# Bootstrap Inference for Matching Estimator under Spatially Correlated Effects

Minghai Mao*

Liaoning University, Li Anmin Institute of Economic Research

Zixin Yang†

Shanghai University of Finance and Economics, School of Statistics and Data Science

January 17, 2026

*Address: Liaoning University, Chongshanzhong Road 66, 110083, Shenyang, China. Email: maominghai@lnu.edu.cn.

†Address: Shanghai University of Finance and Economics, Guoding Road 777, 200433, Shanghai, China. Email: yangzixin@mail.shufe.edu.cn. Corresponding author.

**Abstract**

This paper proposes a dependent wild bootstrap (DWB) procedure for inference on matching estimators in the presence of spatial dependence. The bootstrap algorithm delivers a spatial heteroskedasticity and autocorrelation consistent (HAC) variance estimator that accommodates unknown spatial dependence. By accounting for spatial correlation in the residuals, this approach improves the accuracy of standard error estimation for nonparametric matching estimators. A small Monte Carlo simulation illustrates its performance in a finite sample.

*Keywords*: Matching estimator; Bootstrap inference; Spatial dependence

# 1 Introduction

Matching estimators are widely used in regional policy evaluations, such as those assessing the impact of infrastructure investment, industrial development, and immigration. In such contexts, spatial dependence in outcomes – for instance, due to agglomeration economies – can be substantial and cannot be ignored. For example, Kline and Moretti (2014) investigate the long-run effects of the Tennessee Valley Authority by matching pre-treatment growth patterns between treated and untreated counties, using an Oaxaca-Blinder regression weighted by the inverse propensity score. Dustmann et al. (2017) assess the labor market impacts of a Czech immigration wave into Germany through a differences-in-differences design combined with nearest-neighbor matching. Both studies apply spatial heteroskedasticity and autocorrelation consistent (HAC) variance estimators to account for potential spatial correlation in treatment effect estimation.

Despite their popularity, matching estimators pose challenges for statistical inference, especially under spatial dependence. Abadie and Imbens (2008) show that the standard nonparametric bootstrap fails for nearest-neighbor matching estimators, even under i.i.d. sampling. More recently, Roth et al. (2023) demonstrate that combining matching with regression adjustment invalidates standard bootstrap inference, as the resulting estimator is no longer asymptotically linear. These findings highlight the need for specialized bootstrap methods that remain valid when spatial correlation is present.

This paper proposes a dependent wild bootstrap (DWB) procedure for inference on the average treatment effect on the treated (ATT) estimated via bias-corrected matching. Our method explicitly accounts for spatial dependence in the residuals, delivering a spatial HAC-consistent variance estimator without requiring strong parametric assumptions. The approach builds on the framework of Abadie and Imbens (2011) for bias-corrected matching estimators and extends the bootstrap validity results of Otsu and Rai (2017), who develop a wild bootstrap method for i.i.d. samples. To accommodate spatial correlation, we adopt a flexible spatial HAC structure modeled nonparametrically based on geographic distance.

Our work is closely related to Conley et al. (2023), who apply the DWB to spatial HAC inference in linear regression. We adopt a similar approach by avoiding restrictive spatial mixing conditions by modeling spatial dependence as a linear process—including SAR and Cliff-Ord-type models as special cases (e.g., Kelejian and Prucha, 2007). However, our focus is on the matching estimator, building on existing literature on nonparametric estimation for spatial data (Robinson, 2011; Jenish, 2012; Su, 2012).

The remainder of the paper is organized as follows. Section 2 introduces the matching estimator and our bootstrap inference. Section 3 presents Monte Carlo simulation results for illustration. Section 4 concludes.

# 2 Test statistic

## 2.1 Model setup

Consider a sample composed of $n$ local units, where each unit is assigned a binary treatment. Denote the outcome $Y_n = (Y_{1n}, \ldots, Y_{nn})$ with each $Y_{in} \in \mathcal{Y} \subset \mathbb{R}$, the treatment $D_n = (D_{1n}, \ldots, D_{nn})$ with each $D_{in} \in \{0, 1\}$ and covariates $X_n = (X_{1n}, \ldots, X_{nn})$ with each $X_{in} \in \mathcal{X} \subset \mathbb{R}^k$. We assume the well known Stable Unit Treat Value Assumption (SUTVA) and thus our potential outcomes consist of $(Y_{in}(1), Y_{in}(0))$ such that

$$Y_{in}(d) = \mu(X_{in}, d) + V_{in}(d) \quad \text{for } d = 0, 1, \tag{2.1}$$

where $\mu(x, d) = \mathrm{E}[Y_{in}(d) \mid X_{in} = x]$. The observed outcome $Y_{in} = D_{in}Y_{in}(1) + (1 - D_{in})Y_{in}(0)$. We consider the spatial dependence that occurs in the error term. Specifically, let $V_{in} =$

$Y_{in} - \mu(X_{in}, D_{in})$ and $V_n = (V_{1n}, \ldots, V_{nn})^\top$. We assume that

$$V_n = R_n \epsilon_n, \tag{2.2}$$

where $\epsilon_n$ is an $n \times 1$ vector of i.i.d. innovations with zero mean and unit variance, and $R_n$ is an unknown non-stochastic matrix. The spatial autoregressive (SAR) process is a special case of (2.2) with $R_n = (I_n - \rho W_n)^{-1}$, where $W_n$ is an $n \times n$ spatial weights matrix. We further assume the existence of a meaningful distance measure $\delta_{ijn} = \delta_{jin} \geq 0$ between units $i$ and $j$ (e.g., a Euclidean norm in $\mathbb{R}^r$, $r \geq 1$), together with some technical conditions on the number of neighbors based on this distance. The detailed assumptions are provided in Appendix A.1.

We adopt the bias-corrected matching estimator proposed by Abadie and Imbens (2011) to estimate the average treatment effect on the treated, denoted by $\tau_{ATT}$. Let $\mathcal{J}_M(i)$ denote the set of $M$ nearest neighbors of unit $i$ in the opposite treatment group in terms of covariate values. To correct the potential bias of the matching estimator, we use the nonparametric series estimators $\hat{\mu}(X_{in}, 1)$ and $\hat{\mu}(X_{in}, 0)$ to impute the missing values of $Y_{in}(1)$ and $Y_{in}(0)$, respectively.[1] The imputed potential outcomes for $\tilde{Y}_{in}(1)$ and $\tilde{Y}_{in}(0)$ are defined as

$$\tilde{Y}_{in}(0) = \begin{cases} Y_{in} & \text{if } D_{in} = 0, \\ \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} (Y_{jn} + \hat{\mu}(X_{in}, 0) - \hat{\mu}(X_{jn}, 0)) & \text{if } D_{in} = 1, \end{cases}$$

and

$$\tilde{Y}_{in}(1) = \begin{cases} \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} (Y_{jn} + \hat{\mu}(X_{in}, 1) - \hat{\mu}(X_{jn}, 1)) & \text{if } D_{in} = 0, \\ Y_{in} & \text{if } D_{in} = 1. \end{cases}$$

[1]Note that Assumption 2(i) in Appendix A.1 guarantees that $\mu(x, d) = \mathrm{E}[Y_{in} \mid D_{in} = d, X_{in} = x]$.

The bias-corrected matching estimator for the ATT is given by

$$\begin{aligned}\tilde{\tau}^m_{ATT,bc} &= \frac{1}{n_1}\sum_{i=1}^{n} D_{in}(\tilde{Y}_{in}(1) - \tilde{Y}_{in}(0)) \\ &= \frac{1}{n_1}\sum_{i=1}^{n}\left[D_{in}\hat{V}_{in} - (1-D_{in})\frac{K_M(i)}{M}\hat{V}_{in} + D_{in}(\hat{\mu}(X_{in},1) - \hat{\mu}(X_{in},0))\right],\end{aligned} \tag{2.3}$$

where $n_1$ denotes the sample size of treated units, $\hat{V}_{in} = Y_{in} - \hat{\mu}(X_{in}, D_{in})$, and $K_M(i)$ denotes the number of times unit $i$ is used as a match, that is, $K_M(i) = \sum_{l=1}^{n} \mathbb{1}\{i \in \mathcal{J}_M(l)\}$.

Proposition 1 extends Theorem 2′ in Abadie and Imbens (2011) to the case with a spatially dependent error process.

**Proposition 1.** *Under Assumptions 1–4 in the Appendix, we have*

$$\left(V^{\tau(X),t} + V^{E,t}\right)^{-1/2}\sqrt{n_1}(\tilde{\tau}^m_{ATT,bc} - \tau_{ATT}) \xrightarrow{d} N(0,1),$$

*where* $V^{\tau(X),t} = \mathrm{E}\left[(\mu(X_{in},1) - \mu(X_{in},0) - \tau_{ATT})^2 \mid D_{in} = 1\right]$ *and* $V^{E,t} = n_1\, Var(\tilde{\tau}^m_{ATT,bc} \mid D_n, X_n)$.

