta W'_{ij} \Delta R'_{ij} (\hat{\gamma}_n - \gamma) K'(c_{ij,n}^*)$$

where  $c_{ij,n}^*$  is in between  $\Delta R'_{ij} \gamma$  and  $\Delta R'_{ij} \hat{\gamma}_n$ . Thus, for some constant  $C > 0$ ,

$$\|\hat{S}_{WW} - S_{WW}\| \leq \underbrace{\frac{C}{N} \sum_{i < j} \|\Delta W_{ij}\|^2 \|\Delta R_{ij}\|}_{D_{4,1}} \times h_n^{-2} \|\hat{\gamma}_n - \gamma\|$$

Notice that  $\|\Delta W_{ij}\| = \|w(X_{i1}, X_{j1}) - w(X_{i2}, X_{j2})\|$ ,  $\|\Delta R_{ij}\| = \|r(Z_{i1}, Z_{j1}) - r(Z_{i2}, Z_{j2})\|$  are symmetric in  $i$  and  $j$  by the symmetry of  $w$ ,  $r$ , and  $\|\cdot\|$  so that  $D_{4,1}$  is a second-order U-statistics.

Also,

$$E[\|\Delta W_{12}\|^2 \|\Delta R_{12}\|] < \infty.$$

by Cauchy-Schwartz with Assumption 7. Thus, we can apply the law of large numbers for U-statistics (Hoeffding (1961)):

$$D_{4,1} = O_p(1).$$

By the hypothesis and Assumption 10,

$$\begin{aligned} h_n^{-2} \|\hat{\gamma}_n - \gamma\| &= \frac{\|\sqrt{Nh_n}(\hat{\gamma}_n - \gamma)\|}{\sqrt{Nh_n^5}} \\ &= \frac{\|\sqrt{Nh_n}(\hat{\gamma}_n - \gamma)\|}{\sqrt{Nh_n^{2k+3}}} \times \sqrt{h_n^{2k-2}} \\ &= o_p(1), \end{aligned}$$

as  $\sqrt{Nh_n^{2k+3}}$  is either diverging or  $O(1)$ ,  $\sqrt{h_n^{2k-2}} = o(1)$  for  $k \geq 2$ . Thus,

$$\|\hat{S}_{WW} - S_{WW}\| = O_p(1) \times o_p(1) = o_p(1).$$

This shows that  $\hat{S}_{WW} = S_{WW} + o_p(1)$ . □

*Proof of Lemma 5.*

*Proof.* Expanding  $K(\Delta R'_{ij} \hat{\gamma}_n / h_n)$  around  $\Delta R'_{ij} \gamma$ , for some constant  $C > 0$ , we have

$$\begin{aligned} &\sqrt{Nh_n} \|\hat{S}_{W\lambda} - S_{W\lambda}\| \\ &\leq \underbrace{\frac{1}{Nh_n^2} \sum_{i < j} \|\Delta W_{ij}\| \|\Delta R_{ij}\| |\lambda_{ij}| |K'(\Delta R'_{ij} \gamma / h_n)|}_{D_{5,1}} \times \sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\| \\ &\quad + \underbrace{\frac{1}{Nh_n^2} \sum_{i < j} \|\Delta W_{ij}\| \|\Delta R_{ij}\|^2 |\lambda_{ij}| |K''(\Delta R'_{ij} \gamma / h_n)|}_{D_{5,2}} \times \frac{\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|}{2h_n} \\ &\quad + C \underbrace{\frac{1}{N} \sum_{i < j} \|\Delta W_{ij}\| \|\Delta R_{ij}\|^3 |\lambda_{ij}|}_{D_{5,3}} \times \frac{\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|^3}{6h_n^4} \end{aligned}$$

We follow the following steps to bound the right hand side.

*Step 1*  $D_{5,1}$  and  $D_{5,2}$ .

Observe that

$$\begin{aligned} E[D_{5,1}] &= \frac{1}{h_n^2} \int E[\|\Delta W_{12}\| \|\Delta R_{12}\| \|\Lambda_{12}\| |\Delta R'_{12} \gamma = r|] |r| |K'(r/h_n)| f_{R\gamma}(r) dr \\ &= \int E[\|\Delta W_{12}\| \|\Delta R_{12}\| \|\Lambda_{12}\| |\Delta R'_{12} \gamma = rh_n|] |r| |K'(r)| f_{R\gamma}(rh_n) dr \\ &= O(1), \end{aligned}$$

where the last line holds from Assumptions 4, 6 and 8. Also, writing each summand of  $D_{5,1}$  by  $D_{5,1,ij}$ , we have

$$\begin{aligned} \text{Var}[D_{5,1}] &= \frac{1}{Nh_n^4} \text{Var}[D_{5,1,12}] + \frac{2(n-2)}{Nh_n^4} \text{Cov}[D_{5,1,12}, D_{5,1,13}] \\ &\leq \frac{1}{Nh_n^4} E[D_{5,1,12}^2] + \frac{2(n-2)}{Nh_n^4} E[D_{5,1,12} \times D_{5,1,13}]. \end{aligned}$$

The first term on the far right side is  $O(1/(Nh_n^2))$  because

$$\begin{aligned} E[D_{5,1,12}^2] &= \int E[\|\Delta W_{12}\|^2 \|\Delta R_{12}\|^2 \Lambda_{12}^2 | \Delta R'_{12} \gamma = r] r^2 K'(r/h_n)^2 f_{R\gamma}(r) dr \\ &= h_n^2 \int E[\|\Delta W_{12}\|^2 \|\Delta R_{12}\|^2 \Lambda_{12}^2 | \Delta R'_{12} \gamma = rh_n] r^2 K'(r)^2 f_{R\gamma}(rh_n) dr \\ &= O(h_n^2), \end{aligned}$$

where the last line holds from Assumptions 4, 6, and 8. The second term on the far right side is  $O(1/n)$  because

$$\begin{aligned} E[D_{5,1,12} \times D_{5,1,13}] &= E \left[ \int E[\|\Delta W_{12}\| \|\Delta R_{12}\| \Lambda_{12} | \Delta R'_{12} \gamma = r_1, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{13}\| \|\Delta R_{13}\| \Lambda_{13} | \Delta R'_{13} \gamma = r_2, \xi_1, U_1] \\ &\quad \times |r_1| |r_2| |K'(r_1/h_n)| |K'(r_2/h_n)| f_{R\gamma|\xi_1, U_1}(r_1) f_{R\gamma|\xi_1, U_1}(r_2) dr_1 dr_2 \left. \right] \\ &= h_n^4 E \left[ \int E[\|\Delta W_{12}\| \|\Delta R_{12}\| \Lambda_{12} | \Delta R'_{12} \gamma = r_1 h_n, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{13}\| \|\Delta R_{13}\| \Lambda_{13} | \Delta R'_{13} \gamma = r_2 h_n, \xi_1, U_1] \\ &\quad \times |r_1| |r_2| |K'(r_1)| |K'(r_2)| f_{R\gamma|\xi_1, U_1}(r_1 h_n) f_{R\gamma|\xi_1, U_1}(r_2 h_n) dr_1 dr_2 \left. \right] \\ &= O(h_n^4), \end{aligned}$$

where the last line holds from Assumptions 4, 6, and 8. Thus,

$$\text{Var}[D_{5,1}] = O\left(\frac{1}{Nh_n^2}\right) + O\left(\frac{1}{n}\right) = o(1),$$

