Two-step Series Estimation and Specification Testing of (Partially) Linear Models with Nonparametrically Generated Regressors

Yu-Chin Hsu[*1,2,3], Jen-Che Liao[*1] and Eric S. Lin[4]

[1] *Institute of Economics, Academia Sinica*
[2] *Department of Finance, National Central University*
[3] *Department of Economics, National Chengchi University*
[4] *Department of Economics, National Tsing-Hua University*

This version: July 4, 2018

Abstract

This paper studies three semiparametric models that are useful and frequently encountered in applied econometric work – a linear and two partially linear specifications with generated regressors, i.e., the regressors that are unobserved, but can be nonparametrically estimated from the data. Our framework allows for generated regressors to appear in linear or nonlinear components of partially linear models. We propose two-step series estimators for the finite-dimensional parameters, establish their $\sqrt{n}$-consistency and asymptotic normality, and provide the asymptotic variance formulae that take into account the estimation error of generated regressors. Moreover, we develop a nonparametric specification test for the models considered. Numerical performances of the proposed estimators via simulation experiments illustrate the utility of our approach.

Keywords: *Series estimation, Semiparametric linear model, Partially linear model, Linear and nonlinear generated regressors, Specification tests*

JEL Classification: C01, C13, C14

*Authors of correspondence: Yu-Chin Hsu (ychsu@econ.sinica.edu.tw) and Jen-Che Liao (jcliao@econ.sinica.edu.tw)

1 Introduction

Generated regressors may occur in models where conditional expectations enter a regression model. An example in applied microeconomics is the estimation of simultaneous equation models with a dummy endogenous variable where the conditional expectation that involves the dummy variable is entered as a regressor in the second-step regression (see Heckman (1978)). The generated regressor problem also occurs in the sample selection model with a parametric or nonparametric selection equation, in which cases the added regressor to the main regression for selection correction is a nonlinear transformation of a parametric or nonparametric estimate from the selection equation. Moreover, macroeconomic models with rational expectations could include unknown conditional mean functions as the right-hand side variables (see Barro (1977)). In labor economics, when estimating the wage equation, one may want to control the expected job tenure, which is a function of marital status, education, and other demographic variables and therefore is unobserved.[1] In the area of international trade, one would like to test the trade protection hypothesis that is predicted from the innovative model advanced by Grossman and Helpman (1994); in that setting, the absolute own price elasticity of import demand is viewed as a generated regressor; see Gawande (1998) for detailed discussions.

Econometric issues in the presence of generated regressors have been discussed by Pagan (1984, 1986), who consider the cases in which both the main regression of interest and the auxiliary regression for the generated regressor are parametrically specified. Andrews (1991, 1994) and Newey (1994) examine the cases in which the auxiliary regression can be nonparametrically specified, but the main regression model remains parametrically specified. In general, for this type of models two-step estimators can be proposed; specifically, in the parametric (or nonparametric) first step, the generated regressors are estimated based on the auxiliary regression; while in the second step, the estimated generated regressors are plugged into the main regression for consistent estimation, but the resulting inference is usually invalid if the estimation effect of the generated regressors is not carefully taken into account. On the other hand, one likely suffers inconsistent estimation and invalid inferences owing to misspecification either from the main regression or from the auxiliary regression, as assumed by Pagan (1984, 1986), Andrews (1991, 1994), and Newey (1994). To address this parametric misspecification issue, Stengos and Yan (2001) relax the strong parametric specification in the main regression by considering a partially linear model where the generated regressor appears in a form of a linear index, whereas Rilstone (1996) allows the main regression to

[1] The generated regressor can also be viewed as a technique to reduce dimensionality in the setting of nonparametric regression models with many regressors. See the discussion in Rilstone (1996).

be nonparametrically specified.

In this paper we consider three useful (partially) linear regression models with generated regressors, including a linear model that includes nonparametrically generated regressors and two partially linear models with nonparametrically generated regressors that appear in either the linear or nonlinear components of the models. We refer to the latter models as partially linear models with a linear generated regressor and a nonlinear generated regressor, respectively. For each of the models considered, the respective two-step series estimator for the finite-dimensional parameters is proposed. We provide asymptotic results such as $\sqrt{n}$ consistency and asymptotic normality with correct asymptotic variance of the proposed two-step series estimators. In particular, the asymptotic variance of the two-step series estimator is composed of two sources of error - one is the sampling error term, and the other appears due to the estimation effect of the nonparametrically generated regressor. We further propose a series-based specification test that extends Sun and Li's (2006) test to allow for the presence of generated regressors in partially linear models.

Stengos and Yan (2001) closely relate to our paper, as they propose two two-step kernel estimators for the presence of generated regressors in semiparametric models, including a linear model with generated regressors and a partially linear model with a linear generated regressor, under conditional homoskedastictiy. Compared to their paper, we make the following contributions to the literature. First, our models allow for conditional heteroscedasticity. We also incorporate the potential correlation of the error terms between the main regression and the auxiliary regression. These features are not simultaneously treated in the previous literature. Second, in addition to the models considered in Stengos and Yan (2001), we also consider a partially linear model with nonlinear generated regressors, which is an important and commonly employed model in empirical applications.[2] Third, we propose a specification test for the semiparametric linear models with generated regressors, which has not yet been developed in the literature.

It is worth emphasizing that the series estimation has several theoretical and practical advantages over the kernel estimation. On the one hand, from a theoretical perspective, the rate of the convergence of the series estimator improves upon the kernel counterpart if the regression function is smoother than twice differentiable (see Hansen (2018) for a discussion). On the other hand, from a practical perspective, the advantage of using series estimation is its relative ease of implementation, in the sense that the implementation reduces to a parametric case, and the parametric and nonparametric components of the model can be

[2]For example, Pendakur and Sperlich (2010) consider a semiparametric partially linear model of consumer demand over prices and expenditure, where the uncompensated expenditure-share system is linear in prices and is an unspecified function of log real expenditure. In their context, the problem of a nonparametrically generated regressor arises since real expenditure is not observed, but can be estimated under the model.

estimated simultaneously. It is also known that the series estimator is particularly convenient for the purpose of computing asymptotic standard errors, because parametric standard error formulae apply directly without amendment. Going one step better, Ackerberg, Chen, and Hahn (2012) establish a numerical equivalence between the two-step semiparametric asymptotic variance estimate using the sieve method as the nonparametric first step and the parametric asymptotic variance estimate using the standard parametric two-step procedure.

The paper is organized as follows. Section 2 starts with a simple linear model with a nonparametric generated regressor and presents the asymptotic distribution of the proposed two-step series estimator. Sections 3 and 4 generalize the semiparametric linear model in Section 2 to partially linear models with a linear and nonlinear generated regressor, respectively, and also establish the asymptotic properties of the proposed two-step series estimators. Section 5 suggests a new series-based specification test for the semiparametric linear model with generated regressors. Section 6 performs a small set of Monte Carlo simulations to investigate the finite-sample performance of the proposed methods. Section 7 concludes the paper.

2 Semiparametric linear model with nonparametrically generated regressors

We begin with a semiparametric linear model in which nonparametrically generated regressors enter the model linearly. Specifically, we consider:

$$y = \mathbf{x}'\boldsymbol{\beta} + \boldsymbol{\eta}'\boldsymbol{\alpha} + \varepsilon, \tag{2.1}$$

$$\mathbf{s} = E[\mathbf{s}|\mathbf{z}] + \boldsymbol{u} \equiv \boldsymbol{\eta}(\mathbf{z}) + \boldsymbol{u}, \tag{2.2}$$

where $\mathbf{x}$ are d_x-dimensional observed regressors, $\boldsymbol{\eta}$ are d_η-dimensional unobserved regressors serving as nonparametric components in the model, and (2.2) is the auxiliary regressions that generate the unobserved regressors $\boldsymbol{\eta} \equiv \boldsymbol{\eta}(\mathbf{z})$ with $\mathbf{z}$ being a $d_z \times 1$ vector of auxiliary regressors. For notational simplicity, throughout the paper we focus on the case where $\mathbf{s}$ and therefore $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ are scalar (which from now on are denoted by s, η, and α, respectively), but the theory established below can be easily extended to a multivariate case, where the different conditioning set for each element of $\mathbf{s}$ can be allowed. The nonparametric component η is assumed to take a form of conditional expectation, i.e., $\eta(\mathbf{z}) = E(s|\mathbf{z})$, which is not parametrically specified, so that η must be estimated nonparametricly. The parameters of interest in model (2.1)-(2.2) are $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \alpha)'$.

As a motivating example, let $\eta(\mathbf{z}) = E(s|\mathbf{z})$ be the expected job tenure, where s represents the length of time at the present job, and $\mathbf{z}$ is a vector of exogenous covariates such as age, marital status, the number of children, and other demographic characteristics. The elements of $\mathbf{x}$ and $\mathbf{z}$ are allowed to overlap.[3] We also allow the error term ε_i to be heteroskedastic.

In what follows, let $\mathbf{y} = (y_1, \ldots, y_n)'$ and $\mathbf{X}$ denote the $n \times d_x$ matrix with the i-th row being $\mathbf{x}_i'$. For notational simplicity, we slightly abuse the notation that η may denote either a random variable $\eta = E(s|\mathbf{z})$ or the conditional mean function $\eta(\bar{\mathbf{z}}) = E(s|\mathbf{z} = \bar{\mathbf{z}})$ evaluated at a fixed value $\bar{\mathbf{z}}$ in different contexts when no confusion arises.

2.1 The two-step series estimator

Since the conditional mean function $\eta(\mathbf{z})$ in (2.2) is unknown, we estimate the function $\eta(\mathbf{z})$ nonparametrically by a series-based approximation method. In this paper we focus on power series, but it is not hard to show that, with suitable modifications on the growth rates of the series terms, all the results hold for regression splines.

Assume that we are given an i.i.d. random sample $\{y_i, \mathbf{x}_i, s_i, \mathbf{z}_i\}_{i=1}^n$ of size n. Let $\boldsymbol{p}_K(\mathbf{z}) = (p_1(\mathbf{z}), \ldots, p_K(\mathbf{z}))'$ be a vector of K basis functions of $\mathbf{z}$. We also denote by $\boldsymbol{P}_{zK}$ the $n \times K$ matrix with the i-th row being $\boldsymbol{p}_K(\mathbf{z}_i)$ and by $\mathbf{P}_{zK} = \boldsymbol{P}_{zK}(\boldsymbol{P}_{zK}'\boldsymbol{P}_{zK})^{-1}\boldsymbol{P}_{zK}'$ the associated projection matrix.

The series estimator of the generated regressor $\boldsymbol{\eta}$ is formed by $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \ldots, \hat{\eta}_n)' = \mathbf{P}_{zK}\mathbf{s}$, where $\hat{\eta}_i = \boldsymbol{p}_K(\mathbf{z}_i)'(\boldsymbol{P}_{zK}'\boldsymbol{P}_{zK})^{-1}\boldsymbol{P}_{zK}'\mathbf{s}$ and $\mathbf{s} = (s_1, \ldots, s_n)'$. Denote $\mathbf{w}_a = (\mathbf{x}', \eta)'$ and similarly $\mathbf{w}_{ai} = (\mathbf{x}_i', \eta_i)'$ and $\hat{\mathbf{w}}_{ai} = (\mathbf{x}_i', \hat{\eta}_i)'$. We also denote by $\mathbf{W}_a = (\mathbf{X}, \boldsymbol{\eta})$ and $\hat{\mathbf{W}}_a = (\mathbf{X}, \hat{\boldsymbol{\eta}})$ the $n \times (d_x + 1)$ matrices of observations with their i-th columns being $\mathbf{w}_{ai}'$ and $\hat{\mathbf{w}}_{ai}'$, respectively. We then estimate the finite-dimensional parameters $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \alpha)'$ by regressing y_i on $\hat{\mathbf{w}}_{ai}$, i.e.:

$$\hat{\boldsymbol{\gamma}} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\alpha} \end{pmatrix} = \left(\hat{\mathbf{W}}_a'\hat{\mathbf{W}}_a\right)^{-1}\hat{\mathbf{W}}_a'\mathbf{y}. \tag{2.3}$$

In the next two sections, we extend the proposed two-step series method to more flexible semiparametric regression models.

2.2 Asymptotic results

We make the following assumptions for the first-order asymptotic results. Let $\mathcal{X} \subseteq \mathcal{R}^{d_x}$ and $\mathcal{Z} \subseteq \mathcal{R}^{d_z}$ denote the supports of $\mathbf{x}$ and $\mathbf{z}$, respectively. Define $E(\varepsilon^2|\mathbf{x}, \mathbf{z}) = \sigma_\varepsilon^2(\mathbf{x}, \mathbf{z})$.

[3]More specifically, consider the case of three covariates (x_1, x_2, x_3). One could think of $\mathbf{x}$ as $(x_1, x_2)'$ and $\mathbf{z}$ as $(x_2, x_3)'$ with the overlapping variable x_2.

Assumption 2.1. *Assume that:*

(i) $E(\varepsilon|\mathbf{x},\mathbf{z}) = 0$ *almost surely in* $\mathbf{x}$ *and* $\mathbf{z}$*;*

(ii) $\sup_{\mathbf{x}\in\mathcal{X},\mathbf{z}\in\mathcal{Z}} \sigma_\varepsilon^2(\mathbf{x},\mathbf{z}) < \infty$*;*

(iii) $\sup_{\mathbf{x}\in\mathcal{X},\mathbf{z}\in\mathcal{Z}} E(|\varepsilon|^{2+\delta}|\mathbf{x},\mathbf{z}) < \infty$ *for some* $\delta > 0$*.*

Assumption 2.1 imposes the conditional mean zero and the bounded second and third moments of the error term ε conditioning on $\mathbf{x}$ and $\mathbf{z}$. We note that Assumption 2.1 also allows for conditional heteroskedasticity.

Assumption 2.2. *Assume that:*

(i) $\mathcal{Z}$ *is a Cartesian product of compact connected interval;*

(ii) the probability density function of $\mathbf{z}$ *on* $\mathcal{Z}$ *is bounded away from zero;*

(iii) $\eta(\mathbf{z})$ *is continuously differentiable of order* s_z *on* $\mathcal{Z}$*;*

(iv) $\boldsymbol{p}_K(\cdot)$ *is formed by a power series with the number of the series terms* K *satisfying* $n^{-1}K^2 \to 0$ *and* $nK^{-2s_z/d_z} \to 0$ *as* $n \to \infty$*.*

Assumption 2.2 imposes conditions for the auxiliary regression and restrictions on the growth rate of the number of series terms regarding the covariates $\mathbf{z}$ in the auxiliary regression as the sample size increases. Denote $\mathbf{x}_z(\mathbf{z}) \equiv E(\mathbf{x}|\mathbf{z})$ so that we can write $\mathbf{x} = E(\mathbf{x}|\mathbf{z}) + \mathbf{v} = \mathbf{x}_z(\mathbf{z}) + \mathbf{v}$, where $E(\mathbf{v}|\mathbf{z}) = \mathbf{0}$ almost surely by construction. We also define $E(\mathbf{v}^2|\mathbf{z}) = \sigma_v^2(\mathbf{z})$.

Assumption 2.3. *Assume that:*

(i) $E(\mathbf{v}|\mathbf{z}) = \mathbf{0}$ *almost surely;*

(ii) $\sup_{\mathbf{z}\in\mathcal{Z}} \sigma_v^2(\mathbf{z}) < \infty$*;*

(iii) The d_x*-dimensional vector of functions* $\mathbf{x}_z(\mathbf{z})$ *is continuously differentiable of order* s_z *on* $\mathcal{Z}$*.*

Assumption 2.3 limits the behavior of $\mathbf{x}$ conditioning on $\mathbf{z}$ and also allows for conditional heteroskedasticity.

Assumption 2.4. *Assume that:*

(i) $\boldsymbol{\Phi}_a = E(\mathbf{w}_{ai}\mathbf{w}_{ai}')$ *is positive definite;*

(ii) $E(\|\mathbf{x}\|^{2+\delta}) < \infty$ *for some* $\delta > 0$*;*

Assumption 2.4 imposes the identification condition and moment conditions.

Theorem 2.1. *Suppose Assumptions 2.1, 2.2, 2.3, and 2.4 hold. The asymptotic distribution of the series estimator* $\hat{\boldsymbol{\gamma}}$*, given by* (2.3)*, of the finite-dimensional parameters* $\boldsymbol{\gamma}$ *in model* (2.1)-(2.2) *is given by:*

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_a), \tag{2.4}$$

where $\mathbf{V}_a = \boldsymbol{\Phi}_a^{-1}(\boldsymbol{\Omega}_{a1} + \boldsymbol{\Omega}_{a2} - 2\boldsymbol{\Omega}_{a12})\boldsymbol{\Phi}_a^{-1}$ *with:*

$$\begin{aligned}
\boldsymbol{\Omega}_{a1} &= E\big(\mathbf{w}_a \mathbf{w}_a' \varepsilon^2\big), \\
\boldsymbol{\Omega}_{a2} &= \alpha^2 E\big(\big(\mathbf{x}_z(\mathbf{z})', \eta(\mathbf{z})\big)\big(\mathbf{x}_z(\mathbf{z})', \eta(\mathbf{z})\big)' u^2\big), \\
\boldsymbol{\Omega}_{a12} &= \alpha E\big(\mathbf{w}_a(\mathbf{x}_z(\mathbf{z})', \eta(\mathbf{z}))\varepsilon u\big).
\end{aligned} \tag{2.5}$$

The proof of Theorem 2.1 is in Appendix. Our proof strategy is as follows. We decompose $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ into two terms: $n^{-1/2}(\hat{\mathbf{W}}_a' \hat{\mathbf{W}}_a)^{-1} \hat{\mathbf{W}}_a' \mathbf{u}$ representing sampling errors and $n^{-1/2}(\hat{\mathbf{W}}_a' \hat{\mathbf{W}}_a)^{-1} \hat{\mathbf{W}}_a' (\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})\alpha$ accounting for approximation errors. We then show that the first and second terms converge in distribution to $N(0, \boldsymbol{\Phi}_a^{-1}\boldsymbol{\Omega}_{a1}\boldsymbol{\Phi}_a^{-1})$ and to $N(0, \boldsymbol{\Phi}_a^{-1}\boldsymbol{\Omega}_{a2}\boldsymbol{\Phi}_a^{-1})$, respectively. Moreover, the asymptotic covariance between these two terms is represented by $-2\boldsymbol{\Omega}_{a12}$. When comparing our estimator under the parametric setting with Stengos and Yan's (2001) double kernel method, we find that our results are very similar. Employing the series method we could also establish the $\sqrt{n}$ consistency and asymptotic normality properties. The main difference is that we relax Stengos and Yan's (2001) assumption on conditional homoskedasticity. Donald (1992) considers a similar model using the series method to estimate $\boldsymbol{\gamma}$, assuming that there is no correlation between the error terms ε and u. Therefore, the framework we consider here contains Stengos and Yan's (2001) model and Donald's (1992) model as special cases.

Remark 1. Regarding the required growth rate of power series, here we give a combination of s_z and K such that the conditions in Assumption 2.4 hold. If $s_z = d_z + 1$, then we can pick $K = a \cdot n^{1/2-\delta}$ for $a > 0$ and small $\delta > 0$. We then have $K^2/n = O(n^{-2\delta}) = o(1)$ and $n \cdot K^{-2(d_z+1)/d_z} = O(n^{(2+2/d_z)\delta - 2/d_z}) = o(1)$ when δ is small enough. We also note that the smoother the functions $\eta(\mathbf{z})$ and $\mathbf{x}_z(\mathbf{z})$ are, the slower rate of K that is required. For example, if $s_z = 2d_z$, then we have $K = a \cdot n^b$ for some $a > 0$ and $1/4 < b < 1/2$.