## 2.2 Bootstrap procedure

The conditional variance component $V^{E,t}$ in Proposition 1 depends on the unknown spatial covariance structure of the errors. Direct analytical estimation of $V^{E,t}$ necessitates either the specification of a parametric spatial model or the employment of a nonparametric estimator of the underlying covariance matrix, which can be cumbersome or unreliable in practice. In this section, we propose a dependent wild bootstrap procedure for inference on the matching estimator. We utilize the following decomposition

$$\tilde{\tau}^m_{ATT,bc} - \tau_{ATT} = \frac{1}{n_1}\sum_{i=1}^{2n}\hat{\zeta}_{i,2n}, \tag{2.4}$$

where

$$\hat{\zeta}_{i,2n} = \begin{cases} D_{in}(\hat{\mu}(X_{in}, 1) - \hat{\mu}(X_{in}, 0) - \tilde{\tau}^m_{ATT,bc}) & \text{for } i = 1, \ldots, n \\ D_{in}\hat{V}_{in} - (1 - D_{in})\frac{K_M(i)}{M}\hat{V}_{in} & \text{for } i = n+1, \ldots, 2n. \end{cases}$$

As shown in Abadie and Imbens (2012), in the i.i.d. case, expression (2.4) with the conditional bias eliminated forms a martingale array, which simplifies the derivation of the asymptotic variance of the matching estimator. Although our setting allows for spatial dependence, valid bootstrap inference can still be conducted by resampling from $\{\hat{\zeta}_{i,2n}\}_{i=1}^{2n}$. Specifically, we construct the weighted bootstrap counterpart of $\sqrt{n_1}(\tilde{\tau}^m_{ATT,bc} - \tau_{ATT})$ as

$$\begin{aligned} \sqrt{n_1} T^*_{ATT,bc} &= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{2n} \mathcal{W}^*_{i,2n}\hat{\zeta}_{i,2n} \\ &= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n} \left( \mathcal{W}^*_{i,2n}\hat{\zeta}_{i,2n} + \mathcal{W}^*_{n+i,2n}\hat{\zeta}_{n+i,2n} \right) \\ &= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n} \Bigg[ \mathcal{W}^*_{i,2n} D_{in}(\hat{\mu}(X_{in}, 1) - \hat{\mu}(X_{in}, 0) - \tilde{\tau}^m_{ATT,bc}) \\ &\qquad + \mathcal{W}^*_{n+i,2n}\left( D_{in} - (1 - D_{in})\frac{K_M(i)}{M} \right)\hat{V}_{in} \Bigg]. \end{aligned}$$

For the choice of bootstrap weights $\{\mathcal{W}^*_{i,2n}\}_{i=1}^{2n}$, in the i.i.d. setting, one may adopt the standard wild bootstrap method, such as that proposed by Mammen (1993). The requirements of the weights are given by Assumption W in Otsu and Rai (2017). In our setting, while $\{(D_{in}, X_{in})\}_{i=1}^{n}$ are assumed to be i.i.d., the variables $V_{in}$ are allowed to be spatially dependent. Accordingly, we follow Otsu and Rai (2017) and apply wild bootstrap weights to $\{\mathcal{W}^*_{i,2n}\}_{i=1}^{n}$, but modify the resampling scheme for $\{\mathcal{W}^*_{i,2n}\}_{i=n+1}^{2n}$ to incorporate spatial dependence. Specifically, we generate $\eta_{in} = \mathcal{W}^*_{n+i,2n}$, $i = 1, 2, \ldots, n$, such that

$$\mathrm{Cov}^*(\eta_{in}, \eta_{jn}) = K\left(\frac{\delta_{ijn}}{\delta_n}\right), \tag{2.5}$$

where $K(\cdot)$ denotes a kernel function on $\mathbb{R}$, and $\delta_n$ is a bandwidth parameter. We further impose $\mathrm{E}^*(\eta_{in}) = 0$ and $\mathrm{Var}^*(\eta_{in}) = 1$. The specific requirements on $K(\cdot)$ and $\delta_n$ are given in Assumptions 6 and 7 in the Appendix. Particularly, we require that the matrix $\mathbb{K}_n = [K(\frac{\delta_{ijn}}{\delta_n})]_{i,j=1}^n$ to be positive semidefinite for all $n$. For locations indexed on the line, this condition is equivalent to $\int_{-\infty}^{\infty} K(u)e^{-iux}du \geq 0$ for all $x \in \mathbb{R}$. However, for distance measure corresponding to a Euclidean norm on $\mathbb{R}^r$ with $r \geq 2$, the above condition is not adequate. In this case, the Gaussian kernel, and a polynomial kernel of order $v$, i.e., $K(x) = (1 - |x|)^v \mathbb{1}(|x| \leq 1)$ for some $v \geq (r + 1)/2$, can be verified to satisfy the positive semi-definiteness condition of $\mathbb{K}$; see, e.g., Kelejian and Prucha (2007) and Conley et al. (2023) for more details.

The detailed algorithm to obtain valid bootstrap percentile intervals is given as follows.

**Step 1.** Compute the bias-corrected matching estimator $\tilde{\tau}^m_{ATT,bc}$ as defined in (2.3) using the imputed nonparametric regression estimator $\hat{\mu}(X_{in}, D_{in})$. Obtain the residual $\hat{V}_{in} = Y_{in} - \hat{\mu}(X_{in}, D_{in})$.

**Step 2.** For a given bandwidth $\delta_n$, compute the kernel matrix $\mathbb{K}_n = \left[K\left(\frac{\delta_{ijn}}{\delta_n}\right)\right]_{i,j=1}^n$ and its eigendecomposition $\mathbb{K}_n = \Phi_n \Lambda_n \Phi_n^\top$, where $\Lambda_n = \mathrm{diag}(\lambda_{1n}, \lambda_{2n}, \ldots, \lambda_{nn})$ and $\Phi_n = [\phi_{1n}, \phi_{2n}, \ldots, \phi_{nn}]$ with $\lambda_{kn} \geq 0$ denoting the eigenvalues of $\mathbb{K}_n$ and $\phi_{kn}$ the corresponding orthonormal eigenvectors.

**Step 3.** Generate a $2n \times 1$ vector $(w_{1n}^{*\top}, v_n^\top)^\top \sim (0, I_{2n})$, where $w_{1n}^* = (\mathcal{W}_{1,2n}^*, \ldots, \mathcal{W}_{n,2n}^*)^\top$ follows the two point distribution of Mammen (1993). Then generate the random vector $\eta_n = (\mathcal{W}_{n+1,2n}^*, \ldots, \mathcal{W}_{2n,2n}^*)^\top$ as $\eta_n = \Phi_n \Lambda_n^{1/2} v_n$. We require that the fourth moment of the element of $v_n$ exists. A commonly used example is an $N(0, I_n)$ distributed vector as suggested by Conley et al. (2023).

**Step 4.** Compute the DWB test statistic

$$T_{DWB}^* = \frac{1}{n_1} \sum_{i=1}^{n} \left[ D_{in} \hat{\xi}_{in}^* + D_{in} \hat{V}_{in}^* - (1 - D_{in}) \frac{K_M(i)}{M} \hat{V}_{in}^* \right], \tag{2.6}$$

where $\hat{\xi}^*_{in} = \mathcal{W}^*_{i,2n}\big(\hat{\mu}(X_{in}, 1) - \hat{\mu}(X_{in}, 0) - \tilde{\tau}^m_{ATT,bc}\big)$ and $\hat{V}^*_{in} = \mathcal{W}^*_{n+i,2n}\hat{V}_{in} = \eta_{in}\hat{V}_{in}$.

**Step 5.** Repeat Steps 3–4 $B$ times. The $(1-\alpha)$ confidence interval of $\tau_{ATT}$ is constructed as $[\tilde{\tau}^m_{ATT,bc} - q_{1-\alpha/2}, \tilde{\tau}^m_{ATT,bc} - q_{\alpha/2}]$, where $q_{1-\alpha/2}$ and $q_{\alpha/2}$ denote the $1-\alpha/2$ and $\alpha/2$ quantiles of the distribution of $T^*_{DWB}$, respectively.

**Theorem 1.** *Let $T^*_{DWB}$ be the DWB test statistic obtained from the algorithm described above. Under Assumptions 1–7 in the Appendix, as $n \to \infty$,*

$$\sup_{r\in\mathbb{R}} |Pr\{\sqrt{n_1}T^*_{DWB} \le r|Y_n, D_n, X_n\} - Pr\{\sqrt{n_1}(\tilde{\tau}^m_{ATT,bc} - \tau_{ATT}) \le r\}| \xrightarrow{p} 0.$$

Theorem 1 establishes the consistency of the bootstrap distribution of $\sqrt{n_1}T^*_{DWB}$. It follows by showing the consistency of the bootstrap variance (see (3) in the Appendix), and $(V^{\tau(X),t} + V^{E,t})^{-1/2}\sqrt{n_1}T^*_{DWB} \xrightarrow{d^*} N(0,1)$ in probability; see Appendix A.2 for details.

# 3 Simulation

We evaluate the finite-sample performance of our bootstrap inference methods using a data-generating process (DGP) adapted from Frölich (2004) and Otsu and Rai (2017). Specifically, the outcomes are generated as

$$Y_{in}(1) = \tau + m(\|X_{in}\|) + V_{in}(1), \quad Y_{in}(0) = m(\|X_{in}\|) + V_{in}(0), \tag{3.1}$$

where $X_{in} = (X^{(1)}_{in}, \dots, X^{(k)}_{in})^\top$with $X^{(j)}_{in} = \pi_{in}|\phi^{(j)}_{in}|/\|\phi_{in}\|$ for $j = 1, \dots, k$, $\pi_{in} \backsim U[0,1]$ and $\phi_{in} = (\phi^{(1)}_{in}, \dots, \phi^{(k)}_{in})^\top \backsim N(0, I_k)$. The parameter $\tau$ is fixed at 0 and the dimension $k$ is assigned a value of 5. The data $\{(Y_{in}, X_{in}, D_{in})\}^n_{i=1}$ are generated on a $\sqrt{n} \times \sqrt{n}$ square lattice and the distance $\delta_{ijn}$ is defined as the Euclidean distance between units $i$ and $j$. We focus on the treatment assignment $D_{in} = \mathbb{1}(\delta_{in} \le \delta_c)$, where $\delta_{in}$ represents the unit's distance to the square's upper border.[2]

The error process follows a SAR structure:

$$V_n = \rho W_n V_n + \epsilon_n, \qquad \epsilon_{in} \sim N(0, \varsigma^2/V(\rho)), \tag{3.2}$$

where $W_n$ denotes the row-normalized spatial weights matrix based on the rook-contiguity on the square lattice. We set $\varsigma^2 = 0.04$, and $V(\rho) = \frac{1}{n}\sum_{i=1}^{n} S_{in}(\rho)$, where $S_{in}(\rho)$ is the $i$th diagonal element of $S_n(\rho) = [(I_n - \rho W_n)^\top (I_n - \rho W_n)]^{-1}$. The latter choice ensures that the variance of $V_{in}$ does not vary much across different spatial dependence levels.

