Semiparametric Estimation and Testing of Smooth Coefficient Spatial Autoregressive Models

Malikov, Emir and Sun, Yiguo

Auburn University, University of Guelph

February 2017

Semiparametric Estimation and Testing of Smooth Coefficient Spatial Autoregressive Models*

Emir Malikov[1] Yiguo Sun[2]

[1]Department of Agricultural Economics & Rur. Soc., Auburn University, Auburn, AL
[2]Department of Economics and Finance, University of Guelph, ON

First Draft: December 15, 2014
This Draft: February 25, 2017

Abstract

This paper considers a flexible semiparametric spatial autoregressive (mixed-regressive) model in which unknown coefficients are permitted to be nonparametric functions of some contextual variables to allow for potential nonlinearities and parameter heterogeneity in the spatial relationship. Unlike other semiparametric spatial dependence models, ours permits the spatial autoregressive parameter to meaningfully vary across units and thus allows the identification of a neighborhood-specific spatial dependence measure conditional on the vector of contextual variables. We propose several (locally) nonparametric GMM estimators for our model. The developed two-stage estimators incorporate both the linear and quadratic orthogonality conditions and are capable of accommodating a variety of data generating processes, including the instance of a pure spatially autoregressive semiparametric model with no relevant regressors as well as multiple partially linear specifications. All proposed estimators are shown to be consistent and asymptotically normal. We also contribute to the literature by putting forward two test statistics to test for parameter constancy in our model. Both tests are consistent.

Keywords: Consistent Test, Constrained Estimation, Local Linear Fitting, Nonparametric GMM, Partially Linear, Quadratic Moments, SAR, Spatial Lag

JEL Classification: C13, C14, C21

* *Email*: emalikov@auburn.edu (Malikov), yisun@uoguelph.ca (Sun).

We would like to thank the editor, the associate editor and two anonymous referees for many insightful comments and suggestions that helped improve this article. We are also thankful to conference participants at the NY Camp Econometrics X, 2015 Midwest Econometrics Group Meeting at the St. Louis Fed, 85th Annual Meeting of the Southern Economic Association and the 2015 Tsinghua International Conference for Econometrics in Beijing for many helpful comments and suggestions. Yiguo Sun gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada Insightful Grant 435-2016-0340. Any remaining errors are our own.

1 Introduction

While spatial econometric methods have solidly become part of a standard methodological toolkit of applied researchers in many fields of economics that deal with spatial data which include applications such as land use, hedonic pricing or cross-country growth studies, most empirical work has confined its analysis to linear spatial models only. However, to paraphrase Paelinck & Klaassen (1979), econometric relations in space result more often than not in highly nonlinear specifications (as cited in van Gastel & Paelinck, 1995). For instance, allowing for nonlinearities in the hedonic house price function is often argued to be crucial in order to obtain realistic marginal valuations of housing attributes (e.g., see Parmeter, Henderson & Kumbhakar, 2007). Taking such potential nonlinearities in spatial models for granted is likely to lead to inconsistent parameter estimates and thus misleading conclusions.

This paper proposes a semiparametric method to handling nonlinearity (and parameter heterogeneity) in models of spatial dependence. We extend a particular class of semiparametric models in which parameters of a linear regression are permitted to be unspecified smooth functions of some contextual variables (Hastie & Tibshirani, 1993; Cai, Fan & Li, 2000; Li, Huang, Li & Fu, 2002) to the case of data with spatial dependence. Specifically, we generalize a popular parametric spatial autoregressive mixed-regressive model by allowing its coefficients, including the spatial autoregressive parameter, to be nonparametric functions of unknown form [for concreteness, see eq. (2.1)].

While our "smooth coefficient" spatial autoregressive model closely relates to the family of partially linear semiparametric spatial models recently proposed in the literature (Su & Jin, 2010; Su, 2012; Zhang, 2013; Sun, Hongjia, Zhang & Lu, 2014), its distinct feature is that it permits the spatial autoregressive parameter to meaningfully vary across units. The latter may be highly desirable from a practitioner's point of view since it allows the identification of a neighborhood-specific spatial dependence measure conditional on the vector of contextual variables. For instance, when a spatial autoregressive model is game-theoretically rationalized as a "response function", our model empowers a researcher to estimate heterogeneous "reaction" parameters that can vary with some environmental control factors. Some potential applications of our model, for instance, include the estimation of growth models that explicitly account for technological interdependence between countries in the presence of spillover effects. Such technological interdependence is usually formulated in the form of spatial externalities (e.g., see Ertur & Koch, 2007). However, the intensity of knowledge spillovers is naturally expected to greatly depend on institutional and cultural compatibility of neighboring countries (Kelejian, Murrell & Shepotylo, 2013). Our smooth coefficient spatial autoregressive model presents a practical, easy-to-implement way to allow for such indirect effects of institutions on the degree of spatial dependence in the cross-country conditional convergence regressions.

Our semiparametric treatment of nonlinearities is also relatively more flexible than pioneer nonlinear modeling approaches in spatial dependence models put forward by van Gastel & Paelinck (1995), Baltagi & Li (2001), Pace, Barry, Slawson & Sirmans (2004) and Yang, Li & Tse (2006). At the same time, like in all these studies as well as many others (e.g., Kelejian & Prucha, 1998, 1999, 2010; Lee, 2004, 2007; Su & Jin, 2010; Su, 2012), the consistency of our estimator rests on an admittedly rather restrictive assumption of a correctly pre-specified spatial weighting matrix. Dispensing with this assumption requires either making an alternative assumption of strong spatial mixing along with spatial stationarity or modeling spatial weights as nonparametric functions of the distance between neighbors (see Sun, 2016, and the references therein). The acute disadvantage of both of these alternative approaches is an inability to quantify a spatial autoregressive parameter (a "reaction" parameter) which many empirical studies are specifically interested in. In this paper, we therefore abstract from the issue concerning the correct specification of spatial weights.[1]

We propose several (locally) nonparametric Generalized Method of Moments (GMM) estimators for our model. The developed estimators incorporate both the linear and quadratic orthogonality conditions and are capable of accommodating a variety of data generating processes, including the instance of a pure spatially autoregressive semiparametric model with no relevant regressors as well as multiple partially linear specifications. To this end, we contribute to the literature on four fronts. First, our paper is the first (to the best of our knowledge) attempt in the nonparametric estimation literature to make use of local quadratic orthogonality conditions, which are necessary for the IV identification of spatially autoregressive models in the case when all explanatory covariates are irrelevant in predicting the outcome variable (see Lee, 2007). Second, we propose a two-stage estimation procedure whereby we first obtain an initial estimator of unknown parameter functions using feasible, but likely not so strong, instruments which we then use for the construction of more natural instruments suggested by the model's reduced form. Again, to our knowledge, no prior attempt has been made in the nonparametric econometrics literature to study such a class of estimators which themselves are based on the estimated instruments. Third, we also consider two special cases of our model by allowing some of its parameter functions to be constant thus resulting in a partially linear specification. Our proposed estimators present an alternative to those by Su (2012) and Zhang (2013) who study a similar class of partially linear spatial models. Unlike their estimators, ours however preserves its consistency property if the true model is a pure spatial autoregression. Fourth, we discuss ways of ensuring that the estimated model satisfies the non-singularity condition needed to rule out unstable Nash equilibria. In the instance of a mixed-regressive model, we impose this non-singularity restriction via the "tilting" procedure à la Hall & Huang (2001) whose theoretical results we generalize to the case of GMM estimators in the presence of endogenous regressors. Under fairly mild regularity conditions, we show that all our proposed estimators are consistent and asymptotically normal.

Further, we contribute to the literature by putting forward two test statistics to test for parameter constancy in our model. These model specification tests allow us to discriminate between a standard linear spatial autoregressive and our semiparametric models. The first consistent test utilizes a popular residual-based specification test technique which we extend to spatial data with cross-sectional dependence.[2] Given the well-known poor performance of nonparametric residual-based tests in finite samples, we also suggest a (wild) bootstrap procedure for it which we show to be asymptotically valid in approximating the null distribution of our test statistic regardless of whether the null hypothesis holds true or not. However, our residual-based test is impractical when the spatial model has no regressors. We therefore propose an alternative consistent test statistic à la Henderson, Carroll & Li (2008) which provides a vehicle for testing for parameter constancy in our model even when the model is a pure spatial autoregression.

We investigate the finite sample performance of the proposed estimators and test statistics in a small set of Monte Carlo experiments. The results are encouraging and show that all estimators and tests perform well in finite samples with considerable improvements as the sample size increases. Overall, simulation experiments lend support to our asymptotic results. To showcase our methodology, we then apply it to estimate a spatial hedonic price function using the well-known Harrison & Rubenfeld's house price data from Gilley & Pace (1996), where we let unknown parameter func-

[1]Furthermore, theoretical basis for the widely-believed sensitivity of the estimates to the choice of spatial weights remains rather unclear, as recently argued by LeSage & Pace (2014).

[2]Asymptotic properties of such a test have been studied for independent data (e.g., Zheng, 1996; Li & Wang, 1998; Stengos & Sun, 2001), weakly dependent time series data (e.g., Fan & Li, 1999; Li, 1999) and integrated time series data (e.g., Gao, King, Liu & Tjøstheim, 2009; Wang & Phillips, 2012; Sun, Cai & Li, 2015), to mention few among many contributions.

tions to vary with the $\mathrm{NO_x}$ concentration in the air. We find that spatial dependence between house prices is statistically significant only at higher values of the $\mathrm{NO_x}$ concentration in the air and that the degree of this spatial dependence, on average, increases as the air quality declines. This finding suggests that locational similarity may matter little for house valuations in pollution-free localities.

The rest of the paper proceeds as follows. Section 2 outlines the model. We present our estimators in Section 3, where we also provide their large-sample statistical properties. In Section 4, we consider two different types of partially linear spatial autoregressive models. Section 5 discusses specification tests for parameter constancy. The results of Monte Carlo simulations are described in Section 6. Section 7 concludes.

Throughout the paper, we use M to denote a generic finite constant that can take different values at different appearances and $\mathrm{vec}\{\mathbf{A}\}$ stacks columns of an $n \times m$ matrix $\mathbf{A}$ into an $(nm) \times 1$ vector. Lastly, $\mathrm{tr}\{\mathbf{A}\}$ refers to the trace of a square matrix $\mathbf{A}$.

2 Semiparametric Spatial Autoregressive Model

Consider a semiparametric generalization of the conventional (linear) spatial autoregressive mixed-regressive model, where the coefficients are now permitted to be unknown smooth functions of some relevant exogenous variables, i.e.,

$$y_i = \rho(\mathbf{z}_i) \sum_{j \neq i} w_{ij} y_j + \mathbf{x}_i' \boldsymbol{\beta}(\mathbf{z}_i) + u_i \quad \forall \quad i = 1, \ldots, n, \tag{2.1}$$

where y_i is the (scalar) outcome variable of interest; $\mathbf{x}_i$ and $\mathbf{z}_i$ are $p \times 1$ and $q \times 1$ vectors of exogenous covariates, respectively, and $\mathbf{x}_i$ can contain a constant 1; w_{ij} is the (i, j)-th element of a given $n \times n$ non-stochastic spatial weighting matrix $\mathbf{W}$ such that $w_{ii} = 0$ for all i. Further, $\boldsymbol{\beta}(\mathbf{z}_i)$ is a conformable vector of unknown slope parameter functions of $\mathbf{z}_i$, and $\rho(\mathbf{z}_i)$ is an unknown (scalar) spatial lag parameter function of $\mathbf{z}_i$. The random disturbance u_i is identically and independently distributed over i conditional on $(\mathbf{x}_i, \mathbf{z}_i)$ with zero mean and finite variance, i.e., $u_i|\mathbf{x}_i, \mathbf{z}_i \sim i.i.d.\ (0, \sigma_u^2)$.

We proceed by rewriting model (2.1) in the matrix form as

$$\mathbf{y} = \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}\mathbf{y} + \mathrm{mtx}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\} + \mathbf{u}, \tag{2.2}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$ and $\mathbf{u} = (u_1, \ldots, u_n)'$ are $n \times 1$ vectors; $\boldsymbol{\rho}(\mathbf{Z}) \equiv \mathrm{diag}\{\rho(\mathbf{z}_1), \ldots, \rho(\mathbf{z}_n)\}$ is an $n \times n$ diagonal matrix of spatial autoregressive parameter functions; and $\mathrm{mtx}\{\cdot\}$ is the operator that stacks up $\mathbf{x}_i'\boldsymbol{\beta}(\mathbf{z}_i)$ into an $n \times 1$ vector with the i subscript matching those of $\mathbf{y}$, $\mathbf{u}$ and $\boldsymbol{\rho}(\mathbf{Z})$. Also, $\mathbf{X} = \begin{bmatrix}\mathbf{x}_1 & \ldots & \mathbf{x}_n\end{bmatrix}'$ and $\mathbf{Z} = \begin{bmatrix}\mathbf{z}_1 & \ldots & \mathbf{z}_n\end{bmatrix}'$ are $n \times p$ and $n \times q$ data matrices, respectively.

To facilitate the economic interpretation of $\rho(\mathbf{z}_i)$ as the "reaction" parameter (and to rule out unstable Nash equilibria), we assume the following condition for $\rho(\cdot)$ to ensure the non-singularity of $\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}$:[3]

$$\max_{1 \leq i \leq n} |\lambda_i \{\boldsymbol{\rho}(\mathbf{Z})\mathbf{W}\}| < 1, \tag{2.3}$$

where $\lambda_i\{\mathbf{A}\}$ is the ith eigenvalue of an $n \times n$ matrix $\mathbf{A}$.

[3]See Kelejian & Prucha (2010) for an excellent discussion of the assumptions concerning the parameter space of a spatial autoregressive parameter.

Model (2.1) nests a standard smooth coefficient model with strictly exogenous covariates (Cai et al., 2000; Li et al., 2002) as a special case when $\rho(\mathbf{z}_i) = 0$ for all i, which implies no spatial interaction in the outcome variable y_i. In many applications, it is however imperative to allow for potential spatial dependence in the data. For instance, house prices are widely believed to be spatially autoregressive because residential property values tend to reflect shared local amenities as well as observed and unobserved neighborhood characteristics. While these characteristics can be partly controlled for using locality fixed effects, such an approach may be quite unsatisfactory since it does not let characteristics of neighboring houses affect the price of a given house (Anselin & Lozano-Gracia, 2009). However, by including the spatial lag in a house pricing function, we are able to accommodate such cross-neighbor effects as can be seen from the following expansion of the reduced form of equation (2.2) under the non-singularity condition in (2.3):

$$\mathbb{E}[\mathbf{y}|\mathbf{X},\mathbf{Z}] = [\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}]^{-1}\text{mtx}\left\{\mathbf{X},\boldsymbol{\beta}(\mathbf{Z})\right\} = \sum_{s=0}^{\infty}\left[\boldsymbol{\rho}(\mathbf{Z})\mathbf{W}\right]^{s}\text{mtx}\left\{\mathbf{X},\boldsymbol{\beta}(\mathbf{Z})\right\}. \tag{2.4}$$

From (2.4), it is evident that the conditional mean of y_i depends not only on its own $\mathbf{x}_i$ and $\mathbf{z}_i$ but also on its neighbors' $\mathbf{x}_j$ and $\mathbf{z}_j$ for $j \neq i$. Perhaps more importantly, house prices are likely to be spatially autoregressive because the very process of property valuation at its core relies on sale price information for comparable houses in the local neighborhood that real estate agents base their appraisals on. The latter is known as the "sales comparison approach" to a real estate appraisal which can systematically influence equilibrium prices in the housing market, especially if property owners have limited information about the market (see the reference in Small & Steimetz, 2012). Our model will be able to accommodate this spatial dependence in house prices while also allowing for nonlinearities and parameter heterogeneity in the house valuation function.

Studies of institutional change provide another example of applications where it is crucial to explicitly model spatial dependence in the data. Existing theoretical and empirical work indicate that institutional development in one country affects that of its neighbors (Mukand & Rodrik, 2005), where the institutional diffusion may be driven, say, by lobbying of multinational corporations for the harmonization of the commercial law across countries, requirements for members of a trade pact to standardize regulations or even military conflicts (see Kelejian et al., 2013). The semiparametric model we propose is equipped to examine such institutional diffusions across space while also allowing for nonlinearities and heterogeneity in institutional spillovers.

Lastly, in the instance when $\rho(\mathbf{z}_i)$ is a *non-zero* constant for all i, our model nests a partially linear smooth coefficient spatial autoregressive model as a special case (see Sun et al., 2014).

Remark 1 The differentiation between the two sets of covariates, $\mathbf{x}_i$ and $\mathbf{z}_i$, is a practitioner's prerogative which is likely to change from one empirical application to another. The assignment of relevant regressors into either of the two sets of variables may be done on the basis of practical convenience or, better yet, economic theory suggesting the direct ($\mathbf{X}$) and indirect, contextual ($\mathbf{Z}$) determinants of the spatially correlated outcome variable. From the theory's point of view however, our results do not require the variables in $\mathbf{x}_i$ to be different from or unrelated to those in $\mathbf{z}_i$.

3 Nonparametric GMM Estimator

We propose to estimate the unknown coefficient functions $[\rho(\mathbf{z}_i), \boldsymbol{\beta}(\mathbf{z}_i)']'$ in (2.1) by a local linear regression approach. First, we rewrite equation (2.1) in a compact form as follows:

$$y_i = \mathbf{m}_i'\boldsymbol{\gamma}(\mathbf{z}_i) + u_i \quad \forall \quad i = 1, \ldots, n, \tag{3.1}$$

where $\mathbf{m}_i \equiv \left[\left(\sum_{j\neq i} w_{ij}y_j\right), \mathbf{x}_i'\right]'$ and $\boldsymbol{\gamma}(\mathbf{z}_i) \equiv [\rho(\mathbf{z}_i), \boldsymbol{\beta}(\mathbf{z}_i)']'$ are $(p+1)\times 1$ vectors. Model (3.1) seemingly takes the form of the standard semiparametric smooth coefficient model subject to endogeneity in one of the covariates, namely, the spatial lag term $\sum_{j\neq i} w_{ij}y_j$. We estimate the above model using the (locally) nonparametric GMM estimator along the lines of Cai & Li (2008) subject to the non-singularity condition (2.3).

Under the assumption that smooth parameter functions are twice continuously differentiable in the neighborhood of z, each element in $\boldsymbol{\gamma}(\cdot)$ can then be approximated by its first-order Taylor expansion around z, i.e., $\gamma_s(\mathbf{z}_i) \approx \gamma_s(z) + \nabla\boldsymbol{\gamma}_s(z)'(\mathbf{z}_i - z)$ at point $\mathbf{z}_i$ close to z for $s = 1, \ldots, p+1$, where $\nabla\boldsymbol{\gamma}_s(z) \equiv [\partial\gamma_s(z)/\partial z_1, \ldots, \partial\gamma_s(z)/\partial z_q]'$ is a $q\times 1$ vector of the first-order derivative of $\gamma_s(\cdot)$. Therefore, for $\mathbf{z}_i$ close to z, we can approximate (3.1) by

$$y_i \approx \mathbf{m}_i'\boldsymbol{\theta}(z)\mathcal{Z}_i(z) + u_i = [\mathcal{Z}_i(z)\otimes \mathbf{m}_i]'\,\mathrm{vec}\{\boldsymbol{\theta}(z)\} + u_i \quad \forall \quad i = 1, \ldots, n, \tag{3.2}$$

where $\mathcal{Z}_i(z) \equiv [1, (\mathbf{z}_i - z)'\mathbf{H}^{-1}]'$ is a $(q+1)\times 1$ vector with $\mathbf{H} = \mathrm{diag}\{h_1, \ldots, h_q\}$ being a $q\times q$ diagonal bandwidth matrix, and

$$\boldsymbol{\theta}(z) \equiv \begin{bmatrix} \gamma_1(z) & \nabla\boldsymbol{\gamma}_1(z)'\mathbf{H} \\ \vdots & \vdots \\ \gamma_{p+1}(z) & \nabla\boldsymbol{\gamma}_{p+1}(z)'\mathbf{H} \end{bmatrix}$$

denotes a $(p+1)\times(q+1)$ parameter matrix.

Before introducing our estimator, we first take a closer look at the spatial lag term. Denoting $\mathbf{S}_n(\mathbf{Z}) \equiv [\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}]^{-1}$ and $\mathbf{G}_n(\mathbf{Z}) \equiv \mathbf{W}\mathbf{S}_n(\mathbf{Z})$, we have the reduced form of our model in (2.2):

$$\mathbf{y} = \mathbf{S}_n(\mathbf{Z})\left(\mathrm{mtx}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\} + \mathbf{u}\right),$$

from where it is evident that the spatial lag term $\mathbf{W}\mathbf{y} = \mathbf{G}_n(\mathbf{Z})[\mathrm{mtx}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\} + \mathbf{u}]$ is endogenous because $\mathbb{E}\left[(\mathbf{G}_n(\mathbf{Z})\mathbf{u})'\mathbf{u}\right] = \sigma_u^2\mathbb{E}[\mathrm{tr}\{\mathbf{G}_n(\mathbf{Z})\}] \neq 0$ in general. Now, let $\mathbf{P}_n$ be some $n\times n$ matrix satisfying $\mathbb{E}\left[(\mathbf{P}_n\mathbf{u})'\mathbf{u}\right] = 0$. We then can rewrite $\mathbf{W}\mathbf{y}$ as follows: $\mathbf{W}\mathbf{y} = \mathbf{G}_n(\mathbf{Z})\,\mathrm{mtx}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\} + \mathbf{P}_n\mathbf{u} + [\mathbf{G}_n(\mathbf{Z}) - \mathbf{P}_n]\mathbf{u}$, where both $\mathbf{G}_n(\mathbf{Z})\,\mathrm{mtx}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\}$ and $\mathbf{P}_n\mathbf{u}$ are uncorrelated with $\mathbf{u}$. This observation suggests that $\mathbf{G}_n(\mathbf{Z})\,\mathrm{mtx}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\}$ and $\mathbf{P}_n\mathbf{u}$ can be reasonable valid instruments for $\mathbf{W}\mathbf{y}$. There are multiple choices for the matrix $\mathbf{P}_n$ with a zero-trace $\mathbf{P}_n = \mathbf{G}_n(\mathbf{Z}) - n^{-1}\mathrm{tr}\{\mathbf{G}_n(\mathbf{Z})\}\mathbf{I}_n$ and a zero-diagonal $\mathbf{P}_n = \mathbf{G}_n(\mathbf{Z}) - \mathrm{diag}\{\mathbf{G}_n(\mathbf{Z})\}$ being among the most popular specifications in the parametric spatial regression literature (e.g., Kelejian & Prucha, 1999; Lee, 2007). However, note that neither $\mathbf{G}_n(\mathbf{Z})\,\mathrm{mtx}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\}$ nor $\mathbf{P}_n\mathbf{u}$ are feasible instruments due to the presence of unknown smooth coefficients $\rho(\mathbf{Z})$ and $\boldsymbol{\beta}(\mathbf{Z})$ in $\mathbf{G}_n(\mathbf{Z})$. In this paper, we therefore propose to first obtain an initial consistent estimator of unknown parameter functions using feasible instruments (in Section 3.1) and then to construct a second-stage estimator (in Section 3.2) which instruments for the endogenous spatial lag with $\mathbf{G}_n(\mathbf{Z})\,\mathrm{mtx}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\}$ and $\mathbf{P}_n\mathbf{u}$ constructed using the initial first-stage consistent estimator of $\rho(\cdot)$ and $\boldsymbol{\beta}(\cdot)$. In what follows, we describe these two proposed estimators.

3.1 First-Stage Estimator

The expansion in (2.4) suggests that $\mathbf{WX}$, $\mathbf{WZ}$, $\mathbf{W}^2\mathbf{X}$, $\mathbf{W}^2\mathbf{Z}$, $\ldots$ with linearly dependent columns removed can be valid instruments for the endogenous spatial lag term $\mathbf{W}\mathbf{y}$. Also note that both matrices $\mathbf{P}_{n,l} = \mathbf{W}^l - n^{-1}\mathrm{tr}\left\{\mathbf{W}^l\right\}\mathbf{I}_n$ and $\mathbf{P}_{n,l} = \mathbf{W}^l - \mathrm{diag}\left\{\mathbf{W}^l\right\}$ for $l = 1, 2, \ldots$ satisfy $\mathbb{E}\left[(\mathbf{P}_{n,l}\mathbf{u})'\mathbf{u}\right] = 0$. We thus have identified feasible instruments which can be employed to obtain an initial estimator for our model in (3.2).

Define an $n \times d$ matrix of instruments $\mathbf{Q}_n = [\mathbf{Q}_{1n}, \mathbf{X}] \equiv \begin{bmatrix} \mathbf{Q}_{n,1} & \dots & \mathbf{Q}_{n,n} \end{bmatrix}'$, where $\mathbf{Q}_{1n}$ contains linearly independent instruments taken from $\left\{\mathbf{WX}, \mathbf{WZ}, \mathbf{W}^2\mathbf{X}, \mathbf{W}^2\mathbf{Z}, \dots\right\}$. We then obtain the following kernel-weighted (local) orthogonal conditions:

$$\mathbb{E}\left[\mathbf{Q}(z)'\mathbf{K}_H(z)\Big(\mathbf{y} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}(z)\}\Big)\right] \approx \mathbf{0}_d \tag{3.3}$$

and

$$\mathbb{E}\left[\Big(\mathbf{y} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}(z)\}\Big)'\mathbf{P}_{n,l}\mathbf{K}_H(z)\Big(\mathbf{y} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}(z)\}\Big)\right] \approx 0 \quad \forall \quad l = 1, 2, \dots, m \tag{3.4}$$

for a finite integer m, where $\mathbf{M}(z) = \begin{bmatrix} \mathcal{Z}_1(z) \otimes \mathbf{m}_1 & \dots & \mathcal{Z}_n(z) \otimes \mathbf{m}_n \end{bmatrix}'$ is an $n \times [(p+1)(q+1)]$ data matrix; $\mathbf{Q}(z) = \begin{bmatrix} \mathcal{Z}_1(z) \otimes \mathbf{Q}_{n,1} & \dots & \mathcal{Z}_n(z) \otimes \mathbf{Q}_{n,n} \end{bmatrix}'$ is an $n \times [d(q+1)]$ instrument matrix; $\mathbf{K}_H(z) = \text{diag}\left\{\mathcal{K}_H(\mathbf{z}_1, z), \ \dots, \mathcal{K}_H(\mathbf{z}_n, z)\right\}$ is an $n \times n$ diagonal matrix of kernel weights with $\mathcal{K}_H(\mathbf{z}_i, z) = \mathcal{K}\left(\mathbf{H}^{-1}(\mathbf{z}_i - z)\right)$ being a product kernel.

Denoting

$$\mathbf{g}_n(\boldsymbol{\theta}) = \begin{bmatrix} \Big(\mathbf{y} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}\}\Big)'\mathbf{P}_{n,1}\mathbf{K}_H(z)\Big(\mathbf{y} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}\}\Big) \\ \vdots \\ \Big(\mathbf{y} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}\}\Big)'\mathbf{P}_{n,m}\mathbf{K}_H(z)\Big(\mathbf{y} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}\}\Big) \\ \mathbf{Q}(z)'\mathbf{K}_H(z)\Big(\mathbf{y} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}\}\Big) \end{bmatrix} \tag{3.5}$$

for a $(p+1) \times (q+1)$ vector $\boldsymbol{\theta}$, we construct our initial first-stage nonparametric GMM estimator:

$$\text{vec}\left\{\widehat{\boldsymbol{\theta}}(z)\right\} = \underset{\boldsymbol{\theta}(z)}{\arg\min} \ \mathbf{g}_n(\boldsymbol{\theta}(z))' \, \mathbf{g}_n(\boldsymbol{\theta}(z)) \,. \tag{3.6}$$

Below, we list assumptions used to derive the limiting distribution of our proposed estimator.

Assumption 1 $\{(\mathbf{x}_i, \mathbf{z}_i, u_i)\}$ is *i.i.d.* over index i, y_i is generated according to (2.1) satisfying (2.3). Also, $\mathbb{E}\left(\|\mathbf{x}_i\|^{2(2+\delta_1)}\right) < M$ and $\mathbb{E}\left[\|\mathbf{z}_i\|^{4(2+\delta_1)}\right] < M$ for some $\delta_1 > 0$.

- (i) $\mathbb{E}\left[u_i|\mathbf{x}_i = x, \mathbf{z}_i = z\right] = 0$, $\mathbb{E}\left[u_i^2|\mathbf{x}_i = x, \mathbf{z}_i = z\right] = \sigma_u^2 > 0$ and $\mathbb{E}\left[|u_i|^{4+\delta}\Big|\mathbf{x}_i = x, \mathbf{z}_i = z\right] < M$ for any $x \in \mathbb{R}^p$ and $z \in \mathbb{R}^q$ and some $\delta > 0$;
- (ii) There exists a positive integer N such that both $\mathbf{W}$ and $\left[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}\right]^{-1}$ have finite row- and column-sum matrix norms for all $n > N$;
- (iii) $\mathbf{P}_{n,l}$ is an $n \times n$ matrix with finite row- and column-sum matrix norms for all $n > N$ and $\text{tr}\left\{\mathbf{P}_{n,l}\right\} = 0$ for all $l = 1, \dots, m$, where $m \geq 1$ is a finite positive integer. And, the $n \times d$ instrument matrix $\mathbf{Q}_n$ has a full rank $d \geq p+1$.

Here, the row- and column-sum matrix norms of some $n \times n$ matrix $\mathbf{A}$ are respectively defined as $\|\mathbf{A}\|_{\text{row}} = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$ and $\|\mathbf{A}\|_{\text{column}} = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|$. Assuming u_i has a finite $(4+\delta)$th-order conditional moment in Assumption 1(i) is necessary to apply the central limit theorem derived by Kelejian & Prucha (2001).