2.3 Consistent variance estimation

We now discuss how to estimate the asymptotic variance $\mathbf{V}_a$ of $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ in Theorem 2.1 and establish its consistency. Given the formula for the asymptotic variance provided in Theorem 2.1, we can simply replace the population components with their respective sample counterparts. Specifically, $\hat{\mathbf{\Phi}}_a = n^{-1}\hat{\mathbf{W}}_a'\hat{\mathbf{W}}_a$ and the term $\mathbf{\Omega}_{a1}$ is estimated by White's (1980) method, i.e.:

$$\hat{\mathbf{\Omega}}_{a1} = \frac{1}{n}\hat{\mathbf{W}}_a'\, diag(\hat{\boldsymbol{\varepsilon}}^2)\, \hat{\mathbf{W}}_a,$$

where $diag(\hat{\boldsymbol{\varepsilon}}^2)$ denotes the diagonal matrix with diagonal elements $\hat{\varepsilon}_i^2$, where $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}} - \hat{\eta}_i\hat{\alpha}$. Similarly, $\mathbf{\Omega}_{a2}$ is estimated by $\hat{\mathbf{\Omega}}_{a2}$:

$$\hat{\mathbf{\Omega}}_{a2} = \frac{\hat{\alpha}^2}{n}\left(\hat{\mathbf{X}}_z,\ \hat{\boldsymbol{\eta}}\right)' diag(\hat{\mathbf{u}}^2)\left(\hat{\mathbf{X}}_z,\ \hat{\boldsymbol{\eta}}\right),$$

where $\hat{\mathbf{X}}_z = \mathbf{P}_{zK}\mathbf{X}$ is the estimator for the matrix $\mathbf{X}_z$ with its i-th row being $\mathbf{x}_z(\mathbf{z}_i)'$ and $diag(\hat{\mathbf{u}}^2)$ is the diagonal matrix with diagonal elements $\hat{\mathbf{u}}^2 = (\hat{u}_1^2, \ldots, \hat{u}_n^2)$ with $\hat{u}_i = s_i - \hat{\eta}(\mathbf{z}_i)$. Lastly, the covariance term $\mathbf{\Omega}_{a12}$ is estimated by:

$$\hat{\mathbf{\Omega}}_{a12} = \frac{\hat{\alpha}}{n}\left(\hat{\mathbf{X}}_z,\ \hat{\boldsymbol{\eta}}\right)' diag(\hat{\mathbf{u}})\, diag(\hat{\boldsymbol{\varepsilon}})\, \hat{\mathbf{W}}_a.$$

By the standard argument, it is not hard to show that $\hat{\mathbf{\Phi}}_a \xrightarrow{p} \mathbf{\Phi}_a$, $\hat{\mathbf{\Omega}}_{a1} \xrightarrow{p} \mathbf{\Omega}_{a1}$, $\hat{\mathbf{\Omega}}_{a2} \xrightarrow{p} \mathbf{\Omega}_{a2}$, $\hat{\mathbf{\Omega}}_{a12} \xrightarrow{p} \mathbf{\Omega}_{a12}$, and as a result, $\hat{\mathbf{\Phi}}_a^{-1}(\hat{\mathbf{\Omega}}_{a1} + \hat{\mathbf{\Omega}}_{a2} - 2\hat{\mathbf{\Omega}}_{a12})\hat{\mathbf{\Phi}}_a^{-1} \xrightarrow{p} \mathbf{\Phi}_a^{-1}(\mathbf{\Omega}_{a1} + \mathbf{\Omega}_{a2} - 2\mathbf{\Omega}_{a12})\mathbf{\Phi}_a^{-1}$. We omit the proof for brevity.

3 Partially linear model with linear nonparametrically generated regressors

This section considers a more general semiparametric partially linear model than models (2.1)-(2.2), in the sense that, like model (2.1), the generated regressor appears in the linear part, but, in contrast to (2.2), the observed covariates $\mathbf{x}$ are allowed to enter the model in a nonparametric manner; that is, the parametric specification $\mathbf{x}'\beta$ in model (2.1) is replaced by an unknown but smooth function $\theta(\mathbf{x})$. To be explicit, we state:

$$y_i = \theta(\mathbf{x}) + \eta\alpha + \varepsilon, \tag{3.1}$$

where $E(\varepsilon|\mathbf{x}, \mathbf{z}) = 0$ almost surely and $\eta = \eta(\mathbf{z}) = E(s|\mathbf{z})$ is a generated regressor taking the same form as model (2.2). The parameters to be estimated are the finite-dimensional parameter α and the unknown smooth function θ. To enhance applicability of the model, we also allow for more observable (continuous or discrete) covariates in the linear component.

The conventional partially linear models (where η is observable) were first introduced by Engle, Granger, Rice, and Weiss (1986) and have been extensively studied in econometrics and statistics literature, e.g., Heckman (1986), Speckman (1988), Chen (1988), and Robinson (1988), to name just a few. We refer interested readers to Härdle, Liang, and Gao (2000) for a comprehensive theoretical and applied treatment of partially linear regression techniques. Examples of applications of partially linear specifications in economics include Banerjee and Duflo (2003) to income inequality and economic growth, Blundell and Windmeijer (2000) to demand for health services, and Carneiro, Heckman, and Vytlacil (2011) to estimating marginal return to education, among many others.

Models (3.1) and (2.2) generalize the conventional partially linear models to the case where the linear component includes generated regressors. We are interested in estimating the finite-dimensional parameter α and the nonparametric component θ. To this end, we suggest a two-step series estimation procedure that is described below.

3.1 The two-step series estimator

This subsection proposes a two-step series estimation of α and $\theta(\cdot)$ in model (3.1). In the first step, as in Section 2, the series estimates $\hat{\boldsymbol{\eta}}$ of the generated regressor $\boldsymbol{\eta}$ are given by $\hat{\boldsymbol{\eta}} = \mathbf{P}_{zK}\mathbf{s}$. In the second step, once again, we employ series expansions to approximate the nonparametric component θ in model (3.1). Specifically, denote by $L = L_n$ the number of series terms with respect to $\mathbf{x}$, which is assumed to increase with sample size n. We also denote the associated projection matrix by $\mathbf{P}_{xL} = \boldsymbol{Q}_{xL}(\boldsymbol{Q}_{xL}'\boldsymbol{Q}_{xL})^{-1}\boldsymbol{Q}_{xL}'$, where $\boldsymbol{Q}_{xL}$ is the $n \times L$ matrix whose i-th row is $\boldsymbol{q}_L(\mathbf{x}_i)' = (q_1(\mathbf{x}_i)', \ldots, q_L(\mathbf{x}_i)')$.

Using the formula for the partitioned regression, our two-step series estimator for the finite-dimensional parameter α can be expressed as:

$$\hat{\alpha} = \left(\hat{\boldsymbol{\eta}}'(\mathbf{I} - \mathbf{P}_{xL})\hat{\boldsymbol{\eta}}\right)^{-1} \hat{\boldsymbol{\eta}}'(\mathbf{I} - \mathbf{P}_{xL})\mathbf{y}, \tag{3.2}$$

where $\mathbf{I}$ denotes the conformable identity matrix.

Given $\hat{\alpha}$ obtained from (3.2), it is then straightforward to obtain the series estimator of $\theta(\bar{\mathbf{x}})$ at a fixed value $\mathbf{x} = \bar{\mathbf{x}}$:

$$\hat{\theta}(\bar{\mathbf{x}}) = \boldsymbol{q}_L(\bar{\mathbf{x}}) \left(\boldsymbol{Q}_{xL}'\boldsymbol{Q}_{xL}\right)^{-1} \boldsymbol{Q}_{xL}'(\mathbf{y} - \hat{\boldsymbol{\eta}}\hat{\alpha}). \tag{3.3}$$

The estimated functions of $\boldsymbol{\theta}$ evaluated at the data points are collected in $\hat{\boldsymbol{\theta}}$ $=$

$(\hat{\theta}(\mathbf{x}_1), \ldots, \hat{\theta}(\mathbf{x}_n))'$.

3.2 Asymptotic results

To establish the first-order asymptotic properties of series estimators $\hat{\alpha}$ and $\hat{\theta}(\mathbf{x}_i)$, we make the following assumptions.

Assumption 3.1. *Assume that:*

(i) $\mathcal{X}$ *is a Cartesian product of compact connected interval;*

(ii) the probability density function of x *on* $\mathcal{X}$ *is bounded away from zero;*

(iii) $\theta(\mathbf{x})$ *is continuously differentiable of order* s_x *on* $\mathcal{X}$*.*

Assumption 3.1 imposes conditions on the observed regressors $\mathbf{x}$ and the function θ. This assumption is not needed for the parametric case. When it comes to the semiparametric case, this assumption is required due to the fact that $\theta(\mathbf{x})$ is estimated nonparametrically. We introduce the following notation by letting $\eta = E(\eta|\mathbf{x}) + \zeta \equiv \eta_x + \zeta$, $\zeta = E(\zeta|\mathbf{z}) + \xi \equiv \zeta_z(\mathbf{z}) + \xi$, $\sigma_\zeta^2(\mathbf{x}) = E(\zeta^2|\mathbf{x})$, and $\sigma_\xi^2(\mathbf{z}) = E(\xi^2|\mathbf{z})$.

Assumption 3.2. *Assume that:*

(i) $\sup_{\mathbf{x}\in\mathcal{X}} \sigma_\eta^2(\mathbf{x}) < \infty$ *and* $\sup_{\mathbf{z}\in\mathcal{Z}} \sigma_\xi^2(\mathbf{z}) < \infty$*;*

(ii) $E(\eta|\mathbf{x})$ *is continuously differentiable of order* s_x *on* $\mathcal{X}$ *and* $\zeta_z(\mathbf{z})$ *is continuously differentiable of order* s_z *on* $\mathcal{Z}$*.*

Assumption 3.2 limits the behavior of $\eta(\mathbf{z})$ conditioning on $\mathbf{x}$ and that of ζ conditioning on $\mathbf{z}$.

Assumption 3.3. *Assume* $\Phi_b = E((\eta - E(\eta|\mathbf{x}))^2) > 0$*.*

Assumption 3.4. *Assume that* $\boldsymbol{q}_L(\cdot)$ *is formed by a power series with the number of series terms* $L = L_n$ *satisfying* $n^{-1}L_n^2 \to 0$ *and* $nL_n^{-2s_x/d_x} \to 0$ *as* $n \to \infty$*.*

Assumption 3.3 is the identification condition. Assumption 3.4 restricts the rate at which the number of series terms regarding the covariates x grows as the sample size increases.

Theorem 3.1. *Suppose Assumptions 2.1, 2.2, 3.1, 3.2, 3.3, and 3.4 hold. The asymptotic distribution of the second-step series estimator* $\hat{\alpha}$*, defined by* (3.2)*, of the finite-dimensional parameter* α *in models* (3.1) *and* (2.2) *is given by:*

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, V_b),$$

where $V_b = \Phi_b^{-1}(\Omega_{b1} + \Omega_{b2} - 2\Omega_{b12})\Phi_b^{-1}$ *with:*

$$\begin{aligned}
\Omega_{b1} &= E((\eta - E(\eta|\mathbf{x}))^2\varepsilon^2),\\
\Omega_{b2} &= \alpha^2 E(\zeta_z(\mathbf{z})^2 u^2),\\
\Omega_{b12} &= \alpha E((\eta - E(\eta|\mathbf{x}))\zeta_z(\mathbf{z})\varepsilon u),\\
\zeta_z(\mathbf{z}) &= E(\eta - E(\eta|\mathbf{x})|\mathbf{z}).
\end{aligned}$$

Furthermore, we state:

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}(\mathbf{x}_i) - \theta(\mathbf{x}_i))^2 = O_p\left(\frac{L}{n} + \frac{K}{n}\right).$$

We note that the remark on Theorem 2.1 at the end of Section 2.2 applies here.

3.3 Consistent variance estimation

We now consider the estimation of the asymptotic variance of $\sqrt{n}(\hat{\alpha} - \alpha)$ provided in Theorem 3.1. Following the strategy in the previous section, one could simply replace the population values with their sample counterparts. Specifically, let $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\eta}}\hat{\alpha}$, $\hat{\boldsymbol{\eta}}_x = \mathbf{P}_{xL}\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{u}} = \mathbf{s} - \hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\zeta}} = \hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\eta}}_x$, and $\hat{\boldsymbol{\zeta}}_z = \mathbf{P}_{zK}\hat{\boldsymbol{\zeta}}$. We also define $\hat{\Phi}_b = n^{-1}(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\eta}}_x)'(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\eta}}_x)$, $\hat{\Omega}_{b1} = n^{-1}(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\eta}}_x)' diag(\hat{\boldsymbol{\varepsilon}}^2)(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\eta}}_x)$, $\hat{\Omega}_{b2} = n^{-1}\hat{\alpha}^2\hat{\boldsymbol{\zeta}}_z' diag(\hat{\boldsymbol{u}}^2)\hat{\boldsymbol{\zeta}}_z$, and $\hat{\Omega}_{b12} = n^{-1}\hat{\alpha}(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\eta}}_x)' diag(\hat{\boldsymbol{\varepsilon}}) diag(\hat{\boldsymbol{u}})\hat{\boldsymbol{\zeta}}_z$. It is straightforward to show that $\hat{\Phi}_b \xrightarrow{p} \Phi_b$, $\hat{\Omega}_{b1} \xrightarrow{p} \Omega_{b1}$, $\hat{\Omega}_{b12} \xrightarrow{p} \Omega_{b12}$, and $\hat{\Omega}_{b2} \xrightarrow{p} \Omega_{b2}$. It then follows that $\hat{\Phi}_b^{-1}(\hat{\Omega}_{b1} + \hat{\Omega}_{b2} - 2\hat{\Omega}_{b12})\hat{\Phi}_b^{-1} \xrightarrow{p} \Phi_b^{-1}(\Omega_{b1} + \Omega_{b2} - 2\Omega_{b12})\Phi_b^{-1}$, as desired.

4 Partially linear model with nonlinear nonparametrically generated regressors

In the previous two sections we take on a linear specification and a partially linear specification, in which for both we let generated regressors enter the model in a linear manner. In this section we allow generated regressors to enter the model nonlinearly and nonparametrically to capture nonlinear patterns of generated regressors; specifically, we consider a model of the following form:

$$y = \mathbf{x}'\boldsymbol{\beta} + \theta(\eta) + \varepsilon, \tag{4.1}$$

where once again, we consider only the scalar generated regressor for notational simplicity, i.e., $\eta = \eta(\mathbf{z}) = E(s|\mathbf{z})$. The goal is to estimate the finite-dimensional parameter vector $\boldsymbol{\beta}$ and the unknown function θ.

The partially linear model in the presence of generated regressors has been addressed by several papers. Li and Wooldridge (2002) consider the model (4.1) with a parametrically generated regressor for dependent data and propose a kernel-based efficient estimator for $\boldsymbol{\beta}$. Ahn and Powell (1993) offer a pairwise-differencing estimator of the single-index coefficients in a semiparametric sample selection models where generated regressors take the form of selection probabilities arising from a nonparametric selection mechanism. Zhou and Liang (2009) study a partially linear model with nonparametric components in the form of varying-coefficients when some linear covariates are not observed and develop a profile least-square based estimator.

Most of the aforementioned studies on partially linear models with generated regressors use kernel-based methods. For using series methods to estimate the model with partially linear structures, Donald and Newey (1994) address series estimation in a partially linear model (4.1) under weak smoothness conditions. They also allow for η being multidimensional and discretely distributed. Li (2000) imposes an additive structure on the nonparametric component θ in (4.1) in the absence of generated regressors. Newey (2009) considers a semiparametric sample selection model where the generated regressor η depends on unknown finite-dimensional parameters. Newey, Powell, and Vella (1999) look at a two-step nonparametric series estimator for an additive model where the first-stage residual (as a nonparametric generated regressor) enters the unknown additive component. Das, Newey, and Vella (2003) apply Newey's (2009) estimation procedure to nonparametric sample selection models. Motivated by Newey, Powell, and Vella's (1999) nonparametric triangular mean regression model, Lee (2007) proposes a two-step series estimator for the partially linear quantile regression with a parametrically generated regressor. To our knowledge, the series estimation for the partially linear model in which a nonparametrically generated regressor η enters the nonparametric component appears missing in the literature. This section aims to fill this gap.

4.1 The two-step series estimator

We apply the two-step series estimation method proposed in the present paper to estimate model (4.1). Specifically, in the first step we estimate the generated regressor $\eta_i = E(s_i|\mathbf{z}_i)$ by series approximations and then, in the second stage, approximate the nonparametric function θ by a linear combination of M series functions $\boldsymbol{r}_M(\eta)'\boldsymbol{\alpha}_\eta$ using the series estimator

$\hat{\eta}$ plugged in from the first stage, where $\boldsymbol{r}_M(\eta) = (r_1(\eta), \ldots, r_M(\eta))'$ is a $M \times 1$ vector of basis functions and $\boldsymbol{\alpha}_\eta$ is the corresponding M-dimensional vector of series coefficients. For any given $\eta = \bar{\eta}$, the series approximation to $\theta(\bar{\eta})$ is $\theta_M(\bar{\eta}) = \boldsymbol{r}_M(\bar{\eta})'\boldsymbol{\alpha}_\eta$ with the corresponding approximation error $e_M(\bar{\eta}) = \theta(\bar{\eta}) - \theta_M(\bar{\eta})$. The series estimator of the coefficients $(\boldsymbol{\beta}', \boldsymbol{\alpha}_\eta')'$ is the least squares estimators of regressing y_i on $\hat{\mathbf{w}}_{ci} \equiv (\mathbf{x}_i', \boldsymbol{r}_M(\hat{\eta}_i))'$, i.e.:

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}}_\eta \end{pmatrix} = (\hat{\mathbf{W}}_c'\hat{\mathbf{W}}_c)^{-1}\hat{\mathbf{W}}_c'\mathbf{y},$$

where $\hat{\mathbf{W}}_c = (\mathbf{X}, \hat{\boldsymbol{R}}_{\eta M})$ with $\hat{\boldsymbol{R}}_{\eta M}$ being the $n \times M$ matrix whose i-th row is $\boldsymbol{r}_M(\hat{\eta}_i)' = (r_1(\hat{\eta}_i), \ldots, r_M(\hat{\eta}_i))$ and $\hat{\boldsymbol{\alpha}}_\eta = (\hat{\alpha}_\eta^1, \ldots, \hat{\alpha}_\eta^M)'$ is the series estimator of $\boldsymbol{\alpha}_\eta = (\alpha_\eta^1, \ldots, \alpha_\eta^M)'$.

Using the usual formula for partitioned regression, the series estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}_\eta$ can also alternatively be expressed as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'(\mathbf{I} - \hat{\mathbf{P}}_{\eta M})\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \hat{\mathbf{P}}_{\eta M})\mathbf{y}, \tag{4.2}$$

$$\hat{\boldsymbol{\alpha}}_\eta = (\hat{\boldsymbol{R}}_{\eta M}'\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}_{\eta M}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{4.3}$$

where $\hat{\mathbf{P}}_{\eta M} = \hat{\boldsymbol{R}}_{\eta M}(\hat{\boldsymbol{R}}_{\eta M}'\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}_{\eta M}'$ is the associated projection matrix using the estimated generated regressor $\hat{\eta}_i$. On the other hand, the series estimator of the nonparametric function $\theta(\cdot)$ is given by $\hat{\theta}(\cdot) = \boldsymbol{r}_M(\cdot)'\hat{\boldsymbol{\alpha}}_\eta$. Denote $\hat{\boldsymbol{\theta}} = (\hat{\theta}(\hat{\eta}_1), \ldots, \hat{\theta}(\hat{\eta}_n))' = \hat{\boldsymbol{R}}_{\eta M}\hat{\boldsymbol{\alpha}}_\eta = \hat{\mathbf{P}}_{\eta M}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$.