We set the sample size $n = 400, 900, 2500$ and $\delta_c = 2$. The number of treated units are $n_1 = 40, 60, 100$, respectively. For robustness evaluation, we consider the spatial dependence parameter $\rho \in \{0, 0.25, 0.5\}$. The nonparametric component $m(\cdot)$ is varied in six functional forms to capture a wide range of curvature and smoothness:

DGP 1: $m(z) = 0.15 + 0.7z$,

DGP 2: $m(z) = 0.1 + z/2 + \exp(-200(z - 0.7)^2)/2$,

DGP 3: $m(z) = 0.8 - 2(z - 0.9)^2 - 5(z - 0.7)^2 - 10(z - 0.6)^4$,

DGP 4: $m(z) = 0.2 + \sqrt{1 - z} - 0.6(0.9 - z)^2$,

DGP 5: $m(z) = 0.2 + \sqrt{1 - z} - 0.6(0.9 - z)^2 - 0.1z\cos(30z)$,

DGP 6: $m(z) = 0.4 + 0.25\sin(8z - 5) + 0.4\exp(-16(4z - 2.5)^2)$.

All the shapes of $m(\cdot)$ functions can be found in Otsu and Rai (2017).

To apply the bias-corrected matching estimator, we use $M = 5$ matches and define the imputed regression function $\hat{\mu}(X_{in}, D_{in})$ as a series estimator. Specifically, $\hat{\mu}(X_{in}, D_{in})$ is

[2]We also conduct simulations where $D_{in}$ depends on the covariates $X_{in}$ as in Otsu and Rai (2017). The results show that, in that case, the size distortion of the standard wild bootstrap under spatial dependence weakens. Simulation results are available upon request.

obtained by regressing $Y_{in}$ on polynomial basis functions of $X_{in}$ and $D_{in}$.[3] We employ the DWB using the quadratic triangular kernel $K(x) = (1 - |x|)^2 \mathbb{1}(|x| \leq 1)$. For selecting the bandwidth, we provide an algorithm in the same spirit as Conley et al. (2023), the details of which are presented in Appendix A.3.

Table 1 presents the simulation results of the DWB procedure, which uses 1000 bootstrap replications and selects the bandwidth $\delta_n$ by the provcedure outlined in Appendix A.3. For each setup, we report the bias, the 95% confidence interval (CI) coverage rates, and the average CI lengths across 1000 Monte Carlo replications. As can be seen from Table 1, when there is no spatial dependence ($\rho = 0$), the coverage rates of the DWB are close to the nominal level for all specifications of $m(\cdot)$. With relatively strong spatial dependence ($\rho = 0.5$), the coverage rates decline notably: at $n = 400$, they drop to around 85%. However, as the sample size increases to $n = 2500$, the rates improve to around 90%. Under moderate spatial dependence ($\rho = 0.25$), the results are heterogeneous. For DGPs 2, 3 and 6, the coverage rates approach the nominal level as the sample size increases to $n = 2500$, while for the other cases, they remain at around 90%. This discrepancy reflects uncertainties in bandwidth selection when spatial dependence is relatively weak.

Tables A.1-A.3 in the Appendix show the simulation results for the DWB with bandwidths specified as $\delta_n = c_k \times n^{1/6}$, where $c_k \in \{1, 2, 3\}$. We also report the results for the standard wild bootstrap, which sets $\{\mathcal{W}_{in}\}_{i=1}^{2n}$ as i.i.d. two-point distributed variables. Our findings are as follows. First, when $\rho = 0$, the standard wild bootstrap attains coverage rates close to the 95% nominal level. However, its performance deteriorates greatly when $\rho$ increases: for $\rho = 0.5$, the naive wild bootstrap produces under-coverage even when $n = 2500$. In contrast, the DWB procedure with the specified bandwidth choices delivers satisfactory coverage rates. Despite a decreasing tendency as $\rho$ increases, the coverage rates of the DWB

[3] According to Abadie and Imbens (2011), the polynomial series order $\hat{\mu}(x, d)$ is $L = O(n^{\nu})$ for $\nu \in (0, 2/95)$ when $k = 5$. Hence, we approximate $\mu(x, d)$ by a quadratic function without interaction terms, providing a slightly more flexible specification than the linear model used in Otsu and Rai (2017). This approximation balances between $L = 1$ and 2, which results in a finite sample performance that is slightly better than both the linear and fully quadratic functions.

remain at a reasonable level when the sample size is large. In most cases, the choice of bandwidth $\delta_n$ has a minor effect on the coverage rates and average CI lengths. Notably, when spatial dependence is moderate ($\rho = 0.25$), smaller bandwidths ($c_k = 1, 2$) yield slightly better performance than larger ones ($c_k = 3$) and exhibit a clearer convergence pattern than data-driven bandwidth selection algorithms.

Table 1: Simulation results using the proposed bandwidth selection method.

| | $m(\cdot)$ | $n = 20 \times 20$ | | | $n = 30 \times 30$ | | | $n = 50 \times 50$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | 95% Cov. | CI Length | Bias | 95% Cov. | CI Length | Bias | 95% Cov. | CI Length |
| $\rho = 0$ | 1 | 0.0049 | 0.951 | 0.1405 | 0.0031 | 0.938 | 0.1137 | 0.0017 | 0.947 | 0.0878 |
| | 2 | -0.0002 | 0.965 | 0.1654 | -0.0016 | 0.961 | 0.1349 | 0.0005 | 0.970 | 0.1037 |
| | 3 | 0.0084 | 0.965 | 0.1621 | 0.0052 | 0.968 | 0.1314 | 0.0024 | 0.961 | 0.1015 |
| | 4 | 0.0021 | 0.939 | 0.1408 | 0.0006 | 0.952 | 0.1149 | 0.0001 | 0.941 | 0.0884 |
| | 5 | 0.0027 | 0.935 | 0.1434 | 0.0017 | 0.932 | 0.1159 | 0.0017 | 0.950 | 0.0896 |
| | 6 | -0.0149 | 0.959 | 0.1840 | -0.0129 | 0.957 | 0.1489 | -0.0096 | 0.970 | 0.1150 |
| $\rho = 0.25$ | 1 | 0.0034 | 0.889 | 0.1407 | 0.0019 | 0.894 | 0.1156 | 0.0009 | 0.883 | 0.0891 |
| | 2 | -0.0057 | 0.905 | 0.1658 | -0.0034 | 0.904 | 0.1372 | 0.0013 | 0.944 | 0.1050 |
| | 3 | 0.0083 | 0.906 | 0.1622 | 0.0063 | 0.920 | 0.1334 | 0.0049 | 0.944 | 0.1026 |
| | 4 | 0.0045 | 0.893 | 0.1423 | -0.0002 | 0.900 | 0.1167 | 0.0012 | 0.904 | 0.0902 |
| | 5 | 0.0048 | 0.903 | 0.1445 | 0.0036 | 0.884 | 0.1183 | 0.0034 | 0.899 | 0.0908 |
| | 6 | -0.0140 | 0.929 | 0.1842 | -0.0093 | 0.940 | 0.1499 | -0.0080 | 0.955 | 0.1155 |
| $\rho = 0.50$ | 1 | -0.0008 | 0.832 | 0.1595 | 0.0032 | 0.864 | 0.1400 | -0.0001 | 0.879 | 0.1115 |
| | 2 | -0.0024 | 0.854 | 0.1771 | 0.0006 | 0.878 | 0.1548 | -0.0002 | 0.896 | 0.1212 |
| | 3 | 0.0107 | 0.851 | 0.1730 | 0.0070 | 0.899 | 0.1518 | 0.0019 | 0.889 | 0.1197 |
| | 4 | 0.0011 | 0.833 | 0.1610 | 0.0004 | 0.869 | 0.1398 | 0.0010 | 0.907 | 0.1136 |
| | 5 | 0.0039 | 0.835 | 0.1617 | 0.0032 | 0.871 | 0.1411 | 0.0016 | 0.909 | 0.1139 |
| | 6 | -0.0121 | 0.885 | 0.1920 | -0.0126 | 0.906 | 0.1653 | -0.0082 | 0.915 | 0.1273 |

# 4 Conclusion

This paper proposes a dependent wild bootstrap method for making inferences on the matching estimator in the presence of spatial dependence. We show the consistency of the proposed bootstrap method and conduct a set of Monte Carlo experiments to illustrate its effectiveness. Future work will involve extending the bootstrap inference methods of

Adusumilli (2018) and Bodory et al. (2024) to settings with spatial correlated effects.

# 5 Acknowledgments

This work was supported by the National Science Foundation of China [grant number 72203084] and the Fundamental Research Funds for the Central Universities.

# References

Abadie, A. and Imbens, G. (2002). Simple and bias-corrected matching estimators for average treatment effects.

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Abadie, A. and Imbens, G. W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica*, 76(6):1537–1557.

Abadie, A. and Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, 29(1):1–11.

Abadie, A. and Imbens, G. W. (2012). A martingale representation for matching estimators. *Journal of the American Statistical Association*, 107(498):833–843.

Adusumilli, K. (2018). Bootstrap inference for propensity score matching. Technical report, Working paper.

Bodory, H., Camponovo, L., Huber, M., and Lechner, M. (2024). Nonparametric bootstrap for propensity score matching estimators. *Statistics & Probability Letters*, 208:110069.