Assumption 2 (i) $\rho(z)$, $\boldsymbol{\beta}(z)$, $f(z)$ and $\mathbb{E}\left[\mathbf{Q}_{n,i}\mathbf{m}_i'|\mathbf{z}_i = z\right]$ are all twice continuously differentiable in a neighborhood of z, where $f(z) > 0$ is the probability density function of $\mathbf{z}_i$ evaluated at point z; (ii) $\mathbb{E}\left(\|\mathbf{Q}_{n,i}\|^{2+\delta_1} |\mathbf{z}_i = z\right)$ is continuous and bounded in the neighborhood of z for some $\delta_1 > 0$.

Assumption 3 The kernel function $k(v)$ is a symmetric probability density function with a compact support $[-1,1]$. Also, we define $\mu_{i,j}(k) = \int k^i(v) v^j dv$ and $R_2(K) = \int \mathcal{K}^2(\mathbf{v}) d\mathbf{v}$, where $\mathcal{K}(\mathbf{v}) = \prod_{j=1}^q k(v_j)$ and $\mathbf{v} = [v_1, \ldots, v_q]'$.

Assumption 4 As $n \to \infty$, $\|\mathbf{H}\| \to 0$, $n|\mathbf{H}| \to \infty$, and $\lim_{n\to\infty} n|\mathbf{H}| \|\mathbf{H}\|^4 = c_0 > 0$, a finite constant, where $|\mathbf{H}| = \prod_{j=1}^q h_j$ and $\|\mathbf{H}\|^2 = \sum_{j=1}^q h_j^2$.

Theorem 1 *Under Assumptions 1–4, and if $\varkappa_B(\mathbf{H}, z)$ is nonsingular at an interior point z, we have*

$$\begin{aligned}&\sqrt{n|\mathbf{H}|}\left(\widehat{\boldsymbol{\gamma}}(z) - \boldsymbol{\gamma}(z) - \frac{\mu_{1,2}(k)}{2} Bias(\mathbf{H}, z)\right)\\ &\xrightarrow{d} \mathbb{N}\Big(\mathbf{0}_{p+1},\ f^{-1}(z) R_2(K) \operatorname*{plim}_{n\to\infty} \varkappa_B(\mathbf{H}, z)^{-1} \boldsymbol{\Omega}(z) \operatorname*{plim}_{n\to\infty} \varkappa_B(\mathbf{H}, z)^{-1}\Big),\end{aligned}$$

where $Bias(\mathbf{H}, z) = \mathbf{S}_{p+1} \varkappa_B(\mathbf{H}, z)^{-1} \varkappa_A(\mathbf{H}, z)' = O_p\left(\|\mathbf{H}\|^2\right)$, $\mathbf{S}_{p+1}$ equals the first $p+1$ rows of the identity matrix $\mathbf{I}_{[(p+1)(q+1)]}$, and $\varkappa_A(\mathbf{H}, z)$, $\varkappa_B(\mathbf{H}, z)$, and $\boldsymbol{\Omega}(z)$ are respectively defined in Lemmas 1, 2 and 3 in Appendix A.

Theorem 1 states that the local linear estimator of the varying coefficients $\rho(z)$ and $\boldsymbol{\beta}(z)$ has the conventional bias term of order $O_p\left(\|\mathbf{H}\|^2\right)$ and asymptotic variance of order $O_p\left((n|\mathbf{H}|)^{-1/2}\right)$.

Remark 2 As discussed in Lemma 2, $\varkappa_B(\mathbf{H}, z)$ can be singular if $\mathbf{X}$ is irrelevant in predicting $\mathbf{y}$ or $\boldsymbol{\beta}(z) = \mathbf{0}_p$ holds true over its domain. However, this problem occurs only to the local linear regression approach and not to the local constant approach. Thus, the estimator derived from the nonparametric GMM via a local constant fitting does *not* suffer from the singularity problem even if $\mathbf{X}$ is irrelevant for $\mathbf{y}$.

Next, notice that, since $\widehat{\boldsymbol{\theta}}(z)$ in (3.6) does not have an analytic formula due to the use of the *quadratic* moment (3.4), the computation of our estimator can be significantly slow for relatively large samples. However, if one is reasonably confident that $\boldsymbol{\beta}(z)$ is a non-zero vector over at least one non-empty subset, we are able to construct an alternative consistent estimator of $\boldsymbol{\gamma}(z)$ which has a simple closed-form formula and hence is fast to compute. This alternative estimator uses the *linear* local orthogonal condition in (3.3) only.[4]

Specifically, we are interested in minimizing the following (local) objective function:

$$\min_{\boldsymbol{\theta}(z)} \Big[\mathbf{Q}(z)'\mathbf{K}_H(z)\Big(\mathbf{y} - \mathbf{M}(z)\mathrm{vec}\{\boldsymbol{\theta}(z)\}\Big)\Big]' \mathbf{Q}(z)'\mathbf{K}_H(z)\Big(\mathbf{y} - \mathbf{M}(z)\mathrm{vec}\{\boldsymbol{\theta}(z)\}\Big), \tag{3.7}$$

which yields

$$\mathrm{vec}\left\{\widehat{\boldsymbol{\theta}}(z)\right\} = \Big[\mathbf{M}(z)'\boldsymbol{\Xi}_H(z)\mathbf{M}(z)\Big]^{-1}\mathbf{M}(z)'\boldsymbol{\Xi}_H(z)\mathbf{y}, \tag{3.8}$$

where $\boldsymbol{\Xi}_H(z) \equiv \mathbf{K}_H(z)'\mathbf{Q}(z)\mathbf{Q}(z)'\mathbf{K}_H(z)$. Following the proof of Theorem 1, we obtain the limit result for $\mathrm{vec}\left\{\widehat{\boldsymbol{\theta}}(z)\right\}$ as follows.

[4]Note that $\mathbf{Q}_n$ is *not* a valid instrument matrix if $\mathbf{X}$ and $\mathbf{Z}$ are both irrelevant in predicting $\mathbf{y}$, which occurs if $\boldsymbol{\beta}(z) = \mathbf{0}_p$ and $\rho(z) = \rho_0$ over their domains and our model (2.1) becomes a pure spatial autoregression $y_i = \rho_0 \sum_{j\neq i} w_{ij} y_j + u_i \ \forall\, i$. See Lee (2007) for the detailed discussion in a parametric spatial autoregressive framework.

Corollary 2 *Under Assumptions 1–4 and assuming* $\mathbf{E}_1(z)'\mathbf{E}_1(z)$ *is non-singular, at an interior point* z*, we have*

$$\sqrt{n\,|\mathbf{H}|}\left(\widehat{\boldsymbol{\gamma}}(z)-\boldsymbol{\gamma}(z)-\frac{\mu_{1,2}(k)}{2}Bias\left(\mathbf{H},z\right)\right)\ \xrightarrow{d}\ \mathbb{N}\Big(\mathbf{0}_{p+1},\ \frac{\sigma_u^2}{f(z)}R_2\left(K\right)\boldsymbol{\Sigma}(z)\Big),$$

where $Bias\left(\mathbf{H},z\right)=\left[\mathbf{E}_1(z)'\mathbf{E}_1(z)\right]^{-1}\mathbf{E}_1(z)'\mathbf{E}_2\left(\mathbf{H},z\right)'=O\left(\|\mathbf{H}\|^2\right)$,

$$\boldsymbol{\Sigma}(z)=\left[\mathbf{E}_1(z)'\mathbf{E}_1(z)\right]^{-1}\left[\mathbf{E}_1(z)'\mathbf{E}_3(z)\mathbf{E}_1(z)\right]\left[\mathbf{E}_1'(z)\mathbf{E}_1(z)\right]^{-1}$$

and $\mathbf{E}_1(z)$, $\mathbf{E}_2\left(\mathbf{H},z\right)$ *and* $\mathbf{E}_3(z)$ *are respectively defined in (A.8), (A.9) and (A.24) in Appendix A.*

3.2 Second-Stage Estimator

Having obtained the initial estimator of unknown parameter functions $\rho(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ in (3.6), we can now construct the feasible versions of our originally desired instruments, namely, $\widehat{\mathbf{Q}}_{1n}=\widehat{\mathbf{G}}_n\left(\mathbf{Z}\right)\mathrm{mtx}\left\{\mathbf{X},\widehat{\boldsymbol{\beta}}\left(\mathbf{Z}\right)\right\}$ and, e.g., the zero-trace $\widehat{\mathbf{P}}_n=\widehat{\mathbf{G}}_n\left(\mathbf{Z}\right)-n^{-1}\mathrm{tr}\left\{\widehat{\mathbf{G}}_n\left(\mathbf{Z}\right)\right\}\mathbf{I}_n$, where $\widehat{\mathbf{G}}_n\left(\mathbf{Z}\right)=\mathbf{W}\left[\mathbf{I}_n-\widehat{\boldsymbol{\rho}}\left(\mathbf{Z}\right)\mathbf{W}\right]^{-1}$. With this, we derive our second-stage nonparametric GMM estimator:

$$\mathrm{vec}\left\{\widetilde{\boldsymbol{\theta}}(z)\right\}=\arg\min_{\boldsymbol{\theta}(z)}\mathbf{g}_n\left(\boldsymbol{\theta}(z)\right)'\mathbf{g}_n\left(\boldsymbol{\theta}(z)\right), \tag{3.9}$$

where the moment function is defined as

$$\mathbf{g}_n\left(\boldsymbol{\theta}\right)=\left[\begin{array}{c}\Big(\mathbf{y}-\mathbf{M}(z)\mathrm{vec}\{\boldsymbol{\theta}\}\Big)'\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\Big(\mathbf{y}-\mathbf{M}(z)\mathrm{vec}\{\boldsymbol{\theta}\}\Big)\\ \widehat{\mathbf{Q}}(z)'\mathbf{K}_{H_0}(z)\Big(\mathbf{y}-\mathbf{M}(z)\mathrm{vec}\{\boldsymbol{\theta}\}\Big)\end{array}\right], \tag{3.10}$$

$\mathbf{H}_0$ is the new $q\times q$ diagonal bandwidth matrix; and $\mathbf{M}(z)$, $\widehat{\mathbf{Q}}(z)$ and $\mathrm{vec}\{\boldsymbol{\theta}\}$ are as defined earlier except for $\widehat{\mathbf{Q}}(z)$ being constructed using $\widehat{\mathbf{Q}}_{1n}$ in place of $\mathbf{Q}_{1n}$.

Asymptotic results for our second-stage estimator require the following additional assumptions.

Assumption 5 $\sup_{z\in\mathcal{S}_z}\|\widehat{\boldsymbol{\gamma}}\left(z\right)-\boldsymbol{\gamma}\left(z\right)\|=O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/\left(n\,|\mathbf{H}|\right)}\right)$.

Assumption 6 (i) $\|\mathbf{H}\|\to 0$, $\|\mathbf{H}_0\|\to 0$, $n\,|\mathbf{H}|\to\infty$, $n\,|\mathbf{H}_0|\to\infty$, and $n\,|\mathbf{H}_0|\,\|\mathbf{H}_0\|^4=O\left(1\right)$; (ii) $\|\mathbf{H}\|^4/\left(|\mathbf{H}_0|\,\|\mathbf{H}_0\|^2\right)\to 0$, $n\,|\mathbf{H}|\,|\mathbf{H}_0|\,\|\mathbf{H}_0\|^2/\ln n\to\infty$, and $n\sqrt{|\mathbf{H}|}\,|\mathbf{H}_0|\,\|\mathbf{H}_0\|^3/\ln n\to\infty$.

Assumption 5 strengthens the pointwise consistency result for $\widehat{\boldsymbol{\gamma}}\left(z\right)$ to a uniform convergence result over its domain $\mathcal{S}_z$. This is a reasonable result to be expected if one assumes that $\mathcal{S}_z$ is a compact subset of $\mathbb{R}^q$ and $v^jk\left(v\right)$ satisfies Lipschitz condition for $0\le j\le 3$ by closely following the proof in Masry (1996). If the random variable $\mathbf{z}_i$ contains a component taking values along the real line, Assumption 5 will hold for an expanding subset growing at proper speed (Hansen, 2008).

Assumption 6 regulates the choice of the two bandwidth matrices for the first- and second-stage estimators. For ease of discussion, we set $\mathbf{H}=\mathrm{diag}\left\{h,\ldots,h\right\}$ and $\mathbf{H}_0=\mathrm{diag}\left\{h_0,\ldots,h_0\right\}$ with $h=cn^{-\alpha}$ and $h_0=c_0n^{-\alpha_0}$. Assumption 6(ii) implies $\alpha_0=1/\left(4+q\right)$. Assumption 6(i) requires $\alpha_0\left(2+q\right)/4<\alpha<\min\left\{\left[1-\left(q+3\right)\alpha_0\right]2/q,\ \ \left[1-\left(q+2\right)\alpha_0\right]/q\right\}$ which, when combined with Assumption 6(ii), implies that $q<2$. When $q=1$, Assumption 6 states that $\alpha_0=1/5$ and $0.15<\alpha<0.4$ so that we can set $\alpha_0=\alpha=1/5$ in both steps. However, when $\mathbf{z}_i$ contains more than one continuous variable, one needs to apply a higher-order local polynomial approach in the first-stage estimation to ensure the optimal convergence rate in the second-stage estimation. For instance, if we apply the local rth-order polynomial approach with an odd integer $r \geq 1$, $\|\mathbf{H}\|^4 / \left(|\mathbf{H}_0| \|\mathbf{H}_0\|^2\right) = o(1)$ in Assumption 6(ii) will be replaced by $\|\mathbf{H}\|^{2(r+1)} / \left(|\mathbf{H}_0| \|\mathbf{H}_0\|^2\right) = o(1)$. With $\alpha_0 = 1/(4+q)$, we have $\alpha_0 (2+q) / [2(r+1)] < \alpha < [1-(q+3)\alpha_0] 2/q$ which implies that $q < \sqrt{4r+5} - 1$. So, if $q = 2$, $r > 1$ is required, and hence we can apply the local cubic polynomial approach in the first-stage estimation in that instance.

Theorem 3 *Under Assumptions **1–3**, **5** and **6**, at an interior point z, we have*

$$\sqrt{n|\mathbf{H}_0|}\left(\widetilde{\boldsymbol{\gamma}}(z) - \boldsymbol{\gamma}(z) - \frac{\mu_{1,2}(k)}{2} Bias(\mathbf{H}_0, z)\right)$$
$$\xrightarrow{d} \mathbb{N}\left(\mathbf{0}_{p+1},\ f^{-1}(z) \operatorname*{plim}_{n\to\infty} \varkappa_B(\mathbf{H_0}, z)^{-1} \boldsymbol{\Omega}(z) \operatorname*{plim}_{n\to\infty} \varkappa_B(\mathbf{H}_0, z)^{-1}\right),$$

where $Bias(\mathbf{H}_0, z) = \mathbf{S}_{p+1} \varkappa_B(\mathbf{H}_0, z)^{-1} \varkappa_A(\mathbf{H}_0, z)'$, and $\varkappa_A(\mathbf{H}_0, z)$, $\varkappa_B(\mathbf{H}_0, z)$ and $\boldsymbol{\Omega}(z)$ are respectively defined in Lemmas 4, 5 and 6 in Appendix A.

Theorem 3 states that the second-stage estimator is consistent and has an asymptotic normal distribution at the usual nonparametric convergence rate with the proper order of a local polynomial estimation approach. Intuitively, the second-stage estimator is expected to be asymptotically more efficient than its first-stage counterpart because the instruments employed in the first-stage estimation have lower predictive power for $\mathbf{Wy}$ than those used in the second stage. This intuition is confirmed by our Monte Carlo simulations in Section 6.1. However, we are unable to analytically verify this result since both asymptotic variances in Theorems 1 and 3 take complex forms.

Remark 3 The consistency and asymptotic normality results for the second-stage estimator continue to hold if we use the zero-diagonal $\widehat{\mathbf{P}}_n = \widehat{\mathbf{G}}_n(\mathbf{Z}) - \operatorname{diag}\left\{\widehat{\mathbf{G}}_n(\mathbf{Z})\right\}$ in (3.10). Naturally, the definition of $\varkappa_A(\mathbf{H}_0, z)$, $\varkappa_B(\mathbf{H}_0, z)$ and $\boldsymbol{\Omega}(z)$ referenced in Theorem 3 will change accordingly.

3.3 Non-Singularity Constraint

Until now, we have proceeded with the estimation of (2.1) as if it were a standard smooth coefficient model subject to endogeneity in one of its covariates, namely $\mathbf{Wy}$. However, recall that the said endogenous covariate is a spatially-weighted average of the left-hand-side variable. To ensure that the outcome variable $\mathbf{y}$ is uniquely defined by (2.2), the non-singularity condition in (2.3) needs to be satisfied. We do so using the "tilting" procedure proposed by Hall & Huang (2001) and Du, Parmeter & Racine (2013). Here, we generalize Hall & Huang's (2001) theoretical results to the case of GMM estimators in the presence of endogenous regressors. The procedure essentially mutes or magnifies the impact of any given data point used in the estimation. This allows us to impose the non-singularity restriction *post*-estimation via a quadratic programming technique. The idea is to reweigh observations used in the estimation so that the non-singularity constraint is satisfied in the *local* neighborhood of point z:

$$\max_{1\leq i\leq n} |\lambda_i \{\rho(z)\mathbf{W}\}| < 1. \tag{3.11}$$

Since the estimator derived using not only linear orthogonality condition (3.3) but also quadratic condition (3.4) does not have an analytical solution, the "tilting" procedure proposed by Hall &

Huang (2001) does not apply. We therefore limit our attention to a more practical estimator in (3.8) which makes use of linear moments only and hence is valid when $\boldsymbol{\beta}(z) \neq \mathbf{0}_p$ over at least one non-empty subset.[5]

In order to facilitate the imposition of condition (3.11) in the neighborhood of z, we assume that $\max_{1\leq i\leq n} |\lambda_i\{\mathbf{W}\}| \leq 1$ holds true, which is satisfied if one standardizes a raw spatial weighting matrix by dividing all of its elements by its largest eigenvalue in absolute value. Then, the non-singularity condition (3.11) is satisfied if $|\rho(z)| < 1$ (Kelejian & Prucha, 2004, 2010). To impose the latter condition, we rewrite the kernel estimator of $\rho(z)$ from (3.8) as a weighted average of the outcome variable:

$$\widehat{\rho}(z) = \sum_{i=1}^{n} \omega_i(\mathbf{X}, z) y_i, \tag{3.12}$$

where $\omega_i(\mathbf{X}, z)$ is the ith (column) element in the *first* row of $\left[\mathbf{M}(z)'\mathbf{\Xi}_H(z)\mathbf{M}(z)\right]^{-1}\mathbf{M}(z)'\mathbf{\Xi}_H(z)$. Following Hall & Huang (2001), we can generalize (3.12) as

$$\widetilde{\rho}(z|\boldsymbol{p}) = n\sum_{i=1}^{n} p_i\omega_i(\mathbf{X}, z)\, y_i, \tag{3.13}$$

where $\boldsymbol{p} = (p_1, \ldots, p_n)'$ is the sequence of additional weights such that $\sum_{i=1}^{n} p_i = 1$. Note that p_i equals $1/n$ (i.e., uniform weights) in the case of an *un*constrained estimator in (3.12).

If necessary, we can impose the non-singularity condition by selecting weights $\boldsymbol{p}$ to minimize the L_2-metric $D(\boldsymbol{p}) = (1/n\mathbf{i}_n - \boldsymbol{p})'(1/n\mathbf{i}_n - \boldsymbol{p})$ subject to $\mathbf{i}_n'\boldsymbol{p} = 1$ and $-\mathbf{i}_n < \left[\widetilde{\rho}(z_1|\boldsymbol{p}), \ldots, \widetilde{\rho}(z_n|\boldsymbol{p})\right]' < \mathbf{i}_n$, where $\mathbf{i}_n$ is an $n \times 1$ vector of ones. Here, $D(\boldsymbol{p})$ is the sum of squared deviations of p_i from the unrestricted value of $1/n$. In our choice of the distance metric, we thus follow Du et al. (2013), which allows $\boldsymbol{p}$ to be both positive and negative.[6] The minimization problem is solved via a standard quadratic programming technique. Let $\widehat{\boldsymbol{p}}$ be the solution to this optimization problem. To derive the asymptotic results for $\widetilde{\rho}(z|\widehat{\boldsymbol{p}})$, we need the following additional assumptions.

Assumption 7 (i) The random variable $\mathbf{z}_i$ has a compact support, i.e., $\mathcal{S}_z = [\mathbf{a}, \mathbf{b}] = [a_1, b_1] \times \ldots \times [a_q, b_q]$ is a compact subset of $\mathbb{R}^q$ and has a common Lebesgue probability density $f(z)$ satisfying $\inf_{z\in\mathcal{S}_z} f(z) > 0$; (ii) $\boldsymbol{\beta}(z) \neq \mathbf{0}_p$ over at least one non-empty interval; (iii) the kernel function $k(v)$ satisfies $\left|v^j k(v) - s^j k(s)\right| \leq M|v - s|$ for any $v, s \in \mathbb{R}$ and $0 \leq j \leq 3$.

Assumption 8 Let $\mathcal{A}_{n,\epsilon} = \{j : \widetilde{\rho}(z|\widehat{\boldsymbol{p}}) = \pm(1-\epsilon)\}$ for a given very small $\epsilon \in (0, 1)$ and $\mathcal{W}_n$ be an $n \times |\mathcal{A}_{n,\epsilon}|$ matrix with a typical element $\omega_i(\mathbf{X}, \mathbf{z}_j)y_i$ for $j = 1, \ldots, |\mathcal{A}_{n,\epsilon}|$, where $\mathcal{A}_{n,\epsilon} = \left\{j_1, \ldots, j_{|\mathcal{A}_{n,\epsilon}|}\right\} \subseteq \{1, \ldots, n\}$. Here, $\mathcal{W}_n$ has a full rank of $|\mathcal{A}_{n,\epsilon}|$ with $\lambda_{\min}\{\mathbb{E}[\mathcal{W}_n'\mathcal{W}_n]/n\} \overset{n\to\infty}{\longrightarrow} c_1 > 0$ for a finite constant c_1.

Assumption 7 is required to ensure that $\widehat{\rho}(z)$ converges to $\rho(z)$ uniformly over $z \in \mathcal{S}_z$. Assumption 8 holds if matrix $\mathbf{Z}$ has a full rank and the partial derivative of $\omega_i(\mathbf{X}, \mathbf{z}_j)y_i$ with respect to $\mathbf{z}_j$ does not equal zero for all i. Note that it is not essential to know the value of ϵ since it is introduced to ensure that $\widetilde{\rho}(z|\widehat{\boldsymbol{p}})$ does not reach ± 1. Under these additional assumptions, below we provide the limit result for the *constrained* estimator $\widetilde{\rho}(z|\widehat{\boldsymbol{p}})$.

[5] In practice, when estimating the estimators in (3.6) and (3.9), which incorporate quadratic orthogonality conditions, the non-singularity condition (3.11) may be easily imposed via box constraints on $\rho(z)$ during the numerical optimization.

[6] Hall & Huang (2001) use a power divergence metric which has a rather complicated form and is only valid for non-negative weights.

Theorem 4 *Under Assumptions **1–4**, **7** and **8**, we have:* (i) *there exists* $\{\widehat{p}_i\}_{i\leq n}$ *such that* $\max_{1\leq i\leq n}|\widehat{p}_i| = O_p\left(n^{-1}\right)$ *holds for all* $z\in\mathcal{S}_z$*;* (ii) $\widetilde{\rho}(z|\widehat{\boldsymbol{p}}) = \widehat{\rho}(z) + O_p\left((n\,|\mathbf{H}|)^{-1/2}\right)$ *for an interior point* $z\in\mathcal{S}_z$*.*

By Corollary 2 and Theorem 4, we show that $\widetilde{\rho}(z|\widehat{\boldsymbol{p}}) = \rho(z) + O_p\left(\|\mathbf{H}\|^2 + (n\,|\mathbf{H}|)^{-1/2}\right)$. Hence, $\widetilde{\rho}(z|\widehat{\boldsymbol{p}})$ is a consistent estimator of $\rho(z)$ and has the same convergence rate as $\widehat{\rho}(z)$ does. From the proof of this theorem in Appendix A, we obtain

$$\sqrt{n\,|\mathbf{H}|}\left(\widetilde{\rho}(z|\widehat{\boldsymbol{p}}) - \rho(z) - \mathbf{S}_{1,(p+1)}Bias\left(\mathbf{H},z\right) - (n\,|\mathbf{H}|)^{-1/2}\,\Delta\left(z\right)\right) \xrightarrow{d} \mathbb{N}\left(0,\ \mathbf{S}_{1,(p+1)}\boldsymbol{\Sigma}(z)\mathbf{S}_{1,(p+1)}'\right),$$

where $Bias\left(\mathbf{H},z\right)$ and $\boldsymbol{\Sigma}(z)$ are defined in Corollary 2, $\mathbf{S}_{1,r}$ is the first row of the identity matrix $\mathbf{I}_r$, and $\Delta\left(z\right)$ is some continuous function of z. Thus, compared with $\widehat{\rho}(z)$, $\widetilde{\rho}(z|\widehat{\boldsymbol{p}})$ has an additional vanishing asymptotic bias term of order $O_p\left((n\,|\mathbf{H}|)^{-1/2}\right)$ resulting from the "tilting" procedure. However, the two estimators have the same asymptotic variance.

4 Special Case: Partially Linear Spatial Autoregressive Model

In this section, we consider two special cases of our model in (2.1) by allowing some of the varying coefficients to be constant. That is, we study a partially linear semiparametric spatially autoregressive model. The primary advantage of such a model (over a fully semiparametric specification) is its potential for efficiency gains stemming from the additional information about constancy of some of the parameter functions.

Such partially linear semiparametric models have been extensively studied for sampling with no spatial or cross-sectional dependence by, e.g., Ahmad, Leelahanon & Li (2005), Kai, Li & Zou (2011) and Cai & Xiao (2012). In the spatial autoregression literature, Su (2012) and Zhang (2013) both focus on the case when $\rho\left(\mathbf{z}_i\right) = \rho_0$ over its domain, however, with varying assumptions about $\mathbf{x}_i'\boldsymbol{\beta}\left(\mathbf{z}_i\right)$ (in our notation). Zhang (2013) assumes that $\mathbf{x}_i'\boldsymbol{\beta}\left(\mathbf{z}_i\right) = \mathbf{x}_i'\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2\left(\mathbf{z}_i\right)$, whereas Su (2012) assumes that $\mathbf{x}_i = 1$. The nonparametric GMM estimators proposed in these papers are however inconsistent if the true model is a pure spatial autoregressive model without other regressors. In contrast, the estimators that we put forward here do not suffer from such a problem.

Specifically, we study the case when $\mathbf{x}_i'\boldsymbol{\beta}\left(\mathbf{z}_i\right) = \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \mathbf{x}_{2i}'\boldsymbol{\beta}_2(\mathbf{z}_i)$, where $\mathbf{x}_i = [\mathbf{x}_{1i}', \mathbf{x}_{2i}']'$ is partitioned into two $p_j\times 1$ sub-vectors $\mathbf{x}_{ji}$ for $j=1,2$, and $\boldsymbol{\beta}\left(\mathbf{z}_i\right) = \left[\boldsymbol{\beta}_1', \boldsymbol{\beta}_2(\mathbf{z}_i)'\right]'$ is accordingly split into a $p_1\times 1$ vector of constant parameters $\boldsymbol{\beta}_1$ and $p_2\times 1$ vector of varying parameter functions $\boldsymbol{\beta}_2(\cdot)$. Depending on whether the spatial lag parameter $\rho\left(\mathbf{z}_i\right)$ is also constant or not, we study two alternative partially linear specifications of our model. Section 4.1 treats $\rho\left(\mathbf{z}_i\right)$ as a varying function, while Section 4.2 lets $\rho\left(\mathbf{z}_i\right)$ be constant.

To keep our notation simple and to make a better connection with the results derived earlier, throughout this section we adhere by the notation used in Section 3 wherever possible and redefine variables when necessary.

4.1 Nonlinear in the Spatial Autoregressive Parameter

Consider the following partially linear model:

$$y_i = \rho(\mathbf{z}_i)\sum_{j\neq i} w_{ij}y_j + \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \mathbf{x}_{2i}'\boldsymbol{\beta}_2(\mathbf{z}_i) + u_i \quad \forall \quad i=1,\ldots,n \tag{4.1}$$

for which we redefine $\mathbf{m}_i \equiv \left[\left(\sum_{j\neq i} w_{ij} y_j\right), \mathbf{x}'_{2i}\right]'$, $\boldsymbol{\gamma}(\mathbf{z}_i) \equiv [\rho(z), \boldsymbol{\beta}_2(\mathbf{z}_i)']'$. Also, let $\dot{y}_i \equiv y_i - \mathbf{x}'_{1i}\boldsymbol{\beta}_1$.