since  $Nh_n^2 = Nh^{2k+3} \times h_n^{1-2k}$  diverges by the hypothesis. Thus,

$$D_{5,1} = O_p(1).$$

By a similar calculation, we have

$$D_{5,2} = O_p(1).$$

*Step 2:  $D_{5,3}$ .*

First, observe that

$$E[D_{5,3}] = E[\|\Delta W_{12}\| \|\Delta R_{12}\|^4 |\Lambda_{12}|] < \infty$$



by Hölder's inequality with Assumption 7. Note that by construction,  $\Lambda_{ij}$  is written as a function of  $\xi_i$  and  $\xi_j$  with symmetry with respect to  $i$  and  $j$ , which implies that  $D_{5,3}$  is a second-order U-statistics. Since each summand is non-negative and has a finite mean, we can apply the law of large numbers for U-statistics (Hoeffding (1961)) to show that

$$D_{5,3} = O_p(1).$$

*Step 3: Conclusion.*

Finally, by Assumption 10 and the hypothesis,

$$\begin{aligned}\sqrt{Nh_n}\|\hat{\gamma}_n - \gamma\| &= o_p(1), \\ \frac{\sqrt{Nh_n}\|\hat{\gamma}_n - \gamma\|^2}{2h_n} &= \frac{\|\sqrt{Nh_n}(\hat{\gamma}_n - \gamma)\|^2}{2\sqrt{Nh_n^3}} = o_p(1), \\ \frac{\sqrt{Nh_n}\|\hat{\gamma}_n - \gamma\|^3}{6h_n^4} &= \frac{\|\sqrt{Nh_n}(\hat{\gamma}_n - \gamma)\|^3}{6Nh_n^5} = o_p(1).\end{aligned}$$

Thus,

$$\sqrt{Nh_n}\|\hat{S}_{W\lambda} - S_{W\lambda}\| = o_p(1).$$

This implies that

$$\hat{S}_{W\lambda} = S_{W\lambda} + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

This completes the proof. □

*Proof of Lemma 6.*

*Proof.* By expanding  $K(\Delta R'_{ij}\hat{\gamma}_n/h_n)$  around  $\Delta R'_{ij}\gamma$ , we have

$$\begin{aligned}& \sqrt{Nh_n}(\hat{S}_{W\nu} - S_{W\nu}) \\ &= \underbrace{\frac{1}{Nh_n^2} \sum_{i<j} d_{ij1}d_{ij2}\Delta W_{ij}\Delta R'_{ij}\nu_{ij}K'(\Delta R'_{ij}\gamma/h_n)}_{D_{6,1}} \sqrt{Nh_n}(\hat{\gamma}_n - \gamma) \\ &+ (\hat{\gamma}_n - \gamma)' \underbrace{\frac{1}{Nh_n^2} \sum_{i<j} d_{ij1}d_{ij2}\Delta W_{ij}\Delta R_{ij}\Delta R'_{ij}\nu_{ij}K''(\Delta R'_{ij}\gamma/h_n)}_{D_{6,2}} \sqrt{Nh_n} \frac{(\hat{\gamma}_n - \gamma)}{h_n} \\ &+ \underbrace{\frac{\sqrt{Nh_n}}{Nh_n^4} \sum_{i<j} d_{ij1}d_{ij2}\Delta W_{ij}\nu_{ij}K'''(c_{ij,n}^*/h_n)}_{D_{6,3}} (\Delta R'_{ij}(\hat{\gamma}_n - \gamma))^3\end{aligned}$$

We bound each component by the following steps.

*Step 1:  $D_{6,1}$  and  $D_{6,2}$ .*

Note that  $E[D_{6,1}] = E[D_{6,2}] = 0$  by the conditional mean independence of  $\nu_{ij}$ . Write  $D_{6,1,ij}$

as each summand of  $D_{6,1}$ . Observe that, by the similar calculation as above,

$$\text{Var}[\|D_{6,1}\|] \leq \frac{1}{Nh_n^4} E[\|D_{6,1,12}\|^2] + \frac{2(n-2)}{Nh_n^4} E[\|D_{6,1,12}\| \times \|D_{6,1,13}\|].$$

The first term on the right hand side is  $O(1/(Nh_n^3))$  since

$$\begin{aligned} E[\|D_{6,1,12}\|^2] &\leq \int E[\|\Delta W_{12}\|^2 \|\Delta R_{12}\|^2 \nu_{12}^2 |\Delta R'_{12\gamma} = r| K'(r/h_n)^2 f_{R\gamma}(r) dr] \\ &= h_n \int E[\|\Delta W_{12}\|^2 \|\Delta R_{12}\|^2 \nu_{12}^2 |\Delta R'_{12\gamma} = rh_n| K'(r)^2 f_{R\gamma}(rh_n) dr] \\ &= O(h_n), \end{aligned}$$

where the last line holds from Assumptions 4, 6, and 8. The second term on the right hand side is  $O(1/(nh_n^2))$  since

$$\begin{aligned} E[\|D_{6,1,12}\| \times \|D_{6,1,13}\|] &\leq E \left[ \int E[\|\Delta W_{12}\| \|\Delta R_{12}\| |\nu|_{12} |\Delta R'_{12\gamma} = r_1, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{13}\| \|\Delta R_{13}\| |\nu|_{13} |\Delta R'_{13\gamma} = r_2, \xi_1, U_1] \\ &\quad \times |K'(r_1/h_n)| |K'(r_2/h_n)| f_{R\gamma|\xi_1, U_1}(r_1) f_{R\gamma|\xi_1, U_1}(r_2) dr_1 dr_2 \left. \right] \\ &= h_n^2 E \left[ \int E[\|\Delta W_{12}\| \|\Delta R_{12}\| |\nu|_{12} |\Delta R'_{12\gamma} = r_1 h_n, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{13}\| \|\Delta R_{13}\| |\nu|_{13} |\Delta R'_{13\gamma} = r_2 h_n, \xi_1, U_1] \\ &\quad \times |K'(r_1)| |K'(r_2)| f_{R\gamma|\xi_1, U_1}(r_1 h_n) f_{R\gamma|\xi_1, U_1}(r_2 h_n) dr_1 dr_2 \left. \right] \\ &= O(h_n^2), \end{aligned}$$

where the last line holds from Assumptions 4, 6, and 8. Hence,

$$\text{Var}[\|D_{6,1}\|] = O\left(\frac{1}{Nh_n^3}\right) + O\left(\frac{1}{nh_n^2}\right) = o(1),$$

since both  $Nh_n^3 = Nh_n^{2k+3} \times h_n^{-2k}$  and  $nh_n^2 \sim \sqrt{Nh_n^4} = \sqrt{Nh_n^{2k+3}} \times \sqrt{h_n^{-2k+1}}$  diverge under the hypothesis. This shows that

$$D_{6,1} = o_p(1).$$

A similar calculation shows that

$$D_{6,2} = o_p(1),$$

as well.

Step 2:  $D_{6,3}$ .