Conley, T. G., Gonçalves, S., Kim, M. S., and Perron, B. (2023). Bootstrap inference under cross-sectional dependence. *Quantitative Economics*, 14(2):511–569.

Dustmann, C., Schönberg, U., and Stuhler, J. (2017). Labor supply shocks, native wages, and the adjustment of local employment. *The Quarterly Journal of Economics*, 132(1):435–483.

Frölich, M. (2004). Finite-sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics*, 86(1):77–90.

Jenish, N. (2012). Nonparametric spatial regression under near-epoch dependence. *Journal of Econometrics*, 167(1):224–239.

Kelejian, H. H. and Prucha, I. R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140(1):131–154.

Kline, P. and Moretti, E. (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority. *The Quarterly Journal of Economics*, 129(1):275–331.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1):255–285.

Otsu, T. and Rai, Y. (2017). Bootstrap inference of matching estimators for average treatment effects. *Journal of the American Statistical Association*, 112(520):1720–1732.

Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics*, 165(1):5–19.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244.

Su, L. (2012). Semiparametric GMM estimation of spatial autoregressive models. *Journal of Econometrics*, 167(2):543–560.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*, volume 3. Cambridge university press.

# A Appendix

## A.1 Assumptions

The proof is based on Assumptions 1–7. Our approach integrates the bias-corrected matching estimator framework from Abadie and Imbens (2011) with the spatial dependence models of Kelejian and Prucha (2007) and Conley et al. (2023).

**Assumption 1.** (i) $\{(D_{in}, X_{in}, \epsilon_{in})\}_{i=1}^{n}$ are i.i.d. with $\mathrm{E}[\epsilon_{in}|D_{in}, X_{in}] = 0$, and $\mathrm{Var}(\epsilon_n \mid D_n, X_n) = I_n$. (ii) There exists a constant $C_E$ such that $\mathrm{E}[|\epsilon_{in}|^q|D_{in}, X_{in}] \leq C_E$ for some $q \geq 4$, with $0 < C_E < \infty$. (iii) $X_{in}$ is a random vector of dimension $k$ of continuous covariates distributed on $\mathbb{R}^k$ with compact and convex support $\mathcal{X}$, with density bounded away from zero on its support. (iv) For some $s \geq 1$, $n_1^s/n$ converges to $\theta$ with $0 < \theta < \infty$.

**Assumption 2.** (i) (unconfoundedness) $D_{in} \perp\!\!\!\perp (Y_{in}(1), Y_{in}(0)) \mid X_{in}$ a.s. in $\mathcal{X}$. (ii) (overlap) $c < \Pr(D_{in} = 1|X_{in}) < 1 - c$ for some $c > 0$ and a.s. in $\mathcal{X}$.

**Assumption 3.** For $d \in \{0, 1\}$, (i) $\mu(x, d) = \mathrm{E}[Y_{in}|X_{in} = x, D_{in} = d]$ is Lipschitz continuous in $x$, (ii) the fourth moments of the conditional distribution of $Y_{in}$ given $D_{in} = d$ and $X_{in} = x$ exist and are bounded uniformly in $x$, and (iii) $\sigma_{in}^2(x, d) = \mathrm{Var}(Y_{in}|X_{in} = x, D_{in} = d)$ is bounded away from zero for all $i$.

**Remark.** Assumption 1 imposes the homoskedasticity of $\epsilon_{in}$ given $(D_n, X_n)$ for simplicity. However, our framework can be extended to accommodate heteroskedasticity. Note that under Assumption 1, the conditional variance function $\sigma_{in}^2(x, d)$ is independent of $x$. Assumption 1(iv) permits a general rate of growth for the number of treated units $n_1$ in relation to the total sample size $n$. For simplicity, in the following we concentrate on the specific case $s = 1$, which means that $n_1$ increases proportionally to $n$.

Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ be a $k$-dimensional vector of non-negative integers, with $|\lambda| = \sum_{l=1}^{k} \lambda_l$. Let $\partial^\lambda a(x) = \partial^{|\lambda|} a(x)/\partial x_1^{\lambda_1} \cdots \partial x_k^{\lambda_k}$. For $m \geq 0$, define $|a|_m = \max_{|\lambda| \leq m} \sup_{x \in \mathcal{X}} |\partial^\lambda a(x)|$.

**Assumption 4.** For each $d \in \{0,1\}$ and $\lambda$ satisfying $\sum_{l=1}^{k} \lambda_l = k$, the derivative $\partial^\lambda \mu(x,d)$ exists and satisfies $\sup_{x\in\mathcal{X}} |\partial^\lambda \mu(x,d)| \leq C$ for some $C > 0$. Furthermore, $\hat{\mu}(x,d)$ satisfies $|\hat{\mu}(\cdot,d) - \mu(\cdot,d)|_{k-1} = o_p(n^{-\max(\frac{1}{2}-\frac{1}{k},\frac{1}{4})})$ for each $d \in \{0,1\}$.

**Assumption 5.** The matrix $R_n$ is a non-stochastic matrix. The row and column sums of $R_n$ are uniformly bounded in absolute value by some constant $c_R < \infty$.

**Assumption 6.** Let $\delta_{ijn}$ denote a distance measure between units $i$ and $j$, and $\delta_n$ a bandwidth parameter such that $\delta_n \uparrow \infty$ as $n \to \infty$. For each unit $i$, let $l_{in}$ denote the number of units for which $\delta_{ijn} \leq \delta_n$, i.e.,

$$l_{in} = \sum_{j=1}^{n} \mathbb{1}\{\delta_{ijn} \leq \delta_n\},$$

and $l_n = \max_{1\leq i\leq n} l_{in}$. Then, (i) $l_n = o(n^\tau)$ for some $\tau \leq 1/2$, and (ii) $\max_{1\leq i\leq n} \sum_{j=1}^{n} |\mathrm{E}(V_{in}V_{jn})| \delta_{ijn}^{\rho_S} \leq C_S$ for some $\rho_S \geq 1$ and $0 < C_S < \infty$.

**Assumption 7.** (i) The kernel $K : \mathbb{R} \to [-1,1]$, with $K(0) = 1$, $K(x) = K(-x)$, $K(x) = 0$ for $|x| > 1$, satisfies

$$|K(x) - 1| \leq C_K |x|^{\rho_K}, \quad |x| \leq 1,$$

for some $\rho_K \geq 1$ and $0 < C_K < \infty$. (ii) The matrix $\mathbb{K}_n = \left[K\left(\frac{\delta_{ijn}}{\delta_n}\right)\right]_{i,j=1}^{n}$ is symmetric and positive semidefinite for all $n$.

## A.2 Proof of Theorems

**Lemma 1.** *Suppose that Assumption 1 holds. The number of matches $K_M(i)$ are uniformly bounded in $n$ such that for all $\alpha > 0$,*

$$n^{-\alpha} \max_{1\leq i\leq n} K_M(i) = o_p(1). \tag{1}$$

*Proof.* The statement is located in Abadie and Imbens (2002), which relies on Lemma 3 from Abadie and Imbens (2006). For any $r > 0$, $\mathrm{E}[K_M^r(i)]$ is uniformly bounded in $n$. Let $\varepsilon > 0$.

Consider the probability

$$\Pr\Big(n^{-\alpha} \max_{1\le i\le n} K_M(i) > \varepsilon\Big) = \Pr\Big(\max_{1\le i\le n} K_M(i) > \varepsilon n^{\alpha}\Big).$$

By the union bound, we have $\Pr\big(\max_{1\le i\le n} K_M(i) > \varepsilon n^{\alpha}\big) \le \sum_{i=1}^{n} \Pr\big(K_M(i) > \varepsilon n^{\alpha}\big) = n \Pr\big(K_M(i) > \varepsilon n^{\alpha}\big)$. Now apply Markov's inequality for any integer $r \ge 1$:

$$\Pr\Big(K_M(i) > \varepsilon n^{\alpha}\Big) = \Pr\Big(K_M^r(i) > (\varepsilon n^{\alpha})^r\Big) \le \frac{\mathrm{E}[K_M(i)^r]}{(\varepsilon n^{\alpha})^r}.$$

Hence,

$$\Pr\Big(n^{-\alpha} \max_{1\le i\le n} K_M(i) > \varepsilon\Big) \le \frac{n\,\mathrm{E}[K_M(i)^r]}{(\varepsilon n^{\alpha})^r} = \frac{n\,C_r}{\varepsilon^r n^{\alpha r}},$$

where $C_r \equiv \mathrm{E}[K_M^r(i)] < \infty$ by Lemma 3 in Abadie and Imbens (2006). For any $\alpha > 0$, by choosing $r$ sufficiently large, the right-hand side can be made arbitrarily small as $n \to \infty$. Therefore, for all $\alpha > 0$, $\max_{1\le i\le n} K_M(i) = o_p(n^{\alpha})$. □

For notation simplicity, we slightly abuse the above result by assuming $\max_{1\le i\le n} K_M(i)$ is uniformly bounded as $n$ increases.