Closely following the methodology introduced in Section 3, we have the following kernel-weighted orthogonal moment conditions:

$$\mathbb{E}\left[\mathbf{Q}(z)'\mathbf{K}_H(z)\Big(\dot{\mathbf{y}} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}(z)\}\Big)\right] \approx \mathbf{0}_d \tag{4.2}$$

$$\mathbb{E}\left[\Big(\dot{\mathbf{y}} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}(z)\}\Big)'\mathbf{P}_{n,l}\mathbf{K}_H(z)\Big(\dot{\mathbf{y}} - \mathbf{M}(z)\text{vec}\{\boldsymbol{\theta}(z)\}\Big)\right] \approx 0 \quad \forall \quad l = 1, 2, \ldots, m \tag{4.3}$$

for some finite positive integer m, where $\dot{\mathbf{y}} = [\dot{y}_1, \ldots, \dot{y}_n]'$; $\mathbf{Q}(z)$, $\mathbf{P}_{n,l}$ and $\mathbf{K}_H(z)$ are as defined earlier in Section 3; and both $\mathbf{M}(z)$ and $\boldsymbol{\theta}(z)$ are defined in the same fashion as in Section 3 but using newly redefined $\mathbf{m}_i$ and $\boldsymbol{\gamma}(\cdot)$. Then, our nonparametric GMM estimator is given by

$$\left[\widehat{\boldsymbol{\beta}}_1(z)', \text{vec}\left\{\widehat{\boldsymbol{\theta}}(z)\right\}'\right]' = \arg\min_{\boldsymbol{\beta}_1, \boldsymbol{\theta}(z)} \mathbf{g}_n(\boldsymbol{\theta}(z))' \mathbf{g}_n(\boldsymbol{\theta}(z)), \tag{4.4}$$

where $\mathbf{g}_n(\cdot)$ has the same form as in (3.5) with $\mathbf{y}$ being replaced with $\dot{\mathbf{y}}$.

Since model (4.1) is nested within our model (2.1), the asymptotic normality result shown in Theorem 1 continues to hold for $\widehat{\boldsymbol{\theta}}(z)$ and is thus omitted. Remark 2 applies here too.

Lastly, we estimate constant parameters $\boldsymbol{\beta}_1$ in the second stage by $\widetilde{\boldsymbol{\beta}}_1 = n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{\beta}}_1^{(-i)}(\mathbf{z}_i)$, where $\widehat{\boldsymbol{\beta}}_1^{(-i)}(\mathbf{z}_i)$ is a leave-one-out (first-stage) estimator computed via (4.4) while excluding the ith unit. We make use of the leave-one-out technique in order to remove an asymptotic bias term of order $O_p\left((n|\mathbf{H}|)^{-1/2}\right)$ in the estimation of $\boldsymbol{\beta}_1$. To this end, we make the following assumption.

Assumption 9 $\|\mathbf{H}\| \to 0$, $\ln n/(n|\mathbf{H}|) \to \infty$ and $n\|\mathbf{H}\|^4 \to 0$ as $n \to \infty$.

Theorem 5 *Under Assumptions **1–3**, **5** and **9**, we have*

$$\sqrt{n}\left(\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\right) \overset{d}{\to} \mathbb{N}(\mathbf{0}_{p_1}, \boldsymbol{\Omega}_0),$$

where $\boldsymbol{\Omega}_0$ is defined in Lemma 11 in Supplementary Appendix.

Theorem 5 shows that $\widetilde{\boldsymbol{\beta}}_1$ is a root-n consistent estimator of $\boldsymbol{\beta}_1$ if the bandwidth converges to zero at a faster speed than the optimal bandwidth would suggest.

4.2 Linear in the Spatial Autoregressive Parameter

Next, consider a partially linear model with the constant spatial autoregressive parameter:

$$y_i = \rho_0 \sum_{j\neq i} w_{ij} y_j + \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + \mathbf{x}'_{2i}\boldsymbol{\beta}_2(\mathbf{z}_i) + u_i \quad \forall \quad i = 1, \ldots, n, \tag{4.5}$$

where both the ρ_0 and $\boldsymbol{\beta}_1$ are constant parameters. Incidentally, Sun et al. (2014) study a special case of (4.5) where $\boldsymbol{\beta}_1 = \mathbf{0}_{p_1}$. Extending their estimation method, one can apply a local linear regression method to recover $\boldsymbol{\beta}_2(\cdot)$ and a profile likelihood method to recover ρ_0 and $\boldsymbol{\beta}_1$. Given that a likelihood-based method may not be computationally feasible in many instances with moderate- or large-sized samples (Kelejian & Prucha, 1999), we propose a nonparametric GMM estimator which differs from the one introduced in Sun et al. (2014).

Specifically, we define $\mathbf{m}_{1i} \equiv \left[\left(\sum_{j\neq i} w_{ij}y_j\right), \mathbf{x}'_{1i}\right]'$, $\mathbf{m}_i \equiv \mathbf{x}_{2i}$, $\boldsymbol{\gamma}_0 \equiv \left[\rho_0, \boldsymbol{\beta}'_1\right]'$, $\boldsymbol{\gamma}(z) \equiv \boldsymbol{\beta}_2(z)$ and $\dot{y}_i \equiv y_i - \mathbf{m}'_{1i}\boldsymbol{\gamma}_0$. In addition, we redefine $\mathbf{S}_n \equiv [\mathbf{I}_n - \rho_0\mathbf{W}]^{-1}$ and $\mathbf{G}_n \equiv \mathbf{W}\mathbf{S}_n$. Again, we have the same form of the kernel-weighted orthogonal moment conditions given in (4.2)–(4.3) with the newly redefined $\dot{\mathbf{y}}$, $\mathbf{M}(z)$ and $\boldsymbol{\theta}(z)$. The corresponding nonparametric GMM estimator is

$$\left[\widehat{\boldsymbol{\gamma}}_0(z)', \text{vec}\left\{\widehat{\boldsymbol{\theta}}(z)\right\}'\right]' = \arg\min_{\boldsymbol{\gamma}_0, \boldsymbol{\theta}(z)} \mathbf{g}_n(\boldsymbol{\theta}(z))' \mathbf{g}_n(\boldsymbol{\theta}(z)), \tag{4.6}$$

where $\mathbf{g}_n(\cdot)$ has the same form as in (3.5) with $\mathbf{y}$ being replaced with $\dot{\mathbf{y}}$. As in Section 4.1, we estimate constant parameters $\boldsymbol{\gamma}_0$ in the second stage via $\widetilde{\boldsymbol{\gamma}}_0 = n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{\gamma}}_0^{(-i)}(\mathbf{z}_i)$, where $\widehat{\boldsymbol{\gamma}}_0^{(-i)}(\mathbf{z}_i)$ is a leave-one-out (first-stage) estimator computed via (4.6) while excluding the ith unit.

Theorem 6 *Under Assumptions **1–3**, **5** and **9**, we have*

$$\sqrt{n}\left(\widetilde{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0\right) \xrightarrow{d} \mathbb{N}\left(\mathbf{0}_{p_1+1}, \ \boldsymbol{\Omega}_1\right)$$

where $\boldsymbol{\Omega}_1$ is as defined in Lemma 15 in Supplementary Appendix.

Analogous to the estimator in (4.4), the asymptotic normality result continues to hold for $\widehat{\boldsymbol{\theta}}(z)$ and is thus omitted. Remark 2 carries to this section too. Lastly, Theorem 6 shows that $\widetilde{\boldsymbol{\gamma}}_0$ is a root-n consistent estimator of $\boldsymbol{\gamma}_0$ if we undersmooth in the first stage.

5 Consistent Testing for a Linear Spatial Autoregressive Model

Given that our semiparametric model nests the parametric linear spatial autoregressive model as a special case, one may naturally wish to formally discriminate between the two models. In this section, we propose two test statistics for testing the null hypothesis of a linear spatial autoregressive model against a smooth coefficient spatial autoregressive model defined in (2.1). The proposed are, essentially, specification tests for parameter constancy. Specifically, we consider the following null and alternative hypotheses:

$$\begin{aligned} \mathrm{H}_0: &\quad \Pr\left\{\left[\rho(\mathbf{z}_i), \boldsymbol{\beta}(\mathbf{z}_i)'\right]' = \left[\rho_0, \boldsymbol{\beta}'_0\right]'\right\} = 1 \text{ for some } \left[\rho_0, \boldsymbol{\beta}'_0\right] \in \Theta \subset \mathbb{R}^{1+p} \\ \mathrm{H}_1: &\quad \Pr\left\{\left[\rho(\mathbf{z}_i), \boldsymbol{\beta}(\mathbf{z}_i)'\right]' = \left[\rho, \boldsymbol{\beta}'\right]'\right\} < 1 \text{ for any } \left[\rho, \boldsymbol{\beta}'\right] \in \Theta, \end{aligned}$$

where Θ is a compact subset of $\mathbb{R}^{1+p}$. With model (2.1) as the model under the alternative hypothesis, we are interested in testing whether it is necessary to allow for parameter heterogeneity when studying spatial autoregressive models. In what follows, we adhere by our original notation used in Section 3.

The first specification test we propose is based on the following residual-based test statistic:

$$T_n = \frac{1}{n^2|\mathbf{H}|}\sum_{i=1}^n \sum_{j\neq i}^n \widehat{u}_i \widehat{u}_j \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j), \tag{5.1}$$

where $\widehat{u}_i = y_i - \mathbf{m}'_i\breve{\boldsymbol{\gamma}}$ with $\breve{\boldsymbol{\gamma}} = \left[\breve{\rho}, \breve{\beta}'\right]'$ being a consistent estimator of $\left[\rho_0, \boldsymbol{\beta}'_0\right]'$, and $\mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j) = \mathcal{K}\left(\mathbf{H}^{-1}(\mathbf{z}_i - \mathbf{z}_j)\right)$. Under H_0, model (2.1) becomes a linear spatial autoregressive model, i.e., $y_i = \rho_0 \sum_{j\neq i} w_{ij}y_j + \mathbf{x}'_i\boldsymbol{\beta}_0 + u_i$, $i = 1, \ldots, n$ with an *i.i.d.* zero-mean finite-variance error u_i. Hence, Lee's (2007) GMM estimator can be used to estimate ρ_0 and $\boldsymbol{\beta}_0$. Further, the test does not require $\boldsymbol{\beta}_0 \neq \mathbf{0}_p$ so long as one can obtain a consistent estimator for ρ_0 and $\boldsymbol{\beta}_0$ under H_0. Since our test statistic involves the construction of an estimator of $\left[\rho, \boldsymbol{\beta}'\right]'$, the following assumption is required to regulate the limiting performance of the estimator under the alternative hypothesis.

Assumption 10 (i) Under H_1, there exists $\boldsymbol{\gamma} = \left[\rho, \boldsymbol{\beta}'\right]'$ such that $\breve{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = O_p\left(n^{-1/2}\right)$; (ii) $\sup_{z \in S_z} \mathbb{E}\left(u_i^6 | \mathbf{z}_i = z\right) \leq M$; (iii) $\boldsymbol{\beta}(z) \neq \mathbf{0}_p$ holds over at least one non-empty subset.

With the properly chosen instrumental variables as discussed in Lee (2007), following his proof, one can show that Assumption 10(i) holds. Assumption 10(ii) is required to calculate the stochastic order of the test statistic under H_1. We make Assumption 10(iii) because our test statistic J_n given below has the same limiting distribution under both the null and alternative hypotheses if $\boldsymbol{\beta}(z) = \mathbf{0}_p$. Below we give the limiting distribution of the test statistic and relegate the proof to Appendix B.

Theorem 7 *Under Assumptions **1–4** and **10** , we have, under* H_0,

$$J_n = n\sqrt{|\mathbf{H}|}T_n / \sqrt{\widehat{\sigma}_n^2} \overset{d}{\to} \mathbb{N}(0,1),$$

where

$$\widehat{\sigma}_n^2 = \frac{2}{n^2|\mathbf{H}|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \widehat{u}_i^2 \widehat{u}_j^2 \mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j) \overset{p}{\to} 2\sigma_u^4 R_2(K) \mathbb{E}[f(\mathbf{z})],$$

and under H_1, $\Pr\{|J_n| \geq M_n\} \to 0$ *for any non-stochastic, positive sequence* $M_n = o\left(n\sqrt{|\mathbf{H}|}\right)$.

Theorem 7 indicates that J_n is a consistent test. Since the test statistic is based on the fact that $E\left[\varepsilon_i \mathbb{E}(\varepsilon_i | \mathbf{z}_i)\right]$ equals zero under H_0 and takes a positive value under H_1, where $\varepsilon_i = y_i - \rho \sum_{j \neq i} w_{ij} y_j - \mathbf{x}_i' \boldsymbol{\beta}$ for $i = 1, \ldots, n$, the proposed test is a one-sided test. That is, we reject the null hypothesis if the test statistic J_n is greater than c_α, where c_α is the upper 100α percentile of a standard normal distribution for a given $\alpha \in (0,1)$.

However, it is well-known that the residual-based nonparametric tests perform rather poorly in finite samples leading to a widely popular use of bootstrap methods in order to improve their finite sample performance. Sharing this sentiment, below we propose a bootstrap procedure for our test statistic J_n:

Step 1. Estimate the linear model under the null, i.e., $y_i = \rho \sum_{j \neq i} w_{ij} y_j + \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i$, via Lee's (2007) efficient GMM estimator to obtain residuals $\widehat{u}_i = y_i - \breve{\rho} \sum_{j \neq i} w_{ij} y_j - \mathbf{x}_i' \breve{\boldsymbol{\beta}}$ for all $i = 1, \ldots, n$.

Step 2. Obtain two-point wild bootstrap errors by setting $u_i^* = a\widehat{u}_i$ with probability r and $u_i^* = b\widehat{u}_i$ with probability $1-r$, where $a = \left(1 - \sqrt{5}\right)/2$, $b = \left(1 + \sqrt{5}\right)/2$ and $r = \left(1 + \sqrt{5}\right) / \left(2\sqrt{5}\right)$. Then, compute $\mathbf{y}^* = \left[\mathbf{I}_n - \breve{\rho}\mathbf{W}\right]^{-1}\left(\mathbf{X}\breve{\boldsymbol{\beta}} + \mathbf{u}^*\right)$ and call $\{(\mathbf{x}_i, \mathbf{z}_i, \mathbf{y}_i^*)\}_{i=1}^n$ the bootstrap sample.

Step 3. Reestimate the model under the null via efficient GMM estimator using the constructed bootstrap sample, i.e., $y_i^* = \rho \sum_{j \neq i} w_{ij} y_j^* + \mathbf{x}_i' \boldsymbol{\beta} + u_i^*$ to obtain bootstrap residuals $\widehat{u}_i^* = y_i^* - \breve{\rho}^* \sum_{j \neq i} w_{ij} y_j^* - \mathbf{x}_i' \breve{\boldsymbol{\beta}}^*$ for all $i = 1, \ldots, n$.

Step 4. Compute the bootstrap test statistic $J_n^* = n\sqrt{|\mathbf{H}|}T_n^* / \sqrt{\widehat{\sigma}_n^{*2}}$, where $T_n^* = \left(n^2|\mathbf{H}|\right)^{-1} \sum_{i=1}^n \sum_{j \neq i}^n \widehat{u}_i^* \widehat{u}_j^* \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j)$ and $\widehat{\sigma}_n^{*2} = 2\left(n^2|\mathbf{H}|\right)^{-1} \sum_{i=1}^n \sum_{j \neq i}^n \widehat{u}_i^{*2} \widehat{u}_j^{*2} \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j)$.

Step 5. Repeat steps 1–4 B times.

Step 6. Use the empirical distribution of $B+1$ bootstrap statistics, where the first bootstrap test statistic equals the test statistic calculated from the raw data, to obtain the upper 100α percentile value c_α^*. Use this c_α^* to approximate the upper percentile value of the test statistic J_n under H_0.

Theorem 8 *Under Assumptions **1–4** and **10**, we have*

$$\sup_{s\in\mathbb{R}} \left|\Pr^*\left(J_n^* \leq s\right) - \Phi\left(s\right)\right| = o_p\left(1\right), \tag{5.2}$$

where $\Pr^*(\cdot) = \Pr\left(\cdot\middle|\left\{(\mathbf{x}_i, \mathbf{z}_i, \mathbf{y}_i^*)\right\}_{i=1}^n\right)$, *and* $\Phi(\cdot)$ *is the standard normal cumulative distribution function.*

Theorem 8 shows that the proposed bootstrap method is asymptotically valid in approximating the null distribution of J_n regardless of whether the null hypothesis holds true or not. Specifically, the result in (5.2) means that the bootstrap test statistic J_n^* converges to a standard normal random variable in distribution in probability.

As discussed earlier, our test statistic J_n has its limitations when Assumption 10(iii) is violated. In what follows, we therefore propose an alternative test statistic à la Henderson et al. (2008) which provides a vehicle for discriminating between H_0 and H_1 even when $\boldsymbol{\beta}(z) = \mathbf{0}_p$. The new test statistic is defined as

$$D_n = \frac{1}{n}\sum_{i=1}^n \left[\mathbf{m}_i'\breve{\boldsymbol{\gamma}} - \mathbf{m}_i'\widehat{\boldsymbol{\gamma}}\left(\mathbf{z}_i\right)\right]^2, \tag{5.3}$$

where $\widehat{\boldsymbol{\gamma}}(\mathbf{z}_i)$ is estimated via local constant regression approach following the methodology we propose in Section 3. The below theorem gives the limit property of D_n.

Theorem 9 *Under Assumptions **1–5** and **10(i)**, we have:* $D_n \to 0$ *under* H_0, *and* $D_n = 2n^{-1}\sum_{i=1}^n \left[\mathbf{m}_i'\left(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\right)\right]^2 + o_p(1) = O_e(1)$ *under* H_1.

Theorem 9 states that D_n is a consistent, one-sided test. It is reasonable to expect the proposed test statistic to be asymptotically normal after some proper scaling but we are unable to derive its limit distribution due to the complexity of our estimator $\widehat{\boldsymbol{\gamma}}(\cdot)$. We therefore suggest employing a bootstrap procedure to approximate the finite sample null distribution of D_n. For details on the appropriate bootstrap procedure, see Henderson et al. (2008).

6 Finite Sample Performance

We first study the finite sample performance of our proposed estimators and test statistics in a small set of Monte Carlo simulations. We then showcase our methodology by applying it to estimate a spatial hedonic price function using the well-known Harrison & Rubenfeld's house price data from Gilley & Pace (1996), where we let unknown parameter functions to vary with the $\mathrm{NO_x}$ concentration in the air. To conserve space, we relegate the discussion of empirical application to Supplementary Appendix and, in what follow, we only report the results of Monte Carlo simulations.

6.1 Estimators

We generate the data using the following process:

$$y_i = \rho(z_i)\sum_{j\neq i} w_{ij}y_j + x_i\beta(z_i) + u_i \quad \forall \quad i = 1, \ldots, n, \tag{6.1}$$

where the variables are randomly drawn as follows: $z_i \sim i.i.d.\ \mathbb{U}(0,1)$, $\xi_i \sim i.i.d.\ \mathbb{N}(0,1)$ and $u_i \sim i.i.d.\ \mathbb{N}(0,0.5)$; all are mutually independent. Variables x_i and z_i correlate as follows: $x_i = 0.5z_i + \xi_i$. As in Lee (2007) and Liu, Lee & Bollinger (2010), rather than generating $\{w_{ij}\}$, we instead use the spatial weighting matrix from the crime study for 49 districts in Columbus, OH from Anselin (1988). The spatial weighting matrix is contiguity-based and uses the (first-order) queen definition for Columbus and corresponds to a sample of $n = 49$. To increase the sample size, we generate a block-diagonal spatial matrix with the original 49×49 Columbus matrix used as a diagonal block.

We consider sample sizes $n = \{98,\ 245,\ 490\}$. For each n, we simulate the model 500 times. We use the same Silverman's (1986) rule-of-thumb bandwidth for the smoothing variable z_i in all stages, i.e., $h = h_0 = 1.06 \times \widehat{\sigma}_z n^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of z_i. For each simulation, we compute the root mean squared error (RMSE) and the mean absolute error (MAE) for each coefficient function and report their averages computed over 500 simulations.

In the first stage, we use the following feasible instruments: $\mathbf{Q}_{1n} = (\mathbf{Wx}, \mathbf{Wz})$, where $\mathbf{x} = (x_1, \ldots, x_n)'$ and $\mathbf{z} = (z_1, \ldots, z_n)'$, for linear moment conditions and $\mathbf{P}_{n,l} = \mathbf{W}^l - n^{-1}\mathrm{tr}\left\{\mathbf{W}^l\right\}\mathbf{I}_n$ for $l = 1, 2$ for quadratic moments. For the second-stage estimation, we set $\widehat{\mathbf{Q}}_{1n} = \widehat{\mathbf{G}}_n(\mathbf{z})\,\mathrm{mtx}\{\mathbf{x}, \widehat{\boldsymbol{\beta}}(\mathbf{z})\}$ and $\widehat{\mathbf{P}}_n = \widehat{\mathbf{G}}_n(\mathbf{z}) - n^{-1}\mathrm{tr}\{\widehat{\mathbf{G}}_n(\mathbf{z})\}\mathbf{I}_n$, where $\widehat{\mathbf{G}}_n(\mathbf{z}) = \mathbf{W}\left[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{z})\mathbf{W}\right]^{-1}$; all constructed using the first-stage estimates of $\rho(z_i)$ and $\beta(z_i)$.[7]

We begin by first considering the case when x_i is a *relevant* variable, i.e., $\beta(z_i) \neq 0$. In particular, the coefficient functions are specified as follows: $\beta(z_i) = 1 - z_i^2$ and $\rho(z_i) = 0.75 \times \sin(\pi z_i)$. Here, we use a local linear fitting. Table 1 reports the corresponding results for both the first- and second-stage nonparametric GMM estimators fitted using two sets of orthogonality conditions: *(i)* linear and quadratic moments (left panel) and *(ii)* linear moments only (right panel).

For each of the two stages, we also report the results for the local linear least squares estimator of $\beta(z_i)$ from the equation constructed by moving the endogenous spatial lag term $\rho(z_i)\sum_{j\neq i} w_{ij}y_j$ to the left-hand side of model (2.1) and replacing unknown $\rho(z_i)$ with its estimator $\rho^*(z_i)$, where the latter equals $\widehat{\rho}(z_i)$ in the first stage and $\widetilde{\rho}(z_i)$ in the second stage. It is attractive to explore such a model for likely efficiency gains in finite samples. More concretely, we apply the conventional local linear least squares approach to $[\mathbf{I}_n - \boldsymbol{\rho}^*(\mathbf{Z})\mathbf{W}]\,\mathbf{y} = \mathrm{mtx}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\} + \mathbf{u}$ to obtain, in the local neighborhood of z:

$$\mathbf{y}^* \approx \mathcal{X}(z)\mathrm{vec}\{\boldsymbol{B}(z)\} + \mathbf{u}^*, \tag{6.2}$$

where $\mathbf{y}^* \equiv [\mathbf{I}_n - \boldsymbol{\rho}^*(\mathbf{Z})\mathbf{W}]\,\mathbf{y}$; $\mathcal{X}(z) = \begin{bmatrix}\mathcal{Z}_1(z)\otimes\mathbf{x}_1 & \ldots & \mathcal{Z}_n(z)\otimes\mathbf{x}_n\end{bmatrix}'$ is an $n \times [p(q+1)]$ data matrix; and $\boldsymbol{B}(z) \equiv \begin{bmatrix}\boldsymbol{\beta}(z) & \begin{bmatrix}\nabla\boldsymbol{\beta}_1(z) & \ldots & \nabla\boldsymbol{\beta}_p(z)\end{bmatrix}'\end{bmatrix}$. It is easy to show that the resulting nonparametric least squares estimator is given by

$$\mathrm{vec}\left\{\breve{\boldsymbol{B}}(z)\right\} = \left[\mathcal{X}(z)'\mathbf{K}_H(z)\mathcal{X}(z)\right]^{-1}\mathcal{X}(z)'\mathbf{K}_H(z)\mathbf{y}^*. \tag{6.3}$$

The results in Table 1 indicate that, in all instances, the estimation of both $\rho(\cdot)$ and $\beta(\cdot)$ coefficients becomes more stable as the sample size increases. Both the RMSE and MAE decline significantly as n increases. Regardless of the instrument set, as expected, the second-stage estimator delivers a sizable improvement over its first-stage counterpart, with a greater impact exhibited for the estimation of $\rho(\cdot)$. We also observe that adding quadratic orthogonality conditions leads to an increase in accuracy of the first-stage estimator only. Similarly, the reestimation of $\beta(\cdot)$ via least squares also yields better results in the first stage only.

[7] Throughout, to ensure the non-singularity of $\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z})\mathbf{W}$, we impose (3.11) via either box constraints on $\rho(z)$ during the numerical optimization when employing quadratic moments or the post-estimation "tilting" procedure when making use of linear moments only.

Table 1. Simulation Results for the Estimators when x_i is Relevant

	Linear & Quadratic Moments			*Linear Moments Only*		
	$n = 98$	$n = 245$	$n = 490$	$n = 98$	$n = 245$	$n = 490$
		First Stage			**First Stage**	
$\rho(z_i)$						
RMSE	0.2335	0.1451	0.1067	0.2786	0.2009	0.1638
MAE	0.1844	0.1146	0.0846	0.2096	0.1418	0.1081
$\beta(z_i)$						
RMSE	0.1163	0.0764	0.0564	0.1205	0.0821	0.0598
MAE	0.0884	0.0582	0.0425	0.0917	0.0612	0.0443
$\beta(z_i)$ via LS						
RMSE	0.0992	0.0641	0.0477	0.1025	0.0681	0.0509
MAE	0.0791	0.0515	0.0383	0.0811	0.0538	0.0401
		Second Stage			**Second Stage**	
$\rho(z_i)$						
RMSE	0.2078	0.1299	0.0959	0.2011	0.1203	0.0878
MAE	0.1665	0.1034	0.0761	0.1586	0.0960	0.0699
$\beta(z_i)$						
RMSE	0.1010	0.0660	0.0483	0.0994	0.0643	0.0474
MAE	0.0802	0.0528	0.0388	0.0798	0.0516	0.0381
$\beta(z_i)$ via LS						
RMSE	0.0974	0.0634	0.0470	0.0975	0.0632	0.0468
MAE	0.0777	0.0508	0.0377	0.0774	0.0507	0.0375

Notes: The reported are the averages of respective statistics over 500 simulations. LS stands for least squares.

Next, we turn to the case when x_i is *irrelevant* in explaining y_i, i.e., $\beta(z_i) = 0$. The coefficient functions are specified as follows: $\beta(z_i) = 0$ for all z_i and $\rho(z_i) = 0.75 \times \sin(\pi z_i)$. In this instance, we estimate our first- and second-stage nonparametric GMM estimators via a local constant approach[8] using the following two sets of orthogonality conditions: *(i)* linear and quadratic moments (left panel) and *(ii)* quadratic moments only (right panel). The first set of instruments is meant to simulate the case when the researcher is not aware that $\beta(z_i) = 0$, whereas the second set assumes the researcher knows that the model is a pure spatial autoregressive model with no covariates (and hence, only $\rho(\cdot)$ is estimated).

Table 2 presents the results. Consistent with our theory, all estimators improve with an increase in the sample size. The second-stage estimator offers some gains in the estimation of the $\rho(\cdot)$ coefficient function primarily only when the model incorrectly presumes that $\beta(z) \neq 0$; no gains are exhibited for $\beta(\cdot)$ in this case (left panel of Table 2). In line with one's intuition, the results indicate that the estimator of a correctly specified pure spatial autoregressive model (right panel of Table 2) outperforms that of a "misspecified" model which includes an invalid instrument $\mathbf{Wx}$ in its instrument set.

Overall, simulation experiments lend support to asymptotic results for our proposed estimators.

[8] Consistent with Remark 2.

Table 2. Simulation Results for the Estimators when x_i is Irrelevant

	Linear & Quadratic Moments			*Quadratic Moments Only*		
	$n = 98$	$n = 245$	$n = 490$	$n = 98$	$n = 245$	$n = 490$
	First Stage			**First Stage**		
$\rho(z_i)$						
RMSE	0.2451	0.1841	0.1592	0.2094	0.1751	0.1587
MAE	0.1979	0.1495	0.1320	0.1735	0.1458	0.1335
$\beta(z_i)$						
RMSE	0.0761	0.0524	0.0395			
MAE	0.0638	0.0436	0.0331			
	Second Stage			**Second Stage**		
$\rho(z_i)$						
RMSE	0.2214	0.1759	0.1548	0.2076	0.1737	0.1573
MAE	0.1810	0.1457	0.1324	0.1720	0.1445	0.1321
$\beta(z_i)$						
RMSE	0.0767	0.0524	0.0395			
MAE	0.0645	0.0437	0.0331			

Note: The reported are the averages of respective statistics over 500 simulations.

6.2 Specification Tests

We next examine the small sample performance of our proposed specification tests. As earlier, we consider sample sizes $n = \{98,\ 245,\ 490\}$. For each n, we simulate the model 500 times. Test statistics are bootstrapped 299 times each to obtain the 1%, 5%, 10% and 20% upper percentile (critical) values of their null distributions.

We first study our residual-based statistic J_n. To assess its size and power, we consider the following two experimental designs for the data-generating process given in (6.1):

(i) Linear model: $\beta(z_i) = 0.75$ and $\rho(z_i) = 0.5$ for all z_i.

(ii) Nonlinear model: $\beta(z_i) = 1 - z_i^2$ and $\rho(z_i) = 0.75 \times \sin(\pi z_i)$.

The residuals under the null hypothesis necessary for the construction of J_n are estimated via Lee's (2007) GMM estimator using the same linear and quadratic instruments as the ones we use in the first-stage estimation in Section 6.1 above. To assess the sensitive of the results to the choice of bandwidth for z_i, we try different values of the scale parameter in the Silverman's (1986) rule-of-thumb bandwidth: $c = \{0.80,\ 1.06,\ 1.50\}$. Table 3 reports the estimated size [design (i)] and power [design (ii)] of our test computed as rejection frequencies over 500 simulations. We find that out test statistic J_n exhibits a relatively good size across all considered sample sizes and bandwidth values. From the right panel of Table 3, we also see that the power of the test increases with the sample size as anticipated.