Observe that, for some constant  $C > 0$

$$\|D_{6,3}\| \leq C \underbrace{\frac{1}{N} \sum_{i < j} \|\Delta W_{ij}\| \|\Delta R_{ij}\|^3 |\nu_{ij}|}_{D_{6,4}} \times \frac{\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|^3}{Nh_n^4}.$$

Observe that

$$E[D_{6,4}] = E[\|\Delta W_{12}\| \|\Delta R_{12}\|^3 |\nu_{12}|] < \infty,$$

by Cauchy-Schwartz with Assumption 6. Also, by writing each summand of  $D_{6,4}$  as  $D_{6,4,ij}$ , we have

$$\text{Var}[D_{6,4}] \leq \frac{E[D_{6,4,12}^2]}{N} + \frac{2(n-2)}{N} E[D_{6,4,12} \times D_{6,4,13}].$$

Since

$$\begin{aligned} E[D_{6,4,12}^2] &= E[\|\Delta W_{12}\|^2 \|\Delta R_{12}\|^6 \nu_{12}^2] < \infty \\ E[D_{6,4,12} \times D_{6,4,13}] &= E[\|\Delta W_{12}\| \|\Delta W_{13}\| \|\Delta R_{12}\|^3 \|\Delta R_{13}\|^3 |\nu_{12}| |\nu_{13}|] < \infty \end{aligned}$$

by Hölder's inequality with Assumption 7,

$$\text{Var}[D_{6,4}] = O\left(\frac{1}{N}\right) + O\left(\frac{1}{n}\right) = o(1).$$

This shows that

$$D_{6,4} = O_p(1).$$

Hence, by the previous calculation for the term involving  $\hat{\gamma}_n - \gamma$ ,

$$\|D_{6,3}\| = O_p(1) \times o_p(1) = o_p(1).$$

Step 3: Conclusion.

By the above steps and the hypothesis on  $\hat{\gamma}_n - \gamma$ ,

$$\sqrt{Nh_n} \|\hat{S}_{W\nu} - S_{W\nu}\| = o_p(1).$$

This implies that

$$\hat{S}_{W\nu} = S_{W\nu} + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

This completes the proof. □

*Proof of Lemma 7.*

*Proof.* Define

$$\begin{aligned} S_{ij,1} &\equiv 2d_{ij1}d_{ij2}K_{h_n}(\Delta R'_{ij}\gamma)\Delta W_{ij}\nu_{ij}, \\ S_{ij,2} &\equiv 2d_{ij1}d_{ij2}K_{h_n}(\Delta R'_{ij}\gamma)\Delta W_{ij}\lambda_{ij}, \\ S_{ij,3} &\equiv 2d_{ij1}d_{ij2}K_{h_n}(\Delta R'_{ij}\gamma)\Delta W_{ij}\Delta W'_{ij}(\beta - \hat{\beta}_n). \end{aligned}$$

Since

$$\Delta \hat{\epsilon}_{ij} = \Delta W'_{ij}(\beta - \hat{\beta}_n) + \lambda_{ij} + \nu_{ij},$$

we have

$$S_{ij} = S_{ij,1} + S_{ij,2} + S_{ij,3}.$$

Thus,

$$\begin{aligned} &\tilde{\Sigma}_{W\nu,1} \\ &= \underbrace{\binom{n}{3}^{-1} \sum_{i < j < k} \frac{1}{3} (S_{ij,1}S'_{ik,1} + S_{ij,1}S'_{jk,1} + S_{ik,1}S_{jk,1})}_{D_{7,ijk}} + \mathcal{O}_7, \end{aligned}$$

where  $\mathcal{O}_7$  is the remainder term.

We first show that  $c'D_7c \rightarrow_p c'\Sigma_{W\nu,1}c$ . Note that

$$\begin{aligned} E[c'D_7c] &= E[c'S_{12,1}S_{13,1}c] \\ &= \frac{4}{h_n^2} \int E[d_{121}d_{122}d_{131}d_{132}c'\Delta W_{12}\Delta W'_{13}c\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = s_1, \Delta R'_{13}\gamma = s_2] \\ &\quad \times K(s_1/h_n)K(s_2/h_n)f_{R\gamma,2}(s_1, s_2)ds_1ds_2 \\ &= 4 \int E[d_{121}d_{122}d_{131}d_{132}c'\Delta W_{12}\Delta W'_{13}c\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = s_1h_n, \Delta R'_{13}\gamma = s_2h_n] \\ &\quad \times K(s_1)K(s_2)f_{R\gamma,2}(s_1h_n, s_2h_n)ds_1ds_2 \\ &\rightarrow c'\Sigma_{W\nu,1}c, \end{aligned}$$

as  $n \rightarrow \infty$  by the dominated convergence theorem under Assumptions 4, 6, and 8. Define the third order U-statistics

$$U_{n,1} = \binom{n}{3}^{-1} \sum_{i < j < k} p_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k),$$

where  $\boldsymbol{\xi}_i = (\xi_i, U_i)$  and

$$p_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k) = E[c'D_{7,ijk}c|\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k]$$

By the calculation of Graham et al. (2019) in Appendix B,

$$\begin{aligned}
E[(c'D_7c - U_{n,1})^2] &= \binom{n}{3}^{-1} E[(c'D_{7,123}c - E[c'D_{7,123}c|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3])^2] \\
&\quad + \binom{n}{3}^{-2} \times 3 \binom{n}{2} \binom{n-2}{2} \times E[c'D_{7,123}c - E[c'D_{7,123}c|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3]] \\
&\quad \times E[c'D_{7,124}c - E[c'D_{7,124}c|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_4]] \\
&= O\left(\frac{E[(c'D_{7,123}c)^2]}{n^3}\right).
\end{aligned}$$

Observe that

$$\begin{aligned}
E[(c'D_{7,123}c)^2] &= \frac{1}{9} (3E[(c'S_{12,1}c \times c'S_{13,1}c)^2] + 6E[(c'S_{12,1}c)^2 \times c'S_{13,1}c \times c'S_{23,1}c]) \\
&= O\left(\frac{1}{h_n^2}\right),
\end{aligned}$$

since for some positive constant  $C > 0$ ,

$$\begin{aligned}
E[(c'S_{12,1}c \times c'S_{13,1}c)^2] &\leq \frac{C}{h_n^4} \int E[\|\Delta W_{12}\|^2 \|\Delta W_{13}\|^2 \nu_{12}^2 \nu_{13}^2 | \Delta R'_{12}\gamma = s_1, \Delta R'_{13}\gamma = s_2] \\
&\quad \times K^2(s_1/h_n) K^2(s_2/h_n) f_{R\gamma,2}(s_1, s_2) ds_1 ds_2 \\
&= \frac{C}{h_n} \int E[\|\Delta W_{12}\|^2 \|\Delta W_{13}\|^2 \nu_{12}^2 \nu_{13}^2 | \Delta R'_{12}\gamma = s_1 h_n, \Delta R'_{13}\gamma = s_2 h_n] \\
&\quad \times K^2(s_1) K^2(s_2) f_{R\gamma,2}(s_1 h_n, s_2 h_n) ds_1 ds_2 \\
&= O\left(\frac{1}{h_n^2}\right),
\end{aligned}$$

as  $n \rightarrow \infty$  with the last line coming from Assumption 4, 6, and 8, and,