4.2 Asymptotic results

We impose the following assumptions to establish the $\sqrt{n}$-consistency and asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}$.

Assumption 4.1. *Assume that:*

(i) The nonparametric components $\theta(\eta)$ and $\mathbf{x}_\eta \equiv E(\mathbf{x}|\eta)$ are continuously differentiable in η of order s_η;

(ii) The distribution of η is absolutely continuous with respect to the Lebesgue measure. The density of η is bounded away from zero and infinity on the support of η;

(iii) The errors $\varepsilon = y - \mathbf{x}'\boldsymbol{\beta} - \theta(\eta)$, $\mathbf{v} = \mathbf{x} - E(\mathbf{x}|\eta)$, and $u = s - E(s|\mathbf{z})$ satisfy the following moment conditions: $E(|\varepsilon|^{4(1+\delta)}) < \infty$, $E(|u|^{4(1+\delta)}) < \infty$, and $E(\|\mathbf{v}\|^{\mathbf{4(1+\delta)}}) < \infty$.

Assumption 4.2. *Assume that $\boldsymbol{\Phi}_c = E[(\mathbf{x}_i - E(\mathbf{x}_i|\eta_i))(\mathbf{x}_i - E(\mathbf{x}_i|\eta_i))']$ is positive definite.*

Assumption 4.3. *Assume that $\boldsymbol{r}_M(\cdot)$ is formed by a power series with the number of series terms $M = M_n$ satisfying $n^{1/2}(M_n/n + M_n^{-2s_\eta}) \to 0$, $M_n^3/n \to 0$, $n^{-1/2}M_n^2 K_n^{3/2} \to 0$, and $M_n^2 K_n^{1-s_z/d_z} \to 0$ as $n \to \infty$.*

Assumptions 4.1 (i) and (ii) are standard in the series approximation literature. The smoothness condition imposed in part (i) is required to control the bias of the series estimators of $\theta(\cdot)$ and $\mathbf{x}_\eta$. Take $\theta(\cdot)$ for example, it follows from part (i) and Lorentz (1966) that for power series, there exists a unique series representation $\boldsymbol{r}_M(\eta)'\boldsymbol{\alpha}_\eta$ such that the uniform approximation errors to the unknown function $\theta(\cdot)$ and its first derivative shrink at the rates $\sup_\eta|\theta(\eta)-\boldsymbol{r}_M(\eta)'\boldsymbol{\alpha}_\eta| = O(M^{-s_\eta})$ and $\sup_\eta|\partial\theta(\eta)/\partial\eta - \sum_{m=1}^{M}(\partial r_m(\eta)/\partial\eta)\alpha_\eta^m| = O(M^{-s_\eta+1})$, respectively, as $M \to \infty$. Assumptions 4.1 (iii) and 4.2 are standard moment and identification conditions. Assumption 4.3 requires that the mean square error of the series estimator of θ has an order smaller than $n^{-1/2}$, meaning that the asymptotic bias and variance of the series estimation of θ are sufficiently small so that they do not impact the asymptotic distribution of $\hat{\boldsymbol{\beta}}$. The last two rate conditions associated with the interaction between M_n and K_n in Assumption 4.3 are imposed to control the first-step estimation error of the generated regressor η_i. Theorem 4.1 presents the asymptotic distribution for the second-step series estimator $\hat{\boldsymbol{\beta}}$.

Theorem 4.1. *Suppose that Assumptions 2.1, 2.2, 4.1, 4.2 and 4.3 hold. The asymptotic distribution of $\hat{\boldsymbol{\beta}}$, given by (4.2), of the finite-dimensional parameters $\boldsymbol{\beta}$ in models (4.1) and (2.2) is given by:*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \mathbf{V}_c),$$

where $\mathbf{V}_c = \boldsymbol{\Phi}_c^{-1}(\boldsymbol{\Omega}_{c1} + \boldsymbol{\Omega}_{c2} - 2\boldsymbol{\Omega}_{c12})\boldsymbol{\Phi}_c^{-1}$ with:

$$\begin{aligned}
\boldsymbol{\Omega}_{c1} &= E((\mathbf{x} - E(\mathbf{x}|\eta))(\mathbf{x} - E(\mathbf{x}|\eta)'\varepsilon^2), \\
\boldsymbol{\Omega}_{c2} &= E((\mathbf{x} - E(\mathbf{x}|\eta))(\mathbf{x} - E(\mathbf{x}|\eta)'(d\theta(\eta)/d\eta)^2 u^2), \\
\boldsymbol{\Omega}_{c12} &= E((\mathbf{x} - E(\mathbf{x}|\eta))(\mathbf{x} - E(\mathbf{x}|\eta)'(d\theta(\eta)/d\eta)\varepsilon u).
\end{aligned}$$

4.3 Consistent variance estimation

Based on the formula for the asymptotic variance given in Theorem 4.1, a consistent estimate of the asymptotic variance can be constructed by simply replacing the population values with their sample counterparts. Specifically, let $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{x}'\hat{\boldsymbol{\beta}} - \hat{\theta}(\hat{\boldsymbol{\eta}})$ and $\hat{\boldsymbol{u}} = \mathbf{s} - \hat{\boldsymbol{\eta}}$. We also define $\hat{\boldsymbol{\Phi}}_c = n^{-1}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'(\mathbf{X} - \hat{\mathbf{X}}_\eta)$, $\hat{\boldsymbol{\Omega}}_{c1} = n^{-1}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'diag(\hat{\boldsymbol{\varepsilon}}^2)(\mathbf{X} - \hat{\mathbf{X}}_\eta)$, $\hat{\boldsymbol{\Omega}}_{c2} = n^{-1}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'diag((d\hat{\theta}(\hat{\boldsymbol{\eta}})/d\eta)^2)diag(\hat{\boldsymbol{u}}^2)(\mathbf{X} - \hat{\mathbf{X}}_\eta)$, and $\hat{\boldsymbol{\Omega}}_{c12} = n^{-1}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'diag(\hat{\boldsymbol{\varepsilon}})diag((d\hat{\theta}(\hat{\boldsymbol{\eta}})/d\eta))diag(\hat{\boldsymbol{u}})(\mathbf{X} - \hat{\mathbf{X}}_\eta)$. It is then straightforward to show that $\hat{\mathbf{\Phi}}_c \xrightarrow{p} \mathbf{\Phi}_c$, $\hat{\mathbf{\Omega}}_{c1} \xrightarrow{p} \mathbf{\Omega}_{c1}$, $\hat{\mathbf{\Omega}}_{c12} \xrightarrow{p} \mathbf{\Omega}_{c12}$, $\hat{\mathbf{\Omega}}_{c2} \xrightarrow{p} \mathbf{\Omega}_{c2}$ and therefore $\hat{\mathbf{\Phi}}_c^{-1}(\hat{\mathbf{\Omega}}_{c1} + \hat{\mathbf{\Omega}}_{c2} - 2\hat{\mathbf{\Omega}}_{c12})\hat{\mathbf{\Phi}}_c^{-1} \xrightarrow{p} \mathbf{\Phi}_c^{-1}(\mathbf{\Omega}_{c1} + \mathbf{\Omega}_{c2} - 2\mathbf{\Omega}_{c12})\mathbf{\Phi}_c^{-1}$.

5 Specification testing

We next propose a series-based consistent model specification test for the semiparametric model defined in (2.1). Specifically, we are interested in testing the following null hypothesis:

$$H_0 : E(y|\mathbf{x}, \mathbf{z}) = \mathbf{x}'\boldsymbol{\beta} + \eta(\mathbf{z})\alpha \quad \text{almost surely,} \tag{5.1}$$

which is equivalent to $E(\varepsilon|\mathbf{x}, \mathbf{z}) = 0$ almost surely.

Recall that $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}} - \hat{\eta}(\mathbf{z}_i)\hat{\alpha}$ is the residual obtained from the two-step series regression as described in Section 2. Let $\boldsymbol{S}_{wN}$ denote the $n \times N$ matrix that is formed by the power series of $\mathbf{w} \equiv (\mathbf{x}', \mathbf{z}')'$ with the number of series terms equal to N, and denote by $\boldsymbol{s}_{Ni} \equiv \boldsymbol{s}_N(\mathbf{w}_i)$ the vector of the N approximating functions evaluated at $\mathbf{w}_i \equiv (\mathbf{x}_i', \mathbf{z}_i')'$. Our test statistic is defined as:

$$\hat{\mathcal{T}}_n = n\hat{I}_n/\hat{S}_n, \tag{5.2}$$

where

$$\begin{aligned} \hat{I}_n &= n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \hat{\varepsilon}_i \boldsymbol{s}_{Ni}'(\boldsymbol{S}_{wN}'\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\varepsilon}_j, \\ \hat{S}_n &= 2\sum_{i=1}^{n}\sum_{j\neq i}^{n} \left(\boldsymbol{s}_{Ni}'(\boldsymbol{S}_{wN}'\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\right)^2 \hat{\varepsilon}_i^2\hat{\varepsilon}_j^2. \end{aligned} \tag{5.3}$$

We note that our proposed test extends Sun and Li's (2006) test to allow for the presence of the generated regressor. Similar to Sun and Li (2006) test, our test does not require non-zero center terms, which are needed in the Hong and White (1995) tests.

Under regularity conditions and under H_0, we will show that $\hat{\mathcal{T}}_n \xrightarrow{d} N(0,1)$ and, as a result, for a prespecified significance level ℓ, the rejection rule is: Reject H_0, if $\mathcal{T}_n > c_{1-\ell}$, where $c_{1-\ell}$ is the $(1-\ell)$-th quantile of the standard normal.

Assumption 5.1. *Assume that:*

(i) $\mathcal{X}$ is a Cartesian product of compact connected interval;

(ii) The joint probability density function of $(\mathbf{x}, \mathbf{z})$ on $\mathcal{X} \times \mathcal{Z}$ is bounded away from zero;

(iii) Let $N = N_n$ and $K = K_n$ satisfying $N_n^3/n \to 0$, $K_n^2 N_n^{-1/2} \to 0$, and $nK_n^{-2s_z/d_z} N_n^{-1/2} \to 0$.

Assumptions 5.1(i) and (ii) impose conditions on the support of $\mathbf{x}$ and the product of $\mathcal{X}$ and $\mathcal{Z}$. This condition is not needed for estimation, but is required for specification testing. Assumption 5.1(iii) controls the growth rates of N_n and K_n; the first part is the same as that in Sun and Li (2006), which is needed even when there are no generate regressors in the model. The second and third parts in Assumption 5.1(iii) control the relative growth rates of K_n and N_n. These conditions are imposed here to guarantee that the error from the estimated generated regressor is asymptotically negligible in our test statistic.

Theorem 5.1. *Suppose that Assumptions 2.1-2.4 and 5.1 hold. For a prespecified significance level ℓ, then (i) under H_0, $\lim_{n\to\infty} P(\hat{\mathcal{T}}_n > c_{1-\ell}) = 1 - \ell$ and (ii) under H_1, $\lim_{n\to\infty} P(\hat{\mathcal{T}}_n > c_{1-\ell}) = 1$.*

Theorem 5.1 shows that our specification test can control the size asymptotically very well under the null and that our test is consistent against fixed alternatives. We contribute to the specification test literature by extending existing tests to allow for generated regressors in the model. The proof is similar to that of Theorem 1 of Sun and Li (2006), with the main difference arising from the fact that we need to account for the effect from estimating the generated regressors when forming our test.

We note that it is straightforward to adapt the series-based specification test, as suggested above, to the partially linear models with generated regressors considered in Sections 3 and 4. We omit the details here for brevity.

6 Simulations

This section conducts Monte Carlo simulation experiments to examine the finite-sample performance of our two-step series estimators. The data generating processes (DGPs) considered have the following form:

$$y = \theta_1(x) + \theta_2(\eta) + \varepsilon, \tag{6.1}$$

$$x = z_1 + z_2 + v, \tag{6.2}$$

$$\eta(z) = E(s|z_1, z_2) = (z_1 + z_2)^2, \tag{6.3}$$

$$s = \eta(z_1, z_2) + u, \tag{6.4}$$

where $(z_1, z_2, \varepsilon, u, v)$ are generated from independent standard normal distributions. In our designs, the observed regressor x is allowed to be correlated with the generated regressor η

through z_1 and z_2, as specified in (6.2) and (6.3).

We consider eight DGPs, depending on the specifications of the functions $\theta_1(\cdot)$ and $\theta_2(\cdot)$. Specifically, DGPs 1-4 are designed as the partially linear specification with a linear generated regressor, i.e, $\theta_1(x)$ is an unknown function, and $\theta_2(\eta) = \alpha\eta$, whereas DGPs 5-8 corresponds to partially linear models with a nonlinear generated regressor, i.e., $\theta_1(x) = \beta x$, and $\theta_2(\eta)$ is a nonparametric component. The true values of the finite-dimensional parameters (α, β) are $(1, 1)$. We note that DGP 1 and DGP 5 with either linear $\theta_1(\cdot)$ or linear $\theta_2(\cdot)$ correspond to linear models with a generated regressor.

For the nonparametric components, borrowing from Henderson and Parmeter (2015), the specifications of θ_1 and θ_2 cover several functional shapes, as summarized in Table 1.

Table 1: The functional forms of θ_1 and θ_2 in DGPs 1-8

	$\theta_1(x)$	$\theta_2(\eta)$
DGP 1	$0.8 + 0.7\ x$	$\alpha\eta$
DGP 2	$2.75\frac{\exp(-3x)}{1+\exp(-3x)} - 1$	$\alpha\eta$
DGP 3	$0.7x + 1.4\exp(-16x^2)$	$\alpha\eta$
DGP 4	$\sqrt{x+10}$	$\alpha\eta$
DGP 5	βx	$0.8 + 0.7\ \eta$
DGP 6	βx	$2.75\frac{\exp(-3\eta)}{1+\exp(-3\eta)} - 1$
DGP 7	βx	$0.7\eta + 1.4\exp(-16\eta^2)$
DGP 8	βx	$\sqrt{\eta+10}$

Note: The semiparametric models exhibit the form $y = \theta_1(x) + \theta_2(\eta) + u$, where the linear components are $\theta_2(\eta) = \alpha\eta$ for DGPs 1-4 and $\theta_1(x) = \beta x$ for DGPs 5-8. DGPs 1-4 and 5-8 correspond to the partially linear models with a linear and a nonlinear generated regressor, respectively. The true values of (α, β) are $(1, 1)$.

For estimation, we consider the sample sizes of $n = 250$, 500, $1,000$, and $4,000$. The number of simulation replications is set to $10,000$. We consider the polynomials as basis functions to approximate the unknown functions θ_1, θ_2, and η; specifically, the polynomials of degree 2 are $(1, x, x^2)$ for x, $(1, \eta, \eta^2)$ for η, and $(1, z_1, z_2, z_1^2, z_2^2, z_1z_2)$ for $\mathbf{z} = (z_1, z_2)'$. The polynomial orders we consider are $1, 2, \ldots, 8$, and the number of the series terms is determined by minimizing BIC (Bayesian Information Criterion). To investigate the performance of the estimators, we report their variances (VAR) and Mean Square Error (MSE). For the purpose of comparison, we consider the fully infeasible, partially infeasible, and feasible series estimators, as described below, depending on using true or estimated values of the unknown functions.

1. Fully infeasible series estimator (FISE): the FISE estimator is obtained by regressing $y_i - \theta_1(x_i)$ on η_i in DGPs 1-4 and by regressing $y_i - \theta_2(\eta_i)$ on x_i in DGPs 5-8, assuming all nonparametric components in the models are known, including the unknown function $\theta_1(\cdot)$ and the generated regressor η in DGPs 1-4 and the unknown function $\theta_2(\cdot)$ and the generated regressor η in DGPs 5-8. We denote by $\bar{\alpha}$ and $\bar{\beta}$ the resulting estimators of the finite-dimensional parameters α and β, respectively.

2. Partially infeasible series estimator (PISE): this estimator directly takes the true value of the generated regressor $\eta(\mathbf{z}) = E(s|z_1, z_2)$ instead of estimating it from the auxiliary regression and then uses the series method to estimate the finite-dimensional parameter and the unknown function in the main regression (6.1) simultaneously. We denote by $\tilde{\alpha}$ and $\tilde{\beta}$ the resulting estimators of the finite-dimensional parameters α and β, respectively.

3. The proposed two-step series estimators (2SSE) which treat the nonparametric components η, θ_1, and θ_2 in the models as unknown and estimate these unknown by series methods.

The simulation results are summarized in Tables 2 and 3. Overall, the proposed 2SSEs perform satisfactory in finite samples; the estimation bias and RMSE for the finite- and infinite-dimensional parameters decrease as the sample size increases, indicating consistency of the proposed methods. When compared to infeasible series estimators FISE and PISE, we find that, not surprisingly, the RMSE of feasible 2SSE is uniformly larger across sample sizes and DGPs than that of the infeasible counterparts. In particular, this efficiency loss arising from using series estimates of nonparametric components is more prominent in DGPs 1-4 than that in DGPs 5-8.

We further see from Figure 1 that, interestingly, in most cases the RMSE of the 2SSE $\hat{\alpha}$ of α, the coefficient of the generated regressor, in DGPs 1-4 is largely attributed to the estimation error of the generated regressor relative to the series estimation error of the unknown function $\theta_1(x)$; specifically, it accounts for 94.5% (DGP 1), 75.5% (DGP 2), 74.8% (DGP 3), and 78.5% (DGP 4) in the small sample of size $n = 250$ and 100.0% (DGP 1), 88.1% (DGP 2), 94.0% (DGP 3), and 88.5% (DGP 4) when $n = 4,000$. These figures corresponding to the estimation of the finite-dimensional parameter β in DGPs 5-8 are 92.0% (DGP 5), 32.4% (DGP 6), 65.1% (DGP 7), and 61.5% (DGP 8) under $n = 250$ and 100.0% (DGP 5), 66.7% (DGP 6), 85.7% (DGP 7), and 100.0% (DGP 8) under $n = 4,000$. This pattern clearly reveals that, in our contexts, estimating the generated regressor plays a more important role than estimating the unknown functions (i.e., $\theta_1(\cdot)$ in DGPs 1-4 and $\theta_2(\cdot)$ in DGPs 5-8) in contributing to the RMSE of the estimation of the finite-dimensional parameters.

Figure 2 presents the simulated distributions of the proposed specification test statistic $\hat{\mathcal{T}}_n$ for the sample sizes $n = 250$, 500, and $1,000$. It appears that the simulated distributions are close to the standard normal distribution, which is consistent with the normality result stated in Theorem 5.1.

7 Conclusion

In this paper, we propose the two-step series estimation methods on semiparametric regression models with generated regressors. The semiparametric models considered include linear and partially linear specifications with either linear or nonlinear generated regressors, which are frequently encountered in practice and should be of practical interest. Our two-step series methods allow for conditional heteroskedasticity as well as the possible correlation between error terms in the regression of interest and auxiliary regression, which are not simultaneously investigated in the previous literature. We establish $\sqrt{n}$ consistency and asymptotic normality for our proposed two-step series estimators of the finite-dimensional parameters. Based on the proposed estimation methods, we also provide a new nonparametric specification test. In addition to theoretical justification, a simulation study reveals that our proposed estimators perform well in finite samples.