**Proof of Proposition 1**. As in Abadie and Imbens (2011), we decompose the bias-corrected matching estimator $\tilde{\tau}^m_{ATT,bc}$ as

$$\tilde{\tau}^m_{ATT,bc} - \tau_{ATT} = \big(\overline{\tau_{ATT}(X)} - \tau_{ATT}\big) + E^m_{ATT,bc} + B^m_{ATT,bc} - \hat{B}^m_{ATT,bc},$$

where

$$\begin{aligned}
\overline{\tau_{ATT}(X)} &= \frac{1}{n_1}\sum_{i=1}^{n} D_{in}(\mu(X_{in},1) - \mu(X_{in},0)),\\
E^m_{ATT,bc} &= \frac{1}{n_1}\sum_{i=1}^{n}\left(D_{in} - (1-D_{in})\frac{K_M(i)}{M}\right)V_{in},
\end{aligned}$$

$$B^m_{ATT,bc} = \frac{1}{n_1}\sum_{i=1}^n \frac{D_{in}}{M}\sum_{j\in\mathcal{J}_M(i)}\big(\mu(X_{in},0)-\mu(X_{jn},0)\big),$$
$$\hat{B}^m_{ATT,bc} = \frac{1}{n_1}\sum_{i=1}^n \frac{D_{in}}{M}\sum_{j\in\mathcal{J}_M(i)}\big(\hat{\mu}(X_{in},0)-\hat{\mu}(X_{jn},0)\big).$$

By Assumption 4, we can show that $\sqrt{n_1}(B^m_{ATT,bc} - \hat{B}^m_{ATT,bc}) \xrightarrow{p} 0$. Also, it follows from a standard central limit theorem that $\sqrt{n_1}(\overline{\tau_{ATT}(X)} - \tau_{ATT}) \xrightarrow{d} N(0, V^{\tau(X),t})$. It remains to consider the contribution of $\sqrt{n_1}E^m_{ATT,bc}/\sqrt{V^{E,t}}$. Note that

$$E^m_{ATT,bc} = \frac{1}{n_1}\sum_{i=1}^n\left(D_{in} - (1-D_{in})\frac{K_M(i)}{M}\right)\sum_{j=1}^n r_{ij}\epsilon_{jn} \equiv \frac{1}{n_1}\sum_{i=1}^n \nu_{in}\epsilon_{in},$$

where $\nu_{in} = \sum_{j=1}^n (D_{jn} - (1-D_{jn})K_M(j)/M)r_{ji}$, and $r_{ji}$ is the $(j,i)$th element of $R_n$. Hence, conditional on $D_n$ and $X_n$, the unit-level terms $E^m_i \equiv \nu_{in}\epsilon_{in}$ are independent with zero means and nonidentical distributions. The conditional variance of $E^m_i$ is $\nu^2_{in}\mathrm{Var}(\epsilon_{in}\,|\,X_{in}, D_{in}) = \nu^2_{in}$ under Assumption 1. We will use a Lindeberg–Feller central limit theorem to show that $\sqrt{n_1}E^m_{ATT,bc}/\sqrt{V^{E,t}} \xrightarrow{d} N(0,1)$. For a given $X_n$ and $D_n$, the Lindeberg–Feller condition requires that

$$\frac{1}{n_1 V^{E,t}}\sum_{i=1}^n \mathrm{E}[(\nu_{in}\epsilon_{in})^2 \mathbb{1}\{|\nu_{in}\epsilon_{in}| \geq \eta\sqrt{n_1 V^{E,t}}\}\,|\,X_n, D_n] \to 0 \tag{2}$$

for all $\eta > 0$. Following along the lines of Abadie and Imbens (2006, Theorem 4), we have the left hand side of (2) is bounded by

$$\frac{1}{n_1 V^{E,t}}\sum_{i=1}^n \left(\nu^4_{in}\mathrm{E}[\epsilon^4_{in}\,|\,X_n, D_n]\right)^{1/2}\frac{\nu^2_{in}\mathrm{E}[\epsilon^2_{in}\,|\,X_n, D_n]}{\eta^2 n_1 V^{E,t}} \leq \frac{c}{\eta^2}\frac{1}{n_1}\left(\frac{1}{n_1}\sum_{i=1}^n \nu^4_{in}\right),$$

where $c$ is a finite constant. Now, we verify that $\mathrm{E}\nu^4_{in}$ is uniformly bounded, thus the factor in parentheses is bounded in probability. Denote

$$C_{in} \equiv D_{in} - (1-D_{in})\frac{K_M(i)}{M}.$$

By Hölder's inequality,

$$\mathrm{E}\nu_{in}^4 = \mathrm{E}\left[\sum_{j=1}^{n} C_{jn} r_{ji}\right]^4 \le \mathrm{E}\left[\sum_{j=1}^{n} |C_{jn}|\cdot|r_{ji}|^{1/4}|r_{ji}|^{3/4}\right]^4 \le \mathrm{E}\sum_{j=1}^{n} |C_{jn}|^4 |r_{ji}| \left(\sum_{j=1}^{n} |r_{ji}|\right)^3.$$

Using the fact that $\mathrm{E}|K_M(i)|^4$ is uniformly bounded and Assumption 5, we can show that $\mathrm{E}\nu_{in}^4 \le c < \infty$. Hence, the condition (2) is satisfied for almost all $X_n$ and $D_n$. Finally, $\sqrt{n_1}E_{ATT,bc}^m/\sqrt{V^{E,t}}$ and $\sqrt{n_1}(\overline{\tau_{ATT}(X)} - \tau_{ATT})$ are asymptotically independent, since the central limit theorem for the former holds conditional on $X_n$ and $D_n$. This completes the proof of Proposition 1. □

To prove Theorem 1, we first establish the consistency of bootstrap variance $\hat{V}_{\mathrm{DWB},l} \equiv \mathrm{Var}(\sqrt{n_1}T_{\mathrm{DWB}}^* \,|\, X_n, Y_n, D_n)$. Then the variance can be decomposed as $\hat{V}_{\mathrm{DWB},l} = \hat{V}_{\mathrm{DWB},l}^{\tau(X),t} + \hat{V}_{\mathrm{DWB},l}^{E,t}$, where

$$\begin{aligned}\hat{V}_{\mathrm{DWB},l}^{\tau(X),t} &= \frac{1}{n_1}\mathrm{Var}^*\left(\sum_{i=1}^{n} D_{in}\hat{\xi}_{in}^*\right) = \frac{1}{n_1}\sum_{i=1}^{n} D_{in}\big(\hat{\mu}(X_{in},1) - \hat{\mu}(X_{in},0) - \tilde{\tau}_{ATT,bc}^m\big)^2,\\ \hat{V}_{\mathrm{DWB},l}^{E,t} &= \frac{1}{n_1}\mathrm{Var}^*\left(\sum_{i=1}^{n} C_{in}\hat{V}_{in}^*\right) = \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}\,\mathrm{Cov}^*(\hat{V}_{in}^*, \hat{V}_{jn}^*).\end{aligned} \tag{3}$$

**Lemma 2.** *Suppose that Assumptions 1-7 hold. Then,* $|\hat{V}_{DWB,l}^{\tau(X),t} - V^{\tau(X),t}| = o_p(1)$ *and* $|\hat{V}_{DWB,l}^{E,t} - V^{E,t}| = o_p(1)$*, where* $V^{\tau(X),t}$ *and* $V^{E,t}$ *are defined in Proposition 1.*

*Proof.* Using Assumption 4 and the consistency of $\tilde{\tau}_{ATT,bc}^m$, we can easily show that

$$\left|\hat{V}_{\mathrm{DWB},l}^{\tau(X),t} - \frac{1}{n_1}\sum_{i=1}^{n} D_{in}(\mu(X_{in},1) - \mu(X_{in},0) - \tau_{ATT})^2\right| = o_p(1).$$

The desired convergence of $\hat{V}_{\mathrm{DWB},l}^{\tau(X),t}$ follows from the law of large numbers.

Next, note that $\hat{V}_{\mathrm{DWB},l}^{E,t} = n_1^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}\hat{V}_{in}\hat{V}_{jn}K(\delta_{ijn}/\delta_n)$ according to (2.5). To show the second part of the lemma, we decompose

$$
\begin{aligned}
\hat{V}^{E,t}_{\mathrm{DWB},l} - V^{E,t} &= \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}\hat{V}_{in}\hat{V}_{jn}K(\delta_{ijn}/\delta_n) - \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}\mathrm{E}(V_{in}V_{jn}) \\
&= A_n + B_n + C_n,
\end{aligned} \tag{4}
$$

where

$$
\begin{aligned}
A_n &= \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}(\hat{V}_{in}\hat{V}_{jn} - V_{in}V_{jn})K(\delta_{ijn}/\delta_n), \\
B_n &= \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}\big[V_{in}V_{jn} - \mathrm{E}(V_{in}V_{jn})\big]K(\delta_{ijn}/\delta_n), \\
C_n &= \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}\mathrm{E}(V_{in}V_{jn})[K(\delta_{ijn}/\delta_n) - 1].
\end{aligned}
$$

To show $|\hat{V}^{E,t}_{\mathrm{DWB},l} - V^{E,t}| = o_p(1)$, we prove that each term on the r.h.s of (4) is $o_p(1)$.

First, using the boundedness of $\mathrm{E}|K_M(i)|^q$ for all $q > 0$ uniformly over $n$ (Lemma 3 of Abadie and Imbens (2006)), we can show that $B_n = o_p(1)$ and $C_n = o_p(1)$ under Assumptions 1 and 5-7 using arguments similar to Kelejian and Prucha (2007, Theorem 2). To show $A_n = o_p(1)$, let $\hat{V}_{in} = V_{in} + \Delta_{in}$, where $\Delta_{in}$ denotes the estimation error and satisfies $\max_{1\le i\le n}|\Delta_{in}| = o_p(n^{-1/4})$ according to Assumption 4. Then, $\hat{V}_{in}\hat{V}_{jn} - V_{in}V_{jn} = \Delta_{in}\Delta_{jn} + V_{in}\Delta_{jn} + V_{jn}\Delta_{in}$. Therefore, we can decompose $A_n = A_n^{(1)} + A_n^{(2)} + A_n^{(3)}$ with

$$
\begin{aligned}
A_n^{(1)} &= \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}\Delta_{in}\Delta_{jn}K(\delta_{ijn}/\delta_n), \\
A_n^{(2)} &= \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}\Delta_{in}V_{jn}K(\delta_{ijn}/\delta_n), \\
A_n^{(3)} &= \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}C_{jn}V_{in}\Delta_{jn}K(\delta_{ijn}/\delta_n).
\end{aligned}
$$

For $A_n^{(1)}$, using Assumptions 1(iv) (with $s = 1$) and 4, Lemma 1, and the property that

$|K(x)| \leq C < \infty$ for $x \leq 1$ and $|K(x)| = 0$ for $x > 1$, we have

$$\begin{aligned}
\mathrm{E}\left|A_n^{(1)}\right| &\leq \frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}|C_{in}C_{jn}\Delta_{in}\Delta_{jn}||K(\delta_{ijn}/\delta_n)| \\
&\leq o(n_1^{-1}n^{-1/2})\sum_{i=1}^{n}\sum_{j=1}^{n}|K(\delta_{ijn}/\delta_n)| = o(n^{-1/2})l_n.
\end{aligned}$$

By Assumption 6(i), $l_n = o(n^{\tau})$ for some $\tau \leq 1/2$, which implies that $A_n^{(1)} = o_p(1)$.