$$\begin{aligned}
&E[(c'S_{12,1}c)^2 \times c'S_{13,1}c \times c'S_{23,1}c] \\
&= E[(c'S_{12,1}c)^2 \times E[c'S_{13,1}c|\xi_1, U_1] \times E[c'S_{23,1}c|\xi_2, U_2]] \\
&= E[E[(c'S_{12,1}c)^2 \times c'S_{13,1}c|\xi_1, U_1] \times E[c'S_{23,1}c|\xi_2, U_2]] \\
&= E[(c'S_{12,1}c)^2 \times c'S_{13,1}c] \times E[c'S_{12,1}c] \\
&\leq O(1) \times \frac{C}{h_n^3} \int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12}\gamma = s_1, \xi_1, U_1] \times E[\|\Delta W_{13}\| \nu_{13} | \Delta R'_{13}\gamma = s_2, \xi_1, U_1] \\
&\quad \times K^2(s_1/h_n) K(s_2/h_n) f_{R\gamma|\xi_1, U_1}(s_1) f_{R\gamma|\xi_1, U_1}(s_2) ds_1 ds_2 ds_3 \\
&= O(1) \times \frac{C}{h_n} \int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12}\gamma = s_1 h_n, \xi_1, U_1] \times E[\|\Delta W_{13}\| \nu_{13} | \Delta R'_{13}\gamma = s_2 h_n, \xi_1, U_1] \\
&\quad \times K^2(s_1) K(s_2) f_{R\gamma|\xi_1, U_1}(s_1 h_n) f_{R\gamma|\xi_1, U_1}(s_2 h_n) ds_1 ds_2 ds_3 \\
&= O\left(\frac{1}{h_n}\right),
\end{aligned}$$

where the first to third lines follow from the conditional independence of  $S_{ij,1}$ , the random sampling of  $\xi_i$ , and the conditional independence and exchangeability of  $U_i$  under Assumption 1,

and the last line follows from Assumptions 4, 6, and 8. Observe that, by conditional independence of  $S_{ij,1}$  and  $S_{ik,1}$  given  $\xi_i, U_i$  and  $S_{ij,1} = S_{ji,1}$ , one can show that

$$\begin{aligned} E[c'D_{7,123}c \times c'D_{7,124}c] &= \frac{1}{9} \{ 2E[(c'S_{12,1}c)^2 \times c'S_{13,1}c \times c'S_{14,1}c] \\ &\quad + 2E[(c'S_{12,1}c)^2 \times c'S_{13,1}c] \times E[c'S_{13,1}c] + 5E[c'S_{12,1}c \times c'S_{13,1}c]^2 \} \\ &= \left( \frac{1}{h_n} \right), \end{aligned}$$

where the last line holds since

$$\begin{aligned} &E[(c'S_{12,1}c)^2 \times c'S_{13,1}c \times c'S_{14,1}c] \\ &\leq \frac{C}{h_n^4} E \left[ \int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12}\gamma = s_1, \xi_1, U_1] \times E[\|\Delta W_{12}\| \nu_{12} | \Delta R'_{12}\gamma = s_2, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{14}\| \nu_{14} | \Delta R'_{14}\gamma = s_3, \xi_1, U_1] \times K^2(s_1/h_n) K(s_2/h_n) K(s_3/h_n) \\ &\quad \left. \times f_{R\gamma|\xi_1, U_1}(s_1) f_{R\gamma|\xi_1, U_1}(s_2) f_{R\gamma|\xi_1, U_1}(s_3) ds_1 ds_2 ds_3 \right] \\ &= \frac{C}{h_n} E \left[ \int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12}\gamma = s_1 h_n, \xi_1, U_1] \times E[\|\Delta W_{12}\| \nu_{12} | \Delta R'_{12}\gamma = s_2 h_n, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{14}\| \nu_{14} | \Delta R'_{14}\gamma = s_3 h_n, \xi_1, U_1] \times K^2(s_1) K(s_2) K(s_3) \\ &\quad \left. \times f_{R\gamma|\xi_1, U_1}(s_1 h_n) f_{R\gamma|\xi_1, U_1}(s_2 h_n) f_{R\gamma|\xi_1, U_1}(s_3 h_n) ds_1 ds_2 ds_3 \right] \\ &= O\left(\frac{1}{h_n}\right), \end{aligned}$$

where the last line holds by Assumptions 4, 6, 8,

$$\begin{aligned} &E[(c'S_{12,1}c)^2 \times c'S_{13,1}c] \\ &\leq \frac{C}{h_n^3} E \left[ \int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12}\gamma = s_1, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{13}\| \nu_{13} | \Delta R'_{13}\gamma = s_2, \xi_1, U_1] K^2(s_1/h_n) K(s_2/h_n) f_{R\gamma|\xi_1, U_1}(s_1) f_{R\gamma|\xi_1, U_1}(s_2) ds_1 ds_2 \left. \right] \\ &\leq \frac{C}{h_n} E \left[ \int E[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R'_{12}\gamma = s_1 h_n, \xi_1, U_1] \right. \\ &\quad \times E[\|\Delta W_{13}\| \nu_{13} | \Delta R'_{13}\gamma = s_2 h_n, \xi_1, U_1] K^2(s_1) K(s_2) f_{R\gamma|\xi_1, U_1}(s_1 h_n) f_{R\gamma|\xi_1, U_1}(s_2 h_n) ds_1 ds_2 \left. \right] \\ &= O\left(\frac{1}{h_n}\right), \end{aligned}$$

where the last equality holds from Assumptions 4, 6, and 8, and

$$\begin{aligned}
& E[c'S_{12}c \times c'S_{13}c] \\
& \leq \frac{C}{h_n^2} E \left[ \int E[|\Delta W_{12}| |\nu_{12}| |\Delta R'_{12}\gamma = s_1, \xi_1, U_1] \right. \\
& \quad \times E[|\Delta W_{13}| |\nu_{13}| |\Delta R'_{13}\gamma = s_2, \xi_1, U_1] K(s_1/h_n) K(s_2/h_n) f_{R\gamma|\xi_1, U_1}(s_1) f_{R\gamma|\xi_1, U_1}(s_2) ds_1 ds_2 \left. \right] \\
& \leq CE \left[ \int E[|\Delta W_{12}| |\nu_{12}| |\Delta R'_{12}\gamma = s_1 h_n, \xi_1, U_1] \right. \\
& \quad \times E[|\Delta W_{13}| |\nu_{13}| |\Delta R'_{13}\gamma = s_2 h_n, \xi_1, U_1] K(s_1) K(s_2) f_{R\gamma|\xi_1, U_1}(s_1 h_n) f_{R\gamma|\xi_1, U_1}(s_2 h_n) ds_1 ds_2 \left. \right] \\
& = O(1),
\end{aligned}$$

where the last equality holds from Assumptions 4, 6, and 8. Thus,

$$E[(c'D_7c - U_{n,1})^2] = O\left(\frac{1}{n^3 h_n^2}\right) = o(1).$$

Thus,  $c'D_1$  is well approximated by  $U_n$ . Also, since  $nh_n^2 \rightarrow \infty$  with the stated assumption on  $h_n$ ,

$$E\left[(p_n(\xi_i, \xi_j, \xi_k))^2\right] = O(E[(c'D_{7,123}c)^2]) = O\left(\frac{n}{nh_n^2}\right) = o(1) \times O(n),$$

and by Lemma A.3 of Ahn and Powell (1993), we have

$$U_n = E[U_{n,1}] + o_p(1).$$

This shows that

$$\begin{aligned}
c'D_7c &= E[U_{n,1}] + \underbrace{c'D_7c - U_{n,1}}_{=o_p(1)} + \underbrace{U_{n,1} - E[U_{n,1}]}_{=o_p(1)} \\
&= E[c'D_7c] + o_p(1) \\
&= c'\Sigma_{W\nu,1}c + o_p(1).
\end{aligned}$$

This completes  $c'D_7c \rightarrow_p c'\Sigma_{W\nu,1}c$  as  $n \rightarrow \infty$ .