Several areas are worth exploring for future research. It is possible to generalize the result of this paper (using the series method) to the case where the nonparametric component of the regression function has an additive structure, leading to additive partially linear models with generated regressors. It would also be interesting to investigate the series estimation for the regression function under shape restrictions (such as monotonicity, nonnegativity, or concavity) in the presence of generated regressors. It would be useful to find data-based methods for selecting the number of the series terms in practice, by reflecting the fact that the MSE of the second-stage series estimator is affected by the first stage.

A Appendix

A.1 Proof of Theorem 2.1

In what follows, we use $\eta_i = \eta(z_i)$, $\hat{\eta}_i = \hat{\eta}(z_i)$ and $h_i = h(z_i)$ for notational simplicity. First note that:

$$
\begin{aligned}
\hat{\boldsymbol{\gamma}} &= (\hat{\mathbf{W}}_a'\hat{\mathbf{W}}_a)^{-1}\hat{\mathbf{W}}_a'\mathbf{y} \\
&= \boldsymbol{\gamma} + (\hat{\mathbf{W}}_a'\hat{\mathbf{W}}_a)^{-1}\hat{\mathbf{W}}_a'(\boldsymbol{\varepsilon} + (\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})\alpha).
\end{aligned}
$$

Therefore, we have:

$$
\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = \left(\frac{\hat{\mathbf{W}}_a'\hat{\mathbf{W}}_a}{n}\right)^{-1} \frac{1}{\sqrt{n}}\hat{\mathbf{W}}_a'((\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})\alpha + \boldsymbol{\varepsilon}).
$$

We first want to show that $\hat{\mathbf{W}}_a'\hat{\mathbf{W}}_a/n$ converges in probability to $\boldsymbol{\Phi}_a$. To show this, consider the following decomposition:

$$
\frac{\hat{\mathbf{W}}_a'\hat{\mathbf{W}}_a}{n} = \frac{\mathbf{W}_a'\mathbf{W}_a}{n} + \frac{(\hat{\mathbf{W}}_a - \mathbf{W}_a)'\mathbf{W}_a}{n} + \frac{\mathbf{W}_a'(\hat{\mathbf{W}}_a - \mathbf{W}_a)}{n} + \frac{(\hat{\mathbf{W}}_a - \mathbf{W}_a)'(\hat{\mathbf{W}}_a - \mathbf{W}_a)}{n}. \tag{A.1}
$$

By the law of large number, we have $\mathbf{W}_a'\mathbf{W}_a/n \xrightarrow{p} E(\mathbf{w}_a\mathbf{w}_a') = \boldsymbol{\Phi}_a$. We next claim that the remaining terms on the right-hand side of (A.1) are all $o_p(1)$. It is sufficient to show that $\sum_{i=1}^n x_{ji}(\eta_i - \hat{\eta}_i)/n = o_p(1)$ for all $j = 1, \ldots, d_x$ and $\sum_{i=1}^n(\eta_i - \hat{\eta}_i)^2/n = o_p(1)$. Under the assumptions, we can apply Theorem 4 of Newey (1997) to establish $n^{-1}\sum_{i=1}^n |\eta_i - \hat{\eta}_i|^2 = o_p(1)$. By Cauchy-Schwarz inequality, we then obtain:

$$
\left|\frac{1}{n}\sum_{i=1}^n x_{ji}(\eta_i - \hat{\eta}_i)\right| \leq \left|\frac{1}{n}\sum_{i=1}^n x_{ji}^2\right|^{1/2}\left|\frac{1}{n}\sum_{i=1}^n(\eta_i - \hat{\eta}_i)^2\right|^{1/2} = O_p(1)o_p(1) = o_p(1).
$$

Combining these arguments gives $\hat{\mathbf{W}}_a'\hat{\mathbf{W}}_a/n \xrightarrow{p} E(\mathbf{w}_a\mathbf{w}_a') = \boldsymbol{\Phi}_a$. Next, we wish to establish:

$$
n^{-1/2}\hat{\mathbf{W}}_a'[(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})\alpha + \boldsymbol{\varepsilon}] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_{a1} + \boldsymbol{\Omega}_{a2} - 2\boldsymbol{\Omega}_{a12}).
$$

To do so, we note that:

$$
\begin{aligned}
n^{-1/2}\hat{\mathbf{W}}_a'[(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})\alpha+\boldsymbol{\varepsilon}] &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{x}_i',\hat{\eta}_i)'\varepsilon_i - \frac{\alpha}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{x}_i',\hat{\eta}_i)'(\hat{\eta}_i-\eta_i) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{x}_i',\hat{\eta}_i)'\varepsilon_i - \frac{\alpha}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{x}_i',\eta_i)'(\hat{\eta}_i-\eta_i) - \frac{\alpha}{\sqrt{n}}\sum_{i=1}^{n}(0,\hat{\eta}_i-\eta_i)'(\hat{\eta}_i-\eta_i) \\
&\equiv B_{1n}+B_{2n}+B_{3n}. \qquad \text{(A.2)}
\end{aligned}
$$

It can be shown that $B_{3n}=o_p(1)$ by:

$$
\frac{1}{n}\sum_{i=1}^{n}(\hat{\eta}_i-\eta_i)^2 = O_p\left(\frac{K}{n}+K^{-2s_z/d_z}\right) = o_p(n^{-1/2}),
$$

where the first equality directly follows from Theorem 4 of Newey (1997) and the second equality follows from Assumption 3.4, which requires the mean square error of the first-stage nonparametric series estimator of the generated regressor η_i to have an order smaller than $n^{-1/2}$. We next consider:

$$
\begin{aligned}
B_{2n} &= \frac{-\alpha}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{w}_{ai}(\hat{\eta}_i-\eta_i) \\
&= \frac{-\alpha}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{x}_z(\mathbf{z}_i'),\eta_i)'(\hat{\eta}_i-\eta_i) - \frac{\alpha}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{v}_i',0)'(\hat{\eta}_i-\eta_i) \\
&\equiv C_{1n}+C_{2n}. \qquad \text{(A.3)}
\end{aligned}
$$

To show $C_{2n}=o_p(1)$ in (A.3), it suffices to consider the j-th element of C_{2n}. Denote by v_{zji} the j-th element of $\mathbf{v}_{zi}$ and $\boldsymbol{v}_{zj}=(v_{zj1},\ldots,v_{zjn})'$. The j-th element of C_{2n} can be expressed as:

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_{zji}(\hat{\eta}_i-\eta_i) &= \frac{1}{\sqrt{n}}\boldsymbol{v}_{zj}'\big(\mathbf{P}_{zK}(\boldsymbol{\eta}+\boldsymbol{u})-\boldsymbol{\eta}\big) \\
&= \frac{1}{\sqrt{n}}\boldsymbol{v}_{zj}'\mathbf{P}_{zK}\boldsymbol{u} + \frac{1}{\sqrt{n}}\boldsymbol{v}_{zj}'(\mathbf{P}_{zK}\boldsymbol{\eta}-\boldsymbol{\eta}) \\
&\equiv C_{2n1}+C_{2n2}. \qquad \text{(A.4)}
\end{aligned}
$$

Consider the term C_{2n1}. By Cauchy-Schwarz inequality:

$$
\begin{aligned}
|C_{2n1}| &= \Big|\frac{1}{\sqrt{n}}\boldsymbol{v}'_{zj}\mathbf{P}'_{zK}\mathbf{P}_{zK}\boldsymbol{u}\Big| \\
&\leq \sqrt{n}\Big|\frac{1}{n}\boldsymbol{v}'_{zj}\mathbf{P}'_{zK}\mathbf{P}_{zK}\boldsymbol{v}_{zj}\Big|^{1/2}\Big|\frac{1}{n}\boldsymbol{u}'\mathbf{P}'_{zK}\mathbf{P}_{zK}\boldsymbol{u}\Big|^{1/2} \\
&= \sqrt{n}O_p\left(\sqrt{\frac{K}{n}}\right)O_p\left(\sqrt{\frac{K}{n}}\right) = O_p\left(\frac{K}{\sqrt{n}}\right) = o_p(1),
\end{aligned} \tag{A.5}
$$

where the first equality holds because the projection matrix $\mathbf{P}_{zK}$ is idempotent and symmetric, and the last line follows from Theorem 4 of Newey (1997) and the fact that $E(u|\mathbf{z}) = 0$ and $E(\mathbf{v}|\mathbf{z}) = \mathbf{0}$ by construction. We also establish $C_{2n2} = o_p(1)$ by Cauchy-Schwarz inequality as follows:

$$
\begin{aligned}
|C_{2n2}| &= \Big|\frac{1}{\sqrt{n}}\boldsymbol{v}'_j(\mathbf{P}_{zK}\boldsymbol{\eta} - \boldsymbol{\eta})\Big| \leq \sqrt{n}\Big|\frac{1}{n}\boldsymbol{v}'_j\boldsymbol{v}_j\Big|^{1/2}\Big|\frac{1}{n}(\mathbf{P}_{zK}\boldsymbol{\eta} - \boldsymbol{\eta})'(\mathbf{P}_{zK}\boldsymbol{\eta} - \boldsymbol{\eta})\Big|^{1/2} \\
&= \sqrt{n}O_p(1)O_p\Big(K^{-s_z/d_z}\Big) = o_p(1),
\end{aligned} \tag{A.6}
$$

where the last equality holds, because $n^{-1}(\mathbf{P}_{zK}\boldsymbol{\eta}-\boldsymbol{\eta})'(\mathbf{P}_{zK}\boldsymbol{\eta}-\boldsymbol{\eta}) = O_p(K^{-s_z/d_z})$ by Theorem 4 of Newey (1997). We next consider the term C_{1n}. We denote by x_{zji} the j-th element of $\mathbf{x}_z(\mathbf{z}_i)$ and also denote $\mathbf{X}_{zj} = (x_{zj1}, \ldots, x_{zjn})'$. The j-th element of C_{1n} can be written as:

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_{zji}(\hat{\eta}_i - \eta_i) &= \frac{1}{\sqrt{n}}\mathbf{X}'_{zj}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = \frac{1}{\sqrt{n}}\mathbf{X}'_{zj}(\mathbf{P}_{zK}\boldsymbol{\eta} + \mathbf{P}_{zK}\boldsymbol{u} - \boldsymbol{\eta}) \\
&= \frac{1}{\sqrt{n}}\mathbf{X}'_{zj}\mathbf{P}_{zK}\boldsymbol{u} + \frac{1}{\sqrt{n}}\mathbf{X}'_{zj}(\mathbf{P}_{zK}\boldsymbol{\eta} - \boldsymbol{\eta}) \\
&= \frac{1}{\sqrt{n}}\mathbf{X}'_{zj}\boldsymbol{u} + \frac{1}{\sqrt{n}}(\mathbf{X}'_{zj}\mathbf{P}_{zK} - \mathbf{X}'_{zj})\boldsymbol{u} + \frac{1}{\sqrt{n}}\mathbf{X}'_{zj}(\mathbf{P}_{zK}\boldsymbol{\eta} - \boldsymbol{\eta}) \\
&= \frac{1}{\sqrt{n}}\mathbf{X}'_{zj}\boldsymbol{u} + o_p(1),
\end{aligned}
$$

where the last two terms in the third line above are both $o_p(1)$, using the same argument for (A.6). Similarly, we also show:

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \eta_i(\hat{\eta}_i - \eta_i) = \frac{1}{\sqrt{n}}\boldsymbol{\eta}'\boldsymbol{u} + o_p(1). \tag{A.7}
$$

As a consequence, we end up with:

$$\frac{1}{\sqrt{n}}\hat{\mathbf{W}}_a'\left((\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})\alpha+\boldsymbol{\varepsilon}\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{x}_i',\eta_i)'\varepsilon_i-\frac{\alpha}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbf{x}_z'(\mathbf{z}_i),\eta_i\right)'u_i+o_p(1). \tag{A.8}$$

Putting together (A.8), the established result that $\hat{\mathbf{W}}_a'\hat{\mathbf{W}}_a/n \xrightarrow{p} E(\mathbf{w}_a\mathbf{w}_a') = \boldsymbol{\Phi}_a$, the Lindberg-Levy central limit theorem, and Slutsky's theorem completes the proof.

A.2 Proof of Theorem 3.1

Let $\hat{\boldsymbol{\theta}} = \mathbf{P}_{xL}\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\theta(\mathbf{x}_1), \ldots, \theta(\mathbf{x}_n))'$. We first note that:

$$\begin{aligned}\hat{\alpha} &= (\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\hat{\boldsymbol{\eta}})^{-1}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})(\boldsymbol{\theta}+\boldsymbol{\eta}\alpha+\boldsymbol{u})\\ &= [\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\hat{\boldsymbol{\eta}}]^{-1}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})(\hat{\boldsymbol{\eta}}\alpha+\boldsymbol{\theta}+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})\alpha+\boldsymbol{u})\\ &= \alpha+(\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\hat{\boldsymbol{\eta}})^{-1}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})(\boldsymbol{\theta}+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})\alpha+\boldsymbol{u}).\end{aligned}$$

Therefore,

$$\sqrt{n}(\hat{\alpha}-\alpha) = (\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\hat{\boldsymbol{\eta}})^{-1}\frac{1}{\sqrt{n}}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})(\boldsymbol{\theta}+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})\alpha+\boldsymbol{u}). \tag{A.9}$$

We first show that $n^{-1}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\hat{\boldsymbol{\eta}} \xrightarrow{p} B$. Let $\tilde{\boldsymbol{\eta}}_x = \mathbf{P}_{xL}\boldsymbol{\eta}$ and $\hat{\boldsymbol{\eta}}_x = \mathbf{P}_{xL}\hat{\boldsymbol{\eta}}$. We have that $n^{-1}\|\tilde{\boldsymbol{\eta}}_x - \hat{\boldsymbol{\eta}}_x\|^2 \leq n^{-1}\|\boldsymbol{\eta}-\tilde{\boldsymbol{\eta}}\| = o_p(1)$ and $n^{-1}\|\boldsymbol{\eta}_x - \hat{\boldsymbol{\eta}}_x\|^2 = o_p(1)$. Write:

$$\begin{aligned}\frac{1}{n}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\hat{\boldsymbol{\eta}} &= \frac{1}{n}(\hat{\boldsymbol{\eta}}-\tilde{\boldsymbol{\eta}}_x)'(\hat{\boldsymbol{\eta}}-\tilde{\boldsymbol{\eta}}_x)\\ &= \frac{1}{n}(\boldsymbol{\eta}-\boldsymbol{\eta}_x)'(\boldsymbol{\eta}-\boldsymbol{\eta}_x)+\frac{2}{n}(\boldsymbol{\eta}-\boldsymbol{\eta}_x)'((\hat{\boldsymbol{\eta}}-\tilde{\boldsymbol{\eta}}_x)-(\boldsymbol{\eta}-\boldsymbol{\eta}_x))\\ &\quad+\frac{1}{n}((\hat{\boldsymbol{\eta}}-\tilde{\boldsymbol{\eta}}_x)-(\boldsymbol{\eta}-\boldsymbol{\eta}_x))'((\hat{\boldsymbol{\eta}}-\tilde{\boldsymbol{\eta}}_x)-(\boldsymbol{\eta}-\boldsymbol{\eta}_x)).\end{aligned} \tag{A.10}$$

Similar to the arguments for Theorem 2.1, we have $n^{-1}(\boldsymbol{\eta}-\boldsymbol{\eta}_x)'(\boldsymbol{\eta}-\boldsymbol{\eta}_x) \xrightarrow{p} B$, $n^{-1}(\boldsymbol{\eta}-\boldsymbol{\eta}_x)'((\hat{\boldsymbol{\eta}}-\tilde{\boldsymbol{\eta}}_x)-(\boldsymbol{\eta}-\boldsymbol{\eta}_x)) = o_p(1)$ and $n^{-1}((\hat{\boldsymbol{\eta}}-\tilde{\boldsymbol{\eta}}_x)-(\boldsymbol{\eta}-\boldsymbol{\eta}_x))'((\hat{\boldsymbol{\eta}}-\tilde{\boldsymbol{\eta}}_x)-(\boldsymbol{\eta}-\boldsymbol{\eta}_x)) = o_p(1)$, all of which are sufficient to establish $n^{-1}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\hat{\boldsymbol{\eta}} \xrightarrow{p} B$.

We next show that $n^{-1/2}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})(\boldsymbol{\theta}+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})\alpha+\boldsymbol{u}) \xrightarrow{d} N(0, \boldsymbol{\Omega}_{b1}+\boldsymbol{\Omega}_{b2}-2\boldsymbol{\Omega}_{b12})$. We first claim that $n^{-1/2}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\theta} = o_p(1)$. This holds because:

$$\begin{aligned}\left|\frac{1}{\sqrt{n}}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\theta}\right| &\leq \sqrt{n}\cdot\left\|\frac{1}{\sqrt{n}}\hat{\boldsymbol{\eta}}'\right\|\cdot\left\|\frac{1}{\sqrt{n}}(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\theta}\right\| \\ &= \sqrt{n}\cdot O_p(1)\cdot O_p(L^{-s_x/d_x}) = o_p(1).\end{aligned} \tag{A.11}$$

Secondly, we write:

$$\begin{aligned}&\frac{1}{\sqrt{n}}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\varepsilon} = \frac{1}{\sqrt{n}}\mathbf{s}'\mathbf{P}_{zK}(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\varepsilon} \\ &\quad = \frac{1}{\sqrt{n}}\boldsymbol{\eta}'\mathbf{P}_{zK}(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\varepsilon} + \frac{1}{\sqrt{n}}\boldsymbol{u}'\mathbf{P}_{zK}(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\varepsilon} \\ &\quad = \frac{1}{\sqrt{n}}\boldsymbol{\eta}'(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\varepsilon} - \frac{1}{\sqrt{n}}\boldsymbol{\eta}'(\mathbf{I}-\mathbf{P}_{zK})(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\varepsilon} + \frac{1}{\sqrt{n}}\boldsymbol{u}'\mathbf{P}_{zK}\boldsymbol{\varepsilon} - \frac{1}{\sqrt{n}}\boldsymbol{u}'\mathbf{P}_{zK}\mathbf{P}_{xL}\boldsymbol{\varepsilon} \\ &\quad = \frac{1}{\sqrt{n}}(\boldsymbol{\eta}-\boldsymbol{\eta}_x)'\boldsymbol{\varepsilon} + \frac{1}{\sqrt{n}}\big(\boldsymbol{\eta}_x' - (\boldsymbol{\eta}_x+\boldsymbol{\eta})\big)'\mathbf{P}_{xL}\boldsymbol{\varepsilon} + \frac{1}{\sqrt{n}}\boldsymbol{u}'\mathbf{P}_{zK}\boldsymbol{\varepsilon} - \frac{1}{\sqrt{n}}\boldsymbol{u}'\mathbf{P}_{zK}\mathbf{P}_{xL}\boldsymbol{\varepsilon} \\ &\quad = \frac{1}{\sqrt{n}}(\boldsymbol{\eta}-\boldsymbol{\eta}_x)'\boldsymbol{\varepsilon} + \frac{1}{\sqrt{n}}\boldsymbol{\eta}_x'(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\varepsilon} - \frac{1}{\sqrt{n}}\boldsymbol{\eta}'\mathbf{P}_{xL}\boldsymbol{\varepsilon} + \frac{1}{\sqrt{n}}\boldsymbol{u}\mathbf{P}_{zK}\boldsymbol{\varepsilon} - \frac{1}{\sqrt{n}}\boldsymbol{u}'\mathbf{P}_{zK}\mathbf{P}_{xL}\boldsymbol{\varepsilon} \\ &\quad \equiv D_{11n} + D_{12n} - D_{13n} + D_{14n} - D_{15n}.\end{aligned} \tag{A.12}$$

By the same argument for establishing $C_{21n} = o_p(1)$, we can show that D_{13n}, D_{14n}, and D_{15n} are all $o_p(1)$. By the same argument for establishing $C_{22n} = o_p(1)$, we have $D_{12n} = o_p(1)$. Putting these together yields $n^{-1/2}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})\boldsymbol{\varepsilon} = n^{-1/2}(\boldsymbol{\eta}-\boldsymbol{\eta}_x)'\boldsymbol{\varepsilon} + o_p(1)$.

