Sequentially Estimating the Structural Equation by Power Transformation

JAEDO CHOI

Department of Economics, University of Michigan, 611 Tappan Ave., Ann Arbor, MI 48109, USA

jaedohi@umichi.edu

JIN SEO CHO

School of Humanities and Social Sciences, Beijing Institute of Technology, Haidian, Beijing 100081, China
School of Economics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea

Email: jinseocho@yonsei.ac.kr

HYUNGSIK ROGER MOON

Department of Economics, University of Southern California, 3620 South Vermont, LA, CA 90089, USA
School of Economics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea

moonr@usc.edu

February, 2020.

Abstract

This study provides an econometric methodology to test for a linear structural relationship among economic variables. For this, we propose the so-called distance-difference (DD) test statistic and show that it has omnibus power against arbitrary nonlinear structural relationships. If the DD test statistic rejects the linear model hypothesis, a sequential testing procedure assisted by the DD test statistic can consistently estimate the degree of polynomial function that arbitrarily approximates the nonlinear structural equation. Using extensive Monte Carlo simulations, we confirm the DD test's finite sample properties and compare its performance with the sequential testing procedure assisted by the J-test statistic and moment selection criteria. Finally, we empirically investigate the structural relationship between the log wage and work experience years using Card's (1995) National Longitudinal Survey data and affirm their inferential results by our methodology.

Key Words: GMM estimation; Model linearity testing; Model specification testing; Gaussian stochastic process; Sequential testing procedure; Wage equation.

Subject Classification: C12, C13, C26, C52, J24, J31.

Acknowledgements: The authors are grateful for the helpful discussions they had with Juwon Seo. Cho further acknowledges with gratitude the research grant provided by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A2A01035256).

1 Introduction

"*To climb steep hills requires a slow pace at first.*" — William Shakespeare.

In structural empirical studies, model specification is important because it affects inferences as well as counterfactual experiments to draw important policy implications. An important aim of this study is to develop an efficient and yet easy-to-use method for researchers to test for a linear structural relationship between economic variables. The testing methodology proposed in this paper extends the ones already developed for reduced-form models.

Studies such as Bierens (1990) and Baek, Cho, and Phillips (2015) developed a methodology to test for a linear model hypothesis against general model misspecification in a reduced-form framework. In particular, Baek, Cho, and Phillips (2015) obtained the null limit distribution of the quasi-likelihood ratio (QLR) test statistic by estimating the power coefficient of the economic variable of interest, showing that it has omnibus power. We apply their methodology to the generalized method of moment (GMM) framework and test for a linear structural model hypothesis using a distance-difference (DD) test statistic as in Baek, Cho, and Phillips (2015). We then derive the null limit distribution of the DD test statistic, to show that it has omnibus power against a linear structural model.

Cho and Phillips (2018) further developed a sequential testing procedure using the QLR test statistic to consistently estimate a nonlinear reduced-form equation. In this study, we apply the sequential testing procedure to the DD test statistic as in Cho and Phillips (2018), to find that the unknown polynomial structural model can be consistently estimated using this approach. In case the structural equation differs from any polynomial equation, the polynomial equation estimated using finite samples and our sequential testing approach can be understood as an approximation of the structural equation.

We also compare our testing procedures with some widely used ones in the literature. We first consider the Sargan (1958, 1988) and Hansen (1982)) J-test used for a correctly specified structural model hypothesis and the validity of instrumental variables (*e.g.*, Newey, 1985). We conduct extensive simulations to compare the two test statistics and find that the J- and DD test statistics complement each other. In particular, the DD test statistic outperforms the J-test statistic in sequential testing. Second, we investigate Andrews's procedure (1999) of applying the Akaike (1973), Schwarz (1978), and Hannan and Quinn (1979) information criteria for selecting moment conditions, and thus introduce a procedure to ensure the number of moment conditions that identify unknown parameters. We compare the moment selection criteria (MSCs) and sequential testing procedure using simulations, to find that sequential testing outperforms MSCs.

In the semi/nonparametric literature, studies such as Hong and White (1995), Ai and Chen (2003), Newey and Powel (2003), and Chen and Pouzo (2015) investigated how to estimate and test for unknown structural equations using various semi-nonparametric methods. In contrast to these methods, the DD test statistic is fully parametric. In addition, the DD test statistic can be used as a diagnostic test statistic before applying their methodologies. If the null model is not rejected by the DD test statistic, no need would arise to estimate the structural model using nonparametric methods.

The rest of this paper is structured as follows. Section 2 tests for a nonlinear structural relationship and discusses its motivation and associated problems. This section also examines the linearity condition testing by formally introducing the DD test statistic. The null limit distribution and power properties of the test statistic are also examined. Section 3 extends the linear structure testing to polynomial structures using the same test statistic. Furthermore, sequential testing is applied to estimate the polynomial structural equation. Section 4 reports Monte Carlo simulations and compares the DD test statistics with other methodologies, while Section 5 presents an empirical application. Finally, Section 6 concludes the paper. Mathematical proofs are presented in the appendix.

Before moving to the next section, we introduce some useful mathematical notations. For functions f and $j = 1, 2, \ldots$, let $(d^j/d^jx)f(\bar{x})$ denote $(d^j/dx^j)f(x)|_{x=\bar{x}}$ for notational simplicity. We also assume that $\boldsymbol{\iota}$ is the $n \times 1$ vector of unity.

2 Motivation and Structural Linearity Testing

2.1 Motivation and Heuristics

To motivate this study, we first present a simple model. Assume that Y_t and X_t are dependent and positively valued explanatory variables respectively, such that for the unknown function $m(\cdot)$, their structural relationship is

$$Y_t = m(X_t) + U_t. \tag{1}$$

In (1), let X_t and U_t be correlated. One of the main aims of this study is to test whether the structural relationship between Y_t and X_t, $m(\cdot)$, is linear:

$$H_0 : m(X_t) = \xi_0 + \xi_1 X_t \quad a.s. \tag{2}$$

Some of the economic applications motivating this study include Mincer's (1958) linear model between log wage and education, given the individual's work experience years and their square term, and Balassa and Samuelson's (1964 1964) linear structural model between the ratio of purchasing power parity to exchange rate and the per capita income differentials.