The remainder term  $\mathcal{O}_7$  with each term involving either  $S_{ij,2}$  or (and)  $S_{ij,3}$  is of smaller order than  $D_7$  since  $S_{ij,2}$  and  $S_{ij,3}$  involve  $\|\hat{\beta}_n - \beta\| = O_p(1/\sqrt{n})$  and  $\lambda_{ij} \sim h_n$  for large  $n$ ; By computing in a similar way as before, we can establish that  $E[c'\mathcal{O}_7c] = o(1)$  and  $\text{Var}[c'\mathcal{O}_7c] = o(1)$  so that  $|c'\mathcal{O}_7c| \rightarrow_p 0$ . Hence,

$$|c'\tilde{\Sigma}_{\nu,1}c - c'\Sigma_{W\nu,1}c| \leq |c'D_1c - c'\Sigma_{W\nu,1}c| + |c'\mathcal{O}_7c| = o_p(1),$$

which completes the proof for Lemma 7. □

*Proof of Lemma 8.*

*Proof.* Define

$$\hat{S}_{ij,1} = \frac{2}{h_n^2} d_{ij1} d_{ij2} K' \left( \frac{c_{ij,n}^*}{h_n} \right) \Delta W_{ij} \Delta R'_{ij} \Delta \hat{\epsilon}_{ij}$$

where  $c_{ij,n}^*$  is in between  $\Delta R'_{ij} \gamma$  and  $\Delta R'_{ij} \hat{\gamma}_n$ . In the following argument, we treat  $\Delta \hat{\epsilon}_{ij}$  in  $\hat{S}_{ij,1}$  as  $\nu_{ij}$  because only the existence of higher moments is important and bounding the terms involving  $\nu_{ij}$  suffices. By the expression for  $\Delta \hat{\epsilon}_{ij}$ , we have

$$\hat{S}_{ij} = S_{ij,1} + S_{ij,2} + S_{ij,3} + \hat{S}_{ij,1}(\hat{\gamma}_n - \gamma).$$

By the proof of 7, we know that

$$\binom{n}{3}^{-1} \sum_{i < j < k} \frac{1}{3} (S_{ij,p} S'_{ik,p} + S_{ij,p} S'_{jk,p} + S_{ik,p} S'_{jk,p}) = o_p(1),$$

for  $p = 2, 3$ . Then,

$$\begin{aligned} \|\hat{\Sigma}_{W\nu,1} - \tilde{\Sigma}_{W\nu,1}\| &\leq \underbrace{\sum_{p=1}^3 \binom{n}{3}^{-1} \sum_{i < j < k} \frac{h_n^2}{3} (\|S_{ij,p}\| \|\hat{S}_{ik,1}\| + \|S_{ij,p}\| \|\hat{S}_{jk,1}\| + \|S_{ik,p}\| \|\hat{S}_{jk,1}\|)}_{D_{8,1}^p} \frac{\|\hat{\gamma}_n - \gamma\|}{h_n^2} \\ &\quad + \underbrace{\sum_{p=1}^3 \binom{n}{3}^{-1} \sum_{i < j < k} \frac{h_n^2}{3} (\|\hat{S}_{ij,1}\| \|S_{ik,p}\| + \|\hat{S}_{ij,1}\| \|S_{jk,p}\| + \|\hat{S}_{ik,1}\| \|S_{jk,p}\|)}_{D_{8,2}} \frac{\|\hat{\gamma}_n - \gamma\|}{h_n^2} \\ &\quad + \underbrace{\binom{n}{3}^{-1} \sum_{i < j < k} \frac{h_n^4}{3} (\|\hat{S}_{ij,1}\| \|\hat{S}_{ik,1}\| + \|\hat{S}_{ij,1}\| \|\hat{S}_{jk,1}\| + \|\hat{S}_{ik,1}\| \|\hat{S}_{jk,1}\|)}_{D_{8,3}} \frac{\|\hat{\gamma}_n - \gamma\|^2}{h_n^4} \end{aligned}$$

For  $D_{8,1}^p$ , it suffices to bound  $D_{8,1}^1$  as the similar calculation applies to the other terms. By Assumption 8, for some constant  $C > 0$ ,

$$\begin{aligned} \|S_{ij,1}\| &\leq \frac{2}{h_n} \underbrace{\|\Delta W_{ij}\| |\nu_{ij}| K(\Delta R'_{ij} \gamma / h_n)}_{g_{ij,1}}, \\ \|\hat{S}_{ij,1}\| &\leq \frac{C}{h_n^2} \underbrace{\|\Delta W_{ij}\| \|\Delta R_{ij}\| |\nu_{ij}|}_{g_{ij,2}}. \end{aligned}$$

Thus,

$$D_{8,1}^1 \leq C \binom{n}{3}^{-1} \sum_{i < j < k} \underbrace{(g_{ij,1} g_{ik,2} + g_{ij,1} g_{jk,2} + g_{ik,1} g_{jk,2})}_{g_{ijk,12}}.$$



Observe that

$$\begin{aligned}
E[D_{8,1}^1] &\leq \frac{1}{h_n} E[g_{12,1}g_{13,2}] \\
&= \frac{1}{h_n} E\left[ \int E[\|\Delta W_{12}\| \nu_{12} \|\Delta R'_{12}\gamma = r, \xi_1, U_1] K(r/h_n) f_{R\gamma|\xi_1, U_1}(r) dr \right. \\
&\quad \left. \times E[\|\Delta W_{13}\| \|\Delta R_{13}\| \nu_{13} | \xi_1, U_1] \right] \\
&= E\left[ \int E[\|\Delta W_{12}\| \nu_{12} \|\Delta R'_{12}\gamma = rh_n, \xi_1, U_1] K(r) f_{R\gamma|\xi_1, U_1}(rh_n) dr \right. \\
&\quad \left. \times E[\|\Delta W_{13}\| \|\Delta R_{13}\| \nu_{13} | \xi_1, U_1] \right] \\
&= O(1),
\end{aligned}$$

where the last equality follows from Assumptions 4, 6, 7, and 8. For variance, the leading term involves covariances between variables with one common node, which has  $n \times \binom{n-1}{4}$  elements (up to some constant scale):

$$\text{Var}[D_{8,1}^1] = O\left(\frac{1}{h_n^2} \times \binom{n}{3}^{-2} \times n \times \binom{n-1}{4}\right) \times E[D_{8,1,123}^1 \times D_{8,1,145}^1] = O_p\left(\frac{1}{nh_n^2}\right) = o(1),$$

since  $nh_n^2 \sim n^{(2k-1)/(2k+3)}$  diverges for  $k \geq 2$  and

$$\begin{aligned}
E[D_{8,1,123}^1 \times D_{8,1,145}^1] &\leq E[g_{123,12} \times g_{145,12}] \\
&\leq E[g_{123,12}^2] \\
&\leq CE[\|\Delta W_{12}\|^2 \|\Delta W_{13}\|^2 \|\Delta R_{13}\|^2 \nu_{12}^2 \nu_{13}^2] \\
&= O(1),
\end{aligned}$$

by Cauchy-Schwartz under Assumption 7. Thus,  $D_{8,1} = O_p(1)$ . Similarly,  $D_{8,2} = O_p(1)$  and  $D_{8,3} = O_p(1)$  hold.