Given that the applicability of the J_n statistic does not extend to the case of pure spatial autoregressive models, we next analyze the performance of our second test statistic D_n. To examine its size and power, we consider the following four experimental designs for the data-generating process given in (6.1):

(i) Linear model:

  – Relevant x_i: $\beta(z_i) = 0.75$ and $\rho(z_i) = 0.5$ for all z_i;

Table 3. Simulation Results for the J_n Statistic when x_i is Relevant

	Estimated Size			*Estimated Power*		
Sign. Level	$n = 98$	$n = 245$	$n = 490$	$n = 98$	$n = 245$	$n = 490$
$\mathbf{c = 0.8}$						
1%	0.012	0.006	0.008	0.164	0.548	0.868
5%	0.034	0.080	0.054	0.359	0.756	0.954
10%	0.073	0.134	0.100	0.485	0.813	0.988
20%	0.153	0.247	0.218	0.615	0.898	0.996
$\mathbf{c = 1.06}$						
1%	0.010	0.012	0.006	0.171	0.569	0.870
5%	0.034	0.075	0.054	0.376	0.770	0.966
10%	0.071	0.126	0.092	0.493	0.827	0.986
20%	0.153	0.260	0.192	0.630	0.894	1.00
$\mathbf{c = 1.5}$						
1%	0.010	0.012	0.006	0.141	0.556	0.878
5%	0.042	0.070	0.048	0.363	0.780	0.970
10%	0.080	0.116	0.086	0.510	0.839	0.988
20%	0.158	0.234	0.198	0.640	0.902	1.00

Note: The reported are the rejection frequencies over 500 simulations.

– Irrelevant x_i: $\beta(z_i) = 0$ for all z_i and $\rho(z_i) = 0.5$ for all z_i.

(ii) Nonlinear model:

– Relevant x_i: $\beta(z_i) = 1 - z_i^2$ and $\rho(z_i) = 0.75 \times \sin(\pi z_i)$;

– Irrelevant x_i: $\beta(z_i) = 0$ for all z_i and $\rho(z_i) = 0.75 \times \sin(\pi z_i)$.

We use Silverman's (1986) rule-of-thumb bandwidth with $c = 1.06$ throughout. For the case of a relevant x_i, both the model under H_0 and the model under H_1 are estimated using linear and quadratic moments. However, in the case of a pure spatial autoregressive model, we make use of quadratic moments only thus assuming that the irrelevancy of x_i is an *a priori* knowledge. The results reported in Table 4 show that the D_n test has quite accurate size across all n regardless whether x_i is relevant or not. It exhibits superb power when x_i is relevant in predicting y_i. The power is also decent and rises with the sample size when the true model is a pure spatial autoregression.

7 Conclusion

Most empirical work that deals with spatial data employs standard linear spatial models. These models are however prone to misspecification due to a rather strong assumption of linearity of the spatial relationship. The literature has long ago recognized that econometric relations in space result more often than not in highly nonlinear specifications.

This paper offers a semiparametric method to handling nonlinearity (and parameter heterogeneity) in models of spatial dependence. Specifically, we consider a semiparametric spatial autoregressive (mixed-regressive) model in which unknown coefficients are permitted to be nonparametric functions of some contextual variables to allow for potential nonlinearities and parameter heterogeneity in the spatial relationship. Unlike other semiparametric spatial dependence models, ours permits the spatial autoregressive parameter to meaningfully vary across units. The latter

Table 4. Simulation Results for the D_n Statistic

	Estimated Size			*Estimated Power*		
Sign. Level	$n = 98$	$n = 245$	$n = 490$	$n = 98$	$n = 245$	$n = 490$
Relevant x_i						
1%	0.008	0.012	0.006	0.932	1.00	1.00
5%	0.043	0.053	0.056	0.997	1.00	1.00
10%	0.101	0.117	0.112	1.00	1.00	1.00
20%	0.224	0.207	0.205	1.00	1.00	1.00
Irrelevant x_i **(Pure SAR Model)**						
1%	0.010	0.006	0.014	0.035	0.149	0.392
5%	0.061	0.064	0.064	0.158	0.382	0.652
10%	0.130	0.116	0.132	0.281	0.500	0.768
20%	0.245	0.235	0.235	0.440	0.672	0.892

Notes: The reported are the rejection frequencies over 500 simulations. SAR stands for spatially autoregressive.

may be highly desirable from a practitioner's point of view since it allows the identification of a neighborhood-specific spatial dependence measure conditional on the vector of contextual variables.

We propose several (locally) nonparametric GMM estimators for our model. The developed two-stage estimators incorporate both the linear and quadratic orthogonality conditions and are capable of accommodating a variety of data generating processes, including the instance of a pure spatially autoregressive semiparametric model with no relevant regressors as well as multiple partially linear specifications. All proposed estimators are shown to be consistent and asymptotically normal. We also contribute to the literature by putting forward two test statistics to test for parameter constancy in our model. Both tests are consistent.

Appendix

To simplify notation, we define the following: $\pi_i \equiv \mathcal{K}_H(\mathbf{z}_i, z) = \mathcal{K}(\mathbf{H}^{-1}(\mathbf{z}_i - z))$; $\mathbf{A}_n^s = \mathbf{A}_n + \mathbf{A}_n'$ for any $n \times n$ matrix $\mathbf{A}_n$; $\bigtriangledown^2 g(z) = \partial^2 g(z)/\partial z \partial z'$ is the second-order derivative of a differentiable function $g : \mathbb{R}^q \to \mathbb{R}$; $\mathbf{i}_s$ and $\mathbf{0}_s$ is an $s \times 1$ vector of ones and zeros, respectively; $\mathbf{i}_{s\times t}$ and $\mathbf{0}_{s\times t}$ is an $s \times t$ matrix of ones and zeros, respectively; $\sum_{i\neq j} \equiv \sum_{i=1}^{n}\sum_{j\neq i}^{n}$ and $\sum_{i\neq j\neq i'} \equiv \sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{j\neq i\neq i'}^{n}$. Also, $X_n = O_e(a_n)$ means that $X_n = O_p(a_n)$ but not $X_n = o_p(a_n)$, and $A_n \approx B_n$ indicates that B_n is the leading term of A_n.

A Brief Mathematical Proofs of Theorems 1–4

Proof of Theorem 1. Define $\boldsymbol{\theta}_n = \xi_n \left[\ \widehat{\boldsymbol{\gamma}}(z) - \boldsymbol{\gamma}(z) \quad [\bigtriangledown\widehat{\boldsymbol{\gamma}}(z) - \bigtriangledown\boldsymbol{\gamma}(z)]\,\mathbf{H}\ \right]$, a $(p+1)\times(q+1)$ matrix, $y_i^* = y_i - \mathbf{m}_i'\boldsymbol{\gamma}(z) - \mathbf{m}_i' \bigtriangledown \boldsymbol{\gamma}(z)(\mathbf{z}_i - z)$ and $u_i(\boldsymbol{\theta}) = y_i^* - \xi_n^{-1}\mathbf{m}_i'\boldsymbol{\theta}\mathcal{Z}_i(z)$, where $\{\xi_n\}$ is a sequence of positive constants such that $0 < M_1 < \|\boldsymbol{\theta}_n\| < M_2 < \infty$ for all n. Then, we can rewrite (3.5) as

$$\mathbf{g}_n(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{u}(\boldsymbol{\theta})'\,\mathbf{P}_{n,1}\mathbf{K}_H(z)\mathbf{u}(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{u}(\boldsymbol{\theta})'\,\mathbf{P}_{n,m}\mathbf{K}_H(z)\mathbf{u}(\boldsymbol{\theta}) \\ \mathbf{Q}(z)'\mathbf{K}_H(z)\mathbf{u}(\boldsymbol{\theta}) \end{bmatrix}, \tag{A.1}$$

where $\mathbf{u}(\boldsymbol{\theta})$ is an $n \times 1$ vector with a typical element being equal to $u_i(\boldsymbol{\theta})$. Then, we have

$$\frac{\partial \mathbf{g}_n(\boldsymbol{\theta})}{\partial \text{vec}(\boldsymbol{\theta})'} = -\xi_n^{-1} \begin{bmatrix} \mathbf{u}(\boldsymbol{\theta})' [\mathbf{P}_{n,1}\mathbf{K}_H(z)]^s \mathbf{M}(z) \\ \vdots \\ \mathbf{u}(\boldsymbol{\theta})' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{M}(z) \\ \mathbf{Q}(z)'\mathbf{K}_H(z)\mathbf{M}(z) \end{bmatrix}.$$

Minimizing the GMM objective function referenced in (3.6) is equivalent to minimizing $\Lambda_n(\boldsymbol{\theta}) = \mathbf{g}_n(\boldsymbol{\theta})' \mathbf{g}_n(\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \mathbb{S}$, where the latter is a compact subset of $\mathbb{R}^{p+1} \times \mathbb{R}^{q+1}$. Since $\boldsymbol{\theta}_n$ minimizes $\Lambda_n(\boldsymbol{\theta}) = \mathbf{g}_n(\boldsymbol{\theta})' \mathbf{g}_n(\boldsymbol{\theta})$, we have that

$$\mathbf{0}_{(p+1)(q+1)} = \frac{\partial \mathbf{g}_n(\boldsymbol{\theta}_n)'}{\partial \text{vec}(\boldsymbol{\theta})} \mathbf{g}_n(\boldsymbol{\theta}_n) = \frac{\partial \mathbf{g}_n(\boldsymbol{\theta}_n)'}{\partial \text{vec}(\boldsymbol{\theta})} \left[\mathbf{g}_n(\mathbf{0}) + \frac{\partial \mathbf{g}_n\left(\widetilde{\boldsymbol{\theta}}_n\right)}{\partial \text{vec}(\boldsymbol{\theta})'} \text{vec}(\boldsymbol{\theta}_n) \right],$$

where $\widetilde{\boldsymbol{\theta}}_n$ lies between $\boldsymbol{\theta}_n$ and $\mathbf{0}_{(p+1)(q+1)}$. From above, we obtain

$$\text{vec}(\boldsymbol{\theta}_n) = -\left[\frac{\partial \mathbf{g}_n(\boldsymbol{\theta}_n)'}{\partial \text{vec}(\boldsymbol{\theta})} \frac{\partial \mathbf{g}_n\left(\widetilde{\boldsymbol{\theta}}_n\right)}{\partial \text{vec}(\boldsymbol{\theta})'} \right]^{-1} \frac{\partial \mathbf{g}_n(\boldsymbol{\theta}_n)'}{\partial \text{vec}(\boldsymbol{\theta})} \mathbf{g}_n(\mathbf{0}).$$

Denoting $\boldsymbol{\Xi}_H(z) \equiv \mathbf{K}_H(z)\mathbf{Q}(z)\mathbf{Q}(z)'\mathbf{K}_H(z)$, we decompose the two components of $\text{vec}(\boldsymbol{\theta}_n)$ above as follows

$$\begin{aligned} A_n(z) &\equiv -\frac{\partial \mathbf{g}_n(\boldsymbol{\theta}_n)'}{\partial \text{vec}(\boldsymbol{\theta})} \mathbf{g}_n(\mathbf{0}) \\ &= \frac{1}{2\xi_n} \sum_{l=1}^{m} \mathbf{M}(z)' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{u}(\boldsymbol{\theta}_n) \mathbf{y}^{*\prime} [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{y}^* + \frac{1}{\xi_n} \mathbf{M}(z)'\boldsymbol{\Xi}_H(z)\mathbf{y}^* \end{aligned}$$

and

$$\begin{aligned} B_n(z) &\equiv \frac{\partial \mathbf{g}_n(\boldsymbol{\theta}_n)'}{\partial \text{vec}(\boldsymbol{\theta})} \frac{\partial \mathbf{g}_n\left(\widetilde{\boldsymbol{\theta}}_n\right)}{\partial \text{vec}(\boldsymbol{\theta})'} \\ &= \frac{1}{\xi_n^2} \sum_{l=1}^{m} \left[\mathbf{u}(\boldsymbol{\theta}_n)' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{M}(z) \right]' \mathbf{u}\left(\widetilde{\boldsymbol{\theta}}_n\right)' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{M}(z) + \frac{1}{\xi_n^2} \mathbf{M}(z)'\boldsymbol{\Xi}_H(z)\mathbf{M}(z). \end{aligned}$$

For each i, we define a $(p+1) \times 1$ vector $\boldsymbol{\Pi}(\mathbf{z}_i^*)$ whose lth element equals $\Pi_l(\mathbf{z}_i^*) = (\mathbf{z}_i - z)' \bigtriangledown^2 \gamma_l(\mathbf{z}_i^*)(\mathbf{z}_i - z)$, and $\mathbf{z}_i^*$ lies between $\mathbf{z}_i$ and z for $l = 1, \ldots, p+1$. Then, we have $y_i^* = u_i + \mathbf{m}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)/2$. Further, we define an $n \times 1$ vector $\mathbf{C}(z)$, whose ith term equals $\mathbf{m}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)/2$, along with $\Gamma_{1,l} = \mathbf{u}' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{C}(z)$, $\Gamma_{2,l} = \mathbf{C}(z)' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{C}(z)$, $\Gamma_{3,l} = \mathbf{u}' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{u}$, $\Psi_{1,l} = \mathbf{u}' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{M}(z)$, $\Psi_{2,l} = \mathbf{C}(z)' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{M}(z)$ and $\Psi_{3,l} = \mathbf{M}(z)' [\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s \mathbf{M}(z)$ for $l = 1, \ldots, m$. Then, we have $A_n(z) = A_{n1}(z) + A_{n2}(z) - A_{n3}(z)$ with

$$\begin{aligned} A_{n1}(z) &= \frac{1}{2\xi_n} \sum_{l=1}^{m} \left[\Gamma_{2,l} (\Psi_{1,l} + \Psi_{2,l})' + (2\Gamma_{1,l} + \Gamma_{3,l}) \Psi_{1,l}' \right] + \frac{1}{\xi_n} \mathbf{M}(z)' \boldsymbol{\Xi}_H(z)\mathbf{C}(z) \\ A_{n2}(z) &= \frac{1}{2\xi_n} \sum_{l=1}^{m} (2\Gamma_{1,l} + \Gamma_{3,l}) \Psi_{2,l}' + \frac{1}{\xi_n} \mathbf{M}(z)'\boldsymbol{\Xi}_H(z)\mathbf{u} \end{aligned}$$

$$A_{n3}(z) = \frac{1}{2\xi_n^2}\sum_{l=1}^{m}(2\Gamma_{1,l}+\Gamma_{2,l}+\Gamma_{3,l})\Psi_{3,l}\text{vec}(\boldsymbol{\theta}_n),$$

and

$$B_n(z) = \frac{1}{\xi_n^2}\sum_{l=1}^{m}\left(\Psi_{1,l}+\Psi_{2,l}-\frac{1}{\xi_n}\left[\text{vec}\left(\widetilde{\boldsymbol{\theta}}_n\right)\right]'\Psi_{3,l}\right)'\left(\Psi_{1,l}+\Psi_{2,l}-\frac{1}{\xi_n}\left[\text{vec}\left(\widetilde{\boldsymbol{\theta}}_n\right)\right]'\Psi_{3,l}\right)+ \frac{1}{\xi_n^2}\mathbf{M}(z)'\boldsymbol{\Xi}_H(z)\mathbf{M}(z).$$

By Lemmas 1–3, we have

$$(n|\mathbf{H}|)^{-2}\xi_n A_{n1}(z) = f^2(z)\varkappa_A(\mathbf{H},z)'+o_p\left(\|\mathbf{H}\|^2\right) \tag{A.2}$$
$$(n|\mathbf{H}|)^{-3/2}\xi_n A_{n2}(z) \xrightarrow{d} \mathbb{N}\left(\mathbf{0}_{[(p+1)(q+1)]},\ f^3(z)R_2(K)\boldsymbol{\Omega}(z)\right)$$

and

$$(n|\mathbf{H}|)^{-2}\xi_n^2 A_{n3}(z) = O_p\left(\left(\|\mathbf{H}\|^2+(n|\mathbf{H}|)^{-1/2}\right)\|\mathbf{H}\|^{-1}\right)$$
$$(n|\mathbf{H}|)^{-2}\xi_n^2 B_n(z) = f^2(z)\varkappa_B(\mathbf{H},z)+O_p\left((\xi_n\|\mathbf{H}\|)^{-1}+(\xi_n\|\mathbf{H}\|)^{-2}\right)+o_p(1) \tag{A.3}$$

uniformly over all $\widetilde{\boldsymbol{\theta}}_n\in\mathbb{S}$. Therefore, we have

$$\begin{aligned}&\xi_n\left(\left[\begin{array}{cc}\widehat{\boldsymbol{\gamma}}(z)-\boldsymbol{\gamma}(z) & [\nabla\widehat{\boldsymbol{\gamma}}(z)-\nabla\boldsymbol{\gamma}(z)]\mathbf{H}\end{array}\right]-\left[(n|\mathbf{H}|)^{-2}\xi_n^2B_n(z)\right]^{-1}(n|\mathbf{H}|)^{-2}\xi_nA_{n1}(z)\right)\\ =\ & \frac{\xi_n}{\sqrt{n|\mathbf{H}|}}\left[\frac{\xi_n^2B_n(z)}{(n|\mathbf{H}|)^2}\right]^{-1}\frac{\xi_nA_{n2}(z)}{(n|\mathbf{H}|)^{3/2}}+\left[\frac{\xi_n^2B_n(z)}{(n|\mathbf{H}|)^2}\right]^{-1}\frac{\xi_n^2A_{n3}(z)}{(n|\mathbf{H}|)^2}.\end{aligned} \tag{A.4}$$

Combining (A.2)–(A.4) with the fact that $\|\boldsymbol{\theta}_n\|<M$ for all n, we can deduce that ξ_n must be of order $\sqrt{n|\mathbf{H}|}$. The logic is as follows.

First, if $\xi_n/\sqrt{n|\mathbf{H}|}\to\infty$, it implies that $\sqrt{n|\mathbf{H}|}\left[\begin{array}{cc}\widehat{\boldsymbol{\gamma}}(z)-\boldsymbol{\gamma}(z) & [\nabla\widehat{\boldsymbol{\gamma}}(z)-\nabla\boldsymbol{\gamma}(z)]\mathbf{H}\end{array}\right]\to 0$ as $n\to\infty$. By (A.2)–(A.4) and Assumption 4, we obtain

$$\begin{aligned}&\sqrt{n|\mathbf{H}|}\left(\left[\begin{array}{cc}\widehat{\boldsymbol{\gamma}}(z)-\boldsymbol{\gamma}(z) & [\nabla\widehat{\boldsymbol{\gamma}}(z)-\nabla\boldsymbol{\gamma}(z)]\mathbf{H}\end{array}\right]-\varkappa_B(\mathbf{H},z)^{-1}\varkappa_A(\mathbf{H},z)'\right)\\ \overset{d}{=}\ & \varkappa_B(\mathbf{H},z)^{-1}\mathbb{N}\left(\mathbf{0}_{(p+1)(q+1)},\ f^3(z)R_2(K)\boldsymbol{\Omega}(z)\right)+O_p\left(\frac{\sqrt{n|\mathbf{H}|}}{\xi_n}\left(\|\mathbf{H}\|^2+(n|\mathbf{H}|)^{-1/2}\right)\|\mathbf{H}\|^{-1}\right),\end{aligned}$$

where $A_n\overset{d}{=}B_n$ means that A_n and B_n have the same distribution asymptotically. Since the first term is of order $O_e(1)$, a contradiction occurs.

Now, suppose that $\xi_n/\sqrt{n|\mathbf{H}|}\to 0$ holds true. By (A.2)–(A.4) and Assumption 4, we have $\boldsymbol{\theta}_n=o_p(1)$ which contradicts the fact that $\|\boldsymbol{\theta}_n\|$ is uniformly bounded and positive.

Therefore, applying the exclusion method, we have shown that ξ_n must be of order $\sqrt{n|\mathbf{H}|}$ exactly, which gives

$$\begin{aligned}&\sqrt{n|\mathbf{H}|}\left(\left[\begin{array}{cc}\widehat{\boldsymbol{\gamma}}(z)-\boldsymbol{\gamma}(z) & [\nabla\widehat{\boldsymbol{\gamma}}(z)-\nabla\boldsymbol{\gamma}(z)]\mathbf{H}\end{array}\right]-\varkappa_B(\mathbf{H},z)^{-1}\varkappa_A(\mathbf{H},z)'\right)\\ \overset{d}{=}\ & \varkappa_B(\mathbf{H},z)^{-1}\mathbb{N}\left(\mathbf{0}_{(p+1)(q+1)},\ f^3(z)R_2(K)\boldsymbol{\Omega}(z)\right).\end{aligned}$$

This completes the proof of this theorem. ■

In the following three lemmas, we define $\varkappa_j(\mathbf{H},z) = \varkappa_{j,Q}(\mathbf{H},z) + \varkappa_{j,P}(\mathbf{H},z)$ for $j = \{A,B\}$ and $\mathbf{\Omega}(z) = \mathbf{\Omega}_Q(z) + \mathbf{\Omega}_P(z) + \mathbf{\Omega}_{Q,P}(z)$, where the subscripts Q and P mean that the variable comes from the use of $\mathbf{Q}_n$ and $\{\mathbf{P}_{n,l}\mathbf{u}\}$ as an instrument, respectively, while the "double" subscript Q,P indicates the use of both types of instruments.

Lemma 1 *Under Assumptions 1–4, we obtain* $(n|\mathbf{H}|)^{-2}\xi_n A_{n1}(z) = f^2(z)\varkappa_A(\mathbf{H},z)' + o_p\left(\|\mathbf{H}\|^2\right)$.
(i) If $\boldsymbol{\beta}(z) \neq \mathbf{0}_p$ *holds over at least one non-empty subset, we have*

$$\varkappa_{A,Q}(\mathbf{H},z) = \frac{\mu_{1,2}(k)}{2}\left[\begin{array}{cc}\mathbf{E}_2(\mathbf{H},z)\mathbf{E}_1(z) & \mathbf{0}'_{q(p+1)}\end{array}\right] \tag{A.5}$$

$$\varkappa_{A,P}(\mathbf{H},z) = \left[1,\mathbf{0}'_q\right]\otimes\sum_{l=1}^{m}\Bigg(\mathbf{F}_{3,l}(\mathbf{H},z)\left[\mathbf{F}_{1,l}(\mathbf{H},z)+\frac{1}{4}\mathbf{F}_{2,l}(z)\right] + 2\mathbf{F}_{4,l}(\mathbf{H},z)\mathbf{F}_{1,l}(\mathbf{H},z)\Bigg), \tag{A.6}$$

where $\mathbf{E}_1(z)$, $\mathbf{E}_2(\mathbf{H},z)$, $\mathbf{F}_{1,l}(\mathbf{H},z)$, $\mathbf{F}_2(z)$, $\mathbf{F}_{3,l}(\mathbf{H},z)$ *and* $\mathbf{F}_{4,l}(\mathbf{H},z)$ *are defined below;*
(ii) If $\boldsymbol{\beta}(z) = \mathbf{0}_p$ *holds over its domain and* $\rho(z)$ *is not constant over at least one non-empty subset, we have* $\varkappa_{A,Q}(\mathbf{H},z) = O_p\left(\|\mathbf{H}\|^2(n|\mathbf{H}|)^{-1/2}\right)$ *and* $\varkappa_{A,P}(\mathbf{H},z) = 2\left[1,\mathbf{0}'_q\right]\otimes\sum_{l=1}^m\mathbf{F}_{4,l}(\mathbf{H},z)\mathbf{F}_{1,l}(\mathbf{H},z)$*;*
(iii) If both $\boldsymbol{\beta}(z) = \mathbf{0}_p$ *and* $\rho(z) = \rho_0$ *hold over their domains, we have* $\varkappa_{A,Q}(\mathbf{H},z) = \mathbf{0}_{(p+1)(q+1)}$ *and* $\varkappa_{A,P}(\mathbf{H},z) = O_p\left(\left(n\sqrt{|\mathbf{H}|}\right)^{-1}\right)$.

Proof. We first consider case *(i)*. Applying straightforward calculation, we obtain

$$\begin{aligned}\frac{1}{n|\mathbf{H}|}\mathbf{Q}(z)'\mathbf{K}_H(z)\mathbf{M}(z) &= \frac{1}{n|\mathbf{H}|}\sum_{i=1}^n\pi_i\left[\mathcal{Z}_i(z)\mathcal{Z}_i(z)'\right]\otimes\left(\mathbf{Q}_{n,i}\mathbf{m}'_i\right)\\ &= f(z)\left[\begin{array}{cc}1 & \mathbf{0}'_q\\ \mathbf{0}_q & \mu_{1,2}(k)\mathbf{I}_q\end{array}\right]\otimes\mathbf{E}_1(z) + O_p\left(\|\mathbf{H}\|^2+(n|\mathbf{H}|)^{-1/2}\right)\end{aligned} \tag{A.7}$$

and

$$\begin{aligned}\frac{1}{n|\mathbf{H}|}\mathbf{C}(z)'\mathbf{K}_H(z)\mathbf{Q}(z) &= \frac{1}{2n|\mathbf{H}|}\sum_{i=1}^n\pi_i\mathbf{m}'_i\mathbf{\Pi}(\mathbf{z}^*_i)\left[\mathcal{Z}_i(z)'\otimes\mathbf{Q}'_{n,i}\right]\\ &= \frac{f(z)}{2}\mu_{1,2}(k)\left[\begin{array}{cc}\mathbf{E}_2(\mathbf{H},z) & \mathbf{0}'_{qd}\end{array}\right]\left(1+O_p\left(\|\mathbf{H}\|^2+(n|\mathbf{H}|)^{-1/2}\right)\right),\end{aligned}$$

where we define

$$\mathbf{E}_1(z) \equiv \lim_{n\to\infty}n^{-1}\sum_{i=1}^n\mathbb{E}\left[\mathbf{Q}_{n,i}\mathbf{m}'_i|\mathbf{z}_i=z\right] \tag{A.8}$$

$$\mathbf{E}_2(\mathbf{H},z) \equiv n^{-1}\sum_{i=1}^n\Bigg(\mathrm{tr}\left\{\mathbf{H}\frac{\partial^2\rho(z)}{\partial z\partial z'}\mathbf{H}\right\}\mathbb{E}\left[(\mathbf{W}\mathbf{y})_i\mathbf{Q}'_{n,i}|\mathbf{z}_i=z\right] + \mathbb{E}\left[\mathrm{tr}\left\{\mathbf{H}\frac{\partial^2\left[\boldsymbol{\beta}(z)'\mathbf{x}_i\right]}{\partial z\partial z'}\mathbf{H}\right\}\mathbf{Q}'_{n,i}\middle|\mathbf{z}_i=z\right]\Bigg). \tag{A.9}$$

This gives (A.5).

Next, let the (i,j)th element of $\mathbf{G}_n(\mathbf{Z}) = \mathbf{W}[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}]^{-1}$ be g_{ij}. Since $\{\mathbf{P}_{n,l}\}$ has finite row- and column-sum norms for all l, letting $p_{l,ij}$ be the (i,j)th element of $\mathbf{P}_{n,l}$, we obtain that

$$\begin{aligned}
\frac{\Psi_{1,l}}{n|\mathbf{H}|} &= \frac{1}{n|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j=1}^{n}(p_{l,ij}\pi_j + p_{l,ji}\pi_i)\,u_i\left(\mathcal{Z}_j'(z)\otimes\mathbf{m}_j'\right) \\
&= \frac{2}{n|\mathbf{H}|}\sum_{i=1}^{n}p_{l,ii}\pi_i\mathcal{Z}_i'(z)\otimes\left[g_{ii}u_i^2, u_i\mathbf{x}_i'\right] + O_p\left((n|\mathbf{H}|)^{-1/2}\left[1, \|\mathbf{H}\|^{-1}\mathbf{i}_q'\right]\otimes\mathbf{i}_{p+1}'\right) \\
&= 2f(z)\left[1,\mathbf{0}_q'\right]\otimes\mathbf{F}_{1,l}(\mathbf{H},z) + o_p(1) \qquad \text{(A.10)} \\
\frac{\Psi_{2,l}}{n|\mathbf{H}|} &= \frac{1}{2n|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j=1}^{n}(p_{l,ij}\pi_j + p_{l,ji}\pi_i)\,\mathbf{m}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\left(\mathcal{Z}_j'(z)\otimes\mathbf{m}_j'\right) \\
&= \frac{1}{2n|\mathbf{H}|}\sum_{i\neq j}p_{l,ij}\pi_j\mathbf{m}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\left(\mathcal{Z}_j'(z)\otimes\mathbf{m}_j'\right) + O_p(\|\mathbf{H}\|) \\
&= \frac{1}{2}f(z)\left[1,\mathbf{0}_q'\right]\otimes\mathbf{F}_{2,l}(z) + o_p(1) \qquad \text{(A.11)}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\Gamma_{2,l}}{n|\mathbf{H}|} &= \frac{1}{4n|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j=1}^{n}(p_{l,ij}\pi_j + p_{l,ji}\pi_i)\,\mathbf{m}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\boldsymbol{\Pi}(\mathbf{z}_j^*)'\mathbf{m}_j \\
&= \frac{1}{2n|\mathbf{H}|}\sum_{i\neq j}p_{l,ij}\pi_j\mathbf{m}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\boldsymbol{\Pi}(\mathbf{z}_j^*)'\mathbf{m}_j + O_p\left(\|\mathbf{H}\|^4\right) \\
&= \frac{1}{2}f(z)\,\mathbf{F}_{3,l}(\mathbf{H},z) + O_p\left(\|\mathbf{H}\|^4 + (n|\mathbf{H}|)^{-1/2}\|\mathbf{H}\|^2\right),
\end{aligned}$$

where we define

$$\begin{aligned}
\mathbf{F}_{1,l}(\mathbf{H},z) &= \left[\sigma_u^2\,(n|\mathbf{H}|f(z))^{-1}\sum_{i=1}^{n}p_{l,ii}g_{ii}\pi_i,\ \mathbf{0}_p'\right] \qquad \text{(A.12)} \\
\mathbf{F}_{2,l}(z) &= n^{-1}\sum_{i\neq j}\mathbb{E}\left[p_{l,ij}\mathbf{m}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\mathbf{m}_j' | \mathbf{z}_j = z\right] \qquad \text{(A.13)}
\end{aligned}$$

and

$$\mathbf{F}_{3,l}(\mathbf{H},z) = \frac{1}{n}\sum_{i\neq j}\operatorname{tr}\left\{\mathbf{H}\mathbb{E}\left[p_{l,ij}\mathbf{m}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\left((\mathbf{W}\mathbf{y})_j\frac{\partial^2\rho(z)}{\partial z\partial z'} + \frac{\partial^2\left(\mathbf{x}_j'\boldsymbol{\beta}(z)\right)}{\partial z\partial z'}\right)\right|\mathbf{z}_j = z\right]\mathbf{H}\right\}. \qquad \text{(A.14)}$$

Since $\{\mathbf{P}_{n,l}\}$ has finite row- and column-sum norms for all l, we have $\mathbf{F}_{1,l}(\mathbf{H},z) = O_p(1)$, $\mathbf{F}_{2,l}(z) = O(1)$ and $\mathbf{F}_{3,l}(\mathbf{H},z) = O\left(\|\mathbf{H}\|^2\right)$.