We are motivated to test the null hypothesis from the possibility that the linear model is arbitrarily misspecified. For example, the linear relationship between log wage and education years posited by Mincer (1958) has been questioned in the literature. Mincer (1997) himself obtained a nonlinear education yield function by assuming heterogeneous preferences and earnings opportunities for individuals. As another example, Card and Krueger (1992) obtained a nonlinear return to education along with the so-called credential effect. In such cases, estimating linear models using GMM estimation would introduce an asymptotic bias (*e.g.*, Hall and Inoue, 2003), rendering the asymptotic distribution model dependent.

A number of studies have tested the linearity condition using the semi- or nonparametric method. For example, for X_t as an exogenous variable, Hong and White (1995) estimated $m(\cdot)$ using a sieve series estimation method, to provide an omnibus specification testing. For X_t as an endogenous variable, Chen and Pouzo (2015) estimated $m(\cdot)$ using a penalized semi-parametric minimum distance estimation method and a sieve series under the complete conditional distribution condition of X_t on instrumental variables. They also provided a methodology to test the model hypothesis using the Wald and QLR test statistics, and further showed that $m(\cdot)$ can be consistently estimated by letting the number of sieve series to increase as the sample size increases, introducing a methodology to consistently test the correct model assumption. Nevertheless, for a finite sample size, the testing results would depend on the selected penalty and/or degree of sieves.

In this study, we aim to provide a test statistic that consistently detects arbitrary nonlinearity rather well and overcomes the challenges associated with sieve series estimation, and thus lead to a simple and straightforward testing procedure.

For this, we first extend the approach of Bierens (1990) and Baek, Cho, and Phillips (2015), who estimate $m(\cdot)$ when X_t is exogenous. We then heuristically describe our testing procedure. We specify the parametric model for the structural error U_t as follows:

$$\mathcal{M} := \{m_t(\xi_0, \delta, \beta, \gamma) := Y_t - \xi_0 - \xi_1 X_t - \beta X_t^{\gamma} : (\xi_0, \xi_1, \beta, \gamma) \in \mathbf{\Omega} \subset \mathbb{R}^4\}.$$

We then estimate the unknown parameters using the GMM estimation method. Note that the linear model is nested in $\mathcal{M}$ as a special case. If $\gamma_* = 0, 1$, or $\beta_* = 0$, then Y_t and X_t would be structurally linear, requiring the linear structure hypothesis to be jointly tested via the hypotheses on γ_* and β_*. In this study, we apply the likelihood-ratio (LR) test principle to test the linearity hypothesis. That is, we compare the Sargan (1958, 1988) and Hansen (1982) J-test statistics implied by $\mathcal{M}$ and the linear model, and reject the linearity hypothesis if the difference between the two J-test statistics is sufficiently large. We formally define our test statistic below; this is the DD test statistic.

The DD test statistic based on $\mathcal{M}$ has the following useful properties over other methodologies in the literature. First, the DD test statistic is consistent for general nonlinearity, because the power transform X_t^γ in $\mathcal{M}$ is a sieve basis. In any continuous function $m(\cdot)$, including $(\cdot)^j$ as regressors with $j = 1, 2, 3, \ldots$ would approximate $m(\cdot)$ arbitrarily well (*e.g.*, Chen and Liu, 2014); this means that if $m(\cdot)$ is a linear function, adding any sieve basis to the linear function as regressor would fail to reduce the approximation error measured by the GMM distance. Here, we propose the DD test statistic to compare the GMM distances measured by the linear model and $\mathcal{M}$ in parallel with the LR test statistic. Note that the degree of sieve basis γ is estimated to obtain the optimum sieve that best improves the DD test statistic, instead of including the maximum number of sieve series limited by sample size. Second, we compute the DD test statistic using the GMM estimation method without assuming the complete conditional probability distribution of X_t on instrumental variables; thus, we do not consider the penalty function of $m(\cdot)$ in our estimation. Furthermore, $\mathcal{M}$ is a fully parametric model, making the associated inferences straightforward. Third, the DD test statistic can play the role of diagnostic test statistic before estimating $m(\cdot)$ using other methodologies. Note that Ai and Chen (2003), Newey and Powel (2003), and Chen and Pouzo (2015) estimate the unknown structural equation using the semi-nonparametric minimum distance and nonparametric two-stage least squares estimation methods, respectively; these methods can be computationally demanding. If the DD test statistic does not reject the linear model assumption, no need would arise for their estimations.

A popular trend in the literature is to test the linear model assumption using $\mathcal{M}$ or such other models. First, when X_t is an exogenous variable, Bierens (1990) and Baek, Cho, and Phillips (2015) tested the linear model misspecification using models similar to $\mathcal{M}$, as mentioned above. However, to the best of our knowledge, no study has so far tested the linearity hypothesis when X_t is endogenous. Second, the Sargan (1958, 1988) and Hansen (1982) J-test statistics typically test the structural model misspecification as well as validity of the instrumental variables. Thus, the J-test statistic rejecting the null does not necessarily imply that the linear structural model is misspecified. It may reject the null because the instrumental variables are not valid. However, the DD test statistic presumes valid instrumental variables and focuses on testing the functional form misspecification with omnibus power against general nonlinear functions.