Notice that

$$\frac{\|\hat{\gamma}_n - \gamma\|}{h_n^2} = \frac{\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|}{\sqrt{Nh_n^5}} = o_p(1),$$

since  $1/\sqrt{Nh_n^5}$  diverges by the hypothesis, and similarly,

$$\frac{\|\hat{\gamma}_n - \gamma\|^2}{h_n^4} = \frac{(\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|)^2}{Nh_n^5} = o_p(1).$$

Hence,

$$\|\hat{\Sigma}_{W\nu,1} - \tilde{\Sigma}_{W\nu,1}\| \leq O_p(1) \times o_p(1) + o_p(1) = o_p(1),$$

which completes the proof for Lemma 8. □

*Proof of Lemma 9.*

*Proof.* We only show  $nh_n c'_W \hat{\Sigma} = o_p(1)$  as the other case follows by taking transpose. We also write  $c_W$  as  $c$  for short. The statement is proved by showing that Lemmas 7 and 8 hold even after re-scaled by  $n^{1-\alpha/2}h_n$ . First, we show that  $c'\Sigma_{W\nu,1}c = 0$  implies that  $c'\Sigma_{W\nu,1} = 0$ .

*Step 1: The implication of  $c'\Sigma_{W\nu,1}c = 0$ .*

Note that

$$\begin{aligned}\Sigma_{W\nu,1} &= f_{R\gamma,2}(0,0)Pr(d_{121}d_{122}d_{131}d_{132} = 1|\Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0) \\ &\quad \times E[\Delta W_{12}\Delta W'_{13}\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0].\end{aligned}$$

By Assumption 4,  $f_{R\gamma,2}(0,0) > 0$ . Also, by the conditional independence of  $d_{12t}$  and  $d_{13s}$  under Assumption 1,

$$\begin{aligned}Pr(d_{121}d_{122}d_{131}d_{132} = 1|\Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0) \\ = E\left[\left(Pr(d_{122}d_{122} = 1|\Delta R'_{12}\gamma = 0, \xi_1, U_1)\right)^2 \middle| \Delta R'_{12}\gamma = 0\right].\end{aligned}$$

Since  $Pr(d_{121}d_{122} = 1|\Delta R'_{12}\gamma = 0) > 0$  is implied by Assumption 3, it must be that

$$Pr(d_{121}d_{122}d_{131}d_{132} = 1|\Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0) > 0,$$

as otherwise  $Pr(d_{121}d_{122} = 1|\Delta R'_{12}\gamma = 0, \xi_1, U_1)$  is constant at 0, which contradicts with the locally positive probability of  $d_{121}d_{122} = 1$ . Thus,  $c'\Sigma_{W\nu,1}c = 0$  is equivalent to

$$\begin{aligned}E[\Delta c'W_{12}\Delta W'_{13}c\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0] \\ = E\left[\left(E[c'\Delta W_{12}\nu_{12}|\Delta R'_{12}\gamma = 0, \xi_1, U_1]\right)^2 \middle| \Delta R'_{12}\gamma = 0\right] \\ = 0,\end{aligned}$$

which, in turn, is equivalent to (by the mean independence of  $\nu_{12}$ ),

$$E[c'\Delta W_{12}\nu_{12}|\Delta R'_{12}\gamma = 0, \xi_1, U_1] = 0,$$

almost surely. Thus,  $c'\Sigma_{W\nu,1}c = 0$  implies that

$$\begin{aligned}c'\Sigma_{W\nu,1} &= f_{R\gamma,2}(0,0)Pr(d_{121}d_{122}d_{131}d_{132} = 1|\Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0) \\ &\quad \times E\left[E[c'\Delta W_{12}\nu_{12}|\Delta R'_{12}\gamma = 0, \xi_1, U_1]E[\Delta W'_{13}\nu_{13}|\Delta R'_{13}\gamma, \xi_1, U_1] \middle| \Delta R'_{12}\gamma = \Delta R'_{13}\gamma = 0\right] \\ &= 0.\end{aligned}$$

*Step 2:  $nh_n c'\tilde{\Sigma}_{W\nu,1} = o_p(1)$ .*

Remember that

$$\tilde{\Sigma}_{W\nu,1} = D_7 + \mathcal{O}_7.$$

For  $D_7$ , from the calculation in the proof of Lemma 3, we have that

$$E[c'D_7] = E[c'S_{12,1}S'_{13,1}] = O(h_n^k).$$

, which shows  $nh_n E[c'D_7] = O(nh_n^{k+1}) = o(1)$  under the hypothesis. For any non-zero vector  $a \in \mathbb{R}^{qw}$ , redefining  $U_n$  and  $U_{n,1}$  with the kernel  $p_n(\xi_i, \xi_j, \xi_k) = E[c'D_{7,ijk}a|\xi_i, \xi_j, \xi_k]$ , we can repeat the calculation in the proof of Lemma 7 to get

$$E[(nh_n c'D_7 a - nh_n U_{n,1})^2] = n^2 h_n^2 \times O\left(\frac{1}{n^3 h_n^2}\right) = o(1).$$

Also,  $E[nh_n p_n(\xi_i, \xi_j, \xi_k)] = E[nh_n c'D_7] = o(1)$  and

$$E[(nh_n p_n(\xi_i, \xi_j, \xi_l))^2] = O(n),$$

so that by Lemma A.3 of Ahn and Powell (1993),

$$nh_n U_n = nh_n E[U_{n,1}] + o_p(1) = o_p(1).$$

This shows that, since  $a$  is arbitrary,

$$nh_n c'D_7 = o_p(1).$$

For the remainder term  $\mathcal{O}_7$ , this should again be of smaller order than  $nh_n c'D_7$  since  $\beta - \hat{\beta}_n = O_p(1/\sqrt{n})$  and  $\lambda_{ij} = \Delta R'_{ij} \gamma \Lambda_{ij}$  is locally  $O(h_n^{k+1})$  under the smoothing kernel and smoothness conditions on the density. For example, one of the elements in  $\mathcal{O}_7$  is given by

$$\begin{aligned} & nh_n \binom{n}{3} \sum_{i < j < k} \frac{1}{3} (S_{ij,2} S'_{ik,2} + S_{ij,2} S'_{jk,2} + S_{ik,2} S_{jk,2}) \\ & \leq \binom{n}{3} \sum_{i < j < k} \frac{4}{3h_n^2} \left( |\Delta W_{ij}|^2 |\Delta W_{ik}|^2 |K(\Delta R'_{ij} \gamma / h_n)| |K(\Delta R'_{ik} \gamma / h_n)| \right. \\ & \quad + |\Delta W_{ij}|^2 |\Delta W_{jk}|^2 |K(\Delta R'_{ij} \gamma / h_n)| |K(\Delta R_{jk} / h_n)| \\ & \quad \left. + |\Delta W'_{ik}|^2 |\Delta W_{jk}|^2 |K(\Delta R_{ik} / h_n)| |K(\Delta R_{jk} / h_n)| \right) nh_n \|\beta - \hat{\beta}_n\|^2 \\ & = O_p(h_n) = o_p(1), \end{aligned}$$

where the last line can be shown by the same calculation as before to show  $O_p(1)$  for the summation part and  $\|\beta - \hat{\beta}_n\|^2 = O_p(1/n)$  from Theorem 1. Similarly, we can show the negligibility of the elements in  $\mathcal{O}_7$ . This finishes the step 2.

*Step 3:*  $n^{1-\alpha/2} h_n \|\hat{\Sigma}_{W\nu,1} - \tilde{\Sigma}_{W\nu,1}\| = o_p(1)$ .