In addition, denoting $\bar{\mathbf{u}} \equiv \mathbf{G}_n(\mathbf{Z})\mathbf{u}$ and $\bar{\mathbf{y}} \equiv \mathbf{G}_n(\mathbf{Z})\,\text{mtx}\{\mathbf{X},\boldsymbol{\beta}(\mathbf{Z})\}$, we have $\mathbf{W}\mathbf{y} = \bar{\mathbf{y}} + \bar{\mathbf{u}}$, where the ith elements of $n\times 1$ vectors $\bar{\mathbf{u}}$ and $\bar{\mathbf{y}}$ are respectively denoted by $\bar{u}_i$ and $\bar{y}_i$. Next, define $\bar{\mathbf{m}}_i \equiv [\bar{y}_i, \mathbf{x}_i']'$ and $\eta_i(z) \equiv (\mathbf{z}_i - z)'\bigtriangledown^2\rho(\mathbf{z}_i^*)(\mathbf{z}_i - z)$. With $\bar{\Gamma}_{1,l} \equiv \mathbf{u}'[\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s\bar{\mathbf{C}}(z)$ and $\widetilde{\Gamma}_{1,l} \equiv \mathbf{u}'[\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s\widetilde{\mathbf{C}}(z)$, where the ith elements of $\bar{\mathbf{C}}(z)$ and $\widetilde{\mathbf{C}}(z)$ are respectively equal to $\bar{\mathbf{m}}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)/2$ and $\bar{u}_i\eta_i(z)/2$, and given that $\operatorname{tr}\{\mathbf{P}_{n,l}\} = 0$, we obtain

$$\frac{\widetilde{\Gamma}_{1,l}}{n|\mathbf{H}|} = \frac{1}{2n|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j=1}^{n}(p_{l,ij}\pi_j + p_{l,ji}\pi_i)\,u_i\bar{u}_j\eta_j(z)$$

$$= \quad \frac{1}{n\left|\mathbf{H}\right|}\sum_{i=1}^{n}p_{l,ii}g_{ii}\pi_{i}u_{i}^{2}\eta_{i}\left(z\right)+O_{p}\left(\left(n\sqrt{\left|\mathbf{H}\right|}\right)^{-1}\right)=f\left(z\right)\mathbf{F}_{4,l}\left(\mathbf{H},z\right)+O_{p}\left(\left(n\sqrt{\left|\mathbf{H}\right|}\right)^{-1}\right)$$

and

$$\frac{\bar{\Gamma}_{1,l}\Psi_{1,l}}{\left(n\left|\mathbf{H}\right|\right)^{2}}=O_{p}\left(\frac{\left\|\mathbf{H}\right\|^{2}}{\sqrt{n\left|\mathbf{H}\right|}}\right)\quad\text{and}\quad\frac{\Gamma_{3,l}\Psi_{1,l}}{\left(n\left|\mathbf{H}\right|\right)^{2}}=O_{p}\left(\frac{1}{n\sqrt{\left|\mathbf{H}\right|}}\right),$$

where

$$\mathbf{F}_{4,l}\left(\mathbf{H},z\right)=\frac{\sigma_{u}^{2}}{n\left|\mathbf{H}\right|f\left(z\right)}\sum_{i=1}^{n}p_{l,ii}g_{ii}\pi_{i}\eta_{i}\left(z\right)=O_{p}\left(\left\|\mathbf{H}\right\|^{2}\right). \tag{A.15}$$

Therefore, we verify (A.6), combing all the above results together.

Next, we consider case *(ii)*. In this instance, $\mathbf{m}_i = [\bar{u}_i, \mathbf{x}_i']'$ and $\mathbf{E}_1\left(z\right) \equiv \left[\mathbf{0}_d,\ n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\mathbf{Q}_{n,i}\mathbf{x}_i'|\mathbf{z}_i=z\right]\right]$. In addition, we have

$$\frac{1}{n\left|\mathbf{H}\right|}\mathbf{C}(z)'\mathbf{K}_{H}(z)\mathbf{Q}(z)=\frac{1}{2n\left|\mathbf{H}\right|}\sum_{i=1}^{n}\pi_{i}\bar{u}_{i}\eta_{i}\left(z\right)\left[\mathcal{Z}_{i}'(z)\otimes\mathbf{Q}_{n,i}'\right]=O_{p}\left(\left\|\mathbf{H}\right\|^{2}\left(n\left|\mathbf{H}\right|\right)^{-1/2}\right),$$

$$\begin{aligned}\frac{\Gamma_{2,l}}{n\left|\mathbf{H}\right|} &= \frac{1}{4n\left|\mathbf{H}\right|}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(p_{l,ij}\pi_{j}+p_{l,ji}\pi_{i}\right)\bar{u}_{i}\bar{u}_{j}\eta_{i}(z)\eta_{j}(z)\\ &= \frac{1}{2n\left|\mathbf{H}\right|}\sum_{i\neq j}p_{l,ij}\pi_{j}\bar{u}_{i}\bar{u}_{j}\eta_{i}(z)\eta_{j}(z)+O_{p}\left(\left\|\mathbf{H}\right\|^{4}\right)=O_{p}\left(\left\|\mathbf{H}\right\|^{2}/n+\left\|\mathbf{H}\right\|^{4}\right)\end{aligned}$$

and

$$\frac{\Psi_{2,l}}{n\left|\mathbf{H}\right|}=\frac{1}{2n\left|\mathbf{H}\right|}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(p_{l,ij}\pi_{j}+p_{l,ji}\pi_{i}\right)\eta_{i}(z)\mathcal{Z}_{j}'(z)\otimes\left[\bar{u}_{i}\bar{u}_{j},\bar{u}_{i}\mathbf{x}_{j}'\right]=O_{p}\left(\left\|\mathbf{H}\right\|^{2}\right).$$

Combining all the above results together gives $\left(n\left|\mathbf{H}\right|\right)^{-2}\xi_{n}A_{n1}\left(z\right)\approx\left(n\left|\mathbf{H}\right|\right)^{-2}\sum_{l=1}^{m}\widetilde{\Gamma}_{1,l}\Psi_{1,l}'$, which verifies case *(ii)*.

Lastly, note that if both $\boldsymbol{\beta}\left(z\right)=\mathbf{0}_{p}$ and $\rho\left(z\right)\equiv\rho_{0}$ hold over their domains [i.e., case *(iii)*], we have $\mathbf{C}\left(z\right)=\mathbf{0}_{n}$, which implies $\Gamma_{1,l}=0$, $\Gamma_{2,l}=0$ and $\Psi_{2,l}=\mathbf{0}_{(p+1)(q+1)}'$. Hence, $\left(n\left|\mathbf{H}\right|\right)^{-2}\xi_{n}A_{n1}\left(z\right)=\left(2n\left|\mathbf{H}\right|\right)^{-2}\sum_{l=1}^{m}\Gamma_{3,l}\Psi_{1,l}'=O_{p}\left(\left(n\sqrt{\left|\mathbf{H}\right|}\right)^{-1}\right)$. This completes the proof of this lemma. ■

Remark If diag$\{\mathbf{P}_{n,l}\}=0$ for all l, we have $\varkappa_{A,P}\left(\mathbf{H},z\right)=4^{-1}\left[1,\mathbf{0}_{q}'\right]\otimes\sum_{l=1}^{m}\left[\mathbf{F}_{3,l}\left(\mathbf{H},z\right)\mathbf{F}_{2,l}\left(z\right)\right]$ in case *(i)*, $\varkappa_{A,P}\left(\mathbf{H},z\right)=o_{p}\left(n^{-1/2}+\left\|\mathbf{H}\right\|^{4}\right)$ in case *(ii)*, and $\varkappa_{A,P}\left(\mathbf{H},z\right)=o_{p}\left(\left(n\sqrt{\left|\mathbf{H}\right|}\right)^{-1}\right)$ in case *(iii)*.

Lemma 2 *Under Assumptions 1–4, we obtain*

$$\left(n\left|\mathbf{H}\right|\right)^{-2}\xi_{n}^{2}B_{n}\left(z\right)=f^{2}\left(z\right)\varkappa_{B}\left(\mathbf{H},z\right)+O_{p}\left(\left(\xi_{n}\left\|\mathbf{H}\right\|\right)^{-1}+\left(\xi_{n}\left\|\mathbf{H}\right\|\right)^{-2}\right)+o_{p}\left(1\right), \tag{A.16}$$

where

$$\begin{aligned}\varkappa_{B,Q}\left(\mathbf{H},z\right) &= \begin{bmatrix}1 & \mathbf{0}_{q}'\\ \mathbf{0}_{q} & \mu_{1,2}^{2}\left(k\right)\mathbf{I}_{q}\end{bmatrix}\otimes\left[\mathbf{E}_{1}\left(z\right)'\mathbf{E}_{1}\left(z\right)\right]\\ \varkappa_{B,P}\left(\mathbf{H},z\right) &= \begin{bmatrix}1 & \mathbf{0}_{q}'\\ \mathbf{0}_{q} & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\left\{\sum_{l=1}^{m}\left[2\mathbf{F}_{1,l}\left(\mathbf{H},z\right)+\frac{1}{2}\mathbf{F}_{2,l}\left(z\right)\right]'\left[2\mathbf{F}_{1,l}\left(\mathbf{H},z\right)+\frac{1}{2}\mathbf{F}_{2,l}\left(z\right)\right]\right\}.\end{aligned}$$

Proof. By (A.7), we have

$$\begin{aligned}\frac{1}{(n\,|\mathbf{H}|)^2}\mathbf{M}(z)'\boldsymbol{\Xi}_H(z)\mathbf{M}(z) &= \frac{1}{(n\,|\mathbf{H}|)^2}\mathbf{M}(z)'\mathbf{K}_H(z)\mathbf{Q}(z)\mathbf{Q}(z)'\mathbf{K}_H(z)\mathbf{M}(z)\\ &= \varkappa_{B,Q}\left(\mathbf{H},z\right)\left(1+o_p\left(1\right)\right). \qquad \text{(A.17)}\end{aligned}$$

In addition, we have

$$\begin{aligned}\frac{\Psi_{3,l}}{n\,|\mathbf{H}|} &= \frac{1}{n\,|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(p_{l,ij}\pi_j+p_{l,ji}\pi_i\right)\left[\mathcal{Z}_i(z)\mathcal{Z}_j'(z)\right]\otimes\left(\mathbf{m}_i\mathbf{m}_j'\right)\\ &= \begin{bmatrix} O_p\left(1\right) & O_p\left(\|\mathbf{H}\|^{-1}\ \mathbf{i}_q'\right)\\ O_p\left(\|\mathbf{H}\|^{-1}\ \mathbf{i}_q'\right) & O_p\left(\mathbf{i}_{q\times q}\right)\end{bmatrix}\otimes O_p\left(\mathbf{i}_{(p+1)\times(p+1)}\right).\end{aligned}$$

Combining this result with (A.17), (A.10) and (A.11) yields

$$\begin{aligned}\left(n\,|\mathbf{H}|\right)^{-2}\xi_n^2 B_n\left(z\right) &= \frac{1}{(n\,|\mathbf{H}|)^2}\sum_{l=1}^{m}\left[\left(\Psi_{1,l}+\Psi_{2,l}\right)'\left(\Psi_{1,l}+\Psi_{2,l}\right)\right]+\frac{1}{(n\,|\mathbf{H}|)^2}\mathbf{M}(z)'\boldsymbol{\Xi}_H(z)\mathbf{M}(z)+\\ &\quad O_p\left(\xi_n^{-1}\|\mathbf{H}\|^{-1}+\xi_n^{-2}\|\mathbf{H}\|^{-2}\right)\\ &= \varkappa_{B,P}\left(\mathbf{H},z\right)+\varkappa_{B,Q}\left(\mathbf{H},z\right)+o_p\left(1\right)+O_p\left(\xi_n^{-1}\|\mathbf{H}\|^{-1}+\xi_n^{-2}\|\mathbf{H}\|^{-2}\right).\end{aligned}$$

This completes the proof of this lemma. ■

Remark If $\text{diag}\{\mathbf{P}_{n,l}\}=0$ for all l, we have $\varkappa_{B,P}\left(\mathbf{H},z\right)=4^{-1}\text{diag}\{1,\mathbf{0}_q\}\otimes\left[\sum_{l=1}^{m}\mathbf{F}_{2,l}\left(z\right)'\mathbf{F}_{2,l}\left(z\right)\right]$. Further, $\mathbf{E}_1\left(z\right)\equiv\left[\mathbf{0}_d,\ n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\mathbf{Q}_{n,i}\mathbf{x}_i'|\mathbf{z}_i=z\right]\right]$ if $\boldsymbol{\beta}\left(z\right)=\mathbf{0}_p$ holds over its domain, which leads to a singular $\varkappa_B\left(\mathbf{H},z\right)$ because $\mathbf{E}_1\left(z\right)'\mathbf{E}_1\left(z\right)$ becomes singular. However, this problem vanishes if one uses a local constant estimator (instead of a local linear one) because $\sum_{l=1}^{m}\left[2\mathbf{F}_{1,l}\left(\mathbf{H},z\right)+\frac{1}{2}\mathbf{F}_{2,l}\left(z\right)\right]$ $+\ \mathbf{E}_1\left(z\right)'\mathbf{E}_1\left(z\right)$ can be non-singular.

Lemma 3 *Under Assumptions 1–4, we obtain*

$$\left(n\,|\mathbf{H}|\right)^{-3/2}\xi_n A_{n,2}\left(z\right)\ \xrightarrow{d}\ \mathbb{N}\Big(\mathbf{0},\ f^3\left(z\right)R_2\left(K\right)\boldsymbol{\Omega}\left(z\right)\Big), \qquad \text{(A.18)}$$

where $\boldsymbol{\Omega}\left(z\right)$ is defined in the proof.

Proof. By (A.7) and (A.11), we have

$$\begin{aligned}\left(n\,|\mathbf{H}|\right)^{-3/2}\xi_n A_{n,2}\left(z\right) &= \frac{f\left(z\right)}{4\sqrt{n\,|\mathbf{H}|}}\sum_{l=1}^{m}\left(2\bar{\Gamma}_{1,l}+\Gamma_{3,l}\right)\left[1,\mathbf{0}_q'\right]'\otimes\mathbf{F}_{2,l}\left(z\right)'+\\ &\quad \frac{f\left(z\right)}{\sqrt{n\,|\mathbf{H}|}}\Delta_n\left(z\right)'\mathbf{Q}\left(z\right)'\mathbf{K}_H(z)\mathbf{u}+o_p\left(1\right)\equiv\bar{A}_{n,2}\left(z\right)+o_p\left(1\right),\end{aligned}$$

where $\Delta_n\left(z\right)\equiv\text{diag}\left\{1,\mu_{1,2}\left(k\right)\mathbf{I}_q\right\}\otimes\mathbf{E}_1\left(z\right)$. Letting $\boldsymbol{\alpha}\neq\mathbf{0}$ be a $\left[\left(p+1\right)\left(q+1\right)\right]\times 1$ vector, we define

$$\Lambda_n\left(z\right)\equiv\boldsymbol{\alpha}'\bar{A}_{n,2}\left(z\right)=\mathbf{u}'\mathbf{A}_n\mathbf{u}+\mathbf{b}_n'\mathbf{u}, \qquad \text{(A.19)}$$

where

$$\mathbf{A}_n = \frac{f(z)}{4\sqrt{n|\mathbf{H}|}}\sum_{l=1}^{m}[\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s\left[1,\mathbf{0}_q'\right]\otimes\mathbf{F}_{2,l}(z)\boldsymbol{\alpha}$$

$$\mathbf{b}_n = \frac{f(z)}{\sqrt{n|\mathbf{H}|}}\left(\frac{1}{2}\sum_{l=1}^{m}[\mathbf{P}_{n,l}\mathbf{K}_H(z)]^s\bar{\mathbf{C}}(z)\left[1,\mathbf{0}_q'\right]\otimes\mathbf{F}_{2,l}(z)+\mathbf{K}_H(z)\mathbf{Q}(z)\Delta_n(z)\right)\boldsymbol{\alpha}.$$

Since both $\mathbf{A}_n$ and $\mathbf{b}_n$ are exogenous, we can apply Kelejian & Prucha's (2001) central limit theorem for linear-quadratic forms with minor modification. That is, we need to check the following two conditions:

$$\sum_{i=1}^{n}\mathbb{E}\left[|b_{n,i}|^{2+\delta_1}\right]\to 0 \quad \text{for some } \delta_1>0 \tag{A.20}$$

$$0<c_0<\mathbb{V}ar\left[\Lambda_n(z)\right]<c_1<\infty, \tag{A.21}$$

where $b_{n,i}$ is the ith element of $\mathbf{b}_n$.

First, we verify (A.20). Applying direct calculations gives

$$b_{n,i}=\frac{f(z)}{4\sqrt{n|\mathbf{H}|}}\boldsymbol{\alpha}'\sum_{l=1}^{m}\left[\begin{array}{c}1\\ \mathbf{0}_q\end{array}\right]\otimes\mathbf{F}_{2,l}(z)'\sum_{j=1}^{n}(p_{l,ij}\pi_j+p_{l,ji}\pi_i)\,\bar{\mathbf{m}}_j'\boldsymbol{\Pi}(\mathbf{z}_j^*)+\frac{f(z)}{\sqrt{n|\mathbf{H}|}}\pi_i\left[\mathcal{Z}_i'(z)\otimes\mathbf{Q}_{n,i}'\right]\Delta_n(z)\boldsymbol{\alpha}.$$

Applying Minkoski's inequality gives

$$\mathbb{E}\left(\left|\sum_{j=1}^{n}(p_{l,ij}\pi_j+p_{l,ji}\pi_i)\,\bar{\mathbf{m}}_j'\boldsymbol{\Pi}(\mathbf{z}_j^*)\right|^{2+\delta_1}\right) \le \left[\sum_{j=1}^{n}\left(\mathbb{E}\left|(p_{l,ij}\pi_j+p_{l,ji}\pi_i)\,\bar{\mathbf{m}}_j'\boldsymbol{\Pi}(\mathbf{z}_j^*)\right|^{2+\delta_1}\right)^{1/(2+\delta_1)}\right]^{2+\delta_1}\le M|\mathbf{H}|$$

and

$$\mathbb{E}\left|\pi_i\left[\mathcal{Z}_i'(z)\otimes\mathbf{Q}_{n,i}'\right]\Delta_n(z)\boldsymbol{\alpha}\right|^{2+\delta_1}\le M\mathbb{E}\left[\pi_i\|\mathcal{Z}_i(z)\|\|\mathbf{Q}_{n,i}\|\right]^{2+\delta_1}\le M|\mathbf{H}|$$

if $\mathbb{E}\left[\|\mathbf{Q}_{n,i}\|^{2+\delta_1}|\mathbf{z}_i=z\right]<M$ is continuous and bounded in the neighborhood of z, $\mathbb{E}\left[\|\mathbf{x}_i\|^{2(2+\delta_1)}\right]<M$ and $\mathbb{E}\left[\|\mathbf{z}_i\|^{4(2+\delta_1)}\right]<M$. Hence, we obtain $\mathbb{E}\left[|b_{n,i}|^{2+\delta_1}\right]\le M|\mathbf{H}|(n|\mathbf{H}|)^{-(2+\delta_1)/2}$. This gives (A.20).

Second, we verify (A.21). Note that $\mathbb{E}\left[\Lambda_n(z)\right]=0$ since $\operatorname{tr}\{\mathbf{P}_{n,l}\}=0$ for all l, and

$$\begin{aligned}\mathbb{V}ar\left[\Lambda_n(z)\right] &= 2\sigma_u^4\mathbb{E}\left[\operatorname{tr}\left\{\mathbf{A}_n^2\right\}\right]+\left(\mathbb{E}\left[u_1^4\right]-2\sigma_u^4\right)\sum_{i=1}^{n}\mathbb{E}\left[a_{n,ii}^2\right]+\sigma_u^4\sum_{i\neq j}\mathbb{E}\left(a_{n,ii}a_{n,jj}\right)+\\ &\quad \sigma_u^2\mathbb{E}\left[\mathbf{b}_n'\mathbf{b}_n\right]+2\mathbb{E}\left[u_1^3\right]\sum_{i=1}^{n}\mathbb{E}\left[a_{n,ii}b_{n,i}\right],\end{aligned}\tag{A.22}$$

where $a_{n,ij}$ is the (i,j)th element of $\mathbf{A}_n$. Then, we obtain

$$\sum_{i\neq j}\mathbb{E}\left[a_{n,ii}a_{n,jj}\right]=\frac{f^2(z)}{4n|\mathbf{H}|}\boldsymbol{\alpha}'\sum_{l=1}^{m}\sum_{l'=1}^{m}\left[\begin{array}{cc}1 & \mathbf{0}_q'\\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{array}\right]\otimes\left[\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\right]\boldsymbol{\alpha}\sum_{i\neq j}\mathbb{E}\left[p_{l,ii}p_{l',jj}\pi_i\pi_j\right]=O(|\mathbf{H}|)$$

and

$$\mathbb{E}\left[\operatorname{tr}\left\{\mathbf{A}_n^2\right\}\right] = \frac{f^2(z)}{16n|\mathbf{H}|}\boldsymbol{\alpha}'\sum_{l=1}^{m}\sum_{l'=1}^{m}\begin{bmatrix}1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\left[\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\right]\boldsymbol{\alpha}\mathbb{E}\left[\operatorname{tr}\left\{\left[\mathbf{P}_{n,l}\mathbf{K}_H(z)\right]^s\left[\mathbf{P}_{n,l'}\mathbf{K}_H(z)\right]^s\right\}\right]$$

$$= \frac{f^2(z)}{16n|\mathbf{H}|}\boldsymbol{\alpha}'\sum_{l=1}^{m}\sum_{l'=1}^{m}\begin{bmatrix}1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\left[\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\right]\boldsymbol{\alpha}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[\left(p_{l,ij}\pi_j+p_{l,ji}\pi_i\right)\left(p_{l',ij}\pi_j+p_{l',ji}\pi_i\right)\right]$$

$$= \frac{f^2(z)}{8n|\mathbf{H}|}\boldsymbol{\alpha}'\sum_{l=1}^{m}\sum_{l'=1}^{m}\begin{bmatrix}1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\left[\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\right]\boldsymbol{\alpha}\sum_{i=1}^{n}\mathbb{E}\left[p_{l,ii}p_{l',ii}\pi_i^2\right]+O\left(n^{-1}\right)$$

$$= \sum_{i=1}^{n}\mathbb{E}\left[a_{n,ii}^2\right]/2+O\left(n^{-1}\right)$$

so that $2\sigma_u^4\mathbb{E}\left[\operatorname{tr}\left(\mathbf{A}_n^2\right)\right]+\left(\mathbb{E}\left[u_1^4\right]-2\sigma_u^4\right)\sum_{i=1}^{n}\mathbb{E}\left[a_{n,ii}^2\right]+\sigma_u^4\sum_{i\neq j}\mathbb{E}\left(a_{n,ii}a_{n,jj}\right)=\mathbb{V}ar\left(u_1^2\right)\sum_{i=1}^{n}\mathbb{E}\left[a_{n,ii}^2\right]+O\left(n^{-1}+|\mathbf{H}|\right)$. By definition, we have

$$\sum_{i=1}^{n}\mathbb{E}\left[a_{n,ii}^2\right] = \frac{f^2(z)}{4n|\mathbf{H}|}\boldsymbol{\alpha}'\sum_{l=1}^{m}\sum_{l'=1}^{m}\begin{bmatrix}1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\left[\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\right]\boldsymbol{\alpha}\sum_{i=1}^{n}\mathbb{E}\left[p_{l,ii}p_{l',ii}\pi_i^2\right]$$
$$\approx f^3(z)R_2(K)\boldsymbol{\alpha}'\boldsymbol{\Omega}_1(z)\boldsymbol{\alpha}/4.$$

Hence, we obtain

$$\boldsymbol{\Omega}_1(z)=\begin{bmatrix}1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\sum_{l=1}^{m}\sum_{l'=1}^{m}\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[p_{l,ii}p_{l',ii}|\mathbf{z}_i=z\right]. \tag{A.23}$$

Next, we have

$$\mathbf{b}_n'\mathbf{b}_n = \frac{f^2(z)}{4n|\mathbf{H}|}\sum_{l=1}^{m}\sum_{l'=1}^{m}\bar{\mathbf{C}}(z)'\left[\mathbf{P}_{n,l}\mathbf{K}_H(z)\right]^s\left[\mathbf{P}_{n,l'}\mathbf{K}_H(z)\right]^s\bar{\mathbf{C}}(z)\ \boldsymbol{\alpha}'\begin{bmatrix}1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\left[\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\right]\boldsymbol{\alpha}+$$
$$\frac{f^2(z)}{2n|\mathbf{H}|}\boldsymbol{\alpha}'\sum_{l=1}^{m}\begin{bmatrix}1 \\ \mathbf{0}_q\end{bmatrix}\otimes\mathbf{F}_{2,l}(z)'\bar{\mathbf{C}}(z)'\left[\mathbf{P}_{n,l}\mathbf{K}_H(z)\right]^s\mathbf{K}_H(z)\mathbf{Q}(z)\Delta_n(z)\boldsymbol{\alpha}+$$
$$\frac{f^2(z)}{n|\mathbf{H}|}\left[\Delta_n(z)\boldsymbol{\alpha}\right]'\mathbf{Q}(z)'\mathbf{K}_H^2(z)\mathbf{Q}(z)\Delta_n(z)\boldsymbol{\alpha}$$
$$\equiv f^2(z)\left[\boldsymbol{\alpha}'A_{n1}\boldsymbol{\alpha}/\mathbf{4}+\boldsymbol{\alpha}'A_{n2}\Delta_n(z)\boldsymbol{\alpha}/\mathbf{2}+\left[\Delta_n(z)\boldsymbol{\alpha}\right]'A_{n3}\Delta_n(z)\boldsymbol{\alpha}\right],$$

where the definition of A_{nj} for $j=1,2,3$ is to be clear from the context below.