2.2 Testing Environment and Assumptions

We now formally discuss the model and data structure of interest by generalizing $\mathcal{M}$. Assume that $\{(\mathbf{W}_t', \mathbf{Z}_t', U_t)' := (X_t, \mathbf{D}_t', \mathbf{Z}_t', U_t)' : t = 1, 2, \ldots\}$ is a strictly stationary ergodic process, X_t is a positively valued endogenous variable, $\mathbf{D}_t(\in \mathbb{R}^k)$ is an exogenous variable, and $\mathbf{Z}_t(\in \mathbb{R}^m)$ is an instrumental variable with k and $m \in \mathbb{N}$. Given this data generating process (DGP), we also assume that for some $(\delta_{0*}, \boldsymbol{\delta}_*')'$, Y_t is structurally associated with other variables by

$$Y_t = \xi_{0*} + \mathbf{W}_t'\boldsymbol{\delta}_* + m(X_t) + U_t,$$

such that for the instrumental variable $\mathbf{Z}_t$, $\mathbb{E}[U_t\mathbf{Z}_t] = \mathbf{0}$ and the order condition hold for structural model estimation, namely, $\mathbf{Z}_t \in \mathbb{R}^m$ with $m \geq k+2$. For notational simplicity, we also divide the parameter vector $\boldsymbol{\delta}_*$ into $(\xi_{1*}, \boldsymbol{\eta}_*')'$ such that $\mathbf{W}_t'\boldsymbol{\delta}_* = \xi_{1*}X_t + \mathbf{D}_t'\boldsymbol{\eta}_*$.

Given the DGP condition, we consider a model specified to test the functional form of $m(\cdot)$. In particular, we assume that the empirical researcher is interested in testing the linear structure between Y_t and X_t. To address this, we construct a model attached by a power transform of X_t as follows:

$$\mathcal{M}_1 := \left\{ m_t(\boldsymbol{\omega}) := Y_t - \xi_0 - \mathbf{W}_t'\boldsymbol{\delta} - \beta X_t^{\gamma} : \omega := (\xi_0, \boldsymbol{\delta}', \beta, \gamma)' \in \boldsymbol{\Omega} \subset \mathbb{R}^{k+4} \right\}.$$

Here, we use subscript "1" to test the linearity hypothesis of the structure, and the other subscripts to generalize the linearity hypothesis to other polynomial structures. Note that $\mathcal{M}_1$ could be misspecified under a general nonlinear structure between Y_t and X_t. As Hall and Inoue (2003) have pointed out, in such cases, the power function in $\mathcal{M}_1$ estimated using the GMM method is an approximation for $m(\cdot)$, and so the limit behavior of the estimated parameter can be different from that of a correctly specified model. However, $\mathcal{M}_1$ is correctly specified for the linear structure between Y_t and X_t. By imposing the conditions

$$\mathcal{H}_{0,1} : \beta_* = 0, \qquad \mathcal{H}_{0,2} : \gamma_* = 0, \qquad \text{or} \qquad \mathcal{H}_{0,3} : \gamma_* = 1,$$

we can generate a linear structure between Y_t and X_t and thus hypothesize the researcher's interest as the union of the sub-conditions $\mathcal{H}_0 := \mathcal{H}_{0,1} \cup \mathcal{H}_{0,2} \cup \mathcal{H}_{0,3}$. The negation of $\mathcal{H}_0$ is an alternative hypothesis: $\mathcal{H}_1 : \beta_* \neq 0$, $\gamma_* \neq 0$ or 1. For simplicity, we assume that $\boldsymbol{\Omega}_0 := \{\boldsymbol{\omega} \in \boldsymbol{\Omega} : \beta = 0,\ \gamma = 0,\ \text{or } \gamma = 1\}$ and $\boldsymbol{\Omega}_1 := \boldsymbol{\Omega} \setminus \boldsymbol{\Omega}_0$ are the null and alternative parameter spaces, respectively.

Testing the null hypothesis involves nonstandard problems. Null hypothesis $\mathcal{H}_0$ is associated with an identification problem. If $\beta_* = 0$, γ_* is unidentified, and Davies' (1977, 1987) identification arises under $\mathcal{H}_{0,1}$. Similarly, if $\gamma_* = 0$, only $\xi_{0*} + \beta_*$ is identified, implying that ξ_{0*} and β_* are not separately identified, and Davies' (1977, 1987) identification arises in a different manner under $\mathcal{H}_{0,2}$ from $\mathcal{H}_{0,1}$. Furthermore, Davies' (1977, 1987) identification arises under $\mathcal{H}_{0,3}$, implying that neither ξ_{1*} nor β_* is separately identified. Thus, we find three composite identification problems with $\mathcal{H}_0$. We call this the trifold identification problem following Baek, Cho, and Phillips (2015).

We next test the hypothesis under the following regularity conditions and formalize the above DGP and model conditions.

Assumption 1. *(i)* $\{(\mathbf{W}_t', \mathbf{Z}_t', U_t)' := (X_t, \mathbf{D}_t', \mathbf{Z}_t', U_t)' \in \mathbb{R}^{2+k+m} : t = 1, 2, \ldots\}$ *(k and $m \in \mathbb{N}$) is a strict stationary ergodic (SSE) sequence such that X_t has a positive value with probability 1;*

(iii) for each j, $\{Z_{t,j}U_t, \mathcal{F}_t\}$ is an adapted mixingale of size -1, where $Z_{t,j}$ is the j^{th}-row element, and $\mathcal{F}_t$ is the smallest σ-field generated by $\{U_t, \mathbf{Z}_t, \mathbf{W}_t, U_{t-1}, \mathbf{Z}_{t-1}, \mathbf{W}_{t-1}, \ldots\}$;

(iv) (a) for each j, $\mathbb{E}[Z_{t,j}^4] < \infty$ and $\mathbb{E}[U_t^4] < \infty$;

(b) for each j, $\mathbb{E}[D_{t,j}^2] < \infty$ and $\mathbb{E}[m^2(X_t)] < \infty$, where $D_{t,j}$ is the j^{th}-row element of $\mathbf{D}_t$;

(v) (a) $\mathrm{var}(n^{-1/2}\mathbf{Z}'\mathbf{U})$ *converges to $\mathbf{\Sigma}$ as $n \to \infty$, where n is the sample size;*

(b) $\mathrm{var}(n^{-1/2}\mathbf{Z}'\mathbf{U})$ *is PD uniformly in n, and $\mathbf{\Sigma}$ is finite and PD;*

(vi) (a) $\mathbf{M}_n$ converges to $\mathbf{M}_0$, as $n \to \infty$;

(b) $\mathbf{M}_n$ is symmetric and PD uniformly in n, and $\mathbf{M}_0$ is finite and PD. □