For $A_n^{(2)}$ (and similarly for $A_n^{(3)}$), we have

$$\begin{aligned}
|A_n^{(2)}| &= \left|\frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n}C_{in}C_{jn}\Delta_{in}V_{jn}K(\delta_{ijn}/\delta_n)\right| \\
&\leq \frac{1}{n_1}\sup_{1\leq i\leq n}|\Delta_{in}|\sum_{i=1}^{n}|C_{in}|\left|\sum_{j=1}^{n}C_{jn}V_{jn}K(\delta_{ijn}/\delta_n)\right| \\
&\leq o_p\left(n^{\frac{1}{4}}\right)\left(\frac{1}{n}\sum_{i=1}^{n}|C_{in}|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}C_{jn}V_{jn}\,K\left(\frac{\delta_{ijn}}{\delta_n}\right)\right|^2\right)^{1/2} \\
&= o_p\left(n^{\frac{1}{4}}\right)O_p(\sqrt{l_n/n}),
\end{aligned}$$

because by Markov's inequality

$$\begin{aligned}
&\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}C_{jn}V_{jn}\,K\left(\frac{\delta_{ijn}}{\delta_n}\right)\right|^2 \geq \Delta\right) \\
&\leq \frac{1}{\Delta}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\left|K\left(\frac{\delta_{ij_1n}}{\delta_n}\right)K\left(\frac{\delta_{ij_2n}}{\delta_n}\right)\right||\mathrm{E}(C_{j_1n}C_{j_2n}V_{j_1n}V_{j_2n})| \\
&\leq \frac{1}{\Delta}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j_1=1}^{n}\left|K\left(\frac{\delta_{ij_1n}}{\delta_n}\right)\right|\sup_{k}\sum_{j_2=1}^{n}\left|\mathrm{E}\big(C_{kn}C_{j_2n}V_{kn}V_{j_2n}\big)\right| = O\left(\frac{l_n}{n}\right).
\end{aligned}$$

The last equality follows from the linear representation that $V_{in} = \sum_{l=1}^n r_{il}\epsilon_{ln}$,

$$\sum_{j=1}^n |\mathrm{E}(C_{kn}C_{jn}V_{kn}V_{jn})| = \sum_{j=1}^n |\mathrm{E}\{\mathrm{E}[V_{kn}V_{jn}|X_n, D_n]C_{kn}C_{jn}\}| = \sum_{j=1}^n \left|\sum_{l=1}^n r_{kl}r_{jl}\right| |\mathrm{E}(C_{kn}C_{jn})|$$
$$\leq \max_{1\leq i\leq n} \mathrm{E}|C_{in}|^2 \sum_{l=1}^n |r_{kl}| \sum_{j=1}^n |r_{jl}| \leq \text{Const},$$

where we use Assumptions 1 and 5 and the fact that $\mathrm{E}|C_{in}|^q$ is uniformly bounded over $n$ for all $q > 0$, and $\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\delta_{ijn}}{\delta_n}\right) \leq nl_n$. Therefore, $A_n^{(2)} = o_p(1)$ under Assumption 6(i), which concludes the proof. $\square$

**Proof of Theorem 1.** Decompose the DWB test statistic (2.6) as

$$\sqrt{n_1}T^*_{DWB} = \sqrt{n_1}T^*_n + \sqrt{n_1}R^*_{1n} + \sqrt{n_1}R^*_{2n}, \tag{5}$$

where

$$\sqrt{n_1}T^*_n = \frac{1}{\sqrt{n_1}} \sum_{i=1}^n \left[\mathcal{W}^*_{i,2n} D_{in}\big(\mu(X_{in}, 1) - \mu(X_{in}, 0) - \tau_{ATT}\big) + \mathcal{W}^*_{n+i,2n}\mathcal{V}_{in}\right],$$
$$\sqrt{n_1}R^*_{1n} = \frac{1}{\sqrt{n_1}} \sum_{i=1}^n D_{in}(\hat{\xi}^*_{in} - \xi^*_{in}),$$
$$\sqrt{n_1}R^*_{2n} = \frac{1}{\sqrt{n_1}} \sum_{i=1}^n \mathcal{W}^*_{n+i,2n}\left(D_{in} - (1 - D_{in})\frac{K_M(i)}{M}\right)(\hat{V}_{in} - V_{in}),$$

with $\mathcal{V}_{in} = \big(D_{in} - (1-D_{in})\frac{K_M(i)}{M}\big)V_{in} = C_{in}V_{in}$ and $\xi^*_{in} = \mathcal{W}^*_{i,2n}(\mu(X_{in}, 1) - \mu(X_{in}, 0) - \tau_{ATT})$.

Following Otsu and Rai (2017), we can show that $\Pr\{\sqrt{n_1}|R^*_{1n}| > \epsilon \,|\, Y_n, D_n, X_n\} \xrightarrow{p} 0$ and $\Pr\{\sqrt{n_1}|R^*_{2n}| > \epsilon \,|\, Y_n, D_n, X_n\} \xrightarrow{p} 0$ for any $\epsilon > 0$. It remains to establish that

$$\sup_r \left| \Pr\left\{\sqrt{n_1}T^*_n \leq r \mid Y_n, D_n, X_n\right\} - \Pr\left\{\sqrt{n_1}(\tilde{\tau}^m_{ATT,bc} - \tau_{ATT}) \leq r\right\}\right| \xrightarrow{p} 0. \tag{6}$$

By Pólya's theorem and Proposition 1, it is enough to verify that

$$\Pr\left\{(V^{E,t}+V^{\tau(X),t})^{-1/2}\sqrt{n_1}T_n^* \le r \mid Y_n, D_n, X_n\right\} - \Phi(r) \xrightarrow{p} 0 \quad \text{for all } r\in\mathbb{R},$$

where $\Phi(r)$ is the standard normal distribution function. Recall that the variance matrix of $\eta_n = (\mathcal{W}^*_{n+1,2n},\ldots,\mathcal{W}^*_{2n,2n})^\top$ is given by $\mathbb{K}_n = \left[K\left(\frac{\delta_{ijn}}{\delta_n}\right)\right]_{i,j=1}^n$, which admits a spectral decomposition $\mathbb{K}_n = \Phi_n\Lambda_n\Phi_n^\top$ under Assumption 6(ii). Here, $\Lambda_n = \mathrm{diag}(\lambda_{1,n},\ldots,\lambda_{n,n})$ is a diagonal matrix with the eigenvalues of $\mathbb{K}_n$, and the columns of $\Phi_n = [\phi_{1,n},\ldots,\phi_{n,n}]$ are the associated orthonormal eigenvectors. Hence,

$$\eta_n = \Phi_n\Lambda_n^{1/2}\upsilon_n = [\lambda_{1,n}^{1/2}\phi_{1,n},\ldots,\lambda_{n,n}^{1/2}\phi_{n,n}]\upsilon_n$$

with $\upsilon_n = (\upsilon_{1n},\ldots,\upsilon_{nn})^\top$ and $\upsilon_{in}\sim$ i.i.d.$(0,1)$. Denote $\mathcal{V}_n = (\mathcal{V}_{1n},\ldots,\mathcal{V}_{nn})^\top$. It follows that

$$\frac{1}{\sqrt{n_1}}\sum_{i=1}^n \mathcal{W}^*_{n+i,2n}\mathcal{V}_{in} = \frac{1}{\sqrt{n_1}}\mathcal{V}_n^\top\eta_n = \frac{1}{\sqrt{n_1}}\mathcal{V}_n^\top\Phi_n\Lambda_n^{1/2}\upsilon_n = \frac{1}{\sqrt{n_1}}\sum_{k=1}^n \lambda_{kn}^{1/2}\mathcal{V}_n^\top\phi_{kn}\upsilon_{kn}. \tag{7}$$

Thus,

$$\sqrt{n_1}T_n^* = \frac{1}{\sqrt{n_1}}\sum_{i=1}^{2n}\Upsilon^*_{in},$$

where

$$\Upsilon^*_{in} = \begin{cases} \mathcal{W}^*_{i,2n}D_{in}\big(\mu(X_{in},1)-\mu(X_{in},0)-\tau_{ATT}\big), & \text{for } i=1,\ldots,n,\\[2ex] \lambda_{i-n,n}^{1/2}\,\mathcal{V}_n^\top\phi_{i-n,n}\,\upsilon_{i-n,n}, & \text{for } i=n+1,\ldots,2n. \end{cases} \tag{8}$$