By the proof of Lemma 8, we have

$$n^{1-\alpha/2} h_n \|\hat{\Sigma}_{W\nu,1} - \tilde{\Sigma}_{W\nu,1}\| \leq O_p(1) \left( \frac{n^{1-\alpha/2} \|\hat{\gamma}_n - \gamma\|}{h_n} + \frac{n^{1-\alpha/2} \|\hat{\gamma}_n - \gamma\|^2}{h_n^3} \right).$$

Observe that

$$\begin{aligned}\frac{n^{1-\alpha/2}\|\hat{\gamma}_n - \gamma\|}{h_n} &= O\left(\frac{\sqrt{N}h_n\|\hat{\gamma}_n - \gamma\|}{\sqrt{n^\alpha h_n^3}}\right) = o_p(1), \\ \frac{n^{1-\alpha/2}\|\hat{\gamma}_n - \gamma\|^2}{h_n^3} &= O\left(\frac{\|\sqrt{N}h_n(\hat{\gamma}_n - \gamma)\|^2}{\sqrt{n^{2+\alpha}h_n^7}}\right) = o_p(1),\end{aligned}$$

by Assumption 10 and

$$\begin{aligned}n^\alpha h_n^3 &\sim n^{(\alpha(2k+3)-6)/(2k+3)} \rightarrow \infty, \\ n^{2+\alpha} h_n^7 &\sim n^{((2k+3)\alpha+4k-8)/(2k+3)} \rightarrow \infty,\end{aligned}$$

for  $\alpha \in [6/(2k+3), 1)$ . Hence,

$$n^{1-\alpha/2}h_n\|\hat{\Sigma}_{W\nu,1} - \tilde{\Sigma}_{W\nu,1}\| = o_p(1).$$

Steps 1-3 complete the proof of Lemma 9.  $\square$

*Proof of Lemma 10.*

*Proof.* Write  $c_W$  as  $c$  for short. It suffices to show that  $nh_n\|c'\hat{\Sigma}_{W\nu,1}c - c'\tilde{\Sigma}_{W\nu,1}c\| = o_p(1)$  as we already show, in the proof of Lemma 9, that  $nh_n c'\tilde{\Sigma}_{W\nu,1} = o_p(1)$ , which implies  $nh_n c'\tilde{\Sigma}_{W\nu,1}c = o_p(1)$ . To save the space, in the following argument, we treat  $\Delta\hat{\epsilon}_{ij}$  as  $\nu_{ij}$ ; The other terms are similarly bounded using the properties of  $\beta - \hat{\beta}_n$  and  $\lambda_{ij}$ .

Re-define

$$\begin{aligned}S_{ij} &= \frac{2}{h_n}d_{ij1}d_{ij2}K\left(\frac{\Delta R'_{ij}\gamma}{h_n}\right)c'\Delta W_{ij}\nu_{ij} \\ \hat{S}_{ij} &= \frac{2}{h_n^2}d_{ij1}d_{ij2}K'\left(\frac{\Delta R'_{ij}\gamma}{h_n}\right)c'\Delta W_{ij}\Delta R'_{ij}\nu_{ij} \\ \hat{S}_{ij,2} &= \frac{1}{h_n^3}d_{ij1}d_{ij2}K''\left(\frac{c_{ij,n}^*}{h_n}\right)\Delta R_{ij}c'\Delta W_{ij}\Delta R'_{ij}\nu_{ij},\end{aligned}$$

where  $c_{ij,n}^*$  is in between  $\Delta R'_{ij}\gamma$  and  $\Delta R'_{ij}\hat{\gamma}_n$ . We have that

$$\hat{S}_{ij} = S_{ij} + \hat{S}_{ij,1}(\hat{\gamma}_n - \gamma) + (\hat{\gamma}_n - \gamma)'\hat{S}_{ij,2}(\hat{\gamma}_n - \gamma).$$

Note that, for some constant  $C > 0$ ,  $\|\hat{S}_{ij,1}\| \leq Ch_n^{-2}g_{ij,2}$  as before and

$$\|\hat{S}_{ij,2}\| \leq \frac{C}{h_n^3} \underbrace{\|\Delta W_{ij}\| \|\Delta R_{ij}\|^2}_{g_{ij,3}} |\nu_{ij}|.$$

Observe that

$$\begin{aligned}
& c' \hat{\Sigma}_{W\nu,1} c - c' \tilde{\Sigma}_{W\nu,1} c \\
& \leq \underbrace{\frac{(\hat{\gamma}_n - \gamma)'}{\sqrt{h_n}} \binom{n}{3}^{-1} \sum_{i < j < k} \frac{\sqrt{h_n}}{3} (S_{ij} \hat{S}'_{ik,1} + S_{ij} \hat{S}'_{jk,1} + S_{ik} \hat{S}'_{jk,1})}_{D_{9,1}} \\
& + \underbrace{\binom{n}{3}^{-1} \sum_{i < j < k} \frac{\sqrt{h_n}}{3} (\hat{S}_{ij,1} S_{ik} + \hat{S}_{ij,1} S_{jk} + \hat{S}_{ik,1} S_{jk}) \frac{(\hat{\gamma}_n - \gamma)}{\sqrt{h_n}}}_{D_{10,2}} \\
& + \underbrace{\frac{(\hat{\gamma}_n - \gamma)'}{h_n^{3/2}} \binom{n}{3}^{-1} \sum_{i < j < k} \frac{h_n^3}{3} (S_{ij} \hat{S}'_{ik,2} + S_{ij} \hat{S}'_{jk,2} + S_{ik} \hat{S}'_{jk,2}) \frac{(\hat{\gamma}_n - \gamma)}{h_n^{3/2}}}_{D_{10,3}} \\
& + \underbrace{\frac{(\hat{\gamma}_n - \gamma)'}{h_n^{3/2}} \binom{n}{3}^{-1} \sum_{i < j < k} \frac{h_n^3}{3} (\hat{S}_{ij,2} S_{ik} + \hat{S}_{ij,2} S_{jk} + \hat{S}_{ik,2} S_{jk}) \frac{(\hat{\gamma}_n - \gamma)}{h_n^{3/2}}}_{D_{9,4}} \\
& + C^2 \underbrace{\binom{n}{3}^{-1} \sum_{i < j < k} \frac{1}{3} (g_{ij,2} g_{ik,3} + g_{ij,2} g_{jk,3} + g_{ik,2} g_{jk,3}) \frac{\|\hat{\gamma}_n - \gamma\|^3}{h_n^5}}_{D_{10,5}} \\
& + C^2 \underbrace{\binom{n}{3}^{-1} \sum_{i < j < k} \frac{1}{3} (g_{ij,3} g_{ik,2} + g_{ij,3} g_{jk,2} + g_{jk,3} g_{ik,2}) \frac{\|\hat{\gamma}_n - \gamma\|^3}{h_n^5}}_{D_{10,6}} \\
& + \underbrace{\binom{n}{3}^{-1} \sum_{i < j < k} \frac{h_n^2}{3} (\|\hat{S}_{ij,1}\| \|\hat{S}_{ik,1}\| + \|\hat{S}_{ij,1}\| \|\hat{S}_{jk,1}\| + \|\hat{S}_{jk,1}\| \|\hat{S}_{ik,1}\|) \frac{\|\hat{\gamma}_n - \gamma\|^2}{h_n^2}}_{D_{10,7}} \\
& + C^2 \underbrace{\binom{n}{3}^{-1} \sum_{i < j < k} \frac{1}{3} (g_{ij,3} g_{ik,3} + g_{ij,3} g_{jk,3} + g_{jk,3} g_{ik,3}) \frac{\|\hat{\gamma}_n - \gamma\|^4}{h_n^6}}_{D_{10,8}}.
\end{aligned}$$