Assumption 2. *(i) The structural relationship between Y_t and $\mathbf{W}_t$ is specified as $\mathcal{M}_1 := \{m_t(\boldsymbol{\omega}) := Y_t - \xi_0 - \mathbf{W}_t'\boldsymbol{\delta} - \beta X_t^\gamma : \omega := (\xi_0, \boldsymbol{\delta}', \beta, \gamma)' \in \mathbf{\Omega} \subset \mathbb{R}^{k+4}\}$, where $\mathbf{\Omega} := \mathbf{\Xi} \times \mathbf{\Delta} \times \mathbf{B} \times \mathbf{\Gamma}$ such that $\mathbf{\Xi}$, $\mathbf{\Delta}$, $\mathbf{B}$, and $\mathbf{\Gamma} := [\underline{\gamma}, \overline{\gamma}]$ are convex and compact in $\mathbb{R}$, $\mathbb{R}^{k+1}$, $\mathbb{R}$, and $\mathbb{R}$, respectively, and 0 and 1 are interior elements of $\mathbf{\Gamma}$;*

(ii) for the measurable functions $m(\cdot)$ and $(\xi_{0*}, \boldsymbol{\delta}_*')' \in \mathbb{R}^{2+k}$, $Y_t = \xi_{0*} + \mathbf{W}_t'\boldsymbol{\delta}_* + m(X_t) + U_t$; and

(iii) $\mathbb{E}[\mathbf{V}_t\mathbf{Z}_t']$ and $\sum_{t=1}^n \mathbf{V}_t\mathbf{Z}_t'$ have full row ranks uniformly in n, where $\mathbf{V}_t = (1, \mathbf{W}_t')'$. □

Assumption 3. *An SSE sequence $\{M_t\}$ exists such that (i) $\mathbb{E}[M_t^2] < \infty$ and $\sup_{\gamma\in\mathbf{\Gamma}} |X_t^\gamma| \leq M_t$, and*

(ii) $\mathbb{E}[X_t^4] < \infty$ and $\mathbb{E}[L_t^4] < \infty$. □

Remarks.

(a) Assumptions 1, 2, and 3 impose the DGP, model, and moment conditions, respectively. Assumption 1 is considered throughout this study, whereas Assumptions 2 and 3 are considered only when extending the linear structure testing to polynomial structures.

(b) The DGP and moment conditions are not sufficient to apply the functional central limit theorem (FCLT) as in Baek, Cho, and Phillips (2015). However, the DGP and moment conditions of this study are regular conditions to apply Scott's (1973) mixingale central limit theorem (CLT) to $n^{-1/2}\sum \mathbf{Z}_tU_t$. We can obtain the DD test statistic null limit distribution by applying the CLT differently from Baek, Cho, and Phillips (2015), as detailed below.

(c) The DGP condition allows for a dynamic misspecification. If $\{U_t, \mathcal{F}_t\}$ forms a martingale different array (MDA), $\mathrm{var}(n^{-1/2}\mathbf{Z}'\mathbf{U})$ would be identical uniformly in n.

(d) For power transformation, X_t needs to be positive. Otherwise, X_t would be transformed to other positive variables, but we can allow them to be X_t here. Since this transformation does not substantially modify our theory, we simply assume that X_t has a positive value. $\square$

2.3 Testing Structural Linearity

We next estimate the unknown parameters using the GMM method, assuming the following quadratic distance function:

$$d_n(\boldsymbol{\omega}) := (\mathbf{Y} - \beta\mathbf{X}(\gamma) - \mathbf{V}\boldsymbol{\varsigma})'\mathbf{Z}\mathbf{M}_n\mathbf{Z}'(\mathbf{Y} - \beta\mathbf{X}(\gamma) - \mathbf{V}\boldsymbol{\varsigma}),$$

where $\mathbf{Y} := (Y_1, \ldots, Y_n)'$, $\mathbf{X}(\gamma) := (X_1^\gamma, \ldots, X_n^\gamma)'$, $\mathbf{V}_t := (1, \mathbf{W}_t')'$, $\mathbf{V} := [\mathbf{V}_1', \ldots, \mathbf{V}_n']'$, $\mathbf{Z} := [\mathbf{Z}_1', \ldots, \mathbf{Z}_n']'$, and $\boldsymbol{\varsigma} := (\xi_0, \boldsymbol{\delta}')'$. Note that we obtain the GMM estimator by minimizing the quadratic distance function: $\widehat{\boldsymbol{\omega}}_n := \arg\min_{\boldsymbol{\omega}\in\boldsymbol{\Omega}} d_n(\boldsymbol{\omega})$. We also assume that $\widetilde{\boldsymbol{\omega}}_n := \arg\min_{\boldsymbol{\omega}\in\boldsymbol{\Omega}} d_n(\boldsymbol{\omega})$ such that $\beta = 0$. If $\beta = 0$, γ is a placeholder, with $\widetilde{\boldsymbol{\omega}}_n$ estimating the linear structure between Y_t and X_t.

To test the linear structure using Wald's (1943) test principle is challenging. As Baek, Cho, and Phillips (2015) pointed out, when a multifold identification problem is associated with the null hypothesis, the Wald test statistic can most probably be unbounded under the null, because two parameters belonging to the null parameter space constrained by one of the sub-null hypotheses may belong to the alternative parameter space characterized by another sub-alternative hypothesis.

However, we apply the LR test principle to overcome the multiple identification parameter problem. Specifically, we compare the GMM distances obtained under $\mathcal{H}_0$ and $\mathcal{H}_1$ to test the linearity hypothesis. The DD test statistic is defined as follows:

$$\mathcal{D}_{n,1} := n^{-1}\left\{d_n(\widetilde{\boldsymbol{\omega}}_n) - d_n(\widehat{\boldsymbol{\omega}}_n)\right\}.$$

As earlier, subscript "1" indicates that the DD-statistic tests the linear structure between Y_t and X_t. Note that $\mathcal{M}_1$ approximates the unknown functional form of $m(\cdot)$ by the power transform, and the DD test statistic exploits this approximation to gain the test statistic marginal power; this is exactly the same motivation as that of the LR test statistic.