Note that $\{\mathcal{W}^*_{i,2n}\}_{i=1}^n$ is an i.i.d. sequence, $\lambda_{i-n,n}^{1/2}\,\mathcal{V}_n^\top\phi_{i-n,n}$ is a constant conditional on the data, and $\upsilon_{in}\sim$ i.i.d.$(0,1)$ independent of $\{\mathcal{W}^*_{i,2n}\}_{i=1}^n$. Therefore, $\Upsilon^*_{in}$ is an independent heterogeneous array. We will show that $(V^{E,t}+V^{\tau(X),t})^{-1/2}\sqrt{n_1}T_n^* \xrightarrow{d^*} N(0,1)$ in probability by applying Lyapunov's CLT, see, e.g. Proposition 2.27 of Van der Vaart (1998). First, note that conditionally on the data $\mathrm{E}^*(\Upsilon^*_{in}) = 0$ and

$$\mathrm{Var}^*\left(\sum_{i=1}^{2n} \Upsilon^*_{in}\right) = \sum_{i=1}^{n} D_{in}\big(\mu(X_{in},1) - \mu(X_{in},0) - \tau_{ATT}\big)^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}V_{in}C_{jn}V_{jn}K\left(\frac{\delta_{ijn}}{\delta_n}\right).$$

From the proof of Lemma 2, we have

$$\left|\frac{1}{n_1}\sum_{i=1}^{n} D_{in}\big(\mu(X_{in},1) - \mu(X_{in},0) - \tau_{ATT}\big)^2 - V^{\tau(X),t}\right| \xrightarrow{p} 0,$$
$$\left|\frac{1}{n_1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{in}V_{in}C_{jn}V_{jn}K\left(\frac{\delta_{ijn}}{\delta_n}\right) - V^{E,t}\right| \xrightarrow{p} 0,$$

which implies that $\mathrm{Var}^*(n_1^{-1/2}\sum_{i=1}^{2n}\Upsilon^*_{in}) \xrightarrow{p} V^{\tau(X),t} + V^{E,t}$. It remains to check Lyapunov's condition, which requires that for some $v > 0$,

$$\frac{1}{n_1^{1+v/2}}\sum_{i=1}^{2n} \mathrm{E}^*|\Upsilon^*_{in}|^{2+v} \xrightarrow{p} 0.$$

We will show that the above condition holds for $v = 2$ by showing that

$$\frac{1}{n_1^2}\sum_{i=1}^{n} \mathrm{E}^*|\Upsilon^*_{in}|^4 = \frac{1}{n_1^2}\sum_{i=1}^{n} D_{in}|\mu(X_{in},1) - \mu(X_{in},0) - \tau_{ATT}|^4\,\mathrm{E}^*|\mathcal{W}^*_{i,2n}|^4 \xrightarrow{p} 0, \tag{9}$$
$$\frac{1}{n_1^2}\sum_{i=n+1}^{2n} \mathrm{E}^*|\Upsilon^*_{in}|^4 = \frac{1}{n_1^2}\sum_{k=1}^{n} \mathrm{E}^*\big|\lambda_{kn}^{1/2}\mathcal{V}_n^\top\phi_{kn}v_{kn}\big|^4 \xrightarrow{p} 0. \tag{10}$$

Since $\mathcal{W}^*_{i,2n}$ follows the two-point distribution of Mammen (1993), we have $\mathrm{E}|\mathcal{W}^*_{i,2n}|^4 < c$ for some finite constant $c$. Then (9) follows immediately from the uniform boundedness of $\mu(X_{in},d)$ for $d \in \{0,1\}$ and $i = 1,\ldots,n$, which is guaranteed by Assumption 3 and the compactness of $\mathcal{X}$. To prove (10), observe that

$$\frac{1}{n_1^2}\sum_{k=1}^{n} \mathrm{E}^*|\lambda_k^{1/2}\mathcal{V}_n^\top\phi_{kn}v_{kn}|^4 \le \frac{1}{n_1}\sup_a \lambda_a^2 \times \frac{1}{n_1}\sum_{k=1}^{n}\left(\sum_{i=1}^{n} C_{in}V_{in}\phi_{kn}^{(i)}\right)^4 \mathrm{E}^*|v_{kn}|^4,$$

where $\phi_{kn}^{(i)}$ denotes the $i$th element of $\phi_{kn}$. First, it is straightforward to show that

$n_1^{-1} \sup_a \lambda_a^2 \leq n_1^{-1} \sup_i (\sum_{j=1}^n |K(\delta_{ijn}/\delta_n)|)^2 = O(l_n^2/n_1) = o(1)$ under Assumptions 1(iv) (with $s = 1$) and 6. Next, using the fact that $\mathrm{E}|C_{in}|^q$ is uniformly bounded for any $q > 0$ and similar arguments to those used in Theorem 3.1 in Conley et al. (2023), we can show that

$$\frac{1}{n_1} \sum_{k=1}^n \mathrm{E} \left( \sum_{i=1}^n C_{in} V_{in} \phi_{kn}^{(i)} \right)^4 = \mathrm{E} \left[ \frac{1}{n_1} \sum_{k=1}^n \mathrm{E} \left( \left( \sum_{i=1}^n C_{in} V_{in} \phi_{kn}^{(i)} \right)^4 \middle| X_n, D_n \right) \right] = O(1).$$

Then (10) follows from Markov's inequality and the assumption that $\mathrm{E}^*|v_{kn}|^4 < \infty$. This completes the proof of Theorem 1. □

## A.3 Additional discussion

### A.3.1 Bandwidth selection

In this subsection, we propose a selection procedure for the bandwidth parameter $\delta_n$ in a data-driven way. The proposed algorithm closely follows that of Conley et al. (2023).

**Algorithm (Bandwidth Selection).**

(i) Compute $\hat{V}_{in} = Y_{in} - \hat{\mu}(X_{in}, D_{in})$, $i = 1, \ldots, n$, where $\hat{\mu}(x, d)$ is a nonparametric estimator (e.g., a power series estimator) of $\mu(x, d)$.

(ii) Choose a sequence of potential bandwidths $\delta_0 = (\delta_0^{(1)}, \ldots, \delta_0^{(M)})^\top$ that are ordered in increasing magnitude. For each $k = 1, \ldots, M$, calculate the covariance $\hat{C}(\delta_0^{(k)})$ between residuals $\hat{V}_{in}$ at distance $\delta_0^{(k)}$ nonparametrically through a uniform kernel regression with tolerance of $\epsilon$:

$$\hat{C}(\delta_0^{(k)}) = \frac{\sum_{i,j} \mathbb{1}(|\delta_{ijn} - \delta_0^{(k)}| < \epsilon) \hat{V}_{in} \hat{V}_{jn}}{\sum_{i,j} \mathbb{1}(|\delta_{ijn} - \delta_0^{(k)}| < \epsilon)}.$$

(iii) Use an i.i.d. bootstrap from the empirical distribution of the residuals $\hat{V}_{in}$ to generate

$\{\hat{V}^*_{in}\}_{i=1}^n$. Compute the bootstrap analog of $\hat{C}(\delta_0^{(k)})$ as:

$$\hat{C}^*(\delta_0^{(k)}) = \frac{\sum_{i,j} \mathbb{1}(|\delta_{ijn} - \delta_0^{(k)}| < \epsilon)\hat{V}^*_{in}\hat{V}^*_{jn}}{\sum_{i,j} \mathbb{1}(|\delta_{ijn} - \delta_0^{(k)}| < \epsilon)}.$$

(iv) Repeat step (iii) $\tilde{B}$ times and obtain a bootstrap acceptance region for the null of spatial independence using the 2.5% and 97.5% quantiles of $\hat{C}^*(\delta_0^{(k)})$.

(v) Identify the first candidate bandwidth $\delta_0(1)$ such that $\hat{C}(\delta_0(1))$ falls within the confidence band.

- If $\delta_0(1) = \delta_0^{(1)}$, select the wild bootstrap.
- If $\delta_0(1) = \delta_0^{(k)}$ for some $k \in \{2, \ldots, M\}$, employ the dependent wild bootstrap and set the bandwidth as $\delta_n^* = 3\,\delta_0^{(k-1)}$.

The main distinction between the above procedure and that proposed by Conley et al. (2023) is the adjustment of the bandwidth in step (v). This takes into account the choice of kernel: we employ the quadratic triangular kernel $K(x) = (1 - |x|)^2\,\mathbb{1}(|x| \leq 1)$, which assigns zero weight when $\delta_{ijn}$ reaches the selected bandwidth $\delta_n^*$. In contrast, the Gaussian kernel used by Conley et al. (2023) assigns weights that are nearly zero when $\delta_{ijn}$ reaches three times the bandwidth. To accommodate this, we adjust the selected bandwidth as $\delta_n^* = 3\,\delta_0^{(k-1)}$.

To implement the algorithm, we take the potential bandwidths $\delta_0^{(k)} = c_k \times n^{1/6}$, where $c_k$ ranges from 0.5 to 4 in increments of 0.5, with a tolerance $\epsilon = 0.1 \times n^{1/6}$, and use bootstrap samples $\tilde{B} = 200$. See Section 3 for the simulation results using the above bandwidth selection procedure and related discussion.

### A.3.2 Additional simulation results

In this subsection, we report the simulation results when the bandwidths are selected as $\delta_n = c_k \times n^{1/6}$, where $c_k \in \{1, 2, 3\}$, see Tables A.1-A.3 below. Our findings are as follows.

First, the DWB procedure outperforms the naive wild bootstrap when $\rho \in \{0.25, 0.5\}$ for $n = 900$ and 2500, across all DGPs we consider. The coverage rates of the DWB generally improve as $n$ increases, but those of the standard wild bootstrap do not vary much. Second, the performance of DWB appears to be insensitive to the choice of bandwidths.