Applying standard calculations, we obtain

$$\mathbb{E}\left[A_{n1}\right] = \frac{1}{n|\mathbf{H}|}\begin{bmatrix}1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\sum_{l=1}^{m}\sum_{l'=1}^{m}\left[\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\right]\mathbb{E}\left[\bar{\mathbf{C}}(z)'\left[\mathbf{P}_{n,l}\mathbf{K}_H(z)\right]^s\left[\mathbf{P}_{n,l'}\mathbf{K}_H(z)\right]^s\bar{\mathbf{C}}(z)\right]$$
$$= \frac{1}{4n|\mathbf{H}|}\sum_{l=1}^{m}\sum_{l'=1}^{m}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[\left(p_{l,ij}\pi_j+p_{l,ji}\pi_i\right)\left(p_{l',i'j}\pi_j+p_{l',ji'}\pi_{i'}\right)\bar{\mathbf{m}}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\bar{\mathbf{m}}_{i'}'\boldsymbol{\Pi}(\mathbf{z}_{i'}^*)\right]\times$$
$$\begin{bmatrix}1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\left[\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\right]$$
$$= \frac{1}{4n|\mathbf{H}|}\sum_{l=1}^{m}\sum_{l'=1}^{m}\sum_{i'\neq i\neq j}\mathbb{E}\left[p_{l,ij}p_{l',i'j}\pi_j^2\bar{\mathbf{m}}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\bar{\mathbf{m}}_{i'}'\boldsymbol{\Pi}(\mathbf{z}_{i'}^*)\right]\begin{bmatrix}1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{bmatrix}\otimes\left[\mathbf{F}_{2,l}(z)'\mathbf{F}_{2,l'}(z)\right]+$$

$$
\begin{aligned}
& O\left(\|\mathbf{H}\|^2\right) \\
\approx\ & f(z)\, R_2(K)\, \mathbf{S}_1(z)/4,
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[A_{n2}] &= \frac{1}{n|\mathbf{H}|}\mathbb{E}\left[\sum_{l=1}^{m}\left[\begin{array}{c}1\\ \mathbf{0}_q\end{array}\right]\otimes \mathbf{F}_{2,l}(z)'\,\bar{\mathbf{C}}(z)'\left[\mathbf{P}_{n,l}\mathbf{K}_H(z)\right]^s \mathbf{K}_{\mathbf{H}}(z)\mathbf{Q}(z)\right] \\
&= \frac{1}{2n|\mathbf{H}|}\sum_{l=1}^{m}\left[\begin{array}{c}1\\ \mathbf{0}_q\end{array}\right]\otimes \mathbf{F}_{2,l}(z)'\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[\left(p_{l,ij}\pi_j+p_{l,ji}\pi_i\right)\pi_j\bar{\mathbf{m}}_i'\mathbf{\Pi}(\mathbf{z}_i^*)\mathcal{Z}_j'(z)\otimes \mathbf{Q}_{n,j}'\right] \\
&= \frac{1}{2n|\mathbf{H}|}\sum_{l=1}^{m}\left[\begin{array}{c}1\\ \mathbf{0}_q\end{array}\right]\otimes \mathbf{F}_{2,l}(z)'\sum_{i\neq j}\mathbb{E}\left[p_{l,ij}\pi_j^2\bar{\mathbf{m}}_i'\mathbf{\Pi}(\mathbf{z}_i^*)\left[\mathcal{Z}_j'(z)\otimes \mathbf{Q}_{n,j}'\right]\right]+O\left(\|\mathbf{H}\|^2\right) \\
&\approx f(z)\, R_2(K)\, \mathbf{S}_2(z)/2
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}[A_{n3}] &= \frac{1}{n|\mathbf{H}|}\mathbb{E}\left[\mathbf{Q}(z)'\,\mathbf{K}_H^2(z)\mathbf{Q}(z)\right] \\
&= \frac{1}{n|\mathbf{H}|}\sum_{i=1}^{n}\mathbb{E}\left[\pi_i^2\left[\mathcal{Z}_i(z)\mathcal{Z}_i'(z)\right]\otimes\left(\mathbf{Q}_{n,i}\mathbf{Q}_{n,i}'\right)\right]\approx f(z)\left[\begin{array}{cc}R_2(K) & \mathbf{0}_q'\\ \mathbf{0}_q & \mu_{2,2}(k)\,\mathbf{I}_q\end{array}\right]\otimes \mathbf{E}_3(z),
\end{aligned}
$$

where make use of the following definitions:

$$
\begin{aligned}
\mathbf{S}_1(z) &\equiv \left[\begin{array}{cc}1 & \mathbf{0}_q'\\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{array}\right]\otimes\sum_{l=1}^{m}\sum_{l'=1}^{m}\mathbf{F}_{2,l}(z)'\,\mathbf{F}_{2,l'}(z)\lim_{n\to\infty}\frac{1}{n}\sum_{i'\neq i\neq j}\mathbb{E}\left[p_{l,ij}p_{l',i'j}\bar{\mathbf{m}}_i'\mathbf{\Pi}(\mathbf{z}_i^*)\bar{\mathbf{m}}_{i'}'\mathbf{\Pi}(\mathbf{z}_{i'}^*)|\mathbf{z}_j=z\right] \\
\mathbf{S}_2(z) &\equiv \left[\begin{array}{cc}1 & \mathbf{0}_q'\\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{array}\right]\otimes\sum_{l=1}^{m}\mathbf{F}_{2,l}(z)'\lim_{n\to\infty}\frac{1}{n}\sum_{i\neq j}\mathbb{E}\left[p_{l,ij}\bar{\mathbf{m}}_i'\mathbf{\Pi}(\mathbf{z}_i^*)\mathbf{Q}_{n,j}'|\mathbf{z}_j=z\right] \\
\mathbf{E}_3(z) &\equiv \lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\mathbf{Q}_{n,i}\mathbf{Q}_{n,i}'|\mathbf{z}_i=z\right]. \qquad \text{(A.24)}
\end{aligned}
$$

Moreover, we have

$$
\begin{aligned}
\sum_{i=1}^{n}\mathbb{E}[a_{n,ii}b_{n,i}] &= \frac{f^2(z)}{8n|\mathbf{H}|}\sum_{l=1}^{m}\sum_{l'=1}^{m}\boldsymbol{\alpha}'\left[\begin{array}{cc}1 & \mathbf{0}_q'\\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{array}\right]\otimes\left[\mathbf{F}_{2,l}(z)'\,\mathbf{F}_{2,l'}(z)\right]\boldsymbol{\alpha}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[p_{l,ii}\pi_i\left(p_{l',ij}\pi_j+p_{l',ji}\pi_i\right)\bar{\mathbf{m}}_j'\mathbf{\Pi}(\mathbf{z}_j^*)\right]+ \\
&\quad \frac{f^2(z)}{2n|\mathbf{H}|}\sum_{l=1}^{m}\sum_{l'=1}^{m}\boldsymbol{\alpha}'\left[\begin{array}{c}1\\ \mathbf{0}_q\end{array}\right]\otimes \mathbf{F}_{2,l}(z)'\sum_{i=1}^{n}\mathbb{E}\Big[p_{l',ii}\pi_i^2\left[\mathcal{Z}_i'(z)\otimes \mathbf{Q}_{n,i}'\right]\Big]\Delta_n(z)\,\boldsymbol{\alpha} \\
&= \frac{f^2(z)}{2n|\mathbf{H}|}\sum_{l=1}^{m}\sum_{l'=1}^{m}\boldsymbol{\alpha}'\left[\begin{array}{c}1\\ \mathbf{0}_q\end{array}\right]\otimes \mathbf{F}_{2,l}(z)'\sum_{i=1}^{n}\mathbb{E}\Big[p_{l',ii}\pi_i^2\left[\mathcal{Z}_i'(z)\otimes \mathbf{Q}_{n,i}'\right]\Big]\Delta_n(z)\,\boldsymbol{\alpha}+O\left(\|\mathbf{H}\|^2\right) \\
&= f^3(z)\, R_2(K)\,\boldsymbol{\alpha}'\mathbf{\Omega}_2(z)\,\boldsymbol{\alpha}\mathbf{/2}
\end{aligned}
$$

where

$$
\mathbf{\Omega}_2(z)\equiv\left[\begin{array}{cc}1 & \mathbf{0}_q'\\ \mathbf{0}_q & \mathbf{0}_{q\times q}\end{array}\right]\otimes\left[\sum_{l=1}^{m}\sum_{l'=1}^{m}\mathbf{F}_{2,l}(z)'\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[p_{l',ii}\mathbf{Q}_{n,i}'|\mathbf{z}_i=z\right]\lim_{n\to\infty}\mathbf{E}_1(z)\right]. \qquad \text{(A.25)}
$$

Combining the above results, we get

$$
\mathbf{\Omega}_Q(z) = \left[\begin{array}{cc}1 & \mathbf{0}_q'\\ \mathbf{0}_q & \mu_{1,2}^2(k)\,\mu_{2,2}(k)/R_2(K)\,\mathbf{I}_q\end{array}\right]\otimes\left[\mathbf{E}_1'(z)\,\mathbf{E}_3(z)\,\mathbf{E}_1(z)\right] \qquad \text{(A.26)}
$$

$$
\begin{aligned}
\boldsymbol{\Omega}_P(z) &= \mathbb{V}ar\left(u_1^2\right)\boldsymbol{\Omega}_1(z) + \mathbf{S}_1(z)/4 \\
&= \begin{bmatrix} 1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q} \end{bmatrix} \otimes \sum_{l=1}^{m}\sum_{l'=1}^{m} \mathbf{F}_{2,l}(z)' \mathbf{F}_{2,l'}(z) \left(\mathbb{V}ar\left(u_1^2\right) \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \mathbb{E}\left[p_{l,ii}p_{l',ii}|\mathbf{z}_i = z\right] + \right. \\
&\left. \frac{1}{4}\lim_{n\to\infty}\frac{1}{n}\sum_{i'\neq i\neq j} \mathbb{E}\left[p_{l,ij}p_{l',i'j}\bar{\mathbf{m}}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\bar{\mathbf{m}}_{i'}'\boldsymbol{\Pi}(\mathbf{z}_{i'}^*)|\mathbf{z}_j = z\right]\right) \qquad \text{(A.27)} \\
\boldsymbol{\Omega}_{Q,P}(z) &= \boldsymbol{\Omega}_2(z)/2 + \mathbf{S}_2(z)\Delta_n(z)/2 \\
&= 2^{-1}\begin{bmatrix} 1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mathbf{0}_{q\times q} \end{bmatrix} \otimes \sum_{l=1}^{m} \mathbf{F}_{2,l}(z)' \left(\sum_{l'=1}^{m} \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \mathbb{E}\left[p_{l',ii}\mathbf{Q}_{n,i}'|\mathbf{z}_i = z\right] + \right. \\
&\left. \lim_{n\to\infty}\frac{1}{n}\sum_{i\neq j}\mathbb{E}\left[p_{l,ij}\bar{\mathbf{m}}_i'\boldsymbol{\Pi}(\mathbf{z}_i^*)\mathbf{Q}_{n,j}'|\mathbf{z}_j = z\right]\right) \lim_{n\to\infty}\mathbf{E}_1(z). \qquad \text{(A.28)}
\end{aligned}
$$

Hence, we have $\mathbb{V}ar\left[\Lambda_n(z)\right]$ converges to $f^3(z)R_2(K)\boldsymbol{\alpha}'\boldsymbol{\Omega}(z)\boldsymbol{\alpha} > \mathbf{0}$. This verifies (A.21). Then, applying Theorem 1 in Kelejian & Prucha (2001) and the Cramér's Wold device completes the proof of (A.18). This completes the proof of this lemma. ■

Remark If $\boldsymbol{\beta}(z) = \mathbf{0}_p$ over its domain, we have $\mathbf{F}_{2,l}(z) = O\left(\|\mathbf{H}\|^2/n\right)$ for all l, so that $\boldsymbol{\Omega}_P(z) = O\left(\|\mathbf{H}\|^4/n^2\right)$, $\boldsymbol{\Omega}_{Q,P}(z) = O\left(\|\mathbf{H}\|^2/n\right)$ and

$$
\boldsymbol{\Omega}_Q(z) = \begin{bmatrix} 1 & \mathbf{0}_q' \\ \mathbf{0}_q & \mu_{1,2}^2(k)\mu_{2,2}(k)/R_2(K)\mathbf{I}_q \end{bmatrix} \otimes \begin{bmatrix} 0 & \mathbf{0}_p' \\ \mathbf{0}_p & \mathbf{E}_{13}(z)'\mathbf{E}_3(z)\mathbf{E}_{13}(z) \end{bmatrix}, \qquad \text{(A.29)}
$$

where $\mathbf{E}_{13}(z) = n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\mathbf{Q}_{n,i}\mathbf{x}_i'|\mathbf{z}_i = z\right]$. Hence, $\boldsymbol{\Omega}(z)$ becomes singular.

Proof of Theorem 3. In this theorem, we have $\widehat{\mathbf{G}}_n(\mathbf{Z}) = \mathbf{W}\widehat{\mathbf{S}}_n(\mathbf{Z})$, $\widehat{\mathbf{S}}_n(\mathbf{Z}) = \left[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{Z})\mathbf{W}\right]^{-1} = \mathbf{S}_n(\mathbf{Z}) + \left(\left[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{Z})\mathbf{W}\right]^{-1} - \left[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}\right]^{-1}\right)$,

$$
\begin{aligned}
\widehat{\mathbf{Q}}_{1n} &= \widehat{\mathbf{G}}_n(\mathbf{Z})\,\mathrm{mtx}\left\{\mathbf{X}, \widehat{\boldsymbol{\beta}}(\mathbf{Z})\right\} = \mathbf{Q}_{1n} + \left[\widehat{\mathbf{G}}_n(\mathbf{Z}) - \mathbf{G}_n(\mathbf{Z})\right]\mathrm{mtx}\left\{\mathbf{X}, \widehat{\boldsymbol{\beta}}(\mathbf{Z})\right\} + \\
&\quad \mathbf{G}_n(\mathbf{Z})\,\mathrm{mtx}\left\{\mathbf{X}, \widehat{\boldsymbol{\beta}}(\mathbf{Z}) - \boldsymbol{\beta}(\mathbf{Z})\right\} \qquad \text{(A.30)}
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{\mathbf{P}}_n &= \widehat{\mathbf{G}}_n(\mathbf{Z}) - n^{-1}\mathrm{tr}\left\{\widehat{\mathbf{G}}_n(\mathbf{Z})\right\}\mathbf{I}_n \\
&= \mathbf{P}_n + \widehat{\mathbf{G}}_n(\mathbf{Z}) - \mathbf{G}_n(\mathbf{Z}) - n^{-1}\mathrm{tr}\left\{\widehat{\mathbf{G}}_n(\mathbf{Z}) - \mathbf{G}_n(\mathbf{Z})\right\}\mathbf{I}_n. \qquad \text{(A.31)}
\end{aligned}
$$

In addition, closely following the proof of Theorem 1, we also denote

$$
\begin{aligned}
\widehat{A}_{n1}(z) &= \frac{1}{2\xi_n}\left[\widehat{\Gamma}_2\left(\widehat{\Psi}_1 + \widehat{\Psi}_2\right)' + \left(2\widehat{\Gamma}_1 + \widehat{\Gamma}_3\right)\widehat{\Psi}_1'\right] + \frac{1}{\xi_n}\mathbf{M}(z)'\widehat{\boldsymbol{\Xi}}_{H_0}(z)\mathbf{C}(z) \qquad \text{(A.32)} \\
\widehat{A}_{n2}(z) &= \frac{1}{2\xi_n}\left(2\widehat{\Gamma}_1 + \widehat{\Gamma}_3\right)\widehat{\Psi}_2' + \frac{1}{\xi_n}\mathbf{M}(z)'\widehat{\boldsymbol{\Xi}}_{H_0}(z)\mathbf{u} \\
\widehat{A}_{n3}(z) &= \frac{1}{2\xi_n^2}\left(2\widehat{\Gamma}_1 + \widehat{\Gamma}_2 + \widehat{\Gamma}_3\right)\widehat{\Psi}_3\mathrm{vec}\left(\boldsymbol{\theta}_n\right)
\end{aligned}
$$

and

$$\widehat{B}_n(z) = \frac{1}{\xi_n^2}\left(\widehat{\Psi}_1+\widehat{\Psi}_2-\frac{1}{\xi_n}\left[\mathrm{vec}\left(\widetilde{\boldsymbol{\theta}}_n\right)\right]'\widehat{\Psi}_3\right)\left(\widehat{\Psi}_1+\widehat{\Psi}_2-\frac{1}{\xi_n}\left[\mathrm{vec}\left(\widetilde{\boldsymbol{\theta}}_n\right)\right]'\widehat{\Psi}_3\right)+ \frac{1}{\xi_n^2}\mathbf{M}(z)'\widehat{\boldsymbol{\Xi}}_{H_0}(z)\mathbf{M}(z), \tag{A.33}$$

where $\widehat{\Gamma}_1 \equiv \mathbf{u}'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{C}(z)$, $\widehat{\Gamma}_2 \equiv \mathbf{C}(z)'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{C}(z)$, $\widehat{\Gamma}_3 \equiv \mathbf{u}'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{u}$, $\widehat{\Psi}_1 \equiv \mathbf{u}'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)$, $\widehat{\Psi}_2 \equiv \mathbf{C}(z)'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)$, $\widehat{\Psi}_3 \equiv \mathbf{M}(z)'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)$, $\widehat{\boldsymbol{\Xi}}_{H_0}(z) \equiv \mathbf{K}_{H_0}(z)\widehat{\mathbf{Q}}(z)\widehat{\mathbf{Q}}(z)'\mathbf{K}_{H_0}(z)$. This notation is analogous to that "without the hat" used in the proof of Theorem 1 except for the replacement of $\mathbf{P}_n$ and $\mathbf{Q}_{1n}$ with $\widehat{\mathbf{P}}_n$ and $\widehat{\mathbf{Q}}_{1n}$, respectively.

Closely following the proof of Theorem 1, we have $\xi_n = \sqrt{n|\mathbf{H}_0|}$, and $\varkappa_A(\mathbf{H}_0, z)$, $\varkappa_B(\mathbf{H}_0, z)$ and $\boldsymbol{\Omega}(z)$ are defined in Lemmas 4–6, respectively. This completes the proof of this theorem. ■

Lemma 4 *Under Assumptions **1–3, 5** and **6**, we obtain* $(n|\mathbf{H}_0|)^{-2}\xi_n\widehat{A}_{n1}(z) = \varkappa_A(\mathbf{H}_0, z)' + o_p\left(\|\mathbf{H}_0\|^2\right)$, *where* $\varkappa_{A,Q}(\mathbf{H}_0, z)$ *and* $\varkappa_{A,P}(\mathbf{H}_0, z)$ *are the same as in Lemma 1 with the* l *subscript and* $\sum_{l=1}^m$ *dropped from* $\varkappa_{A,P}(\mathbf{H}_0, z)$.

Proof. Once we show $\widehat{A}_{n1}(z) = A_{n1}(z)[1+o_p(1)]$, this lemma's result follows from Lemma 1.

First, we have

$$\begin{aligned}\left\|\widehat{\mathbf{S}}_n(\mathbf{Z}) - \mathbf{S}_n(\mathbf{Z})\right\|_{sp} &= \left\|[\mathbf{I}_n - \widehat{\rho}(\mathbf{Z})\mathbf{W}]^{-1} - [\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}]^{-1}\right\|_{sp} \\ &= \left\|[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{Z})\mathbf{W}]^{-1}[\widehat{\boldsymbol{\rho}}(\mathbf{Z}) - \boldsymbol{\rho}(\mathbf{Z})]\mathbf{W}[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}]^{-1}\right\|_{sp} \\ &\le \left\|[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{Z})\mathbf{W}]^{-1}\right\|_{sp}\|[\widehat{\boldsymbol{\rho}}(\mathbf{Z}) - \boldsymbol{\rho}(\mathbf{Z})]\mathbf{W}\|_{sp}\left\|[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}]^{-1}\right\|_{sp}\end{aligned}$$

where we denote $\|\mathbf{A}\|_{sp} \equiv \lambda_{\max}^{1/2}(\mathbf{AA}')$. Note that $\|\mathbf{A}\|_{sp} \le \|\mathbf{A}\|$, where $\|\mathbf{A}\| = \sqrt{\mathrm{tr}\{\mathbf{AA}'\}}$. By Weyl's theorem (Seber, 2008, p.117), we have

$$\left\|[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{Z})\mathbf{W}]^{-1}\right\|_{sp} = \|\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}\|_{sp} + O(\|[\widehat{\boldsymbol{\rho}}(\mathbf{Z}) - \boldsymbol{\rho}(\mathbf{Z})]\mathbf{W}\|),$$

where we obtain, under Assumption 5, that

$$\|[\widehat{\boldsymbol{\rho}}(\mathbf{Z}) - \boldsymbol{\rho}(\mathbf{Z})]\mathbf{W}\| = \left(\sum_{i=1}^n\sum_{j\neq i}[\widehat{\rho}(\mathbf{z}_i) - \rho(\mathbf{z}_i)]^2[\widehat{\rho}(\mathbf{z}_i) - \rho(\mathbf{z}_j)]^2 w_{ij}^2\right)^{1/2} = O_p\left(\|\mathbf{H}\|^2 + \sqrt{\frac{\ln n}{n|\mathbf{H}|}}\right).$$

In addition, $\|\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{Z})\mathbf{W}\|_{sp} \le 1$ almost surely under Assumption 1. Therefore, we obtain

$$\left\|\widehat{\mathbf{S}}_n(\mathbf{Z}) - \mathbf{S}_n(\mathbf{Z})\right\|_{sp} = O_p(\|[\widehat{\boldsymbol{\rho}}(\mathbf{Z}) - \boldsymbol{\rho}(\mathbf{Z})]\mathbf{W}\|) = O_p\left(\|\mathbf{H}\|^2 + \sqrt{\frac{\ln n}{n|\mathbf{H}|}}\right). \tag{A.34}$$

It then follows that

$$\left|n^{-1}\mathrm{tr}\left\{\widehat{\mathbf{G}}_n(\mathbf{Z}) - \mathbf{G}_n(\mathbf{Z})\right\}\right| = \left|n^{-1}\sum_{i=1}^n\mathbf{e}_i'\mathbf{W}\left[\widehat{\mathbf{S}}_n(\mathbf{Z}) - \mathbf{S}_n(\mathbf{Z})\right]\mathbf{e}_i\right|$$

$$
\begin{aligned}
&\leq n^{-1}\sum_{i=1}^{n}\left\|\mathbf{e}_i'\mathbf{W}\right\|\left\|\widehat{\mathbf{S}}_n(\mathbf{Z})-\mathbf{S}_n(\mathbf{Z})\right\|_{sp}\|\mathbf{e}_i\| \\
&= O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right), \qquad (A.35)
\end{aligned}
$$

where $\mathbf{e}_i$ denotes the ith column of the identify matrix $\mathbf{I}_n$.

Next, we work on each component of $(n|\mathbf{H}_0|)^{-2}\xi_n\widehat{A}_{n1}(z)$ given in (A.32).

(i) By Assumption 5 and (A.34) we obtain

$$
\begin{aligned}
\frac{1}{n|\mathbf{H}_0|}\widehat{\mathbf{Q}}(z)'\mathbf{K}_{H_0}(z)\mathbf{M}(z) &= \frac{1}{n|\mathbf{H}_0|}\mathbf{Q}(z)'\mathbf{K}_{H_0}(z)\mathbf{M}(z)+O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right) \\
\frac{1}{n|\mathbf{H}_0|}\mathbf{C}(z)'\mathbf{K}_{H_0}(z)\widehat{\mathbf{Q}}(z) &= \frac{1}{n|\mathbf{H}_0|}\mathbf{C}(z)'\mathbf{K}_{H_0}(z)\mathbf{Q}(z)+o_p\left(\|\mathbf{H}_0\|^2\right).
\end{aligned}
$$

(ii) By (A.31), we have

$$
\begin{aligned}
\frac{\widehat{\Psi}_1}{n|\mathbf{H}_0|} &= \frac{1}{n|\mathbf{H}_0|}\mathbf{u}'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z) \\
&= \frac{\Psi_1}{n|\mathbf{H}_0|}+\frac{1}{n|\mathbf{H}_0|}\mathbf{u}'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)- \\
&\quad \frac{2}{n^2|\mathbf{H}_0|}\operatorname{tr}\left\{\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right\}\mathbf{u}'\mathbf{K}_{H_0}(z)\mathbf{M}(z)=\frac{\Psi_1}{n|\mathbf{H}_0|}+o_p(1) \qquad (A.36)
\end{aligned}
$$

since applying (A.34) gives

$$
\begin{aligned}
&\frac{1}{n|\mathbf{H}_0|}\left\|\mathbf{u}'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)\right\| \\
\leq\ &\frac{1}{n|\mathbf{H}_0|}\left\|\widehat{\mathbf{S}}_n(\mathbf{Z})-\mathbf{S}_n(\mathbf{Z})\right\|_{sp}\left(\left\|\mathbf{u}'\mathbf{W}\right\|\|\mathbf{K}_{H_0}(z)\mathbf{M}(z)\|+\left\|\mathbf{u}'\mathbf{K}_{H_0}(z)\right\|\left\|\mathbf{W}'\mathbf{M}(z)\right\|\right) \\
=\ &O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right)O_p\left(\left(\sqrt{|\mathbf{H}_0|}\,\|\mathbf{H}_0\|\right)^{-1}\right)=o_p(1)
\end{aligned}
$$

under Assumption 6.

Similarly, we have

$$
\begin{aligned}
\frac{\widehat{\Psi}_2}{n|\mathbf{H}_0|} &= \frac{1}{n|\mathbf{H}_0|}\mathbf{C}(z)'\left[\widehat{\mathbf{P}}_{n,l}\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z) \\
&= \frac{\Psi_2}{n|\mathbf{H}_0|}+\frac{1}{n|\mathbf{H}_0|}\mathbf{C}(z)'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)- \\
&\quad \frac{2}{n^2|\mathbf{H}_0|}\operatorname{tr}\left\{\widehat{\mathbf{G}}_{\mathbf{n}}(\mathbf{Z})-\mathbf{G}_{\mathbf{n}}(\mathbf{Z})\right\}\mathbf{C}(z)'\mathbf{K}_{H_0}(z)\mathbf{M}(z) \\
&= \frac{\Psi_2}{n|\mathbf{H}_0|}+O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right)O_p\left(|\mathbf{H}_0|^{-1/2}\right)+ \\
&\quad O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right)O_p\left(\|\mathbf{H}_0\|^2\right)=\frac{\Psi_2}{n|\mathbf{H}_0|}+o_p(1) \qquad (A.37)
\end{aligned}
$$

by (A.35) under Assumption 6, since applying (A.34) gives

$$
\frac{1}{n|\mathbf{H}_0|}\left\|\mathbf{C}(z)'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)\right\|
$$

$$
\begin{aligned}
&\leq \frac{1}{n\left|\mathbf{H}_0\right|}\left\|\widehat{\mathbf{S}}_n(\mathbf{Z})-\mathbf{S}_n(\mathbf{Z})\right\|_{sp}\left(\left\|\mathbf{C}(z)'\mathbf{W}\right\|\left\|\mathbf{K}_{H_0}(z)\mathbf{M}(z)\right\|+\left\|\mathbf{C}(z)'\mathbf{K}_{H_0}(z)\right\|\left\|\mathbf{W}'\mathbf{M}(z)\right\|\right)\\
&= O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right)O_p\left(|\mathbf{H}_0|^{-1/2}\right).
\end{aligned}
$$

Furthermore, following the proof of (A.36), we obtain

$$
\begin{aligned}
\frac{\widehat{\Gamma}_2}{n\left|\mathbf{H}_0\right|} &= \frac{1}{n\left|\mathbf{H}_0\right|}\mathbf{C}(z)'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{C}(z)\\
&= \frac{\Gamma_2}{n\left|\mathbf{H}_0\right|}+\frac{1}{n\left|\mathbf{H}_0\right|}\mathbf{C}(z)'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\mathbf{C}(z)-\\
&\quad \frac{2}{n^2\left|\mathbf{H}_0\right|}\mathrm{tr}\left\{\widehat{\mathbf{G}}_{\mathbf{n}}(\mathbf{Z})-\mathbf{G}_{\mathbf{n}}(\mathbf{Z})\right\}\mathbf{C}(z)'\mathbf{K}_{H_0}(z)\mathbf{C}(z)\\
&= \frac{\Gamma_2}{n\left|\mathbf{H}_0\right|}+O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right)O_p\left(\|\mathbf{H}_0\|^2\left|\mathbf{H}_0\right|^{-1/2}\right)+\\
&\quad O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right)O_p\left(\|\mathbf{H}_0\|^4\right)=\frac{\Gamma_2}{n\left|\mathbf{H}_0\right|}+o_p\left(\|\mathbf{H}_0\|^2\right)
\end{aligned}
$$

under Assumption 5.