The DD and QLR test statistics are defined similarly, but have different structures. The GMM distance is defined by the weighted distance of the orthogonality conditions, and not by the prediction error, to obtain a null limit distribution different from that of the QLR test statistic.

We specifically examine how this aspect is associated with the null limit distribution by first deriving the null limit approximations of the test under the sub-hypotheses ($\mathcal{H}_{0,1}$, $\mathcal{H}_{0,2}$, and $\mathcal{H}_{0,3}$), and then combining them into a single statistic to yield the null limit distribution of the DD test statistic.

In our first step, we examine the limit approximation under $\mathcal{H}_{0,1} : \beta_* = 0$. Note that since γ_* is not identified under $\mathcal{H}_{0,1}$, we conduct GMM optimization with respect to γ in a later stage compared to for any other parameter: $\min_\gamma \min_\beta \min_\varsigma d_n(\boldsymbol{\omega})$. If $\mathbf{Q}_1 := \ddot{\mathbf{Z}}(\mathbf{I} - \ddot{\mathbf{Z}}'\mathbf{V}(\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V})^{-1}\mathbf{V}'\ddot{\mathbf{Z}})\ddot{\mathbf{Z}}'$, $\ddot{\mathbf{Z}} := \mathbf{Z}\mathbf{M}_n^{1/2}$, and $\mathbf{U} := (U_1, \ldots, U_n)'$, then we have

$$\mathcal{D}_{n,1}^{(\beta=0)}(\epsilon) := -\inf_{\gamma\in\boldsymbol{\Gamma}(\epsilon)}\inf_\beta n^{-1}\{d_n(\beta;\gamma) - d_n(0;\gamma)\} = \sup_{\gamma\in\boldsymbol{\Gamma}(\epsilon)} \frac{1}{n}\frac{\{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)}, \tag{3}$$

where $\boldsymbol{\Gamma}(\epsilon) := \boldsymbol{\Gamma}\backslash\{(-\epsilon,\epsilon)\cup(1-\epsilon,1+\epsilon)\}$, and $\mathcal{D}_{n,1}^{(\beta=0)}(\epsilon)$ denotes the DD test statistic designed to test $\mathcal{H}_{0,1}$. Here, the γ space is modified to $\boldsymbol{\Gamma}$ from $\boldsymbol{\Gamma}(\epsilon)$ to exclude 0 and 1. If $\gamma = 0$ or 1, the model would introduce the identification problems under $\mathcal{H}_{0,2}$ and $\mathcal{H}_{0,3}$ and complicate the derivation. We relax this restriction, as shown below, to derive the DD test statistic limit distribution under $\mathcal{H}_0$.

Thus far, we provide the limit distribution of $\mathcal{D}_{n,1}^{(\beta=0)}(\epsilon)$ under $\mathcal{H}_{0,1}$:

Lemma 1. *Given Assumptions 1, 2, 3, and $\mathcal{H}_{0,1}$, for each $\epsilon > 0$, we have $\mathcal{D}_{n,1}^{(\beta=0)}(\epsilon) \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}(\epsilon)} \mathcal{Z}_1^2(\gamma)$, where for each $\gamma \in \boldsymbol{\Gamma}(\epsilon)$, $\mathcal{Z}_1(\gamma) \sim N(0, \rho(\gamma,\gamma))$ such that for each pair (γ,γ'), $\mathbb{E}[\mathcal{Z}_1(\gamma)\mathcal{Z}_1(\gamma')] = \rho_1(\gamma,\gamma') := \kappa_1(\gamma,\gamma')/\{\sigma_1^2(\gamma)\sigma_1^2(\gamma')\}^{1/2}$, $\kappa_1(\gamma,\gamma') := \mathbb{E}[X_t^\gamma\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\widetilde{\boldsymbol{\Sigma}}\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_tX_t^{\gamma'}]$, $\sigma_1^2(\gamma) := \mathbb{E}[X_t^\gamma\widetilde{\mathbf{Z}}_t']\mathbf{J}_1$ $\mathbb{E}[\widetilde{\mathbf{Z}}_tX_t^\gamma]$, $\widetilde{\mathbf{Z}}_t := \mathbf{M}_0^{1/2}\mathbf{Z}_t$, $\widetilde{\boldsymbol{\Sigma}} := \mathbf{M}_0^{1/2}\boldsymbol{\Sigma}\mathbf{M}_0^{1/2}$, and $\mathbf{J}_1 := \mathbf{I} - \mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}_t'](\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t']\mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}_t'])^{-1}\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t']$.* $\Box$

Remarks.

(a) Although Lemma 1 represents the null limit distribution as a Gaussian stochastic process function, the associated Gaussian process is essentially the product of a deterministic γ function and a multivariate normal random variable. If for each γ, $\widetilde{\mathcal{Z}}_1(\gamma) := \boldsymbol{\pi}_1(\gamma)'\mathbf{G}$, where $\boldsymbol{\pi}_1(\gamma) := \mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_tX_t^\gamma]/\sigma_1^2(\gamma)^{1/2}$ and $\mathbf{G} \sim N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}})$, then the covariance kernel structure of $\widetilde{\mathcal{Z}}_1(\cdot)$ is identical to that of $\mathcal{Z}_1(\cdot)$, implying that the nonlinearity of $\mathcal{Z}_1(\cdot)$ stems from $\boldsymbol{\pi}_1(\cdot)$.