First we stochastically bound  $D_{10,1}$  and  $D_{10,2}$ . For any vector  $a \in \mathbb{R}^{q_r}$  and some constant  $C > 0$ ,

$$\begin{aligned}
E[a' D_{10,1}] &= \frac{\sqrt{h_n}}{h_n^3} E \left[ E[c' S_{12} | \xi_1, U_1] E[a' \hat{S}_{13} c | \xi_1, U_1] \right] \\
&\leq \frac{1}{h_n^{3/2}} \left\{ E \left[ \int E[d_{121} d_{122} c' \Delta W_{\nu 12} | \Delta R_{12} = s_1 h_n, \xi_1, U_1]^2 K(s_1) f_{R\gamma | \xi_1, U_1}(s_1 h_n) \right] \right\}^{1/2} \\
&\times C E[\|\Delta W_{12}\|^2 \|\Delta R_{13}\|^2 \nu_{13}^2]^{1/2} \\
&= O(h_n^{(2k-1)/2}) = o(1),
\end{aligned}$$

where the first line hold from the conditional independence, the inequality follows from Cauchy-Schwartz and Assumption 8, and the final line holds from the implication from  $c'\Sigma_{W\nu,1}c = 0$  and Assumptions 4, 6, 7, and 8. Repeating the calculation in the proof for Lemma 7 and adjusting for covariances with one index in common,

$$\text{Var}[a'D_{10,1}] = O\left(\frac{1}{n^2h_n^3}\right) + O\left(\frac{h_n E[S_{12}a'\hat{S}'_{13,1}S_{14}a'\hat{S}'_{15,1}]}{n}\right) = o(1),$$

where the last equality holds from the stated assumption on  $h_n$  for  $k \geq 1$  and

$$\begin{aligned} & E[S_{12}a'\hat{S}'_{13,1}S_{14}a'\hat{S}'_{15,1}] \\ &= \frac{16}{h_n^2} E\left[ E[d_{121}d_{122}c'\Delta W_{12}\nu_{12}|\Delta R'_{12}\gamma = h_n s_1, \boldsymbol{\xi}_1] \right. \\ & \times E[d_{131}d_{132}c'\Delta W_{13}a'\Delta R_{13}\nu_{13}|\Delta R'_{13}\gamma = h_n s_2, \boldsymbol{\xi}_1] \\ & \times E[d_{141}d_{142}c'\Delta W_{14}\nu_{14}|\Delta R'_{14}\gamma = h_n s_3, \boldsymbol{\xi}_1] \\ & \times E[d_{151}d_{152}c'\Delta W_{15}a'\Delta R_{15}\nu_{15}|\Delta R'_{15}\gamma = h_n s_4, \boldsymbol{\xi}_1] \\ & \left. \times K(s_1)K'(s_2)K(s_3)K'(s_4)\prod_{i=1}^4 f_{R\gamma|\xi_1, U_1}(s_i h_n) \right] \\ &= O\left(\frac{1}{h_n^2}\right) \end{aligned}$$

by Assumptions 4, 6, and 8. Thus,  $D_{1,1} = o_p(1)$ . Similarly, we have  $D_{1,2} = o_p(1)$ .

Next, we stochastically bound  $D_{10,3}$  and  $D_{9,4}$ . For any finite  $a, b \in \mathbb{R}^{q_r}$ , for some  $C > 0$ ,

$$\begin{aligned} & E[a'D_{10,3}b] \\ & \leq \frac{C}{h_n} E[|S_{12}|g_{13,3}] \\ &= CE\left[ \int E[d_{121}d_{122}|c'\Delta W_{12}||\nu_{12}||\Delta R'_{12}\gamma = h_n s_1, \boldsymbol{\xi}_1] \right. \\ & \times E[g_{13}|\boldsymbol{\xi}_1] \\ & \left. \times K(s_1)f_{R\gamma|\xi_1, U_1}(s_1 h_1)f_{R\gamma|\xi_1, U_1}(s_2) \right] \\ &= O(1), \end{aligned}$$

where the last equality holds from Assumptions 4, 6, 7 and 8. The variance is calculated similarly as before:

$$\text{Var}[a'D_{10,3}b] = O\left(\frac{1}{n^2h_n}\right) + O\left(\frac{h_n}{n}\right) = o(1).$$

Thus,  $D_{10,3} = O_p(1)$ . Similarly,  $D_{10,4} = O_p(1)$ .

$D_{10,5}, D_{10,6}$ , and  $D_{10,8}$  are all  $O_p(1)$  by the similar computation as in Lemma 7.

$D_{9,7}$  is stochastically bounded as follows. Observe that

$$\begin{aligned}
E[D_{10,7}] &= \frac{4}{h_n^2} E[\|\hat{S}_{12,1}\| \|\hat{S}_{13,1}\|] \\
&\leq 4 \int E[|c' \Delta W_{12}| |c' \Delta W_{12}| \|\Delta R_{12}\| \|\Delta R_{13}\| |\nu_{12}| |\nu_{13}| |\Delta R'_{12} \gamma = h_n s_1, \Delta R'_{13} \gamma = h_n s_2|] \\
&\quad \times K'(s_1) K'(s_2) f_{R\gamma,2}(h_n s_1, h_n s_2) \\
&= O(1),
\end{aligned}$$

where the last line holds from Assumption 4, 6, and 8. The variance is calculated similarly as before:

$$Var[D_{10,7}] = O\left(\frac{1}{n^2 h_n^2}\right) + O\left(\frac{1}{n}\right) = o(1).$$

Thus,  $D_{10,7} = O_p(1)$ .

Finally, the above implies

$$\begin{aligned}
&nh_n |c' \hat{\Sigma}_{W\nu,1} c - c' \tilde{\Sigma}_{W\nu,1} c| \\
&\leq O_p(1) \times \left( \frac{nh_n \|\hat{\gamma}_n - \gamma\|}{h_n} + \frac{nh_n \|\hat{\gamma}_n - \gamma\|^2}{h_n^3} + \frac{nh_n \|\hat{\gamma}_n - \gamma\|^3}{h_n^5} + \frac{nh_n \|\hat{\gamma}_n - \gamma\|^4}{h_n^6} \right) = o_p(1),
\end{aligned}$$

because

$$\begin{aligned}
\frac{nh_n \|\hat{\gamma}_n - \gamma\|}{h_n} &= O(\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|) = o_p(1), \\
\frac{nh_n \|\hat{\gamma}_n - \gamma\|^2}{h_n^3} &= O\left(\frac{(\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|)^2}{nh_n^3}\right) = o_p(1), \\
\frac{nh_n \|\hat{\gamma}_n - \gamma\|^3}{h_n^5} &= O\left(\frac{(\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|)^3}{n^2 h_n^{11/2}}\right) = o_p(1), \\
\frac{nh_n \|\hat{\gamma}_n - \gamma\|^4}{h_n^6} &= O\left(\frac{(\sqrt{Nh_n} \|\hat{\gamma}_n - \gamma\|)^4}{n^3 h_n^7}\right) = o_p(1),
\end{aligned}$$

by Assumption 10,  $nh_n^3 = O(n^{(2k-3)/(2k+3)})$ ,  $n^2 h_n^{11/2} = O(n^{(4k-5)/(2k+3)})$ , and  $n^3 h_n^7 = O(n^{(6k-5)})$  all diverging for  $k \geq 2$ . This completes the proof for Lemma 10.  $\square$