Next, following the proof of (A.37), we obtain

$$
\begin{aligned}
\frac{\widehat{\widetilde{\Gamma}}_1}{n\left|\mathbf{H}_0\right|} &= \frac{1}{n\left|\mathbf{H}_0\right|}\mathbf{u}'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\widetilde{\mathbf{C}}(z)\\
&= \frac{\widetilde{\Gamma}_1}{n\left|\mathbf{H}_0\right|}+\frac{1}{n\left|\mathbf{H}_0\right|}\mathbf{u}'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\widetilde{\mathbf{C}}(z)-\\
&\quad \frac{2}{n^2\left|\mathbf{H}_0\right|}\mathrm{tr}\left\{\widehat{\mathbf{G}}_{\mathbf{n}}(\mathbf{Z})-\mathbf{G}_{\mathbf{n}}(\mathbf{Z})\right\}\mathbf{u}'\mathbf{K}_{H_0}(z)\widetilde{\mathbf{C}}(z)\\
&= \frac{\widetilde{\Gamma}_1}{n\left|\mathbf{H}_0\right|}+O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right)O_p\left(\|\mathbf{H}_0\|^2+\left(n\left|\mathbf{H}_0\right|\right)^{-1/2}\right)=\frac{\widetilde{\Gamma}_1}{n\left|\mathbf{H}_0\right|}+o_p\left(\|\mathbf{H}_0\|^2\right),
\end{aligned}
$$

where

$$
\begin{aligned}
&\left|\frac{1}{n\left|\mathbf{H}_0\right|}\mathbf{u}'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\widetilde{\mathbf{C}}(z)\right|\\
&\leq \frac{1}{n\left|\mathbf{H}_0\right|}\left\|\widehat{\mathbf{S}}_n(\mathbf{Z})-\mathbf{S}_n(\mathbf{Z})\right\|_{sp}\left(\left\|\mathbf{u}'\mathbf{W}\right\|\left\|\mathbf{K}_{H_0}(z)\widetilde{\mathbf{C}}(z)\right\|+\left\|\mathbf{u}'\mathbf{K}_{H_0}(z)\right\|\left\|\mathbf{W}\widetilde{\mathbf{C}}(z)\right\|\right)\\
&= O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right)O_p\left(\left(n\left|\mathbf{H}_0\right|\right)^{-1/2}\right).
\end{aligned}
$$

(iii) Under Assumptions 5 and 6, we have

$$
\begin{aligned}
\frac{\widehat{\bar{\Gamma}}_1\Psi_1}{(n|\mathbf{H}|)^2} &= \frac{\bar{\Gamma}_1\Psi_1}{\left(n\left|\mathbf{H}_0\right|\right)^2}+\left(\frac{1}{\left(n\left|\mathbf{H}_0\right|\right)^2}\mathbf{u}'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\bar{\mathbf{C}}(z)+\right.\\
&\quad \left.\frac{1}{n^3\left|\mathbf{H}_0\right|^2}\mathrm{tr}\left\{\widehat{\mathbf{G}}_{\mathbf{n}}(\mathbf{Z})-\mathbf{G}_{\mathbf{n}}(\mathbf{Z})\right\}\mathbf{u}'\mathbf{K}_{H_0}(z)\bar{\mathbf{C}}(z)\right)\mathbf{u}'\left[\mathbf{P}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)=\frac{\bar{\Gamma}_1\Psi_1}{\left(n\left|\mathbf{H}_0\right|\right)^2}\left(1+o_p(1)\right)
\end{aligned}
$$

by (A.35), and

$$
\frac{1}{\left(n\left|\mathbf{H}_0\right|\right)^2}\mathbf{u}'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z})-\mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\bar{\mathbf{C}}(z)\mathbf{u}'\left[\mathbf{P}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)
$$

$$
\begin{aligned}
&= \frac{1}{(n|\mathbf{H}_0|)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(\widehat{g}_{ij}-g_{ij}\right)\pi_j+\left(\widehat{g}_{ji}-g_{ji}\right)\pi_i\right]u_i\bar{\mathbf{m}}_j'\mathbf{\Pi}(\mathbf{z}_j^*)\times \\
&\quad \sum_{i'=1}^{n}\sum_{j'=1}^{n}\left(p_{i'j'}\pi_{j'}+p_{j'i'}\pi_{i'}\right)u_{i'}\mathcal{Z}_{j'}(z)'\otimes\mathbf{m}_{j'}' \\
&\approx \frac{1}{(n|\mathbf{H}_0|)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(\widehat{g}_{ij}-g_{ij}\right)\pi_j+\left(\widehat{g}_{ji}-g_{ji}\right)\pi_i\right]u_i^2\bar{\mathbf{m}}_j'\mathbf{\Pi}(\mathbf{z}_j^*)\times \\
&\quad \sum_{j'=1}^{n}\left(p_{ij'}\pi_{j'}+p_{j'i}\pi_i\right)\mathcal{Z}_{j'}(z)'\otimes\mathbf{m}_{j'}' \\
&\approx \frac{1}{(n|\mathbf{H}_0|)^2}\sum_{i=1}^{n}\sum_{j'=1}^{n}p_{j'i}\sum_{j=1}^{n}\left[\left(\widehat{g}_{ji}-g_{ji}\right)\pi_i^2\right]u_i^2\bar{\mathbf{m}}_j'\mathbf{\Pi}(\mathbf{z}_j^*)\mathcal{Z}_{j'}(z)'\otimes\mathbf{m}_{j'}' \\
&= O_p\left(\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}\right)O_p\left(\frac{\mathbf{1}}{n|\mathbf{H}_0|^{3/2}\|\mathbf{H}_0\|}\right) \\
&= O_p\left(\frac{\|\mathbf{H}\|^2+\sqrt{\ln n/(n|\mathbf{H}|)}}{\sqrt{n}|\mathbf{H}_0|\|\mathbf{H}_0\|^3}\right)O_p\left(\frac{\|\mathbf{H}_0\|^2}{\sqrt{n|\mathbf{H}_0|}}\right)=o_p\left(\frac{\|\mathbf{H}_0\|^2}{\sqrt{n|\mathbf{H}_0|}}\right)
\end{aligned}
$$

along with

$$
\begin{aligned}
&\frac{1}{(n|\mathbf{H}_0|)^2}\mathbf{u}'\mathbf{K}_{H_0}(z)\bar{\mathbf{C}}(z)\mathbf{u}'\left[\mathbf{P}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z) \\
&= \frac{1}{(n|\mathbf{H}_0|)^2}\sum_{i=1}^{n}\pi_i u_i\bar{\mathbf{m}}_i'\mathbf{\Pi}(\mathbf{z}_i^*)\sum_{i'=1}^{n}\sum_{j'=1}^{n}\left(p_{i'j'}\pi_{j'}+p_{j'i'}\pi_{i'}\right)u_{i'}\bar{\mathbf{m}}_{j'}'\mathbf{\Pi}(\mathbf{z}_{j'}^*) \\
&\approx \frac{1}{(n|\mathbf{H}_0|)^2}\sum_{i=1}^{n}\sum_{j'=1}^{n}\left(p_{ij'}\pi_{j'}\pi_i+p_{j'i}\pi_i^2\right)u_i^2\bar{\mathbf{m}}_i'\mathbf{\Pi}(\mathbf{z}_i^*)\bar{\mathbf{m}}_{j'}'\mathbf{\Pi}(\mathbf{z}_{j'}^*) \\
&\approx \frac{2}{(n|\mathbf{H}_0|)^2}\sum_{i=1}^{n}p_{ii}\pi_i^2u_i^2\left[\bar{\mathbf{m}}_i'\mathbf{\Pi}(\mathbf{z}_i^*)\right]^2+\frac{1}{(n|\mathbf{H}_0|)^2}\sum_{i=1}^{n}\sum_{j'\neq i}p_{j'i}\pi_i^2u_i^2\bar{\mathbf{m}}_i'\mathbf{\Pi}(\mathbf{z}_i^*)\bar{\mathbf{m}}_{j'}'\mathbf{\Pi}(\mathbf{z}_{j'}^*) \\
&= O_p\left((n|\mathbf{H}_0|)^{-1}\|\mathbf{H}_0\|^4\right)+O_p\left((n|\mathbf{H}_0|)^{-1}\|\mathbf{H}_0\|^2\right)=O_p\left((n|\mathbf{H}_0|)^{-1}\|\mathbf{H}_0\|^2\right).
\end{aligned}
$$

This completes the proof of this lemma. ■

Lemma 5 *Under Assumptions **1**–**3**, **5** and **6**, we obtain*

$$
\frac{\xi_n}{(n|\mathbf{H}_0|)^2}\widehat{B}_n(z)=f^2(z)\varkappa_B(\mathbf{H}_0,z)+o_p(1), \tag{A.38}
$$

where $\varkappa_B(\mathbf{H}_0,z)$ is as defined in Lemma 2 except for the l subscript and $\sum_{l=1}^{m}$ being removed.

Proof. By Lemma 2, we only need to show that $(n|\mathbf{H}_0|)^{-2}\xi_n\left(\widehat{B}_n(z)-B_n(z)\right)=o_p(1)$. Making use of the results in the proof of Lemma 4, we only need to show the following

$$
\frac{\widehat{\Psi}_3}{n|\mathbf{H}_0|} = \frac{1}{n|\mathbf{H}_0|}\mathbf{M}(z)'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{M}(z)
$$

$$
\begin{aligned}
&= \frac{\Psi_3}{n|\mathbf{H}_0|} + \frac{1}{n|\mathbf{H}_0|}\mathbf{M}(z)' \left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z}) - \mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s \mathbf{M}(z) - \\
&\quad \frac{2}{n^2|\mathbf{H}_0|}\mathrm{tr}\left\{\widehat{\mathbf{G}}_{\mathbf{n}}(\mathbf{Z}) - \mathbf{G}_{\mathbf{n}}(\mathbf{Z})\right\}\mathbf{M}(z)'\mathbf{K}_{H_0}(z)\mathbf{M}(z) \\
&= \frac{\Psi_3}{n|\mathbf{H}_0|} + O_p\left(\|\mathbf{H}\|^2 + \sqrt{\ln n/(n|\mathbf{H}|)}\right) O_p\left(|\mathbf{H}_0|^{-1/2}\right)
\end{aligned}
$$

by (A.35) under Assumption 6, since applying (A.34) gives

$$
\begin{aligned}
&\frac{1}{n|\mathbf{H}_0|}\left\|\mathbf{M}(z)'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z}) - \mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s \mathbf{M}(z)\right\| \\
\leq\ & \frac{2}{n|\mathbf{H}_0|}\left\|\widehat{\mathbf{S}}_n(\mathbf{Z}) - \mathbf{S}_n(\mathbf{Z})\right\|_{sp}\left(\|\mathbf{M}(z)'\mathbf{W}\|\,\|\mathbf{K}_{H_0}(z)\mathbf{M}(z)\|\right) \\
=\ & O_p\left(\|\mathbf{H}\|^2 + \sqrt{\ln n/(n|\mathbf{H}|)}\right) O_p\left(|\mathbf{H}_0|^{-1/2}\right) = o_p(1)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\widehat{\Gamma}_3}{n|\mathbf{H}_0|} &= \frac{1}{n|\mathbf{H}_0|}\mathbf{u}'\left[\widehat{\mathbf{P}}_n\mathbf{K}_{H_0}(z)\right]^s\mathbf{u} \\
&= \frac{\Gamma_3}{n|\mathbf{H}_0|} + \frac{1}{n|\mathbf{H}_0|}\mathbf{u}'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z}) - \mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\mathbf{u} - \\
&\quad \frac{2}{n^2|\mathbf{H}_0|}\mathrm{tr}\left\{\widehat{\mathbf{G}}_{\mathbf{n}}(\mathbf{Z}) - \mathbf{G}_{\mathbf{n}}(\mathbf{Z})\right\}\mathbf{u}'\mathbf{K}_{H_0}(z)\mathbf{u} = \frac{\Gamma_3}{n|\mathbf{H}_0|}(1 + o_p(1))
\end{aligned}
$$

because we have

$$
\begin{aligned}
\frac{1}{n|\mathbf{H}_0|}\left\|\mathbf{u}'\left[\left(\widehat{\mathbf{G}}_n(\mathbf{Z}) - \mathbf{G}_n(\mathbf{Z})\right)\mathbf{K}_{H_0}(z)\right]^s\mathbf{u}\right\| &\leq \frac{2}{n|\mathbf{H}_0|}\left\|\widehat{\mathbf{S}}_n(\mathbf{Z}) - \mathbf{S}_n(\mathbf{Z})\right\|_{sp}\|\mathbf{u}'\mathbf{W}\|\,\|\mathbf{K}_{H_0}(z)\mathbf{u}\| \\
&= O_p\left(\|\mathbf{H}\|^2 + \sqrt{\ln n/(n|\mathbf{H}|)}\right) O_p\left((n|\mathbf{H}_0|)^{-1/2}\right).
\end{aligned}
$$

This completes the proof of this lemma. ■

Lemma 6 *Under Assumptions **1–3**, **5** and **6**, we obtain*

$$
\frac{\xi_n \widehat{A}_{n,2}(z)}{(n|\mathbf{H}_0|)^{3/2}} \overset{d}{\to} \mathbb{N}\left(\mathbf{0},\ f^3(z)\,R_2(K)\,\mathbf{\Omega}(z)\right), \tag{A.39}
$$

where $\mathbf{\Omega}(z)$ is as defined in Lemma 3 except for the l subscript and $\sum_{l=1}^m$ being removed.

Proof. By means of the results in Lemmas 8 and 9 (see Supplementary Appendix), we can show that

$$
\frac{\xi_n \widehat{A}_{n,2}(z)}{(n|\mathbf{H}_0|)^{3/2}} = \frac{\xi_n A_{n,2}(z)}{(n|\mathbf{H}_0|)^{3/2}} + o_p(1).
$$

Then, applying Lemma 3 completes the proof of this lemma. ■

Proof of Theorem 4. We first discuss the existence of $\widehat{\boldsymbol{p}}$, where $\widehat{\boldsymbol{p}} = \arg\min_{\boldsymbol{p}\in\mathbb{R}^n}(1/n\mathbf{i}_n - \boldsymbol{p})'(1/n\mathbf{i}_n - \boldsymbol{p})$ subject to $\mathbf{i}_n'\boldsymbol{p} = 1$ and $-\mathbf{i}_n < \left[\widetilde{\rho}(z_1|\boldsymbol{p}), \ldots, \widetilde{\rho}(z_n|\boldsymbol{p})\right]' < \mathbf{i}_n$.

For some very small constant $\epsilon \in (0,1)$, our restrictions can be equivalently written as $|\widetilde{\rho}(z|\boldsymbol{p})| \leq 1-\epsilon$ or $1-\epsilon+\widetilde{\rho}(z|\boldsymbol{p}) \geq 0$ and $1-\epsilon-\widetilde{\rho}(z|\ \boldsymbol{p}) \geq 0$. Since $\sum_{i=1}^{n} p_i = 1$, we have $R_1(z) \equiv 1-\epsilon+\widetilde{\rho}(z|\boldsymbol{p}) \equiv \sum_{i=1}^{n} p_i r_{1,i}(z)$ and $R_2(z) \equiv 1-\epsilon-\widetilde{\rho}(z|\boldsymbol{p}) \equiv \sum_{i=1}^{n} p_i r_{2,i}(z)$, where $r_{1,i}(z) = n\omega_i(\mathbf{X},z)y_i + 1 - \epsilon$ and $r_{2,i}(z) = -n\omega_i(\mathbf{X},z)y_i + 1 - \epsilon$ are both continuous functions of $z \in \mathbb{R}^q$ for all $i \in \{1,\ldots,n\}$.

Let $d_1(z)$ and $d_2(z)$ be the numerator and denominator of $\widehat{\rho}(z)$, respectively, where the former can be written as

$$d_1(z) = \sum_{i=1}^{n} \pi_i y_i \sum_{j=1}^{n} \pi_j (\mathbf{W}\mathbf{y})_j \mathcal{Z}_j(z)'\mathcal{Z}_i(z) Q_{nj}' Q_{ni} \equiv \sum_{i=1}^{n} d_{1,i}(\mathbf{X},z) y_i.$$

Hence, we have $\omega_i(\mathbf{X},z) \equiv d_{1,i}(\mathbf{X},z)/d_2(z)$ for all i. Under Assumption 3, we have $\pi_i = 0$ and $r_{1,i}(z) = r_{2,i}(z) = 1-\epsilon > 0$ if $|z_{l,i} - z_l| > h_l$ for some $l \in \{1,\ldots,q\}$, where $z_{l,i}$ and z_l are the lth element of the $q \times 1$ vector $\mathbf{z}_i$ and z, respectively. Consequently, for a given bandwidth matrix $\mathbf{H}$, we can construct a sequence of overlapping open subsets $\mathcal{O}_{i_j} = \left(o_{1,i_j}^L, o_{1,i_j}^U\right) \times \cdots \times \left(o_{q,i_j}^L, o_{q,i_j}^U\right) \subset \mathbb{R}^q$ such that (i) for each j, both $r_{1,i_j}(z) > 0$ and $r_{2,i_j}(z) > 0$ hold for $z \in \mathcal{O}_{i_j}$; (ii) $\mathcal{S}_z \subseteq \cup_{j=1}^{k} \mathcal{O}_{i_j}$ for $\{i_1,\ldots,i_k\} \subseteq \{1,\ldots,n\}$; (iii) for any give $z \in \mathcal{S}_z$, z is contained in at least one of the subsets. Then, following the induction method used in the proof of Theorem 4.1 in Hall & Huang (2001), we can show that there exists a solution $\widehat{\boldsymbol{p}}$ such that $\mathbf{i}_n'\widehat{\boldsymbol{p}} = 1$, $\widehat{p}_i \in (0,1)$ for all i and $R_j(z) > 0$ for all $z \in \mathcal{S}_z$ and $j = 1,2$.

Next, we study the property of $\widehat{\boldsymbol{p}}$. We define the Lagrangian function of our optimization problem as follows:

$$\mathscr{L}(\boldsymbol{p},\zeta,\boldsymbol{\varphi},\boldsymbol{\psi}) = \sum_{i=1}^{n} \left(p_i - n^{-1}\right)^2 + \zeta\left(\sum_{i=1}^{n} p_i - 1\right) - \sum_{j=1}^{n} \varphi_j\left(\widetilde{\rho}(\mathbf{z}_j|\boldsymbol{p}) + 1 - \epsilon\right) - \sum_{j=1}^{n} \psi_j\left(1 - \epsilon - \widetilde{\rho}(\mathbf{z}_j|\boldsymbol{p})\right),$$

where $\widetilde{\rho}(z|\boldsymbol{p}) = n\sum_{i=1}^{n} p_i\omega_i(\mathbf{X},z)y_i$, $\boldsymbol{\varphi} = [\varphi_1,\ldots,\varphi_n]'$ and $\boldsymbol{\psi} = [\psi_1,\ldots,\psi_n]'$. Under Assumption 6 and the proof of the existence of $\widehat{\boldsymbol{p}}$, the Karush-Kuhn-Tucker theorem holds so that there exist ζ, $\boldsymbol{\varphi}$ and $\boldsymbol{\psi}$ such that

$$2\left(\widehat{p}_i - n^{-1}\right) + \zeta - n\sum_{j=1}^{n} (\varphi_j - \psi_j)\,\omega_i(\mathbf{X},\mathbf{z}_j)y_i = 0 \quad \text{(A.40)}$$

$$\varphi_j\left(\widetilde{\rho}(\mathbf{z}_j|\widehat{\boldsymbol{p}}) + 1 - \epsilon\right) = 0 \ \forall\ j \quad \text{(A.41)}$$

$$\psi_j\left(1 - \epsilon - \widetilde{\rho}(\mathbf{z}_j|\ \widehat{\boldsymbol{p}})\right) = 0 \ \forall\ j \quad \text{(A.42)}$$

$$\sum_{i=1}^{n} \widehat{p}_i = 1 \quad \text{(A.43)}$$

and $\varphi_j \geq 0$, $\psi_j \geq 0$ and $\epsilon - 1 \leq \widetilde{\rho}(z_j|\widehat{\boldsymbol{p}}) \leq 1 - \epsilon$ for all $1 \leq j \leq n$. Solving (A.40) yields

$$\widehat{p}_i = \frac{1}{n} + \frac{n}{2}\sum_{j=1}^{n} (\varphi_j - \psi_j)\,\omega_i(\mathbf{X},\mathbf{z}_j)y_i - \frac{\zeta}{2}. \quad \text{(A.44)}$$

Combining (A.43) with (A.44), we have $\zeta = \sum_{j=1}^{n} (\varphi_j - \psi_j) \sum_{i=1}^{n} \omega_i(\mathbf{X},\mathbf{z}_j)y_i = \sum_{j=1}^{n} (\varphi_j - \psi_j)\,\widehat{\rho}(\mathbf{z}_j)$, which implies that

$$\widehat{p}_i = \frac{1}{n} + \frac{n}{2}\sum_{j=1}^{n} (\varphi_j - \psi_j)\,\omega_i(\mathbf{X},\mathbf{z}_j)y_i - \frac{1}{2}\sum_{j=1}^{n} (\varphi_j - \psi_j)\,\widehat{\rho}(\mathbf{z}_j)$$

$$= \frac{1}{n} + \frac{1}{2}\sum_{j=1}^{n}(\varphi_j - \psi_j)\left[n\omega_i(\mathbf{X},\mathbf{z}_j)y_i - \widehat{\rho}(\mathbf{z}_j)\right]. \quad (A.45)$$

It then follows that

$$\begin{aligned}\widetilde{\rho}(\mathbf{z}_j|\widehat{\boldsymbol{p}}) &= \widehat{\rho}(\mathbf{z}_j) + \frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{n}(\varphi_l - \psi_l)\left[n\omega_i(\mathbf{X},\mathbf{z}_l)y_i - \widehat{\rho}(\mathbf{z}_l)\right]n\omega_i(\mathbf{X},\mathbf{z}_j)y_i \\ &= \widehat{\rho}(\mathbf{z}_j) + \frac{1}{2}\sum_{l=1}^{n}(\varphi_l - \psi_l)\sum_{i=1}^{n}\left[n\omega_i(\mathbf{X},\mathbf{z}_l)y_i - \widehat{\rho}(\mathbf{z}_l)\right]\left[n\omega_i(\mathbf{X},\mathbf{z}_j)y_i - \widehat{\rho}(\mathbf{z}_j)\right]. \quad (A.46)\end{aligned}$$

Let $\mathcal{A}_{n1} = \{j : \varphi_j > 0\} = \{j_s, s = 1, \ldots, |\mathcal{A}_{n1}|\}$ and $\mathcal{A}_{n2} = \{j : \psi_j > 0\} = \{j'_s, s = 1, \ldots, |\mathcal{A}_{n2}|\}$, where $|\mathcal{A}_{n1}|$ and $|\mathcal{A}_{n2}|$ define the number of elements in $\mathcal{A}_{n1}$ and $\mathcal{A}_{n2}$, respectively. By (A.41), we have $\widetilde{\rho}(\mathbf{z}_j|\widehat{\boldsymbol{p}}) = \epsilon - 1$ for $j \in \mathcal{A}_{n1}$. By (A.42), we have $\widetilde{\rho}(\mathbf{z}_j|\widehat{\boldsymbol{p}}) = 1 - \epsilon$ for $j \in \mathcal{A}_{n2}$. Clearly, $\mathcal{A}_{n1}$ and $\mathcal{A}_{n2}$ share no common elements since $\epsilon \in (0,1)$, hence $|\mathcal{A}_{n1}| + |\mathcal{A}_{n2}| \le n$. Therefore, for $j_t \in \mathcal{A}_{n1}$, having denoted $\chi_{i,l} = n\omega_i(\mathbf{X},\mathbf{z}_l)y_i - \widehat{\rho}(\mathbf{z}_l)$, we obtain

$$\widehat{\rho}(\mathbf{z}_{j_t}) + 1 - \epsilon = \frac{1}{2}\left(\sum_{s=1}^{|\mathcal{A}_{n2}|}\psi_{j'_s}\sum_{i=1}^{n}\chi_{i,j_t}\chi_{i,j'_s} - \sum_{s=1}^{|\mathcal{A}_{n1}|}\varphi_{j_s}\sum_{i=1}^{n}\chi_{i,j_t}\chi_{i,j_s}\right). \quad (A.47)$$

For $j'_t \in \mathcal{A}_{n2}$, we obtain

$$\widehat{\rho}\left(\mathbf{z}_{j'_t}\right) - (1-\epsilon) = \frac{1}{2}\left(\sum_{s=1}^{|\mathcal{A}_{n2}|}\psi_{j'_s}\sum_{i=1}^{n}\chi_{i,j'_t}\chi_{i,j'_s} - \sum_{s=1}^{|\mathcal{A}_{n1}|}\varphi_{l_s}\sum_{i=1}^{n}\chi_{i,j'_t}\chi_{i,j_s}\right). \quad (A.48)$$

Let $\boldsymbol{\delta} = \left[-\phi_{j_1}, \ldots, -\phi_{j_{|\mathcal{A}_{n1}|}}, \psi_{j'_1}, \ldots, \psi_{j'_{|\mathcal{A}_{n2}|}}\right]'$, $\mathbf{b} = [\widehat{\rho}(\mathbf{z}_{j_1}) + 1 - \epsilon, \ldots, \widehat{\rho}\left(\mathbf{z}_{j_{|\mathcal{A}_{n1}|}}\right) + 1 - \epsilon, \widehat{\rho}\left(\mathbf{z}_{j'_1}\right) - (1-\epsilon), \ldots, \widehat{\rho}\left(\mathbf{z}_{j'_{|\mathcal{A}_{n2}|}}\right) - (1-\epsilon)]'$. Further, let $\boldsymbol{\alpha}'_i = \left[\chi_{i,j_1}, \ldots, \chi_{i,j_{|\mathcal{A}_{n1}|}}, \chi_{i,j'_1}, \ldots, \chi_{i,j'_{|\mathcal{A}_{n2}|}}\right]$ be the ith row of an $n \times (|\mathcal{A}_{n1}| + |\mathcal{A}_{n2}|)$ matrix $\mathbb{A}_n$. Then, we can rewrite (A.47) and (A.48) in the matrix form: $\mathbb{A}'_n\mathbb{A}_n\boldsymbol{\delta} = 2\mathbf{b}$, where $\mathbb{A}'_n\mathbb{A}_n$ equals n times the sample covariance of $\{\boldsymbol{\alpha}_i\}_{i=1}^n$ and is non-singular under Assumption 6. Hence, $\boldsymbol{\delta} = 2\left(\mathbb{A}'_n\mathbb{A}_n\right)^{-1}\mathbf{b}$ and by (A.45) we then obtain

$$\widehat{p}_i = \frac{1}{n} - \frac{1}{2}\boldsymbol{\alpha}'_i\boldsymbol{\delta} = \frac{1}{n} - \boldsymbol{\alpha}'_i\left(\mathbb{A}'_n\mathbb{A}_n\right)^{-1}\mathbf{b}. \quad (A.49)$$

Therefore, we obtain $\sum_{i=1}^n\left(\widehat{p}_i - n^{-1}\right)^2 = \mathbf{b}'\left(\mathbb{A}'_n\mathbb{A}_n\right)^{-1}\sum_{i=1}^n\boldsymbol{\alpha}_i \quad \boldsymbol{\alpha}'_i\left(\mathbb{A}'_n\mathbb{A}_n\right)^{-1}\mathbf{b} = \mathbf{b}'\left(\mathbb{A}'_n\mathbb{A}_n\right)^{-1}\mathbf{b}$.

By Corollary 2 and Assumption 5, we have $\sup_{z\in\mathcal{S}_z}|\widehat{\rho}(z) - \rho(z)| = O_p\left(\|\mathbf{H}\|^2 + \sqrt{\ln n/(n\mathbf{H})}\right)$ by, say, Masry (1996), so $\mathbf{b}'\mathbf{b} = \sum_{s=1}^{|\mathcal{A}_{n1}|}\left(\widehat{\rho}(\mathbf{z}_{j_s}) + 1 - \epsilon\right)^2 + \sum_{s=1}^{|\mathcal{A}_{n2}|}\left(\widehat{\rho}\left(\mathbf{z}_{j'_s}\right) - (1-\epsilon)\right)^2 = O_p(|\mathcal{A}_{n1}| + |\mathcal{A}_{n2}|)$ $= O_p(n)$. Then, we obtain $\mathbf{b}'\left(\mathbb{A}'_n \quad \mathbb{A}_n/n^3\right)^{-1}\mathbf{b} \le \lambda_{\min}^{-1}\left(\mathbb{A}'_n\mathbb{A}_n/n^3\right)\mathbf{b}'\mathbf{b} = O_p(n)$ since $\lambda_{\min}\left\{\mathbb{A}'_n\mathbb{A}_n/n^3\right\} =$ $\lambda_{\min}\{\mathcal{W}'_n\mathcal{W}_n/n\} + o_p(1) = \lambda_{\min}\{\mathbb{E}\left[\mathcal{W}'_n\mathcal{W}_n/n\right]\} + o_p(1)$ under Assumption 6. Hence, $\sum_{i=1}^n\left(\widehat{p}_i - n^{-1}\right)^2$ $= O_p\left(n^{-2}\right)$, which gives $\max_{1\le i\le n}\left|\widehat{p}_i - n^{-1}\right| = O_p\left(n^{-1}\right)$. This proves Theorem 4(i).

Lastly, we study the distance between $\widetilde{\rho}(z|\,\widehat{\boldsymbol{p}})$ and $\widehat{\rho}(z)$. By Hölder's inequality, we have

$$\widetilde{\rho}(z|\widehat{\boldsymbol{p}}) - \widehat{\rho}(z) = \sum_{i=1}^{n}(n\widehat{p}_i - 1)\,\omega_i(\mathbf{X},z)y_i \le \sqrt{\sum_{i=1}^{n}(n\widehat{p}_i - 1)^2}\sqrt{\sum_{i=1}^{n}\omega_i^2(\mathbf{X},z)y_i^2}$$

$$= O_p(1)\, O_p\left((n|\mathbf{H}|)^{-1/2}\right) = O_p\left((n|\mathbf{H}|)^{-1/2}\right),$$

where $\sum_{i=1}^n \omega_i^2(\mathbf{X},z)y_i^2 = d_2^{-2}(z)\sum_{i=1}^n d_{1,i}^2(\mathbf{X},z)y_i^2$, $d_2(z) = O_p\left((n|\mathbf{H}|)^{-2}\right)$ by (A.17), and $(n|\mathbf{H}|)^{-4} \times \sum_{i=1}^n d_{1,i}^2(\mathbf{X},z)y_i^2 = O_p\left((n|\mathbf{H}|)^{-1}\right)$ following the proof of Theorem 1. ■

B Brief Mathematical Proofs of Theorems 7–9

Proof of Theorem 7. The proposed test statistic is $T_n = \left(n^2|\mathbf{H}|\right)^{-1}\sum_{i=1}^n\sum_{j\neq i}^n \widehat{u}_i\widehat{u}_j\mathcal{K}_H(\mathbf{z}_i,\mathbf{z}_j)$, where $\widehat{u}_i = y_i - \mathbf{m}_i'\breve{\boldsymbol{\gamma}}$ and $\breve{\boldsymbol{\gamma}} = \left[\breve{\rho},\breve{\boldsymbol{\beta}}'\right]'$.