Lastly,

$$\begin{aligned}&\frac{1}{\sqrt{n}}\hat{\boldsymbol{\eta}}'(\mathbf{I}-\mathbf{P}_{xL})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}) = \frac{1}{\sqrt{n}}\boldsymbol{\eta}'(\mathbf{I}-\mathbf{P}_{xL})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}) + \frac{1}{\sqrt{n}}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})'(\mathbf{I}-\mathbf{P}_{xL})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}) \\ &\quad = \frac{1}{\sqrt{n}}\boldsymbol{\eta}'(\mathbf{I}-\mathbf{P}_{xL})(\boldsymbol{\eta}-\mathbf{P}_{zK}\boldsymbol{\eta}) - \frac{1}{\sqrt{n}}\boldsymbol{\eta}'(\mathbf{I}-\mathbf{P}_{xL})\mathbf{P}_{zK}\boldsymbol{u} + o_p(1) \\ &\quad = o_p(1) - \frac{1}{\sqrt{n}}\boldsymbol{h}'(\mathbf{I}-\mathbf{P}_{xL})\mathbf{P}_{zK}\boldsymbol{u} - \frac{1}{\sqrt{n}}\boldsymbol{\eta}'(\mathbf{I}-\mathbf{P}_{xL})\mathbf{P}_{zK}\boldsymbol{u} + o_p(1) \\ &\quad = -\frac{1}{\sqrt{n}}\boldsymbol{\eta}'\mathbf{P}_{zK}\boldsymbol{u} + \frac{1}{\sqrt{n}}\boldsymbol{\eta}'\mathbf{P}_{xL}\mathbf{P}_{zK}\boldsymbol{u} + o_p(1) \\ &\quad = -\frac{1}{\sqrt{n}}{\boldsymbol{\eta}_z}'\mathbf{P}_{zK}\boldsymbol{u} - \frac{1}{\sqrt{n}}{\boldsymbol{\eta}_x}'\mathbf{P}_{zK}\boldsymbol{u} + o_p(1) \\ &\quad = -\frac{1}{\sqrt{n}}{\boldsymbol{\eta}_z}'\boldsymbol{u} - \frac{1}{\sqrt{n}}{\boldsymbol{\eta}_z}'(\mathbf{I}-\mathbf{P}_{zK})\boldsymbol{u} + o_p(1) \\ &\quad = -\frac{1}{\sqrt{n}}{\boldsymbol{\eta}_z}'\boldsymbol{u} + o_p(1),\end{aligned} \tag{A.13}$$

where the equalities follow from the same arguments for establishing $C_{21n} = o_p(1)$ and

$C_{22n} = o_p(1)$, i.e., those terms we drop in the lines above are all $o_p(1)$. As a result, we have shown that $\sqrt{n}(\hat{\alpha} - \alpha) = n^{-1/2}(\boldsymbol{\eta} - \boldsymbol{h})'\boldsymbol{u} - \alpha^{-1/2}\boldsymbol{\eta}_z{}'\boldsymbol{u} + o_p(1)$, which is sufficient to establish the first part of Theorem 3.1.

On the other hand, given $\hat{\alpha}$, $\hat{\boldsymbol{\theta}}$ is obtained by:

$$
\begin{aligned}
\hat{\boldsymbol{\theta}} &= \mathbf{P}_{xL}(\boldsymbol{y} - \hat{\boldsymbol{\eta}}\hat{\alpha}) && \text{(A.14)} \\
&= \mathbf{P}_{xL}(\boldsymbol{\theta} + \boldsymbol{\eta}\alpha + \boldsymbol{\varepsilon} - \hat{\boldsymbol{\eta}}\hat{\alpha}) \\
&= \mathbf{P}_{xL}\boldsymbol{\theta} + \mathbf{P}_{xL}\boldsymbol{\varepsilon} + \mathbf{P}_{xL}\hat{\boldsymbol{\eta}}(\alpha - \hat{\alpha}) + \mathbf{P}_{xL}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}). && \text{(A.15)}
\end{aligned}
$$

By the standard arguments used for series estimation, we have $\|n^{-1}\mathbf{P}_{xL}\boldsymbol{\theta}\|^2 = O_p(L^{-2s_x/d_x})$ and $\|n^{-1}\mathbf{P}_{xL}\boldsymbol{u}\|^2 = O_p(K_x/n)$. Moreover, $\|n^{-1}\mathbf{P}_{xL}\hat{\boldsymbol{\eta}}(\alpha - \hat{\alpha})\|^2 = O_p(n^{-1})$, because $(\alpha - \hat{\alpha}) = O_p(n^{-1/2})$, and $\|n^{-1}\mathbf{P}_{xL}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})\|^2 \leq \|n^{-1}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})\|^2 = O_p(K^{-2s_z/d_z} + O_p(K/n))$. By the assumption, we have that L^{-2s_x/d_x} is of a smaller order of L/n and that K^{-2s_z/d_z} is of a smaller order of K/n. These established results are sufficient to show the second part of Theorem 3.1.

A.3 Proof of Theorem 4.1

Let $\mathbf{P}_{\eta M} = \boldsymbol{R}_{\eta M}(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})^{-1}\boldsymbol{R}'_{\eta M}$ and $\hat{\mathbf{P}}_{\eta M} = \hat{\boldsymbol{R}}_{\eta M}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M}$ be the $n \times n$ projection matrices based on the true conditioning variable $\boldsymbol{\eta}$ and its estimate $\hat{\boldsymbol{\eta}}$, respectively. We also define $\tilde{\boldsymbol{A}}_\eta = \mathbf{P}_{\eta M}\boldsymbol{A}$ and $\hat{\boldsymbol{A}}_\eta = \hat{\mathbf{P}}_{\eta M}\boldsymbol{A}$ for any $n \times 1$ or $n \times d_x$ matrix $\boldsymbol{A}$ and $\mathbf{Q}_{\eta M} = \mathbf{I} - \mathbf{P}_{\eta M}$ and similarly $\hat{\mathbf{Q}}_{\eta M} = \mathbf{I} - \hat{\mathbf{P}}_{\eta M}$. To be clear, we emphasize here that $\tilde{\boldsymbol{A}}_\eta$ is the series estimate of the conditional expectation of $\boldsymbol{A}$ using *true* values $\boldsymbol{\eta}$ as conditioning variables as well as evaluated at $\boldsymbol{\eta}$, whereas $\hat{\boldsymbol{A}}_\eta$ is the series estimate of the conditional expectation of $\boldsymbol{A}$ given $\boldsymbol{\eta}$ using *estimated* values $\hat{\boldsymbol{\eta}}$ as conditioning variables as well as evaluated at $\hat{\boldsymbol{\eta}}$. We also denote $\boldsymbol{\theta} = \theta(\boldsymbol{\eta})$ and $\hat{\boldsymbol{\theta}} = \theta(\hat{\boldsymbol{\eta}})$. Let C be a generic positive constant. We first express the second-step series estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ in model (4.1) as:

$$
\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= (\mathbf{X}'\hat{\mathbf{Q}}_{\eta M}\mathbf{X}/n)^{-1}\mathbf{X}'\hat{\mathbf{Q}}_{\eta M}(\boldsymbol{\theta} + \boldsymbol{\varepsilon})/\sqrt{n} \\
&= (\mathbf{X}'\hat{\mathbf{Q}}_{\eta M}\mathbf{X}/n)^{-1}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'\hat{\mathbf{Q}}_{\eta M}(\boldsymbol{\theta} + \boldsymbol{\varepsilon})/\sqrt{n} \\
&= (\mathbf{X}'\hat{\mathbf{Q}}_{\eta M}\mathbf{X}/n)^{-1}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + (\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}))/\sqrt{n} \\
&= (\mathbf{X}'\hat{\mathbf{Q}}_{\eta M}\mathbf{X}/n)^{-1}(T_{1n} + T_{2n}),
\end{aligned}
$$

where the second equality uses the fact that $\hat{\mathbf{Q}}_{\eta M}$ is idempotent and $\hat{\mathbf{X}}_\eta = \hat{\mathbf{P}}_{\eta M}\mathbf{X}$, and we define:

$$T_{1n} = (\mathbf{X} - \hat{\mathbf{X}}_\eta)'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})/\sqrt{n}, \tag{A.16}$$
$$T_{2n} = (\mathbf{X} - \hat{\mathbf{X}}_\eta)'(\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}})/\sqrt{n} = (\mathbf{X} - \hat{\mathbf{X}}_\eta)'\boldsymbol{\varepsilon}/\sqrt{n} - (\mathbf{X} - \hat{\mathbf{X}}_\eta)'\hat{\boldsymbol{\varepsilon}}/\sqrt{n} \equiv T_{2n1} + T_{2n2}. \tag{A.17}$$

We will prove $T_{2n2} = o_p(1)$, and as a result the asymptotic behavior of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is driven by the terms T_{1n} and T_{2n1}, where T_{1n} accounts for the error from estimating the generated regressor. To do so, we present some useful lemmas that will be utilized to examine the terms T_{1n} and T_{2n}.

Lemma A.1. $\frac{1}{\sqrt{n}}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'\boldsymbol{\varepsilon} = \frac{1}{\sqrt{n}}\boldsymbol{v}_\eta'\boldsymbol{\varepsilon} + o_p(1)$.

Proof. Recall $\mathbf{x} = E(\mathbf{x}|\eta) + \mathbf{v}_\eta \equiv \mathbf{x}_\eta + \mathbf{v}_\eta$. Denote by $\mathbf{X}_\eta$ the $n \times d_x$ matrix with the i-th row being $\mathbf{x}_{\eta i} = E(\mathbf{x}_i|\eta_i)$, the series estimator $\hat{\mathbf{x}}_{\eta i}' = \boldsymbol{r}_M(\hat{\eta}_i)'(\hat{\boldsymbol{R}}_{\eta M}'\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}_{\eta M}'\mathbf{X}$ using the estimate $\hat{\boldsymbol{\eta}}$ in place of $\boldsymbol{\eta}$, $\hat{\mathbf{X}}_\eta = \hat{\mathbf{P}}_{\eta M}\mathbf{X}$ the $n \times d_x$ matrix with the i-th row being $\hat{\mathbf{x}}_{\eta i}'$, and $\boldsymbol{v}_\eta = (\mathbf{v}_{\eta 1}, \ldots, \mathbf{v}_{\eta n})'$. We then write:

$$\begin{aligned}
\frac{1}{\sqrt{n}}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'\boldsymbol{\varepsilon} &= \frac{1}{\sqrt{n}}(\mathbf{X}_\eta + \boldsymbol{v}_\eta - \hat{\mathbf{X}}_\eta)'\boldsymbol{\varepsilon} \\
&= \frac{1}{\sqrt{n}}\boldsymbol{v}_\eta'\boldsymbol{\varepsilon} + \frac{1}{\sqrt{n}}(\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta)'\boldsymbol{\varepsilon} \\
&= T_{2n11} + T_{2n12}.
\end{aligned} \tag{A.18}$$

For the term T_{2n11}, it is straightforward to show $n^{-1/2}\boldsymbol{v}_\eta'\boldsymbol{\varepsilon} = n^{-1/2}\sum_{i=1}^n v_{\eta i}\varepsilon_i \xrightarrow{d} N(0, \Omega_{c1})$, where $\Omega_{c1} = E(\varepsilon_i^2 \mathbf{v}_{\eta i}\mathbf{v}_{\eta i}')$ by virtue of the Lindeberg-Levy central limit theorem. Next, for the term T_{2n12}, note that:

$$\begin{aligned}
E\left[\left\|\frac{1}{n}(\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta)'\boldsymbol{\varepsilon}\right\|^2 |\eta\right] &= n^{-2}\mathrm{tr}\left[(\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta)(\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta)'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\eta)\right] \\
&\leq Cn^{-1}\mathrm{tr}[(\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta)(\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta)'/n] = o_p(1)
\end{aligned} \tag{A.19}$$

by Lemma A.3 below. Hence, $T_{2n12} = n^{-1/2}(\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta)'\boldsymbol{\varepsilon} = o_p(1)$.

□

We denote by $\xi_0(K)$ and $\zeta_1(M)$, respectively, two sequences of constants satisfying $\sup_{\mathbf{z}}\|\boldsymbol{p}_K(\mathbf{z})\| \leq \xi_0(K)$ and $\sup_\eta\|d\boldsymbol{r}_M(\eta)/d\eta\| \leq \zeta_1(M)$. As established by Newey (1997, Theorem 1), the maximum error of the first-step series estimation $\hat{\eta}_i = \boldsymbol{p}_K(\mathbf{z}_i)(\boldsymbol{P}_{zK}'\boldsymbol{P}_{zK})^{-1}\boldsymbol{P}_{zK}'\mathbf{s}$ can be shown to be $\max_i|\hat{\eta}_i - \eta_i| = O_p(\xi_0(K)\Delta_\eta)$, where $\Delta_\eta \equiv \sqrt{K/n} + K^{-s_z/d_z}$.

Lemma A.2. *Denote by* $\hat{\boldsymbol{\gamma}}_\eta = (\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M}\mathbf{X}$ *the* $M \times d_x$ *matrix of the series coefficient estimates for the conditional expectation* $E(\mathbf{x}|\eta)$ *using the estimate* $\hat{\eta}_i$ *of* η_i*. We then have* $\|\hat{\boldsymbol{\gamma}}_\eta - \boldsymbol{\gamma}_\eta\| = O_p(M^{-s_\eta} + \zeta_1(M)\xi_0(K)\Delta_\eta)$*, where* $\boldsymbol{\gamma}_\eta$ *is a vector of series coefficients such that* $\boldsymbol{r}_M(\cdot)'\boldsymbol{\gamma}_\eta$ *is the series approximation to* $\mathbf{x}_\eta \equiv E(\mathbf{x}|\eta)$*.*

Proof. Consider the scalar case of $\mathbf{x}$, i.e., $d_x = 1$. First note:

$$
\begin{aligned}
&(\mathbf{X} - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta)'(\mathbf{X} - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta)/n \\
&\quad = (\mathbf{X} - \boldsymbol{R}_{\eta M}\boldsymbol{\gamma}_\eta + (\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\boldsymbol{\gamma}_\eta)'(\mathbf{X} - \boldsymbol{R}_{\eta M}\boldsymbol{\gamma}_\eta + (\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\boldsymbol{\gamma}_\eta)/n \\
&\quad = (\mathbf{X} - \boldsymbol{R}_{\eta M}\boldsymbol{\gamma}_\eta)'(\mathbf{X} - \boldsymbol{R}_{\eta M}\boldsymbol{\gamma}_\eta)/n \\
&\qquad + (\mathbf{X} - \boldsymbol{R}_{\eta M}\boldsymbol{\gamma}_\eta)'(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\boldsymbol{\gamma}_\eta/n + \boldsymbol{\gamma}'_\eta(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})'(\mathbf{X} - \boldsymbol{R}_{\eta M}\boldsymbol{\gamma}_\eta)/n \\
&\qquad + \boldsymbol{\gamma}'_\eta(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})'(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\boldsymbol{\gamma}_\eta/n \\
&\quad = O_p(M^{-2s_\eta} + n^{-1}\zeta_1^2(M)\xi_0^2(K)\Delta_\eta^2),
\end{aligned} \tag{A.20}
$$

where the last equality follows from the fact that the second and third terms are dominated by the last term on the right-hand side of the second equality and by the second-order Taylor expansion:

$$
\begin{aligned}
\left\|\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M}\right\|^2 &= \sum_{m=1}^{M}\left[r_m(\hat{\eta}_i) - r_m(\eta_i)\right]^2 = \sum_{m=1}^{M}\left(\frac{dr_m(\bar{\eta})}{d\eta}\right)^2(\hat{\eta}_i - \eta_i)^2 \\
&\le \max_{1\le i\le n}(\hat{\eta}_i - \eta_i)^2\sum_{m=1}^{M}\left(\frac{dr_m(\bar{\eta})}{d\eta}\right)^2 = O_p(\zeta_1^2(M)\xi_0^2(K)\Delta_\eta^2),
\end{aligned}
$$

where $\bar{\eta}_i$ is between $\hat{\eta}_i$ and η_i. As a consequence:

$$
\begin{aligned}
\|\hat{\boldsymbol{\gamma}}_\eta - \boldsymbol{\gamma}_\eta\| &= \|(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M}(\mathbf{X} - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta)\| \\
&= \left\{(\mathbf{X} - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta)'\hat{\boldsymbol{R}}_{\eta M}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M}/n)^{-1}\hat{\boldsymbol{R}}'_{\eta M}(\mathbf{X} - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta)/n\right\}^{1/2} \\
&= O_p(1)\left\{(\mathbf{X} - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta)'\hat{\boldsymbol{R}}_{\eta M}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M}(\mathbf{X} - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta)/n\right\}^{1/2} \\
&= O_p(1)\left\{(\mathbf{X} - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta)'(\mathbf{X} - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta)/n\right\}^{1/2} \\
&= O_p(M^{-s_\eta} + n^{-1/2}\zeta_1(M)\xi_0(K)\Delta_\eta),
\end{aligned}
$$

where the last equality follows from (A.20). □

Lemma A.3.

$$\frac{1}{n}(\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta)'(\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta) = O_p(M^{-2s_\eta}) = o_p(n^{-1/2}).$$

Proof. It suffices to consider $\mathbf{x}_\eta$ as a univariate function. We thus write:

$$\begin{aligned}
n^{-1}\|\mathbf{X}_\eta - \hat{\boldsymbol{R}}_{\eta M}\hat{\boldsymbol{\gamma}}_\eta\|^2 &\leq n^{-1}\left\{\|\mathbf{X}_\eta - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\gamma}_\eta\|^2 + \|\hat{\boldsymbol{R}}_{\eta M}(\boldsymbol{\gamma}_\eta - \hat{\boldsymbol{\gamma}}_\eta)\|^2\right\} \\
&\leq n^{-1}\left\{\|\mathbf{X}_\eta - \boldsymbol{R}_{\eta M}\boldsymbol{\gamma}_\eta\|^2 + \|(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\boldsymbol{\gamma}_\eta\|^2 + \|\hat{\boldsymbol{R}}_{\eta M}(\boldsymbol{\gamma}_\eta - \hat{\boldsymbol{\gamma}}_\eta)\|^2\right\} \\
&= O(M^{-2s_\eta}) + n^{-1}\|(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\boldsymbol{\gamma}_\eta\|^2 + (\boldsymbol{\gamma}_\eta - \hat{\boldsymbol{\gamma}}_\eta)'\left(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M}/n\right)(\boldsymbol{\gamma}_\eta - \hat{\boldsymbol{\gamma}}_\eta) \\
&= O(M^{-2s_\eta}) + (\boldsymbol{\gamma}_\eta - \hat{\boldsymbol{\gamma}}_\eta)'\left(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M}/n + o_p(1)\right)(\boldsymbol{\gamma}_\eta - \hat{\boldsymbol{\gamma}}_\eta) \\
&= o_p(n^{-1/2}),
\end{aligned}$$

where the first equality follows from the fact that $n^{-1}\|(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\boldsymbol{\gamma}_\eta\|^2 \leq Cn^{-1}\|\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M}\|^2 \leq C\zeta_1^2(M)\max_i(\hat{\eta}_i - \eta_i)^2$ and the second equality follows since under Assumption 4.3, we have:

$$\begin{aligned}
\|\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M}/n - \hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M}/n\| &\leq \|\hat{\boldsymbol{R}}_{\eta M} - \boldsymbol{R}_{\eta M}\|^2/n + 2\|\boldsymbol{R}'_{\eta M}(\hat{\boldsymbol{R}}_{\eta M} - \boldsymbol{R}_{\eta M})\|/n \\
&\leq C\|\hat{\boldsymbol{R}}_{\eta M} - \boldsymbol{R}_{\eta M}\|^2/n \\
&= O_p(n^{-1}\zeta_1(M)^2\xi_0(K)^2\Delta_\eta^2) \\
&= o_p(1).
\end{aligned}$$

□

Lemma A.4. $n^{-1/2}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'\hat{\boldsymbol{\varepsilon}} = o_p(1)$.