(b) The covariance kernel of $\mathcal{Z}_1(\cdot)$ depends on the form of $\mathbf{M}_n$. If $\mathbf{M}_n$ consistently estimates $\boldsymbol{\Sigma}^{-1}$, then $\widetilde{\boldsymbol{\Sigma}} = \mathbf{I}$ and $\kappa_1(\gamma,\gamma') = \mathbb{E}[X_t^\gamma\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_tX_t^{\gamma'}]$, because $\mathbf{J}_1$ is an idempotent matrix, and so for each γ, $\rho_1(\gamma,\gamma) = 1$, and

$$\rho_1(\gamma,\gamma') = \frac{\kappa_1(\gamma,\gamma')}{\sqrt{\kappa_1(\gamma,\gamma)}\sqrt{\kappa_1(\gamma',\gamma')}}.$$

(c) The DD test statistic null limit distribution can be obtained through simulation. If $\widehat{\boldsymbol{\pi}}_{n,1}(\cdot)$ and $\widehat{\boldsymbol{\Sigma}}_n$ consistently estimate $\boldsymbol{\pi}_1(\cdot)$ and $\widetilde{\boldsymbol{\Sigma}}$, respectively, the limit distribution of $\sup_{\gamma\in\boldsymbol{\Gamma}(\epsilon)}(\widehat{\boldsymbol{\pi}}_{n,1}(\gamma)'\widehat{\mathbf{Z}}_n)^2$ would estimate the null limit distribution of $\mathcal{D}_{n,1}^{(\beta=0)}(\epsilon)$, where $\widehat{\mathbf{Z}}_n \sim N(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_n)$. Hansen's (1996) weighted bootstrap can also be applied to obtain the null limit distribution. $\square$

We next examine the limit distribution of $\mathcal{D}_{n,1}$ under $\mathcal{H}_{0,2}$. If $\gamma_* = 0$, ξ_{0*} and β_* are not separately identifiable. We therefore first assume that β_* is unidentified, to obtain the null approximation, then reverse the order by allowing ξ_{0*} to be unidentified, and finally compare them under $\mathcal{H}_{0,2}$. Since β_* (resp. ξ_{0*}) is not identified, we optimize $d_n(\cdot)$ with respect to β (resp. ξ_0) in a later stage compared to for any other parameter, to obtain

$$\mathcal{D}_{n,1}^{(\gamma=0;\beta)} := -\inf_{\beta}\inf_{\gamma} n^{-1}\{d_n(\gamma;\beta) - d_n(0;\beta)\} = \sup_{\beta}\frac{1}{n}\frac{\{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0} + o_{\mathbb{P}}(1), \tag{4}$$

$$\mathcal{D}_{n,1}^{(\gamma=0;\xi_0)} := -\inf_{\xi_0}\inf_{\gamma} n^{-1}\{d_n(\gamma;\xi_0) - d_n(0;\xi_0)\} = \sup_{\xi_0}\frac{1}{n}\frac{\{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0} + o_{\mathbb{P}}(1) \tag{5}$$

by applying a second-order Taylor expansion, where $\mathbf{C}_0 := [L_1, \ldots, L_n]'$, $L_t := \log(X_t)$, and $\mathcal{D}_{n,1}^{(\gamma=0;\beta)}$ (resp. $\mathcal{D}_{n,1}^{(\gamma=0;\xi_0)}$) denotes the DD test statistic designed to test $\mathcal{H}_{0,2}$ by treating β_* (resp. ξ_{0*}) as an unidentified parameter. Here, the right-hand side (RHS) parameters of (4) and (5) are asymptotically free of β and ξ_0, respectively, under our regularity conditions. Thus, the maximization with respect to β and ξ_0 in (4) and (5) respectively is an innocuous process relative to the null limit distribution. Furthermore, the same asymptotic approximations in (4) and (5) imply the uniquely determined limit distribution of $\mathcal{D}_{n,1}$ irrespective of the optimization order. We assume that $\mathcal{D}_{n,1}^{(\gamma=0)}$ denotes the DD test statistic testing $\mathcal{H}_{0,2}$ and contains the null limit distribution in the following lemma:

Lemma 2. *Given Assumptions 1, 2, 3, and $\mathcal{H}_{0,2}$, $\mathcal{D}_{n,1}^{(\gamma=0)} = \{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2/\{n\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0\} + o_{\mathbb{P}}(1) \overset{A}{\sim} \mathcal{Z}_0^2$, where $\mathcal{Z}_0 \overset{\text{A}}{\sim} N(0, \kappa_0^2)$ and $\kappa_0^2 := \mathbb{E}[L_t\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\widetilde{\boldsymbol{\Sigma}}\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t L_t]/\mathbb{E}[L_t\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t L_t]$.* $\square$

Remarks.

(a) The DD test statistic null limit distribution is a noncentral chi-square distribution, unlike the limit distribution under $\mathcal{H}_{0,1}$. This is mainly because the null limit approximations in (4) and (5) are free of nuisance parameters β and ξ_0, respectively.

(b) As for the case under $\mathcal{H}_{0,1}$, if $\mathbf{M}_0 = \boldsymbol{\Sigma}^{-1}$, then $\kappa_0^2 = 1$, and so $\mathcal{D}_{n,1}^{(\gamma=0)} \overset{\text{A}}{\sim} \mathcal{X}_1^2$ under $\mathcal{H}_{0,2}$.

(c) The weak limits of the DD test statistic under $\mathcal{H}_{0,1}$ and $\mathcal{H}_{0,2}$ are not independent. We examine their joint distribution along with the weak limits under $\mathcal{H}_{0,3}$. $\square$

Next, we examine the limit distribution of $\mathcal{D}_{n,1}$ under $\mathcal{H}_{0,3} : \gamma_* = 1$. The process is similar to that under $\mathcal{H}_{0,2}$. That is, if $\gamma_* = 1$, ξ_{1*} and β_* are not separately identifiable. We therefore treat one of them as unidentified and identify the other one similarly to that under $\mathcal{H}_{0,2}$. If we treat β_* or ξ_{1*} as the unidentified parameter, the corresponding null approximation is obtained as

$$\mathcal{D}_{n,1}^{(\gamma=1;\beta)} := -\inf_{\beta}\inf_{\gamma\in\mathbf{\Gamma}} n^{-1}\{d_n(\gamma;\beta) - d_n(1;\beta)\} = \sup_{\beta}\frac{1}{n}\frac{\{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1} + o_{\mathbb{P}}(1), \tag{6}$$

$$\mathcal{D}_{n,1}^{(\gamma=1;\xi_1)} := -\inf_{\xi_1}\inf_{\gamma} n^{-1}\{d_n(\gamma;\xi_1) - d_n(1;\xi_1)\} = \sup_{\xi_0}\frac{1}{n}\frac{\{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1} + o_{\mathbb{P}}(1) \tag{7}$$