Table A.1: Simulation results with bandwidth $\delta_n = c_k \times n^{1/6}$ ($n = 20 \times 20$).

| | $m(\cdot)$ | Bias | Wild | | DWB ($c_k = 1$) | | DWB ($c_k = 2$) | | DWB ($c_k = 3$) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 95% Cov. | CI Length | 95% Cov. | CI Length | 95% Cov. | CI Length | 95% Cov. | CI Length |
| $\rho = 0$ | 1 | 0.0026 | 0.937 | 0.1399 | 0.929 | 0.1373 | 0.911 | 0.1318 | 0.900 | 0.1263 |
| | 2 | -0.0016 | 0.957 | 0.1664 | 0.953 | 0.1632 | 0.936 | 0.1567 | 0.926 | 0.1508 |
| | 3 | 0.0088 | 0.961 | 0.1611 | 0.954 | 0.1582 | 0.939 | 0.1523 | 0.929 | 0.1466 |
| | 4 | 0.0001 | 0.940 | 0.1421 | 0.940 | 0.1398 | 0.923 | 0.1340 | 0.909 | 0.1282 |
| | 5 | 0.0030 | 0.931 | 0.1433 | 0.924 | 0.1403 | 0.898 | 0.1343 | 0.884 | 0.1287 |
| | 6 | -0.0156 | 0.967 | 0.1838 | 0.959 | 0.1808 | 0.940 | 0.1737 | 0.930 | 0.1674 |
| $\rho = 0.25$ | 1 | 0.0027 | 0.885 | 0.1401 | 0.893 | 0.1469 | 0.891 | 0.1467 | 0.875 | 0.1422 |
| | 2 | -0.0038 | 0.912 | 0.1654 | 0.913 | 0.1702 | 0.911 | 0.1685 | 0.897 | 0.1636 |
| | 3 | 0.0068 | 0.905 | 0.1615 | 0.911 | 0.1667 | 0.908 | 0.1654 | 0.896 | 0.1611 |
| | 4 | 0.0007 | 0.895 | 0.1402 | 0.898 | 0.1466 | 0.889 | 0.1458 | 0.869 | 0.1418 |
| | 5 | 0.0031 | 0.890 | 0.1422 | 0.901 | 0.1486 | 0.892 | 0.1483 | 0.877 | 0.1444 |
| | 6 | -0.0141 | 0.923 | 0.1831 | 0.929 | 0.1875 | 0.922 | 0.1857 | 0.916 | 0.1812 |
| $\rho = 0.50$ | 1 | 0.0028 | 0.792 | 0.1399 | 0.842 | 0.1590 | 0.852 | 0.1687 | 0.854 | 0.1676 |
| | 2 | -0.0018 | 0.819 | 0.1649 | 0.863 | 0.1791 | 0.865 | 0.1854 | 0.859 | 0.1826 |
| | 3 | 0.0133 | 0.827 | 0.1620 | 0.859 | 0.1779 | 0.870 | 0.1858 | 0.856 | 0.1836 |
| | 4 | 0.0040 | 0.789 | 0.1407 | 0.821 | 0.1598 | 0.837 | 0.1698 | 0.830 | 0.1687 |
| | 5 | 0.0003 | 0.781 | 0.1426 | 0.841 | 0.1601 | 0.854 | 0.1686 | 0.844 | 0.1669 |
| | 6 | -0.0172 | 0.869 | 0.1835 | 0.888 | 0.1961 | 0.892 | 0.2020 | 0.886 | 0.2002 |

Table A.2: Simulation results with bandwidth $\delta_n = c_k \times n^{1/6}$ ($n = 30 \times 30$).

| | $m(\cdot)$ | Bias | Wild | | DWB ($c_k = 1$) | | DWB ($c_k = 2$) | | DWB ($c_k = 3$) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 95% Cov. | CI Length | 95% Cov. | CI Length | 95% Cov. | CI Length | 95% Cov. | CI Length |
| $\rho = 0$ | 1 | 0.0028 | 0.948 | 0.1138 | 0.947 | 0.1117 | 0.933 | 0.1080 | 0.920 | 0.1041 |
| | 2 | -0.0007 | 0.970 | 0.1351 | 0.968 | 0.1326 | 0.960 | 0.1284 | 0.947 | 0.1243 |
| | 3 | 0.0053 | 0.969 | 0.1321 | 0.957 | 0.1297 | 0.953 | 0.1256 | 0.941 | 0.1217 |
| | 4 | 0.0015 | 0.939 | 0.1152 | 0.930 | 0.1131 | 0.919 | 0.1092 | 0.907 | 0.1053 |
| | 5 | 0.0010 | 0.950 | 0.1167 | 0.943 | 0.1145 | 0.926 | 0.1107 | 0.919 | 0.1072 |
| | 6 | -0.0108 | 0.968 | 0.1498 | 0.967 | 0.1481 | 0.955 | 0.1435 | 0.946 | 0.1394 |
| $\rho = 0.25$ | 1 | 0.0011 | 0.891 | 0.1136 | 0.908 | 0.1210 | 0.910 | 0.1221 | 0.901 | 0.1200 |
| | 2 | -0.0009 | 0.924 | 0.1352 | 0.929 | 0.1408 | 0.927 | 0.1402 | 0.909 | 0.1371 |
| | 3 | 0.0059 | 0.931 | 0.1315 | 0.938 | 0.1381 | 0.930 | 0.1384 | 0.925 | 0.1356 |
| | 4 | 0.0021 | 0.882 | 0.1148 | 0.901 | 0.1224 | 0.894 | 0.1231 | 0.889 | 0.1209 |
| | 5 | 0.0040 | 0.888 | 0.1161 | 0.908 | 0.1235 | 0.901 | 0.1243 | 0.885 | 0.1219 |
| | 6 | -0.0124 | 0.934 | 0.1488 | 0.939 | 0.1540 | 0.934 | 0.1541 | 0.933 | 0.1512 |
| $\rho = 0.50$ | 1 | 0.0013 | 0.795 | 0.1138 | 0.856 | 0.1328 | 0.865 | 0.1408 | 0.850 | 0.1400 |
| | 2 | -0.0011 | 0.849 | 0.1354 | 0.897 | 0.1519 | 0.901 | 0.1585 | 0.901 | 0.1569 |
| | 3 | 0.0033 | 0.853 | 0.1321 | 0.878 | 0.1483 | 0.895 | 0.1548 | 0.891 | 0.1535 |
| | 4 | 0.0038 | 0.810 | 0.1150 | 0.875 | 0.1344 | 0.899 | 0.1431 | 0.890 | 0.1434 |
| | 5 | 0.0004 | 0.797 | 0.1166 | 0.855 | 0.1354 | 0.868 | 0.1435 | 0.861 | 0.1433 |
| | 6 | -0.0114 | 0.873 | 0.1494 | 0.905 | 0.1637 | 0.909 | 0.1702 | 0.903 | 0.1693 |

Table A.3: Simulation results with bandwidth $\delta_n = c_k \times n^{1/6}$ ($n = 50 \times 50$).

| | $m(\cdot)$ | Bias | Wild | | DWB ($c_k = 1$) | | DWB ($c_k = 2$) | | DWB ($c_k = 3$) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 95% Cov. | CI Length | 95% Cov. | CI Length | 95% Cov. | CI Length | 95% Cov. | CI Length |
| $\rho = 0$ | 1 | 0.0014 | 0.944 | 0.0877 | 0.946 | 0.0860 | 0.929 | 0.0836 | 0.919 | 0.0813 |
| | 2 | 0.0003 | 0.971 | 0.1037 | 0.967 | 0.1025 | 0.962 | 0.1006 | 0.954 | 0.0985 |
| | 3 | 0.0027 | 0.964 | 0.1018 | 0.963 | 0.1002 | 0.957 | 0.0978 | 0.946 | 0.0954 |
| | 4 | -0.0003 | 0.945 | 0.0885 | 0.937 | 0.0872 | 0.932 | 0.0850 | 0.918 | 0.0829 |
| | 5 | 0.0000 | 0.951 | 0.0897 | 0.949 | 0.0886 | 0.936 | 0.0864 | 0.931 | 0.0847 |
| | 6 | -0.0086 | 0.975 | 0.1149 | 0.976 | 0.1137 | 0.966 | 0.1114 | 0.962 | 0.1092 |
| $\rho = 0.25$ | 1 | 0.0005 | 0.887 | 0.0877 | 0.906 | 0.0950 | 0.907 | 0.0964 | 0.898 | 0.0953 |
| | 2 | -0.0014 | 0.944 | 0.1045 | 0.956 | 0.1103 | 0.953 | 0.1112 | 0.941 | 0.1097 |
| | 3 | 0.0040 | 0.921 | 0.1017 | 0.929 | 0.1077 | 0.933 | 0.1083 | 0.927 | 0.1074 |
| | 4 | -0.0001 | 0.917 | 0.0885 | 0.934 | 0.0959 | 0.928 | 0.0969 | 0.926 | 0.0958 |
| | 5 | 0.0002 | 0.888 | 0.0897 | 0.914 | 0.0967 | 0.917 | 0.0978 | 0.904 | 0.0968 |
| | 6 | -0.0091 | 0.945 | 0.1147 | 0.948 | 0.1203 | 0.946 | 0.1203 | 0.941 | 0.1186 |
| $\rho = 0.50$ | 1 | 0.0026 | 0.788 | 0.0881 | 0.882 | 0.1072 | 0.892 | 0.1138 | 0.884 | 0.1140 |
| | 2 | 0.0000 | 0.862 | 0.1041 | 0.900 | 0.1206 | 0.913 | 0.1265 | 0.906 | 0.1263 |
| | 3 | 0.0022 | 0.856 | 0.1026 | 0.910 | 0.1190 | 0.919 | 0.1245 | 0.914 | 0.1243 |
| | 4 | 0.0013 | 0.807 | 0.0890 | 0.881 | 0.1079 | 0.901 | 0.1146 | 0.900 | 0.1152 |
| | 5 | 0.0010 | 0.815 | 0.0898 | 0.884 | 0.1079 | 0.898 | 0.1142 | 0.898 | 0.1144 |
| | 6 | -0.0092 | 0.890 | 0.1156 | 0.924 | 0.1305 | 0.928 | 0.1351 | 0.922 | 0.1354 |