Proof. Similar to Lemma A.1, we write:

$$\begin{aligned}
\frac{1}{\sqrt{n}}(\mathbf{X} - \hat{\mathbf{X}}_\eta)'\hat{\boldsymbol{\varepsilon}} &= \frac{1}{\sqrt{n}}\left[\boldsymbol{v}'_\eta\hat{\boldsymbol{\varepsilon}} + (\mathbf{X}_\eta - \hat{\mathbf{X}}_\eta)'\hat{\boldsymbol{\varepsilon}} + \hat{\boldsymbol{v}}'_\eta\hat{\boldsymbol{\varepsilon}}\right] \\
&= \frac{1}{\sqrt{n}}\Big[\boldsymbol{v}'_\eta\tilde{\boldsymbol{\varepsilon}} + \boldsymbol{v}'_\eta(\hat{\boldsymbol{\varepsilon}} - \tilde{\boldsymbol{\varepsilon}}) + (\mathbf{X}_\eta - \tilde{\mathbf{X}}_\eta + \tilde{\mathbf{X}}_\eta - \hat{\mathbf{X}}_\eta)'(\tilde{\boldsymbol{\varepsilon}} + \hat{\boldsymbol{\varepsilon}} - \tilde{\boldsymbol{\varepsilon}}) + \hat{\boldsymbol{v}}'_\eta\hat{\boldsymbol{\varepsilon}}\Big].
\end{aligned}$$

The results that $n^{-1/2}\boldsymbol{v}'_\eta\tilde{\boldsymbol{\varepsilon}} = o_p(1)$, $n^{-1/2}(\mathbf{X}_\eta - \tilde{\mathbf{X}}_\eta)'\tilde{\boldsymbol{\varepsilon}} = o_p(1)$, and $n^{-1/2}\tilde{\boldsymbol{v}}'_\eta\tilde{\boldsymbol{\varepsilon}} = o_p(1)$ can be established by the same proof in Li and Racine (2007, p. 485). It remains to show that the terms that involve the estimation error of the generated regressor η_i are $o_p(1)$, i.e.:

$$\frac{1}{\sqrt{n}}\boldsymbol{v}_\eta'(\hat{\boldsymbol{\varepsilon}}-\tilde{\boldsymbol{\varepsilon}}) = o_p(1), \tag{A.21}$$

$$\frac{1}{\sqrt{n}}(\mathbf{X}_\eta - \tilde{\mathbf{X}}_\eta)'(\hat{\boldsymbol{\varepsilon}}-\tilde{\boldsymbol{\varepsilon}}) = o_p(1), \tag{A.22}$$

$$\frac{1}{\sqrt{n}}(\tilde{\mathbf{X}}_\eta - \hat{\mathbf{X}}_\eta)'\hat{\boldsymbol{\varepsilon}} = o_p(1), \tag{A.23}$$

$$\frac{1}{\sqrt{n}}\hat{\boldsymbol{v}}_\eta'\hat{\boldsymbol{\varepsilon}} = o_p(1). \tag{A.24}$$

We verify (A.21)-(A.24) separately below. First of all, to show (A.21), we first note that $n^{-1/2}\boldsymbol{v}_\eta'(\hat{\boldsymbol{\varepsilon}}-\tilde{\boldsymbol{\varepsilon}}) = n^{-1/2}\boldsymbol{v}_\eta'(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\boldsymbol{\varepsilon}$ and consider:

$$\begin{aligned} E\left[\left\|\boldsymbol{v}_\eta'(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\boldsymbol{\varepsilon}/\sqrt{n}\right\|^2|\eta\right] &= \mathrm{tr}\left(\boldsymbol{v}_\eta'(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\eta)(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\boldsymbol{v}_\eta\right)/n \\ &\leq C\mathrm{tr}\left(\boldsymbol{v}_\eta'(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\boldsymbol{v}_\eta\right)/n. \end{aligned}$$

It then follows similarly to (A.3) in Newey (2009) that both $\boldsymbol{v}_\eta'(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\hat{\mathbf{P}}_{\eta M}\boldsymbol{v}_\eta/n$ and $\boldsymbol{v}_\eta'(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\mathbf{P}_{\eta M}\boldsymbol{v}_\eta/n$ are $o_p(1)$, and hence $n^{-1/2}\boldsymbol{v}_\eta'(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\boldsymbol{\varepsilon} = o_p(1)$, as desired.

We next turn to (A.22). By the Cauchy-Schwarz inequality, $n^{-1/2}(\mathbf{X}_\eta - \tilde{\mathbf{X}}_\eta)'(\hat{\boldsymbol{\varepsilon}}-\tilde{\boldsymbol{\varepsilon}}) \leq n^{1/2}((\mathbf{X}_\eta - \tilde{\mathbf{X}}_\eta)'(\mathbf{X}_\eta - \tilde{\mathbf{X}}_\eta)/n)^{1/2}((\hat{\boldsymbol{\varepsilon}}-\tilde{\boldsymbol{\varepsilon}})'(\hat{\boldsymbol{\varepsilon}}-\tilde{\boldsymbol{\varepsilon}})/n)^{1/2} = O_p(M^{-s_\eta})O_p(n^{-1}\zeta_1(M)^2\Delta_\eta)$, where the first equality follows from Lemma A.3 and $n^{-2}E(\|(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\boldsymbol{\varepsilon}\|^2|\eta) = n^{-2}\mathrm{tr}((\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})E(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}|\eta)(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})) \leq Cn^{-2}\|(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\|^2 = O(n^{-1}\zeta_1(M)^2\Delta_\eta^2) = o(n^{-1})$, implying $n^{-1}\|(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\boldsymbol{\varepsilon}\| = o_p(n^{-1/2})$. This establishes (A.22).

For (A.23), we write $n^{-1/2}(\tilde{\mathbf{X}}_\eta - \hat{\mathbf{X}}_\eta)'\hat{\boldsymbol{\varepsilon}} = n^{-1/2}((\mathbf{P}_{\eta M}-\hat{\mathbf{P}}_{\eta M})\mathbf{X})'\hat{\boldsymbol{\varepsilon}} = n^{-1/2}\mathbf{X}'(\mathbf{P}_{\eta M}-\hat{\mathbf{P}}_{\eta M})\hat{\boldsymbol{\varepsilon}} = n^{-1/2}(\mathbf{X}_\eta+\boldsymbol{v}_\eta)'(\mathbf{P}_{\eta M}-\hat{\mathbf{P}}_{\eta M})\hat{\boldsymbol{\varepsilon}}$. We have previously shown $n^{-1/2}\boldsymbol{v}_\eta'(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\boldsymbol{\varepsilon} = o_p(1)$. By similar arguments, we can also show $n^{-1/2}\mathbf{X}_\eta'(\hat{\mathbf{P}}_{\eta M}-\mathbf{P}_{\eta M})\boldsymbol{\varepsilon} = o_p(1)$, completing the proof of (A.23).

To show $n^{-1/2}\hat{\boldsymbol{v}}_\eta'\hat{\boldsymbol{\varepsilon}} = o_p(1)$ as stated in (A.24), we consider $n^{-1/2}E(\hat{\boldsymbol{v}}_\eta'\hat{\boldsymbol{v}}_\eta)|\eta) = n^{-1/2}E(\boldsymbol{v}_\eta'\hat{\mathbf{P}}_{\eta M}\hat{\mathbf{P}}_{\eta M}\boldsymbol{v}_\eta|\eta) = n^{-1/2}\mathrm{tr}(\hat{\mathbf{P}}_{\eta M}E(\boldsymbol{v}_\eta\boldsymbol{v}_\eta'|\eta)\hat{\mathbf{P}}_{\eta M}) \leq Cn^{-1/2}\mathrm{tr}(\hat{\mathbf{P}}_{\eta M}) = O(Mn^{-1/2})$, which implies $n^{-1/2}\hat{\boldsymbol{v}}_\eta'\hat{\boldsymbol{v}}_\eta = O_p(Mn^{-1/2})$. Similarly, following the same proof as above, one can show $n^{-1/2}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = n^{-1/2}\boldsymbol{\varepsilon}'\hat{\mathbf{P}}_{\eta M}\boldsymbol{\varepsilon} = O_p(Mn^{-1/2})$. By the Cauchy-Schwarz inequality, we then have $n^{-1/2}\hat{\boldsymbol{v}}_\eta'\hat{\boldsymbol{\varepsilon}} \leq (n^{-1/2}\hat{\boldsymbol{v}}_\eta'\hat{\boldsymbol{v}}_\eta)^{1/2}(n^{-1/2}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}})^{1/2} = O_p(Mn^{-1/2}) = o_p(1)$ under Assumption 4.3. □

Lemma A.5.

$$n^{-1}\sum_{i=1}^n(\theta_i-\hat{\theta}_i)^2 = n^{-1}\sum_{i=1}^n(\theta_i-\tilde{\theta}_i)^2 + o_p(1).$$

Proof. By adding and subtracting $\tilde{\theta}_i$, we have:

$$\frac{1}{n}\sum_{i=1}^{n}(\theta_i-\hat{\theta}_i)^2=\frac{1}{n}\sum_{i=1}^{n}(\theta_i-\tilde{\theta}_i)^2+\frac{1}{n}\sum_{i=1}^{n}(\tilde{\theta}_i-\hat{\theta}_i)^2+\frac{2}{n}\sum_{i=1}^{n}(\theta_i-\tilde{\theta}_i)(\tilde{\theta}_i-\hat{\theta}_i).$$

The second term above can be expressed as $n^{-1}\|(\mathbf{P}_{\eta M}-\hat{\mathbf{P}}_{\eta M})\boldsymbol{\theta}\|^2=n^{-1}\boldsymbol{\theta}'(\mathbf{P}_{\eta M}-\hat{\mathbf{P}}_{\eta M})(\mathbf{P}_{\eta M}-\hat{\mathbf{P}}_{\eta M})\boldsymbol{\theta}=O(1)\mathrm{tr}((\mathbf{P}_{\eta M}-\hat{\mathbf{P}}_{\eta M})(\mathbf{P}_{\eta M}-\hat{\mathbf{P}}_{\eta M}))=O_p(\zeta_1(M)^2\Delta_\eta^2)=o_p(M^{-s_\eta})=o_p(n^{-1/2})$. Lastly, we have $n^{-1}\sum_{i=1}^{n}(\theta_i-\tilde{\theta}_i)(\tilde{\theta}_i-\hat{\theta}_i)\leq n^{-1}(\sum_{i=1}^{n}(\theta_i-\tilde{\theta}_i)^2+\sum_{i=1}^{n}(\tilde{\theta}_i-\hat{\theta}_i)^2)=O_p(M^{-s_\eta})=o_p(n^{-1/2})$. This completes the lemma. $\square$

We now move to the term T_{1n} in (A.16), which can be rewritten as:

$$\begin{aligned}T_{1n}&=(\mathbf{X}_\eta-\hat{\mathbf{X}}_\eta+\boldsymbol{v}_\eta)'(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})/\sqrt{n}\\&=(\mathbf{X}_\eta-\hat{\mathbf{X}}_\eta)'(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})/\sqrt{n}+\boldsymbol{v}_\eta'(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})/\sqrt{n} &\text{(A.25)}\\&\equiv T_{1n1}+T_{1n2}. &\text{(A.26)}\end{aligned}$$

Consider the term T_{1n1}. Note that:

$$\begin{aligned}n^{-1}(\mathbf{X}_\eta-\hat{\mathbf{X}}_\eta)'(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\leq&\left\{\left[n^{-1}\sum_{i=1}^{n}(x_{\eta i}-\hat{x}_{\eta_i})^2\right]\left[n^{-1}\sum_{i=1}^{n}(\theta_i-\hat{\theta}_i)^2\right]\right\}^{1/2}\\=&O_p(M^{-2s_\eta})\\=&o_p(n^{-1/2}),\end{aligned}$$

where the first inequality follows the Cauchy-Schwarz inequality, and the last two equalities follow Lemma A.5 and Lemma 15.7 in Li and Racine (2007, p. 487), where we only consider the scalar case for $\mathbf{x}_{\eta i}$, i.e., $\mathbf{x}_{\eta i}=x_{\eta i}$, because it is sufficient to get probability order for T_{1n1}. Hence, $T_{1n1}=o_p(1)$.

We next turn to the term T_{1n2} that accounts for the estimation error of the generated regressor $\hat{\eta}_i-\eta_i$. We write:

$$\frac{1}{\sqrt{n}}\boldsymbol{v}_\eta'(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})=\frac{1}{\sqrt{n}}\left[\boldsymbol{v}_\eta'(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})+\boldsymbol{v}_\eta'(\tilde{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}})\right]=T_{1n21}+T_{1n22},$$

where we recall $\tilde{\boldsymbol{\theta}}=\mathbf{P}_{\eta M}\boldsymbol{\theta}$.

The term T_{1n21} can be shown to be $O_p(n^{-1}M^{-2s_\eta})=o_p(1)$ by the same proof in Li and Racine (2007, p. 485). For the term T_{1n22}, define $\boldsymbol{A}_{\eta M}=\boldsymbol{R}_{\eta M}(\boldsymbol{R}_{\eta M}'\boldsymbol{R}_{\eta M})^{-1}$ and similarly

$\hat{\boldsymbol{A}}_{\eta M} = \hat{\boldsymbol{R}}_{\eta M}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}$. We note that:

$$\begin{aligned}
\mathrm{P}_{\eta M} - \hat{\mathrm{P}}_{\eta M} &= \mathrm{P}_{\eta M} - \boldsymbol{R}_{\eta M}(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M} + \boldsymbol{R}_{\eta M}(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M} \\
&\quad - \boldsymbol{R}_{\eta M}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M} + \boldsymbol{R}_{\eta M}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M} - \hat{\mathrm{P}}_{\eta M} \\
&= \boldsymbol{R}_{\eta M}(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})^{-1}(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})' + \boldsymbol{R}_{\eta M}(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})^{-1}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M} \\
&\quad - \boldsymbol{R}_{\eta M}(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})^{-1}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M} + (\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M} \\
&= \boldsymbol{R}_{\eta M}(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})^{-1}(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})' \\
&\quad + \boldsymbol{R}_{\eta M}(\boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})^{-1}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M} - \boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M} \\
&\quad + (\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M})^{-1}\hat{\boldsymbol{R}}'_{\eta M} \\
&= \boldsymbol{A}_{\eta M}(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})' + \boldsymbol{A}_{\eta M}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M} - \boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})\hat{\boldsymbol{A}}'_{\eta M} + (\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\hat{\boldsymbol{A}}'_{\eta M}
\end{aligned}$$

and recall that $\|\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M}\|^2 = O_p(\zeta_1(M)^2\xi_0(K)^2\Delta_\eta^2)$. We therefore have:

$$\begin{aligned}
\left\|\frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})\right\| &= \left\|\frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta(\mathrm{P}_{\eta M} - \hat{\mathrm{P}}_{\eta M})\boldsymbol{\theta}\right\| \\
&\leq \left\|\frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta\boldsymbol{A}_{\eta M}(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})'\boldsymbol{\theta}\right\| + \left\|\frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta\boldsymbol{A}_{\eta M}(\hat{\boldsymbol{R}}'_{\eta M}\hat{\boldsymbol{R}}_{\eta M} - \boldsymbol{R}'_{\eta M}\boldsymbol{R}_{\eta M})\hat{\boldsymbol{A}}'_{\eta M}\boldsymbol{\theta}\right\| \\
&\quad + \left\|\frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\hat{\boldsymbol{A}}'_{\eta M}\boldsymbol{\theta}\right\|.
\end{aligned}$$

By expressing $\boldsymbol{\theta} = \boldsymbol{R}_{\eta M}\boldsymbol{\alpha}_\eta + \boldsymbol{\varepsilon}_\eta$ with the approximation error $\boldsymbol{\varepsilon}_\eta$, we then have:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) &= \frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta(\boldsymbol{R}_{\eta M}\boldsymbol{\alpha}_\eta + \boldsymbol{\varepsilon}_\eta - \hat{\boldsymbol{R}}_{\eta M}\hat{\boldsymbol{\alpha}}_\eta) \\
&= \frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta(\boldsymbol{R}_{\eta M}\boldsymbol{\alpha}_\eta - \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\alpha}_\eta + \hat{\boldsymbol{R}}_{\eta M}\boldsymbol{\alpha}_\eta - \hat{\boldsymbol{R}}_{\eta M}\hat{\boldsymbol{\alpha}}_\eta + \boldsymbol{\varepsilon}_\eta) \\
&= \frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\boldsymbol{\alpha}_\eta + \frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta\hat{\boldsymbol{R}}_{\eta M}(\boldsymbol{\alpha}_\eta - \hat{\boldsymbol{\alpha}}_\eta) + \frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta\boldsymbol{\varepsilon}_\eta \\
&= \frac{1}{\sqrt{n}}\boldsymbol{v}'_\eta(\boldsymbol{R}_{\eta M} - \hat{\boldsymbol{R}}_{\eta M})\boldsymbol{\alpha}_\eta + o_p(1), \qquad \text{(A.27)}
\end{aligned}$$

where the last equality follows, because $E(\|\boldsymbol{v}'_\eta\boldsymbol{\varepsilon}_\eta\|^2|z)/n = \boldsymbol{\varepsilon}'_\eta E(\boldsymbol{v}_\eta\boldsymbol{v}'_\eta|z)\boldsymbol{\varepsilon}_\eta/n \leq \boldsymbol{\varepsilon}'_\eta\boldsymbol{\varepsilon}_\eta/n = O_p(n^{-1}M^{-2s_\eta}) = o_p(1)$, resulting in $\boldsymbol{v}'_\eta\boldsymbol{\varepsilon}_\eta/\sqrt{n} = o_p(1)$. We now consider the first term on the right-hand side of (A.27). We note that by a Taylor series expansion:

$$
\begin{aligned}
(\boldsymbol{r}_M(\eta_i)-\boldsymbol{r}_M(\hat{\eta}_i))'\boldsymbol{\alpha}_\eta &= \sum_{m=1}^{M}[r_m(\eta_i)-r_m(\hat{\eta}_i)]\boldsymbol{\alpha}_\eta^m = \sum_{m=1}^{M}\frac{dr_m(\bar{\eta}_i)}{d\eta}(\eta_i-\hat{\eta}_i)\boldsymbol{\alpha}_\eta^m \\
&= \left(\sum_{m=1}^{M}\frac{dr_m(\bar{\eta}_i)}{d\eta}\boldsymbol{\alpha}_\eta^m - \frac{d\theta(\bar{\eta}_i)}{d\eta}\right)(\eta_i-\hat{\eta}_i) + \left(\frac{d\theta(\bar{\eta}_i)}{d\eta} - \frac{d\theta(\eta_i)}{d\eta}\right)(\eta_i-\hat{\eta}_i) \\
&\quad + \frac{d\theta(\eta_i)}{d\eta}(\eta_i-\hat{\eta}_i) \\
&= O_p\left(M^{-(s_\eta-1)}\xi_0(K)\Delta_\eta\right) + o_p(1) + \frac{d\theta(\eta_i)}{d\eta}(\eta_i-\hat{\eta}_i) \\
&= \frac{d\theta(\eta_i)}{d\eta}(\eta_i-\hat{\eta}_i) + o_p(1),
\end{aligned}
$$

where $\bar{\eta}_i$ is between η_i and $\hat{\eta}_i$, and the last equality follows from $\sup_\eta|(d\boldsymbol{r}_M(\eta)'/d\eta)\boldsymbol{\alpha}_\eta - d\theta(\eta)/d\eta| = O_p(M^{-(s_\eta-1)})$ and θ is continuously differentiable. As a result, we have:

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\boldsymbol{v}_\eta'(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) &= \sqrt{n}\frac{1}{n}\boldsymbol{v}_\eta'\begin{bmatrix}\frac{d\theta(\eta_1)}{d\eta}(\eta_1-\hat{\eta}_1)\\ \vdots \\ \frac{d\theta(\eta_n)}{d\eta}(\eta_n-\hat{\eta}_n)\end{bmatrix} + o_p(1) \\
&= \sqrt{n}\frac{1}{n}\boldsymbol{v}_\eta'\left(\frac{d\theta(\boldsymbol{\eta})}{d\eta} * (\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})\right) + o_p(1) \\
&= \sqrt{n}\frac{1}{n}\boldsymbol{v}_\eta'\left(\frac{d\theta(\boldsymbol{\eta})}{d\eta} * (\boldsymbol{\eta}-\mathbf{P}_{zK}\boldsymbol{\eta}-\mathbf{P}_{zK}\boldsymbol{u})\right) + o_p(1) \\
&= \sqrt{n}\frac{1}{n}\boldsymbol{v}_\eta'\left(\frac{d\theta(\boldsymbol{\eta})}{d\eta} * (\boldsymbol{\eta}-\mathbf{P}_{zK}\boldsymbol{\eta}) - \frac{d\theta(\boldsymbol{\eta})}{d\eta} * \mathbf{P}_{zK}\boldsymbol{u})\right) + o_p(1) \\
&= -\sqrt{n}\frac{1}{n}\boldsymbol{v}_\eta'\left(\frac{d\theta(\boldsymbol{\eta})}{d\eta} * \mathbf{P}_{zK}\boldsymbol{u}\right) + o_p(1) \\
&= -\sqrt{n}\frac{1}{n}\boldsymbol{v}_\eta'\left(\frac{d\theta(\boldsymbol{\eta})}{d\eta} * (\boldsymbol{u}+\mathbf{P}_{zK}\boldsymbol{u}-\boldsymbol{u})\right) + o_p(1) \\
&= -\sqrt{n}\frac{1}{n}\boldsymbol{v}_\eta'\left(\frac{d\theta(\boldsymbol{\eta})}{d\eta} * \boldsymbol{u}\right) + o_p(1),
\end{aligned}
$$

where $*$ denotes the element-by-element product operator, and the fourth to last equalities are based on the similar arguments of (A.6). We then end up with establishing:

$$
\frac{1}{\sqrt{n}}\hat{\mathbf{X}}'\hat{\mathbf{Q}}_{\eta M}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}+\boldsymbol{\varepsilon}-\hat{\boldsymbol{\varepsilon}}) = \frac{1}{\sqrt{n}}\boldsymbol{v}_\eta'\boldsymbol{\varepsilon} - \frac{1}{\sqrt{n}}\boldsymbol{v}_\eta'\left(\frac{d\theta(\boldsymbol{\eta})}{d\eta} * \boldsymbol{u}\right) + o_p(1). \tag{A.28}
$$

Putting together (A.28), the established result that $\mathbf{X}'\hat{\mathbf{Q}}_{\eta M}\mathbf{X}/n \xrightarrow{p} E((\mathbf{x}_i - E(\mathbf{x}_i|\eta_i))(\mathbf{x}_i -$

$E(\mathbf{x}_i|\eta_i))') \equiv \Phi_c$, the Lindberg-Levy central limit theorem, and thus Slutsky's theorem completes the proof.

A.4 Proof of Theorem 5.1

Our proof is similar to that of Theorem 1 of Sun and Li (2006). To begin with, let:

$$\begin{aligned}I_n &= n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\varepsilon_i \boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\varepsilon_j,\\ S_n &= 2\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\right)^2\varepsilon_i^2\varepsilon_j^2.\end{aligned}$$

To show the first part, it is sufficient to present that (a) $\hat{\mathcal{T}}_n = nI_n/S_n + o_p(1)$ and (b) $nI_n/S_n \xrightarrow{d} N(0,1)$. Part (b) has been shown in the literature, and so we only need to show (a). The main difference is that we need to account for the estimation effect from estimating the generated regressors in our test. To show (a), it is sufficient to show that $n(\hat{I}_n - I_n) = o_p(K^{-1/2})$ and $\hat{S}_n - S_n = o_p(K^{-1/2})$.

Note that:

$$\begin{aligned}\hat{I}_n &= n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\varepsilon}_i \boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\varepsilon}_j\\ &= n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}} - \hat{\eta}(\mathbf{z}_i)\hat{\alpha})\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}(y_j - \mathbf{x}'_j\hat{\boldsymbol{\beta}} - \hat{\eta}(\mathbf{z}_j)\hat{\alpha})\\ &= n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\varepsilon_i + \mathbf{x}'_i(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) + (\eta(\mathbf{z}_i)(\alpha-\hat{\alpha}) + \hat{\alpha}(\eta(\mathbf{z}_i) - \hat{\eta}(\mathbf{z}_i)))\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\\ &\qquad\qquad (\varepsilon_j + \mathbf{x}'_j(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) + (\eta(\mathbf{z}_j)(\alpha-\hat{\alpha}) + \hat{\alpha}(\eta(\mathbf{z}_j) - \hat{\eta}(\mathbf{z}_j)))\\ &\equiv I_n + I_{2n} + I_{3n} + I_{4n} + 2I_{12n} + 2I_{13n} + 2I_{14n} + 2I_{23n} + 2I_{24n} + 2I_{34n},\end{aligned}$$

where we will define I_{jn} terms later. We want to show that $n(\hat{I}_n - I_n)/\sqrt{N} = o_p(1)$, and it is sufficient to show that $nI_{jn}/\sqrt{N} = o_p(1)$ for all terms. Note that:

$$I_{2n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \mathbf{x}_i(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\mathbf{x}'_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}),$$
$$I_{3n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \eta(\mathbf{z}_i)(\alpha - \hat{\alpha})\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\eta(\mathbf{z}_j)(\alpha - \hat{\alpha}),$$
$$I_{23n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \mathbf{x}'_i(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\eta(\mathbf{z}_j)(\alpha - \hat{\alpha}),$$

are similar to that of I_{2n} in Sun and Li (2006). Therefore, by the same argument of Sun and Li (2006), we have $I_{2n} = O_p(n^{-1})$, $I_{3n} = O_p(n^{-1})$, and $I_{23n} = O_p(n^{-1})$, and so $nI_{2n}/\sqrt{N} = o_p(n^{-1})$, $nI_{3n}/\sqrt{N} = o_p(n^{-1})$ and $nI_{23n}/\sqrt{N} = o_p(n^{-1})$ since $N \to \infty$. Next, we have:

$$\begin{aligned}
I_{4n} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \hat{\alpha}(\eta(\mathbf{z}_i) - \hat{\eta}(\mathbf{z}_i))\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}(\eta(\mathbf{z}_j) - \hat{\eta}(\mathbf{z}_j)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i}^{n} \hat{\alpha}(\eta(\mathbf{z}_i) - \hat{\eta}(\mathbf{z}_i))\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}(\eta(\mathbf{z}_j) - \hat{\eta}(\mathbf{z}_j)) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n} \hat{\alpha}(\eta(\mathbf{z}_i) - \hat{\eta}(\mathbf{z}_i))\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}'_{Ni}\hat{\alpha}(\eta(\mathbf{z}_i) - \hat{\eta}(\mathbf{z}_i)) \\
&\equiv I_{4n,1} - I_{4n,2}.
\end{aligned}$$

First,

$$\begin{aligned}
I_{4n,1} &= \frac{1}{n}\hat{\alpha}^2(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})'\boldsymbol{S}_N(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{S}'_N(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}) \\
&\leq \hat{\alpha}^2\frac{1}{n}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})'(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}) = O_p\Big(\frac{K}{n} + K^{-2s_z/d_z}\Big),
\end{aligned}$$

where the first equality holds when we rewrite the $I_{4n,1}$ in matrix form, the first inequality holds, because $\boldsymbol{S}_N(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{S}'_N$ is idempotent, and the last equality holds by standard argument in a series estimation such as Theorem 1 of Newey (1997). By Assumption 5.1, we have $nI_{4n,1}/\sqrt{N} = o_p(1)$. For the term $I_{4n,2}$, we thus have:

$$
\begin{aligned}
|I_{4n,2}| &= \Big|\frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}(\eta(\mathbf{z}_i)-\hat{\eta}(\mathbf{z}_i)))\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Ni}\hat{\alpha}(\eta(\mathbf{z}_i)-\hat{\eta}(\mathbf{z}_i)))\Big| \\
&\le \hat{\alpha}^2\sup_i\{\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Ni}\}\frac{1}{n}\sum_{i=1}^{n}(\eta(\mathbf{z}_i)-\hat{\eta}(\mathbf{z}_i))^2 \\
&= o_p(1)\cdot O_p\Big(\frac{K}{n}+K^{-2s_z/d_z}\Big).
\end{aligned}
$$

It then follows that $nI_{4n,2}/\sqrt{N} = o_p(1)$. Next, we note that:

$$
\begin{aligned}
I_{12n} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\varepsilon_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\mathbf{x}'_j(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}), \\
I_{13n} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\varepsilon_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\eta(\mathbf{z}_j)(\alpha-\hat{\alpha}).
\end{aligned}
$$

The terms I_{12n} and I_{13n} terms are similar to I_{2n} of Sun and Li (2006). Therefore, by the same argument of Sun and Li (2006), we have $I_{12n} = O_p(n^{-1})$ and $I_{13n} = O_p(n^{-1})$, and so $nI_{12n}/\sqrt{N} = o_p(1)$ and $nI_{13n}/\sqrt{N} = o_p(1)$.

Under regularity conditions, we know that for any K, there exists a $K\times 1$ vector of ϑ such that $\sup_{\mathbf{z}\in\mathcal{Z}}|\eta(\mathbf{z})-\boldsymbol{p}_K(\mathbf{z})\vartheta| = O(K^{-s_z/d_z})$ for all K. Let $\hat{\vartheta}_n = (\boldsymbol{P}'_{zK}\boldsymbol{P}_{zK})^{-1}\boldsymbol{P}'_{zK}\mathbf{s}$. For the term I_{14n}, we write:

$$
\begin{aligned}
I_{14n} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\varepsilon_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}(\eta(\mathbf{z}_j)-\boldsymbol{p}_K(\mathbf{z}_j)'\hat{\vartheta}_n) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\varepsilon_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}(\eta(\mathbf{z}_j)-\boldsymbol{p}_K(\mathbf{z}_j)'\vartheta+\boldsymbol{p}_K(\mathbf{z}_j)'\vartheta-\boldsymbol{p}_K(\mathbf{z}_j)'\hat{\vartheta}_n)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\varepsilon_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}(\eta(\mathbf{z}_j)-\boldsymbol{p}_K(\mathbf{z}_j)'\vartheta) \\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\varepsilon_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}\boldsymbol{p}_K(\mathbf{z}_j)'(\vartheta-\hat{\vartheta}_n)) \\
&\quad-\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}(\eta(\mathbf{z}_i)-\boldsymbol{p}_K(\mathbf{z}_i)'\vartheta) \\
&\quad-\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}\boldsymbol{p}_K(\mathbf{z}_i)'(\vartheta-\hat{\vartheta}_n)) \\
&= \hat{\alpha}(I_{14n,1}+I_{14n,2}(\vartheta-\hat{\vartheta}_n)-I_{14n,3}-I_{14n,4}(\vartheta-\hat{\vartheta}_n)).
\end{aligned}
$$

Let $C = \sup_{\mathbf{x}\in\mathcal{X},\mathbf{z}\in\mathcal{Z}} \sigma^2(\mathbf{x},\mathbf{z})$ be a finite number. Following the similar argument for the term $I_{2n,1}$ of Sun and Li (2006) and by the fact that $\boldsymbol{S}_N(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{S}'_N$ is idempotent, we have:

$$E\|I^2_{14n,1}\|^2 \leq C \sup_{\mathbf{z}\in\mathcal{Z}} \left(\eta(\mathbf{z}_j) - \boldsymbol{p}_K(\mathbf{z}_j)'\vartheta\right)^2 n^{-2} E\Big[\sum_{i=1}^{n} 1\Big] = O(K^{-2s_z/d_z})O(n^{-1}).$$

By Chebyshev's inequality, we now have $I_{14n,1} = O_p(K^{-s_z/d_z}n^{-1/2})$, and it implies that $nI_{14n,1}/\sqrt{K} = o_p(1)$. Similarly, we have:

$$E|I^2_{14n,2}| \leq Cn^{-2}E\Big[\sum_{i=1}^{n} p_k^2(\mathbf{z}_i)\Big] = O(Kn^{-1}),$$

since we can always normalize the series such that $E[\boldsymbol{p}_K(\mathbf{z})\boldsymbol{p}_K(\mathbf{z})'] = \mathbf{I}$, and this implies that $E[p_k^2(\mathbf{z})] = 1$ for all k. As in (A.2) of Newey (1997), we have $\|\vartheta - \hat{\vartheta}_n\| = O_p(\sqrt{K}n^{-1/2} + K^{-s_z/d_z})$.

As a result, we have:

$$\frac{nI_{14n,2}(\vartheta - \hat{\vartheta}_n)}{\sqrt{N}} = n \cdot O_p(\sqrt{K}n^{-1/2}) \cdot O_p(\sqrt{K}n^{-1/2} + K^{-s_z/d_z})O(K^{-1/2}) = o_p(1).$$

For the term $I_{14n,3}$, by the same argument for $I_{2n,2}$ in Sun and Li (2006), we have:

$$E(I^2_{14n,3}) \leq Cn^{-2}\sup_{\mathbf{z}\in\mathcal{Z}}(\eta(\mathbf{z}) - \boldsymbol{p}_K(\mathbf{z})'\vartheta)^2 \sum_{i=1}^{n} E\Big[\sup_i \left(\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Ni}\right)\Big] = OK^{-2s_z/d_z}o(n^{-1}).$$

We then have $I_{14n,3} = o_p(K^{-s_z/d_z}n^{-1/2})$, and this implies that $nI_{14n,1}/\sqrt{N} = o_p(1)$. Similarly, we can show that:

$$E[I^2_{14n,4}] \leq Cn^{-2}E\Big[\sum_{i=1}^{n} p_k^2(\mathbf{z}_i)\Big] = O_p(Kn^{-1})\sum_{i=1}^{n} E\Big[p_k^2(\mathbf{z}_i)\sup_i \left(\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Ni}\right)\Big] = o(Kn^{-1})$$

and this implies that $nI_{14n,4}(\vartheta - \hat{\vartheta}_n)/\sqrt{N} = o_p(1)$. Next,

$$\begin{aligned}
I_{24n} = & \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{x}_i\boldsymbol{s}_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}(\eta(\mathbf{z}_j) - \hat{\eta}(\mathbf{z}_j)) \\
= & \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{x}_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Nj}\hat{\alpha}(\eta(\mathbf{z}_j) - \hat{\eta}(\mathbf{z}_j)) \\
& - \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{x}_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Ni}\hat{\alpha}(\eta(\mathbf{z}_i) - \hat{\eta}(\mathbf{z}_i)) \\
= & \hat{\alpha}(I_{24n,1} - I_{24n,2}).
\end{aligned}$$

First, we note:

$$I_{24n,1} = \frac{1}{n}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{P}_{zK}\mathbf{P}_{zK}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}),$$

and

$$\begin{aligned}
|I_{24n,1}|^2 &\leq n^{-2}\|(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{P}_{zK}\|^2\|\mathbf{P}_{zK}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})\|^2 \\
&\leq n^{-2}\|(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\|^2\|(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})\|^2.
\end{aligned}$$

We have $n^{-1}\|(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\|^2 = O_p(n^{-1})$ and $n^{-1}\|(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})\|^2 = O_p(Kn^{-1} + K^{-2s_z/d_z})$. It then follows that: $I_{24n,1} = O_p(\sqrt{K}n^{-1} + K^{-s_z/d_z}n^{-1/2})$ and $nI_{24n,1}/\sqrt{N} = o_p(1)$. For the term $I_{24n,2}$, we write:

$$\begin{aligned}
|I_{24n,2}|^2 &= \Big|\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{x}_i\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Ni}\hat{\alpha}(\eta(\mathbf{z}_i) - \hat{\eta}(\mathbf{z}_i))\Big|^2 \\
&\leq \sup_i(\boldsymbol{s}'_{Ni}(\boldsymbol{S}'_{wN}\boldsymbol{S}_{wN})^{-1}\boldsymbol{s}_{Ni})^2\Big|\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{x}_i\hat{\alpha}(\eta(\mathbf{z}_i) - \hat{\eta}(\mathbf{z}_i))\Big|^2 \\
&= o_p(1)\cdot O_p(Kn^{-2} + K^{-2s_z/d_z}n^{-2}) = o_p(Kn^{-2} + K^{-2s_z/d_z}n^{-2}).
\end{aligned}$$

It follows that $nI_{24n,2}/\sqrt{N} = o_p(1)$. The term I_{34n} is similar to I_{24n}, and so it holds that $nI_{34n,2}/\sqrt{N} = o_p(1)$. All these results are sufficient to show that $n(\hat{I}_n - I_n) = o_p(K^{-1/2})$.

The arguments to show that $(\hat{S}_n - S_n) = o_p(K^{-1/2})$ are similar, and therefore we omit the details. It has been shown that there exists $0 < c < 1$ such that $cN \leq S_n^2 \leq c^{-1}N$. The conditions $(\hat{S}_n - S_n) = o_p(N^{-1/2})$ and $cN \leq S_n^2 \leq c^{-1}N$ imply that $\hat{S}_n/S_n \xrightarrow{p} 1$ and

$n(\hat{I}_n - I_n)/S_n = o_p(1)$. Therefore, we conclude:

$$\begin{aligned}
&\frac{n\hat{I}_n}{\hat{S}_n} - \frac{nI_n}{S_n} = \frac{n\hat{I}_n S_n/\hat{S}_n - nI_n}{S_n} \\
&= \frac{(n\hat{I}_n - nI_n) + o_p(1)\cdot n\hat{I}_n}{S_n} = \frac{(n\hat{I}_n - nI_n)}{S_n} + o_p(1)\frac{n\hat{I}_n}{S_n} \\
&= o_p(1) + o_p(1) = o_p(1),
\end{aligned}$$

which establishes part (a) of Theorem 5.1. To show part (b), let $(\boldsymbol{\beta}^*, \alpha^*)$ denote the probability limit of $(\hat{\boldsymbol{\beta}}, \hat{\alpha})$ under the alternative, which is also called the pseudo-true parameters. We can show that $\hat{I}_n \xrightarrow{p} E[(E[y|\mathbf{x},\mathbf{z}] - ((\boldsymbol{\beta}^*, \alpha^*)(\mathbf{x}', \eta(\mathbf{z}))'))^2] > 0$ and $\hat{S}_n = O_p(\sqrt{N})$, and so we have $n\hat{I}_n/\hat{S}_n \to \infty$ at the rate $n/\sqrt{N}$. This is sufficient to show the consistency of our specification test.