by applying the second-order Taylor approximation to $d_n(\cdot)$, where for $j = 1, 2, ...$, we let $\mathbf{C}_j := [X_t^j L_1, \dots, X_n^j L_n]$. Here, $\mathcal{D}_{n,1}^{(\gamma=1;\beta)}$ (resp. $\mathcal{D}_{n,1}^{(\gamma=1;\xi_1)}$) denotes the DD test statistic designed to test $\mathcal{H}_{0,3}$ obtained by treating β_* (resp. ξ_{1*}) as the unidentified parameter, letting $d_n(\cdot)$ be optimized with respect to β (resp. ξ_1) in the final stage. As earlier, the RHS parameters of (6) and (7) are asymptotically free of β and ξ_1, respectively, under our regularity conditions. Furthermore, the null approximation in (6) is identical to that in (7), implying that the limit distribution under $\mathcal{H}_{0,3}$ is identical to that under $\mathcal{H}_{0,2}$ irrespective of the optimization order. We assume that $\mathcal{D}_{n,1}^{(\gamma=1)}$ denotes the DD test statistic and contains its null limit distribution in the following lemma:

Lemma 3. *Given Assumptions 1, 2, 3, and $\mathcal{H}_{0,3}$, $\mathcal{D}_{n,1}^{(\gamma=1)} = \{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2/\{n\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1\} + o_{\mathbb{P}}(1) \overset{A}{\sim} \mathcal{Z}_1^2$, where $\mathcal{Z}_1 \sim N(0, \kappa_1^2)$ and $\kappa_1^2 := \mathbb{E}[C_t\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\widetilde{\mathbf{\Sigma}}\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_tC_t]/\mathbb{E}[C_t\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_tC_t]$.* □

Finally, we derive the DD test statistic limit distribution under $\mathcal{H}_0$ using all the three null approximations under $\mathcal{H}_{0,1}$, $\mathcal{H}_{0,2}$, and $\mathcal{H}_{0,3}$. Note that regular relationships exist among null approximations. For this examination, we first assume that $N_n(\gamma) := \{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2$ and $D_n(\gamma) := n\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)$. These are the numerator and denominator of (3) respectively, and we examine the probability limits when γ converges to 0 or 1, to thereby remove the restriction to $\mathbf{\Gamma}$ by ϵ. Note that $\text{plim}_{\gamma\to 0}N_n(\gamma) = 0$ and $\text{plim}_{\gamma\to 0}D_n(\gamma) = 0$, because $\gamma \to 0$, implying that the probability limit of the ratio has to be obtained by the L'Hôpital rule. We observe the same aspect when γ converges to 1. The following lemma contains the probability limits of $N_n^{(j)} := (\partial^j/\partial\gamma^j)N_n(\gamma)$ and $D_n^{(j)} := (\partial^j/\partial\gamma^j)\, D_n(\gamma)$ for $j = 1$ and 2:

Lemma 4. *Given Assumptions 1 and 2,*

(i) $\text{plim}_{\gamma\to 0}N_n^{(1)}(\gamma) = 0$ *and* $\text{plim}_{\gamma\to 0}D_n^{(1)}(\gamma) = 0$*;*

(ii) $\text{plim}_{\gamma\to 1}N_n^{(1)}(\gamma) = 0$ *and* $\text{plim}_{\gamma\to 1}D_n^{(1)}(\gamma) = 0$*;*

(iii) $\text{plim}_{\gamma\to 0}N_n^{(2)}(\gamma) = 2\{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2$ *and* $\text{plim}_{\gamma\to 0}D_n^{(2)}(\gamma) = 2n\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0$*; and*

(iv) $\text{plim}_{\gamma\to 1}N_n^{(2)}(\gamma) = 2\{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2$ *and* $\text{plim}_{\gamma\to 1}D_n^{(2)}(\gamma) = 2n\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1$. □

By Lemma 4, the L'Hôpital rule has to be applied twice for the ratio probability limits. That is,

$$\text{plim}_{\gamma\to 0}\frac{N_n(\gamma)}{D_n(\gamma)} = \frac{\{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2}{n\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0} \quad \text{and} \quad \text{plim}_{\gamma\to 1}\frac{N_n(\gamma)}{D_n(\gamma)} = \frac{\{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{n\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1}, \tag{8}$$

which are in fact the null limit approximations given in Lemmas 2 and 3. This also implies that

$$\mathcal{D}_{n,1}^{(\beta=0)} := \sup_{\gamma\in\mathbf{\Gamma}} \frac{1}{n}\frac{\{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)} \geq \max\left[\frac{\{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2}{n\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0}, \frac{\{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{n\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1}\right] = \max\left[\mathcal{D}_{n,1}^{(\gamma=0)}, \mathcal{D}_{n,1}^{(\gamma=1)}\right] + o_{\mathbb{P}}(1).$$

Therefore, the biggest GMM distance is obtained under $\mathcal{H}_{0,1}$. This implies that the DD test statistic limit distribution under $\mathcal{H}_0$ has to be represented as a functional of $\mathcal{Z}_1(\cdot)$ derived under $\mathcal{H}_{0,1}$. We summarize the key result in the following theorem:

Theorem 1. *Given Assumptions 1, 2, 3, and* $\mathcal{H}_0$, $\mathcal{D}_{n,1} \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}} \mathcal{Z}_1^2(\gamma)$. □

2.4 Testing for Structural Nonlinearity

The DD test statistic has a consistent and nontrivial local power against general nonlinearity when valid instrumental variables are employed, to lead to omnibus power. To examine this omnibus power, we assume the possibly of no (β, γ) such that $m(X_t) = \beta X_t^\gamma$ with probability 1, and examine the omnibus power property of the DD test statistic.