First, under H$_0$, we have $\widehat{u}_i = \mathbf{m}_i'(\boldsymbol{\gamma}_0 - \breve{\boldsymbol{\gamma}}) + u_i$. Hence, we have

$$\begin{aligned} T_n = {} & \frac{1}{n^2|\mathbf{H}|}(\boldsymbol{\gamma}_0-\breve{\boldsymbol{\gamma}})'\sum_{i=1}^n\sum_{j\neq i}^n \mathbf{m}_i\mathbf{m}_j'\mathcal{K}_H(\mathbf{z}_i,\mathbf{z}_j)(\boldsymbol{\gamma}_0-\breve{\boldsymbol{\gamma}}) + \\ & \frac{2}{n^2|\mathbf{H}|}(\boldsymbol{\gamma}_0-\breve{\boldsymbol{\gamma}})'\sum_{i=1}^n\sum_{j\neq i}^n \mathbf{m}_i u_j\mathcal{K}_H(\mathbf{z}_i,\mathbf{z}_j) + \frac{1}{n^2|\mathbf{H}|}\sum_{i=1}^n\sum_{j\neq i}^n u_iu_j\mathcal{K}_H(\mathbf{z}_i,\mathbf{z}_j) \\ \equiv {} & T_{n1} + 2T_{n2} + T_{n3}, \end{aligned} \tag{B.1}$$

where the definitions of T_{nj}, $j = 1,2,3$, should be apparent from the following context. Lee (2007) showed that $\boldsymbol{\gamma}_0 - \breve{\boldsymbol{\gamma}} = O_p(n^{-1/2})$, and it is straightforward to show that, under H$_0$, $T_{n1} = O_p(n^{-1})$, $T_{n2} = O_p(n^{-1})$ and $T_{n3} = O_p\left(\left(n\sqrt{|\mathbf{H}|}\right)^{-1}\right)$. Hence, T_{n3} is the leading term of T_n under H$_0$ and $\mathbb{E}[T_{n3}] = 0$. Applying Hall's (1984, Th.1) central limit theorem for a second-order degenerate U-statistic, one can show that

$$n\sqrt{|\mathbf{H}|}T_{n3} \xrightarrow{d} \mathbb{N}\left(0,\ 2\sigma_u^4 R_2(K)\,\mathbb{E}[f(\mathbf{z})]\right) \tag{B.2}$$

since, denoting $\boldsymbol{\chi}_i \equiv [u_i,\mathbf{z}_i']'$, $\mathbb{H}_n(\boldsymbol{\chi}_i,\boldsymbol{\chi}_j) \equiv |\mathbf{H}|^{-1/2}u_iu_j\mathcal{K}_H(\mathbf{z}_i,\mathbf{z}_j)$ and $G_n(\boldsymbol{\chi}_1,\boldsymbol{\chi}_2) \equiv \mathbb{E}[\mathbb{H}_n(\boldsymbol{\chi}_1,\boldsymbol{\chi}_i)\mathbb{H}_n(\boldsymbol{\chi}_2,\boldsymbol{\chi}_i)|\boldsymbol{\chi}_1,\boldsymbol{\chi}_2]$, for $i\neq j$ we have

$$\frac{\mathbb{E}\left[G_n^2(\boldsymbol{\chi}_1,\boldsymbol{\chi}_2)\right] + n^{-1}\mathbb{E}\left[\mathbb{H}_n^4(\boldsymbol{\chi}_1,\boldsymbol{\chi}_2)\right]}{\left[\mathbb{E}\left[\mathbb{H}_n^2(\boldsymbol{\chi}_1,\boldsymbol{\chi}_2)\right]\right]^2} = \frac{O\left(\|\mathbf{H}\|^2\right) + O\left((n|\mathbf{H}|)^{-1}\right)}{O(1)} = o(1).$$

Second, under H$_1$, we have $\widehat{u}_i = \mathbf{m}_i'[\boldsymbol{\gamma}(\mathbf{z}_i) - \breve{\boldsymbol{\gamma}}] + u_i$. We decompose T_n as

$$\begin{aligned} T_n = {} & \frac{1}{n^2|\mathbf{H}|}\sum_{i=1}^n\sum_{j\neq i}^n[\boldsymbol{\gamma}(\mathbf{z}_i)-\breve{\boldsymbol{\gamma}}]'\ \mathbf{m}_i\mathbf{m}_j'[\boldsymbol{\gamma}(\mathbf{z}_i)-\breve{\boldsymbol{\gamma}}]\mathcal{K}_H(\mathbf{z}_i,\mathbf{z}_j) + \\ & \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^n\sum_{j\neq i}^n[\boldsymbol{\gamma}(\mathbf{z}_i)-\breve{\boldsymbol{\gamma}}]'\mathbf{m}_iu_j\mathcal{K}_H(\mathbf{z}_i,\mathbf{z}_j) + \frac{1}{n^2|\mathbf{H}|}\sum_{i=1}^n\sum_{j\neq i}^n u_iu_j\mathcal{K}_H(\mathbf{z}_i,\mathbf{z}_j) \\ \equiv {} & T_{n1}^a + 2T_{n2}^a + T_{n3}, \end{aligned}$$

where T_{n3} is as defined in (B.1). Given the proof above, we have $T_{n3} = O_p\left(\left(n\sqrt{|\mathbf{H}|}\right)^{-1}\right)$. In addition, applying straightforward calculations, we obtain

$$T_{n2}^a = \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^n\sum_{j\neq i}^n[\boldsymbol{\gamma}(\mathbf{z}_i)-\breve{\boldsymbol{\gamma}}]'\ \mathbf{m}_iu_j\mathcal{K}_H(\mathbf{z}_i,\mathbf{z}_j)$$

$$= \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\boldsymbol{\gamma}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}\right]'\mathbf{m}_i u_j \mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right)+\frac{2}{n^2|\mathbf{H}|}\left(\boldsymbol{\gamma}-\breve{\boldsymbol{\gamma}}\right)'\sum_{i=1}^{n}\sum_{j\neq i}^{n}\mathbf{m}_i u_j \mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right)$$

$$= O_p\left(n^{-1/2}\right)+O_p\left(n^{-1}\right)=O_p\left(n^{-1/2}\right).$$

Next, we show that $T_{n1}^{a}=O_p\left(1\right)$. Under H_1, Assumption 10(i) states that there exists $\boldsymbol{\gamma}\in\Theta$ such that $\breve{\boldsymbol{\gamma}}-\boldsymbol{\gamma}=O_p\left(n^{-1/2}\right)$. Then, we have

$$\begin{aligned}
T_{n1}^{a} &= \frac{1}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\boldsymbol{\gamma}\left(\mathbf{z}_i\right)-\breve{\boldsymbol{\gamma}}\right]'\ \mathbf{m}_i\mathbf{m}_j'\left[\boldsymbol{\gamma}\left(\mathbf{z}_i\right)-\breve{\boldsymbol{\gamma}}\right]\mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right) \\
&= \frac{1}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\boldsymbol{\gamma}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}\right]'\mathbf{m}_i\ \mathbf{m}_j'\left[\boldsymbol{\gamma}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}\right]\mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right)+ \\
&\quad \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\boldsymbol{\gamma}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}\right]'\mathbf{m}_i\ \mathbf{m}_j'\mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right)\left(\boldsymbol{\gamma}-\breve{\boldsymbol{\gamma}}\right)+ \\
&\quad \frac{2}{n^2|\mathbf{H}|}\left(\boldsymbol{\gamma}-\breve{\boldsymbol{\gamma}}\right)\sum_{i=1}^{n}\sum_{j\neq i}^{n}\mathbf{m}_i\mathbf{m}_j'\mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right)\left(\boldsymbol{\gamma}-\breve{\boldsymbol{\gamma}}\right) \\
&= T_{n1,1}^{a}+O_p\left(n^{-1/2}\right)+O_p\left(n^{-1}\right), \qquad \text{(B.3)}
\end{aligned}$$

where, denoting $\varsigma_i\equiv\bar{y}_i\left[\rho\left(\mathbf{z}_i\right)-\rho\right]+\mathbf{x}_i'\left[\boldsymbol{\beta}\left(\mathbf{z}_i\right)-\boldsymbol{\beta}\right]$, we have $\mathbf{m}_i'\left[\boldsymbol{\gamma}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}\right]=\varsigma_i+\bar{u}_i\left[\rho\left(\mathbf{z}_i\right)-\rho\right]$ and

$$\begin{aligned}
T_{n1,1}^{a} &= \frac{1}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\boldsymbol{\gamma}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}\right]'\mathbf{m}_i\mathbf{m}_j'\left[\boldsymbol{\gamma}\left(\mathbf{z}_j\right)-\boldsymbol{\gamma}\right]\mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right) \\
&= \frac{1}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\varsigma_i\varsigma_j\mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right)+O_p\left(n^{-1/2}\right) \\
&= \frac{1}{n^2|\mathbf{H}|}\boldsymbol{\varsigma}'\mathcal{K}_H\left(\mathbf{Z}\right)\boldsymbol{\varsigma}+O_p\left(n^{-1/2}\right) \\
&\leq \frac{1}{n^2|\mathbf{H}|}\boldsymbol{\varsigma}'\boldsymbol{\varsigma}\boldsymbol{\lambda}_{\max}\left\{\mathcal{K}_H\left(\mathbf{Z}\right)\right\}+O_p\left(n^{-1/2}\right),
\end{aligned}$$

where $\boldsymbol{\varsigma}$ is an $n\times 1$ vector with a typical element being equal to ς_i, and $\mathcal{K}_H\left(\mathbf{Z}\right)$ is an $n\times n$ matrix with zero elements on the principal diagonal, the (i,j)th element being equal to $\mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right)$ for $i\neq j$ and the largest eigenvalue $\lambda_{\max}\left\{\mathcal{K}_H\left(\mathbf{Z}\right)\right\}=\max_{\|\varpi\|=1}\varpi'\mathcal{K}_H\left(\mathbf{Z}\right)\varpi$ for any $n\times 1$ vector $\varpi\neq\mathbf{0}_n$. Under Assumption 1, we have $n^{-1}\boldsymbol{\varsigma}'\boldsymbol{\varsigma}=O_p\left(1\right)$, in addition to which we can show that $\lambda_{\max}\left\{\mathcal{K}_H\left(\mathbf{Z}\right)\right\}=O_p\left(n\left|\mathbf{H}\right|\right)$ since

$$\begin{aligned}
\mathbb{E}\left[\left|\lambda_{\max}\left\{\mathcal{K}_H\left(\mathbf{Z}\right)\right\}\right|\right] &\leq \max_{\|\varpi\|=1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left|\varpi_i\right|\left|\varpi_j\right|\mathbb{E}\left[\mathcal{K}_H\left(\mathbf{z}_i,\mathbf{z}_j\right)\right] \\
&= \left|\mathbf{H}\right|\left[\mathbb{E}\left[f\left(\mathbf{z}_1\right)\right]+O\left(\left\|\mathbf{H}\right\|^2\right)\right]\max_{\|\varpi\|=1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left|\varpi_i\right|\left|\varpi_j\right| \\
&\leq Mn\left|\mathbf{H}\right|\left[\mathbb{E}\left[f\left(\mathbf{z}_1\right)\right]+O\left(\left\|\mathbf{H}\right\|^2\right)\right]
\end{aligned}$$

and applying Hőlder's inequality gives $\left[\sum_{i=1}^{n}\left|\varpi_i\right|\right]^2\leq n\left\|\varpi\right\|^2$. Hence, we obtain $T_{n1,1}^{a}=O_p\left(1\right)$.

Furthermore, redefining $\boldsymbol{\chi}_i \equiv [\mathbf{x}_i', \mathbf{z}_i']'$ and $\mathbb{H}_n(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j) \equiv |\mathbf{H}|^{-1} [\boldsymbol{\beta}(\mathbf{z}_i) - \boldsymbol{\beta}]' \mathbf{x}_i \mathbf{x}_j' [\boldsymbol{\beta}(\mathbf{z}_j) - \boldsymbol{\beta}] \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j)$ and because $\mathbb{E}[\mathbb{H}_n^2(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)] = O\left(|\mathbf{H}|^{-1}\right) = o(n)$, applying Lemma 3.1 in Powell, Stock & Stoker (1989), we have

$$\begin{aligned}
\frac{1}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}[\boldsymbol{\beta}(\mathbf{z}_i) - \boldsymbol{\beta}]' \mathbf{x}_i \mathbf{x}_j' [\boldsymbol{\beta}(\mathbf{z}_j) - \boldsymbol{\beta}] \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j) &= \mathbb{E}[\mathbb{H}_n(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)] + o_p(1) \\
&\xrightarrow{p} \mathbb{E}\left[f(\mathbf{z}_1)\left(\mathbb{E}[\mathbf{x}_1'|\mathbf{z}_1][\boldsymbol{\beta}(\mathbf{z}_1) - \boldsymbol{\beta}]\right)^2\right] > 0.
\end{aligned}$$

However, we are unable to show that $T_{n1,1}^a$ converges to a positive constant since $\{\bar{y}_i\}$ is not an independent sequence.

Lastly, if $\boldsymbol{\beta}(z) = \mathbf{0}_p$ holds over its domain, we have $T_{nj}^a = O_p(n^{-1})$ for $j = 1, 2$ and $T_{n3} = O_p\left(\left(n\sqrt{|\mathbf{H}|}\right)^{-1}\right)$ under H_1. This means $n\sqrt{|\mathbf{H}|}T_n$ has the same distribution under both H_0 and H_1. Therefore, our test fails to differentiate the null hypothesis from the alternative hypothesis. Combing the above results with Lemma 7 completes the proof of this theorem. ■

Lemma 7 *Under* H_0, $\widehat{\sigma}^2 \xrightarrow{p} 2\sigma_u^4 R_2(K)\mathbb{E}[f(\mathbf{z})]$; *under* H_1, $\widehat{\sigma}^2 = O_p(1)$, *where*

$$\widehat{\sigma}^2 = \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\widehat{u}_i^2\widehat{u}_j^2\mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j).$$

Proof. Under H_0, following the proof of Theorem 7 , we can show that the leading term of $\widehat{\sigma}^2$ is given by

$$\vec{\sigma}^2 = \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}u_i^2u_j^2\mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j),$$

which is a standard second-order U-statistic with $\mathbb{H}_n(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j) \equiv 2|\mathbf{H}|^{-1} u_i^2u_j^2\mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j)$ and $\boldsymbol{\chi}_i \equiv (u_i, \mathbf{z}_i')'$. Given that $\mathbb{E}[\mathbb{H}_n^2(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)] = O(1) = o(n)$, applying Lemma 3.1 of Powell et al. (1989) gives $\vec{\sigma}^2 = \mathbb{E}[\vec{\sigma}^2] + o_p(1)$, where it is easy to show that $\mathbb{E}[\vec{\sigma}^2] = 2\sigma_u^4 R_2(K)\mathbb{E}[f(\mathbf{z})] + o(1)$.

Under H_1, we can show that the leading term of $\widehat{\sigma}^2$ is given by

$$\begin{aligned}
\breve{\sigma}^2 &= \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\{u_i + \mathbf{m}_i'[\gamma(\mathbf{z}_i) - \gamma]\}^2\{u_j + \mathbf{m}_j'[\gamma(\mathbf{z}_j) - \gamma]\}^2\mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j) \\
&= \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\{u_iu_j + \mathbf{m}_i'[\gamma(\mathbf{z}_i) - \gamma]u_j + \mathbf{m}_j'[\gamma(\mathbf{z}_j) - \gamma]u_i + \mathbf{m}_i'[\gamma(\mathbf{z}_i) - \gamma]\mathbf{m}_j'[\gamma(\mathbf{z}_j) - \gamma]\}^2\mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j) \\
&= \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}u_i^2u_j^2\mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j) + \frac{4}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\{\mathbf{m}_i'[\gamma(\mathbf{z}_i) - \gamma]\}^2u_j^2\mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j) + \\
&\quad \frac{2}{n^2|\mathbf{H}|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\{\mathbf{m}_i'[\gamma(\mathbf{z}_i) - \gamma]\mathbf{m}_j'[\gamma(\mathbf{z}_j) - \gamma]\}^2\mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j) + O_p\left((n|\mathbf{H}|)^{-1/2}\right) = O_p(1)
\end{aligned}$$

under Assumptions 1–4 and 10 (ii), following steps in the proof of Theorem 7. This completes the proof of this lemma. ■

Proof of Theorem 8. Since $\Phi(s)$ is a continuous c.d.f., applying Pola's Theorem (Bhattacharya & Rao, 1986), we only need to show that $\Pr^*(J_n^* \leq s) - \Phi(s) = o_p(1)$ for any given value of s.

The bootstrap test statistic $T_n^* = \left(n^2|\mathbf{H}|\right)^{-1} \sum_{i=1}^n \sum_{j\neq i}^n \widehat{u}_i^* \widehat{u}_j^* \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j)$, where $\widehat{u}_i^* = y_i^* - \breve{\rho}^* \sum_{j\neq i} w_{ij} y_j^* - \mathbf{x}_i' \breve{\boldsymbol{\beta}}^* = \mathbf{m}_i^{*\prime}(\breve{\boldsymbol{\gamma}} - \breve{\boldsymbol{\gamma}}^*) + u_i^*$ and $\mathbf{m}_i^* = \left[\sum_{j\neq i} w_{ij} y_j^*, \mathbf{x}_i'\right]'$. Hence, we have

$$\begin{aligned} T_n^* =& \frac{1}{n^2|\mathbf{H}|}(\breve{\boldsymbol{\gamma}} - \breve{\boldsymbol{\gamma}}^*)' \sum_{i=1}^n \sum_{j\neq i}^n \mathbf{m}_i^* \mathbf{m}_j^{*\prime} \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j)(\breve{\boldsymbol{\gamma}} - \breve{\boldsymbol{\gamma}}^*) + \\ & \frac{2}{n^2|\mathbf{H}|}(\breve{\boldsymbol{\gamma}} - \breve{\boldsymbol{\gamma}}^*)' \sum_{i=1}^n \sum_{j\neq i}^n \mathbf{m}_i^* u_j^* \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j) + \frac{1}{n^2|\mathbf{H}|} \sum_{i=1}^n \sum_{j\neq i}^n u_i^* u_j^* \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j) \\ \equiv& T_{n1}^* + 2T_{n2}^* + T_{n3}^*, \end{aligned} \tag{B.4}$$

where T_{nj}^* for $j = 1, 2, 3$ are defined in the order of their appearance in the first equality.

Let $\mathbb{E}^*[\cdot] = \mathbb{E}\left[\cdot|\{(\mathbf{x}_i, \mathbf{z}_i, y_i^*)\}_{i=1}^n\right]$. The wild bootstrap method implies that $\mathbb{E}^*[u_i^*] = 0$, $\mathbb{E}^*\left[u_i^{*2}\right] = \widehat{u}_i^2$, and $\mathbb{E}^*\left[u_i^{*3}\right] = \widehat{u}_i^3$ for all i, where $\widehat{u}_i = \mathbf{m}_i'(\boldsymbol{\gamma} - \breve{\boldsymbol{\gamma}}) + \varepsilon_i$ is the estimated residual from a parametric linear spatial autoregressive model. Since $\breve{\boldsymbol{\gamma}} - \breve{\boldsymbol{\gamma}}^* = O_p(n^{-1/2})$ and $\mathbb{E}^*[u_i^*] = 0$, it is easy to show that $T_{n1}^* = O_p(n^{-1})$, $T_{n2}^* = O_p\left(n^{-1}\right)$ and $T_{n3}^* = O_p\left(\left(n\sqrt{|\mathbf{H}|}\right)^{-1}\right)$. Hence, T_{n3}^* is the leading term of T_n^*. Since $\{u_i^*\}$ is independent but not identically distributed, we apply de Jong's (1987) central limit theorem for a second-order degenerate U-statistic to show the asymptotic result for $n\sqrt{|\mathbf{H}|}T_{n3}^*$.

Specifically, denoting $\boldsymbol{\chi}_i^* \equiv [u_i^*, \mathbf{z}_i']'$ and $\mathbb{H}_n^*\left(\boldsymbol{\chi}_i^*, \boldsymbol{\chi}_j^*\right) \equiv 2\left(n^2|\mathbf{H}|\right)^{-1} u_i^* u_j^* \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j)$, we have

$$T_{n3}^* = \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{H}_n^*\left(\boldsymbol{\chi}_i^*, \boldsymbol{\chi}_j^*\right),$$

where $\mathbb{E}^*\left[\mathbb{H}_n^*\left(\boldsymbol{\chi}_i^*, \boldsymbol{\chi}_j^*\right)|\boldsymbol{\chi}_j^*\right] = 0$ for all $i \neq j$. In addition, we have

$$\begin{aligned} \sigma_n^{*2} &= \mathbb{E}^*\left[T_{n3}^{*2}\right] = \frac{4}{n^4|\mathbf{H}|^2} \sum_{i=1}^n \sum_{j=i+1}^n \sum_{i'=1}^n \sum_{j'=i'+1}^n \mathbb{E}^*\left[u_i^* u_j^* u_{i'}^* u_{j'}^*\right] \mathcal{K}_H(\mathbf{z}_i, \mathbf{z}_j) \mathcal{K}_H\left(\mathbf{z}_{i'}, \mathbf{z}_{j'}\right) \\ &= \frac{4}{n^4|\mathbf{H}|^2} \sum_{i=1}^n \sum_{j=i+1}^n \widehat{u}_i^2 \widehat{u}_j^2 \mathcal{K}_H^2(\mathbf{z}_i, \mathbf{z}_j) = \frac{2\widehat{\sigma}^2}{n^2|\mathbf{H}|}, \end{aligned}$$

and $\sigma_n^{*2} = O_p\left(\left(n^2|\mathbf{H}|\right)^{-1}\right)$ by Lemma 7.

Further, defining $\omega_{ij} \equiv \mathbb{H}_n^*\left(\boldsymbol{\chi}_i^*, \boldsymbol{\chi}_j^*\right)$, we can easily show that $G_I^* \equiv \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{E}^*\left[\omega_{ij}^4\right] = O_p\left(\left(n^4|\mathbf{H}|^3\right)^{-1}\right)$, $G_{II}^* \equiv \sum_{i=1}^n \sum_{j=i+1}^n \sum_{l=j+1}^n \mathbb{E}^*\left[\omega_{ij}^2\omega_{il}^2 + \omega_{ji}^2\omega_{jl}^2 + \omega_{li}^2\omega_{lj}^2\right] = O_p\left(\left(n^5|\mathbf{H}|^2\right)^{-1}\right)$ and $G_{IV}^* \equiv \sum_{i=1}^n \sum_{j=i+1}^n \sum_{l=j+1}^n \sum_{t=l+1}^n \mathbb{E}^*\left[\omega_{ij}\omega_{il}\omega_{tj}\omega_{tl} + \omega_{ij}\omega_{it}\omega_{lj}\omega_{lt} + \omega_{il}\omega_{it}\omega_{jl}\omega_{jt}\right] = O_p\left(\left(n^4|\mathbf{H}|\right)^{-1}\right)$ because both u_i and $\mathbf{m}_i$ have finite fourth moments. Hence, G_I^*, G_{II}^* and G_{IV}^* are all of order $o_p\left(\sigma_n^{*4}\right)$. By Proposition 3.2 in de Jong (1987), we then obtain $n\sqrt{|\mathbf{H}|}T_{n3}^*/\sqrt{\sigma_n^{*2}} \xrightarrow{d} \mathbb{N}(0,1)$. This completes the proof of this theorem. ■

Proof of Theorem 9. Under H_0, we have

$$D_n = \frac{1}{n}\sum_{i=1}^n \left[\mathbf{m}_i'\breve{\boldsymbol{\gamma}} - \mathbf{m}_i'\widehat{\boldsymbol{\gamma}}(\mathbf{z}_i)\right]^2$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\mathbf{m}_i'\left[\widehat{\boldsymbol{\gamma}}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}_0\right]\right\}^2 - \frac{2}{n}\sum_{i=1}^{n}\mathbf{m}_i'\left[\widehat{\boldsymbol{\gamma}}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}_0\right]\left(\breve{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right)'\mathbf{m}_i + \frac{2}{n}\sum_{i=1}^{n}\left[\mathbf{m}_i'\left(\breve{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right)\right]^2 \\
&= O_p\left(\|\mathbf{H}\|^4 + \frac{\ln n}{n|\mathbf{H}|}\right)
\end{aligned}
$$

under Assumptions 5. Under H_1, we have

$$
\begin{aligned}
D_n &= \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{m}_i'\breve{\boldsymbol{\gamma}} - \mathbf{m}_i'\widehat{\boldsymbol{\gamma}}\left(\mathbf{z}_i\right)\right]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\mathbf{m}_i'\left[\widehat{\boldsymbol{\gamma}}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}_0\right]\right\}^2 - \frac{2}{n}\sum_{i=1}^{n}\mathbf{m}_i'\left[\widehat{\boldsymbol{\gamma}}\left(\mathbf{z}_i\right)-\boldsymbol{\gamma}_0\right]\left(\breve{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right)'\mathbf{m}_i + \frac{2}{n}\sum_{i=1}^{n}\left[\mathbf{m}_i'\left(\breve{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right)\right]^2 \\
&= \frac{2}{n}\sum_{i=1}^{n}\left[\mathbf{m}_i'\left(\breve{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right)\right]^2 + O_p\left(\|\mathbf{H}\|^2 + \sqrt{\frac{\ln n}{n|\mathbf{H}|}}\right) \approx \frac{2}{n}\sum_{i=1}^{n}\left[\mathbf{m}_i'\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\right)\right]^2 = O_e\left(1\right).
\end{aligned}
$$

This completes the proof of this theorem. ■

C Proofs of Theorems 5–6 and Empirical Application

Supplementary material related to this article can be found online.

References

Ahmad, I., Leelahanon, S., & Li, Q. (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *Annals of Statistics*, *33*, 258–283.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.

Anselin, L. & Lozano-Gracia, N. (2009). Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics*, *34*, 5–34.

Baltagi, B. & Li, D. (2001). LM tests for functional form and spatial correlation. *International Regional Science Review*, *24*, 194–225.

Bhattacharya, R. N. & Rao, R. R. (1986). *Normal Approximations and Asymptotic Expansions*. Krieger.

Cai, Z., Fan, J., & Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, *95*(451), 888–902.

Cai, Z. & Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory*, *24*(5), 1321–1342.

Cai, Z. & Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, *167*, 413–425.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, *75*, 261–277.

Du, P., Parmeter, C. F., & Racine, J. S. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica*, *23*(3), 1347–1371.

Ertur, C. & Koch, W. (2007). Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics*, *22*, 1033–1062.

Fan, Y. & Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics*, *10*, 245–271.

Gao, J., King, M., Liu, Z., & Tjøstheim, D. (2009). Nonparametric specification testing for non-linear time series with nonstationarity. *Econometric Theory*, *25*, 1869–1892.

Gilley, O. W. & Pace, R. K. (1996). On the Harrison and Rubenfeld data. *Journal of Environmental Economics and Management*, *31*, 403–405.

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

Hall, P. & Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics*, *29*(3), 624–647.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, *24*, 726–748.

Harrison, D. & Rubenfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, *5*, 81–102.

Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, *55*(4), 757–796.

Henderson, D. J., Carroll, R. J., & Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, *144*(1), 257–275.

Kai, B., Li, R., & Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics*, *39*, 305–332.

Kelejian, H. H., Murrell, P., & Shepotylo, O. (2013). Spatial spillovers in the development of institutions. *Journal of Development Economics*, *101*, 297–315.

Kelejian, H. H. & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive models with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, *17*, 99–121.

Kelejian, H. H. & Prucha, I. R. (1999). A generalized moment estimator for the autoregressive parameter in a spatial model. *International Economic Review*, *40*, 509–533.

Kelejian, H. H. & Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, *104*, 219–257.

Kelejian, H. H. & Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics*, *118*, 27–50.

Kelejian, H. H. & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, *157*, 53–67.

Lee, L.-f. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica*, *72*, 1899–1926.

Lee, L.-f. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, *137*, 489–514.

LeSage, J. & Pace, R. K. (2014). The biggest myth in spatial econometrics. *Econometrics*, *2*, 217–249.

Li, Q. (1999). Consistent model specification tests for time series econometric models. *Journal of Econometrics*, *92*, 101–147.

Li, Q., Huang, C. J., Li, D., & Fu, T.-T. (2002). Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics*, *20*(3), 412–422.

Li, Q. & Wang, S. (1998). A simple consistent bootstrap test for a parametric regression functional form. *Journal of Econometrics*, *87*, 145–165.

Liu, X., Lee, L.-f., & Bollinger, C. T. (2010). An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics*, *159*, 303–319.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis*, *17*, 571–599.

Mukand, S. W. & Rodrik, D. (2005). In search of the holy grail: Policy convergence, experimentation, and economic performance. *American Economic Review*, *95*, 374–383.

Pace, P. K., Barry, R., Slawson, V. C., & Sirmans, C. F. (2004). Simultaneous spatial and functional form transformation. In L. Anselin, R. Florax, & S. J. Rey (Eds.), *Advances in Spatial Econometrics*. Berlin: Springer-Verlag.

Paelinck, J. H. P. & Klaassen, L. H. (1979). *Spatial Econometrics*. Farnborough: Saxon House.

Parmeter, C. F., Henderson, D. J., & Kumbhakar, S. C. (2007). Nonparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, *22*, 695–699.

Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, *57*, 1403–1430.

Seber, G. A. F. (2008). *A Matrix Handbook for Statisticians*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.

Silverman, B. W. (1986). *Density Estimation*. London: Chapman and Hall.

Small, K. A. & Steimetz, S. S. C. (2012). Spatial hedonics and the willingness to pay for residential amenities. *Journal of Regional Science*, *52*, 635–647.

Stengos, T. & Sun, Y. (2001). Consistent model specification test for a regression function based on nonparametric wavelet estimation. *Econometric Reviews*, *20*, 41–60.

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

Su, L. & Jin, S. (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics*, *157*, 18–33.

Sun, Y. (2016). Functional-coefficient spatial autoregressive models with nonparametric spatial weights. Working Paper, University of Guelph.

Sun, Y., Cai, Z., & Li, Q. (2015). A consistent nonparametric test on semiparametric smooth coefficient models with integrated time series. *Econometric Theory*. forthcoming.

Sun, Y., Hongjia, Y., Zhang, W., & Lu, Z. (2014). A semiparametric spatial dynamic model. *Annals of Statistics*, *42*, 700–727.

van Gastel, R. A. J. J. & Paelinck, J. H. P. (1995). Computation of Box-Cox transform parameters: A new method and its application to spatial econometrics. In L. Anselin & R. Florax (Eds.), *New Directions in Spatial Econometrics*. Berlin: Springer-Verlag.

Wang, Q. & Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, *40*, 727–758.

Yang, Z., Li, C., & Tse, Y. K. (2006). Functional form and spatial dependence in dynamic panels. *Economic Letters*, *91*, 138–145.

Zhang, Z. (2013). A pairwise difference estimator for partially linear spatial autoregressive models. *Spatial Economic Analysis*, *8*, 176–194.

Zheng, J. X. (1996). A consistent test of function form via nonparametric estimation techniques. *Journal of Econometrics*, *75*, 263–289.