A.5 Tables and Figures

Table 2: Simulation results for the estimation of the parametric component α and nonparametric component θ_1 in DGPs 1-4

		BIAS	SD	RMSE	RMSE.N	BIAS	SD	RMSE	RMSE.N
		DGP 1				DGP 2			
$n = 250$	FISE	-0.0003	0.0187	0.0187	-	-0.0002	0.0188	0.0188	-
	PISE	-0.0001	0.0194	0.0194	0.0547	0.0004	0.0234	0.0234	0.2636
	2SSE	-0.0153	0.0276	0.0315	0.0799	-0.0200	0.0318	0.0376	0.3120
$n = 500$	FISE	-0.0001	0.0130	0.0130	-	0.0002	0.0124	0.0124	-
	PISE	-0.0001	0.0131	0.0131	0.0370	0.0006	0.0155	0.0155	0.2097
	2SSE	-0.0079	0.0204	0.0219	0.0569	-0.0070	0.0244	0.0254	0.2848
$n = 1000$	FISE	-0.0004	0.0093	0.0093	-	0.0000	0.0093	0.0093	-
	PISE	-0.0005	0.0095	0.0095	0.0260	-0.0002	0.0113	0.0113	0.1828
	2SSE	-0.0036	0.0157	0.0161	0.0441	-0.0024	0.0199	0.0201	0.2775
$n = 4000$	FISE	0.0001	0.0044	0.0044	-	0.0001	0.0046	0.0046	-
	PISE	0.0001	0.0044	0.0044	0.0126	0.0002	0.0054	0.0054	0.1812
	2SSE	0.0003	0.0080	0.0080	0.0171	-0.0004	0.0113	0.0113	0.1929
		DGP 3				DGP 4			
$n = 250$	FISE	0.0006	0.0195	0.0195	-	-0.0002	0.0191	0.0191	-
	PISE	-0.0096	0.0228	0.0247	0.3398	-0.0095	0.0212	0.0232	0.1005
	2SSE	-0.0274	0.0293	0.0401	0.3459	-0.0252	0.0287	0.0382	0.1132
$n = 500$	FISE	0.0001	0.0135	0.0135	-	-0.0004	0.0133	0.0133	-
	PISE	-0.0089	0.0177	0.0198	0.3335	-0.0086	0.0163	0.0185	0.0892
	2SSE	-0.0198	0.0228	0.0302	0.3406	-0.0182	0.0217	0.0283	0.1043
$n = 1000$	FISE	0.0004	0.0094	0.0094	-	0.0005	0.0090	0.0090	-
	PISE	-0.0049	0.0140	0.0148	0.3258	-0.0062	0.0129	0.0143	0.0771
	2SSE	-0.0174	0.0194	0.0261	0.3395	-0.0158	0.0168	0.0230	0.0960
$n = 4000$	FISE	-0.0002	0.0045	0.0045	-	-0.0001	0.0046	0.0046	-
	PISE	-0.0003	0.0056	0.0056	0.2886	-0.0007	0.0064	0.0064	0.0305
	2SSE	-0.0175	0.0146	0.0228	0.3333	-0.0169	0.0113	0.0203	0.0817

Notes: FISE, PISE, and 2SSE refer to the Fully Infeasible Series Estimator (assuming the nonparametric components are known), Partially Infeasible Series Estimator (assuming the generated regressor is known), and the proposed feasible Two-step Series Estimators, respectively. RMSE.N stands for root meam squared errors of the series estimates of the nonparametric functions evaluated at data points.

Table 3: Simulation results for the estimation of the parametric component β and nonparametric component θ_2 in DGPs 5-8

		BIAS	SD	RMSE	RMSE.N	BIAS	SD	RMSE	RMSE.N
		DGP 5				DGP 6			
$n = 250$	FISE	-0.0006	0.0373	0.0373	-	-0.0014	0.0362	0.0362	-
	PISE	-0.0010	0.0375	0.0375	0.0537	-0.0014	0.0385	0.0385	0.2104
	2SSE	-0.0009	0.0397	0.0398	0.0806	-0.0011	0.0396	0.0396	0.3103
$n = 500$	FISE	-0.0033	0.0257	0.0259	-	0.0008	0.0264	0.0264	-
	PISE	-0.0032	0.0257	0.0259	0.0363	0.0010	0.0275	0.0275	0.1560
	2SSE	-0.0027	0.0271	0.0273	0.0555	0.0013	0.0283	0.0284	0.2324
$n = 1000$	FISE	0.0000	0.0190	0.0190	-	-0.0001	0.0179	0.0179	-
	PISE	0.0000	0.0190	0.0190	0.0263	-0.0001	0.0183	0.0183	0.1245
	2SSE	0.0003	0.0200	0.0200	0.0394	0.0000	0.0190	0.0190	0.1791
$n = 4000$	FISE	-0.0001	0.0090	0.0090	-	-0.0001	0.0091	0.0091	-
	PISE	-0.0001	0.0090	0.0090	0.0129	-0.0002	0.0093	0.0093	0.1264
	2SSE	-0.0002	0.0096	0.0096	0.0193	-0.0002	0.0097	0.0097	0.1406
		DGP 7				DGP 8			
$n = 250$	FISE	-0.0007	0.0376	0.0376	-	0.0010	0.0371	0.0371	-
	PISE	-0.0011	0.0391	0.0391	0.2730	0.0011	0.0381	0.0381	0.1132
	2SSE	-0.0013	0.0419	0.0419	0.2823	0.0012	0.0397	0.0397	0.1270
$n = 500$	FISE	0.0009	0.0251	0.0252	-	0.0005	0.0252	0.0252	-
	PISE	0.0011	0.0258	0.0258	0.2100	0.0007	0.0254	0.0254	0.0900
	2SSE	0.0012	0.0275	0.0275	0.2183	0.0005	0.0270	0.0270	0.1002
$n = 1000$	FISE	0.0008	0.0179	0.0179	-	0.0000	0.0181	0.0181	-
	PISE	0.0009	0.0185	0.0185	0.1569	-0.0001	0.0183	0.0183	0.0604
	2SSE	0.0007	0.0199	0.0199	0.1628	0.0001	0.0194	0.0194	0.0682
$n = 4000$	FISE	-0.0003	0.0091	0.0091	-	0.0004	0.0093	0.0093	-
	PISE	-0.0002	0.0092	0.0092	0.1269	0.0004	0.0093	0.0093	0.0308
	2SSE	-0.0003	0.0098	0.0098	0.1285	0.0002	0.0098	0.0098	0.0340

Notes: FISE, PISE, and 2SSE refer to the Fully Infeasible Series Estimator (assuming that the nonparametric components are known), Partially Infeasible Series Estimator (assuming that the generated regressor is known), and the proposed feasible Two-step Series Estimators, respectively. RMSE.N stands for root meam squared errors of the series estimates of the nonparametric functions evaluated at data points.

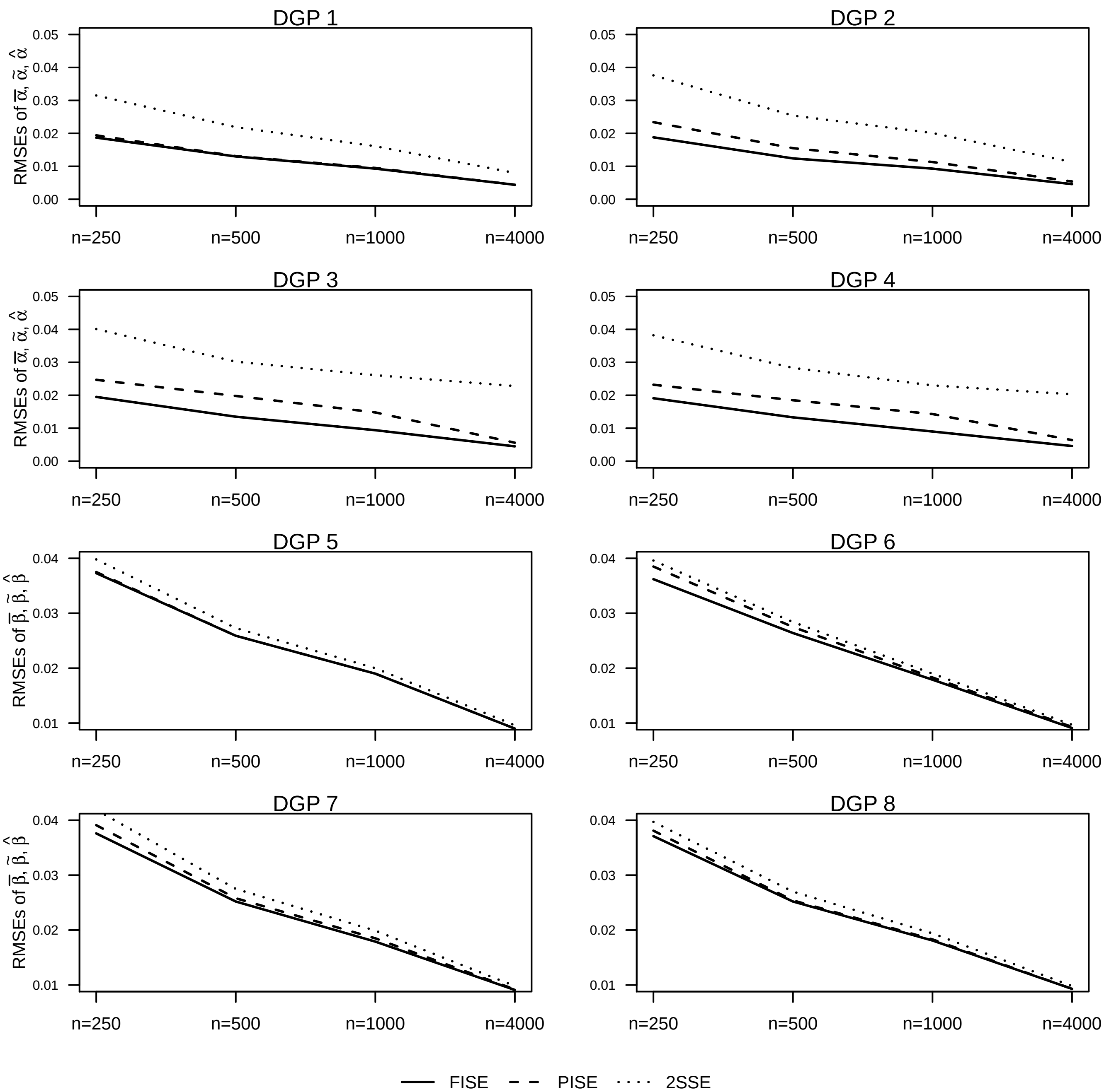

Figure 1: Comparisons of RMSEs for finite-dimenstional parameter estimation

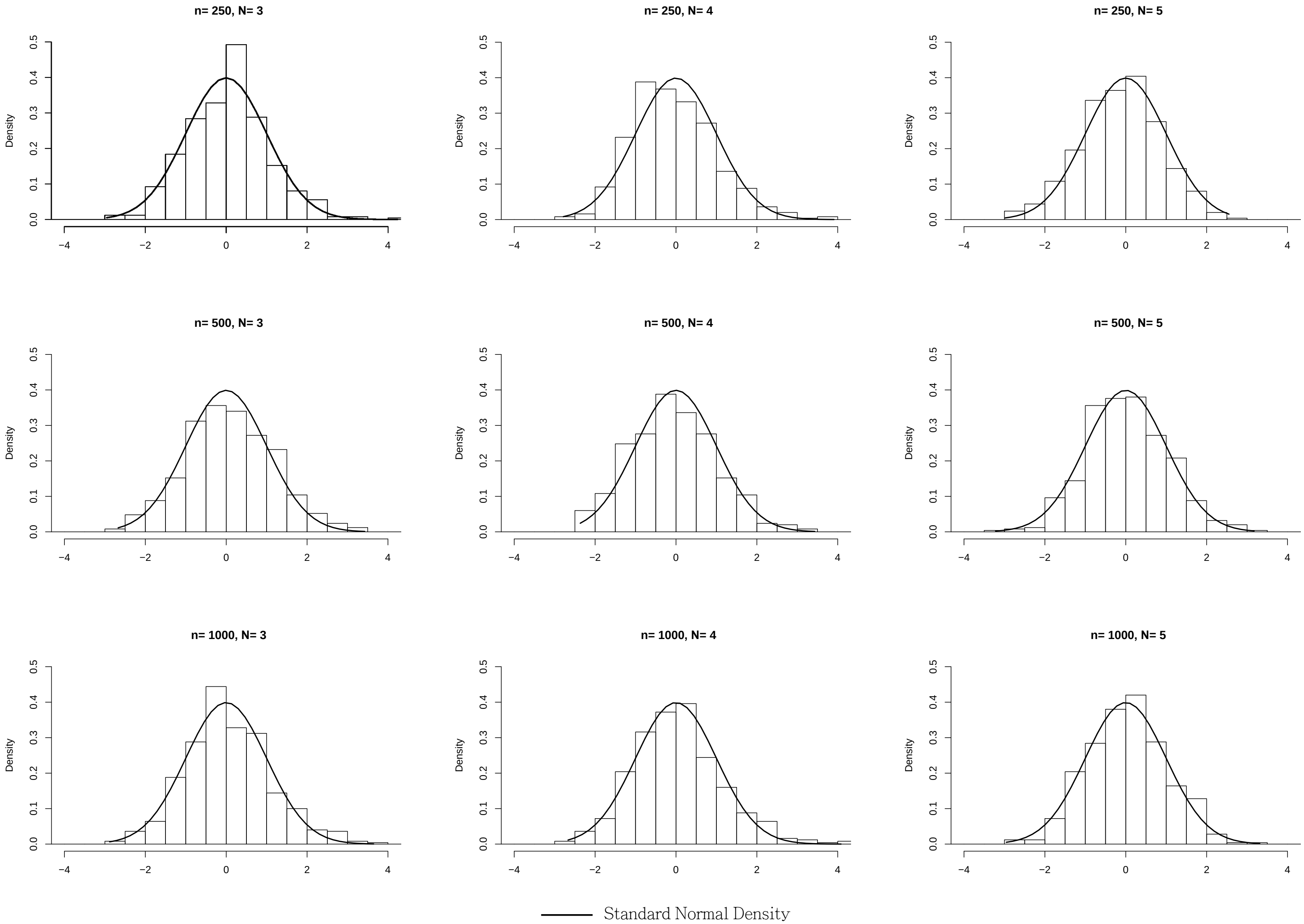

Figure 2: The simulated distributions of the test statistic $\hat{\mathcal{T}}_n$

References

Ackerberg, D., X. Chen, and J. Hahn (2012): "A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators," *Review of Economics and Statistics*, 94, 481–498.

Ahn, H., and J. L. Powell (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58(1), 3–29.

Andrews, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59(2), 307–345.

——— (1994): "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity," *Econometrica*, 62(1), 43–72.

Banerjee, A. V., and E. Duflo (2003): "Inequality and Growth: What Can the Data Say," *Journal of Economic Growth*, 8, 267–299.

Barro, R. J. (1977): "Unanticipated Money Growth and Unemployment in the United States," *The American Economic Review*, 67(2), 101–115.

Blundell, R., and F. Windmeijer (2000): "Identifying Demand for Health Resources Using Waiting Times Information," *Health Economics*, 9(6), 465–474.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101, 2754–2781.

Chen, H. (1988): "Convergence Rates for Parametric Components in a Partly Linear Model," *Annals of Statistics*, 16(1), 136–146.

Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric Estimation of Sample Selection Models," *The Review of Economic Studies*, 70(1), 33–58.

Donald, S., and W. Newey (1994): "Series Estimation of Semilinear Models," *Journal of Multivariate Analysis*, 50(1), 30–40.

Donald, S. G. (1992): "Some Asymptotic Results for Averages of Series Based Nonparametric Estimates," Working paper, Department of Economics, University of Florida.

Engle, R. F., C. W. J. Granger, J. Rice, and A. Weiss (1986): "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81(394), 310–320.

Gawande, K. (1998): "Generated Regressors in Linear and Nonlinear Models," *Economic Letters*, 54, 119–126.

Grossman, G. M., and E. Helpman (1994): "Protection for Sale," *American Economic Review*, 84(4), 833–850.

Hansen, B. E. (2018): "Econometrics," Lecture Notes, The University of Wisconsin.

Härdle, W., H. Liang, and J. Gao (2000): *Partially Linear Models*. Physica-Verlag Heidelberg.

Heckman, J. J. (1978): "Dummy Endogenous Variables in A Simultaneous Equation System," *Econometrica*, 46, 931–959.

Heckman, N. E. (1986): "Spline Smoothing in a Partly Linear Model," *Journal of the Royal Statistical Society, Series B*, 48(2), 244–248.

Henderson, D. J., and C. F. Parmeter (2015): "Single-step Estimation of a Partially Linear Model," Working paper.

Hong, Y., and H. White (1995): "Consistent Specification Testing Via Nonparametric Series Regression," *Econometrica*, 63(5), 1133–1159.

Lee, S. (2007): "Endogeneity in Quantile Regression Models: A Control Function Approach," *Journal of Econometrics*, 141(2), 1131–1158.

Li, Q. (2000): "Efficient Estimation of Additive Partially Linear Models," *International Economic Review*, 41(4), 1073–1092.

Li, Q., and S. Racine (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton.

Li, Q., and J. M. Wooldridge (2002): "Semiparametric Estimation of Partially Linear Models for Dependent Data with Generated Regressors," *Econometric Theory*, 18(3), 625–645.

Lorentz, G. G. (1966): *Approximation of Functions*. New York: Chelsea.

Newey, W. K. (1994): "Series Estimation of Regression Functionals," *Econometric Theory*, 10, 1–28.

——— (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79(1), 147–168.

——— (2009): "Two-Step Series Estimation of Sample Selection Models.," *Econometrics Journal*, 12(1), 217–229.

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67(3), 565–603.

Pagan, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25(1), 221–247.

——— (1986): "Two Stage and Related Estimators and Their Applications," *The Review of Economic Studies*, 53(4), 517–538.

Pendakur, K., and S. Sperlich (2010): "Semiparametric Estimation of Consumer Demand Systems in Real Expenditure," *Journal of Applied Econometrics*, 25(3), 420–457.

Rilstone, P. (1996): "Nonparametric Estimation of Models with Generated Regressors," *International Economic Review*, 37(2), 299–313.

Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931–954.

Speckman, P. (1988): "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society, Series B*, 50(3), 413–436.

Stengos, T., and B. Yan (2001): "Double Kernel Nonparametric Estimation in Semiparametric Econometric Models," *Journal of Nonparametric Statistics*, 13(6), 883–906.

Sun, Y., and Q. Li (2006): "An Alternative Series Based Consistent Model Specification Test," *Economics Letters*, 93(1), 37–44.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

Zhou, Y., and H. Liang (2009): "Statistical Inference for Semiparametric Varying-coefficient Partially Linear Models with Error-prone Linear Covariates," *Annals of Statistics*, 37(1), 427–458.