For this, we first derive the GMM distance limits under the null and alternative models and then examine their difference. We examine the null distance at the limit, which we denote as $d_0 := \text{plim}_{n\to\infty} n^{-2} d_n(\widetilde{\boldsymbol{\omega}}_n)$, and obtain it by the ergodic theorem:

$$d_0 = \min_{\boldsymbol{\varsigma}} \mathbb{E}[(Y_t - \mathbf{V}_t'\boldsymbol{\varsigma})\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[(Y_t - \mathbf{V}_t'\boldsymbol{\varsigma})\mathbf{Z}_t] = \mathbb{E}[m(X_t)\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)].$$

Here, if $\boldsymbol{\varsigma}_0$ is the argument for d_0, then $\boldsymbol{\varsigma}_0 = \boldsymbol{\varsigma}_* + (\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t']\mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}_t'])^{-1}\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t']\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)]$, implying that the GMM estimator is asymptotically biased, as pointed out by Hall and Inoue (2003). We then derive the alternative GMM distance at the limit: for each γ, if $d(\gamma) := \min_{\boldsymbol{\varsigma},\beta} \text{plim}_{n\to\infty} n^{-2} d_n(\boldsymbol{\varsigma}, \beta, \gamma)$,

$$d(\gamma) = \min_{\boldsymbol{\varsigma},\beta} \mathbb{E}[(Y_t - \mathbf{V}_t'\boldsymbol{\varsigma} - \beta X_t^\gamma)\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t(Y_t - \mathbf{V}_t'\boldsymbol{\varsigma} - \beta X_t^\gamma)] = \mathbb{E}[m(X_t)\widetilde{\mathbf{Z}}_t']\mathbf{J}_1(\gamma)\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)],$$

where for each $\gamma \in \mathbf{\Gamma}$, $\mathbf{J}_1(\gamma) := \mathbf{I} - \mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}_t(\gamma)'](\mathbb{E}[\mathbf{V}_t(\gamma)\widetilde{\mathbf{Z}}_t']\mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}_t(\gamma)'])^{-1}\mathbb{E}[\mathbf{V}_t(\gamma)\widetilde{\mathbf{Z}}_t']$, and $\mathbf{V}_t(\gamma) :=$

$(\mathbf{V}_t', X_t^{\gamma}) = (1, \mathbf{W}_t' X_t^{\gamma})$, so that

$$d_0 - d(\gamma) = \frac{\{\mathbb{E}[m(X_t)\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]\}^2}{\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]}. \tag{9}$$

Note that $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]$ is the projection error of $\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]$ against $\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t']$, and $\mathbf{J}_1$ is an idempotent matrix, so that $\{\mathbb{E}[m(X_t)\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]\}^2 > 0$, unless $\mathbb{E}[m(X_t)\widetilde{\mathbf{Z}}_t]$ and $\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t]$ are sub-vectors of $\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t]$. Therefore, for each γ, $d_0 - d(\gamma) > 0$. Here, even for $\gamma = 0$ or 1, $d_0 - d(\gamma) > 0$. If $\gamma = 0$ or 1, then $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}] = \mathbf{0}$, because $\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t]$ is a sub-vector of $\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t]$. Nevertheless, the RHS of (9) is obtained by the L'Hôpital rule if $\gamma = 0$ or 1, because $\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t]$ is present in both the numerator and denominator. Therefore,

$$\lim_{\gamma\to 0} d_0 - d(\gamma) = \frac{\{\mathbb{E}[m(X_t)\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t L_t]\}^2}{\mathbb{E}[L_t\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t L_t]} \quad \text{and} \quad \lim_{\gamma\to 1} d_0 - d(\gamma) = \frac{\{\mathbb{E}[m(X_t)\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t C_t]\}^2}{\mathbb{E}[C_t\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t C_t]}.$$

Note that the two limits are still strictly positive.

The DD test statistic gains power from the difference between d_0 and $d(\cdot)$. Note that $n^{-1}\mathcal{D}_{n,1} = d_0 - \inf_{\gamma\in\mathbf{\Gamma}} d(\gamma) + o_{\mathbb{P}}(1) = \sup_{\gamma\in\mathbf{\Gamma}} \mu_1^2(\gamma) + o_{\mathbb{P}}(1)$, where

$$\mu_1(\cdot) := \frac{\mathbb{E}[X_t^{(\cdot)}\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)]}{\{\mathbb{E}[X_t^{(\cdot)}\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}]\}^{1/2}}.$$

Indeed, $\sup_{\gamma\in\mathbf{\Gamma}} \mu_1^2(\gamma)$ is strictly positive, to obtain a consistent power for the DD test statistic. We include this result in the following theorem:

Theorem 2. *Given Assumptions 1, 2, and 3,*

(i) if $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] \neq \mathbf{0}$ and there is no (β,γ) such that $m(X_t) = \beta X_t^{\gamma}$ with probability 1, then for some $\widetilde{\gamma} \in \mathbf{\Gamma}$, $d(\widetilde{\gamma}) \in (0, d_0)$ and $n^{-1}\mathcal{D}_{n,1} = d_0 - d(\widetilde{\gamma}) + o_{\mathbb{P}}(1)$; and

(ii) if for the measurable function $s(\cdot)$, $m(X_t) = n^{-1/2}s(X_t)$ with probability 1, $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t s(X_t)] \neq \mathbf{0}$ and there is no (β,γ) such that $s(X_t) = \beta X_t^{\gamma}$ with probability 1, then $\mathcal{D}_{n,1} \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}}\{\mathcal{Z}_1(\gamma) + \nu_1(\gamma)\}^2$, where $\nu_1(\cdot) := \mathbb{E}[X_t^{(\cdot)}\widetilde{\mathbf{Z}}_t]\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t s(X_t)]/\sigma_1(\cdot)$. □

Remarks.

(a) For consistent power, we need to select valid instrumental variables for $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] \neq \mathbf{0}$, as presumed for the proper application of the DD test statistic.

(b) The DD test statistic draws consistent power from a factor different from that for the J-test statistic. The J-test statistic directly tests whether or not $\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] = \mathbf{0}$ and asymptotically rejects the linear structure condition if the instrumental variables are not valid. In contrast, the DD test statistic draws its power from the correlation between $\widetilde{\mathbf{Z}}_t'\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]$ and $\widetilde{\mathbf{Z}}_t'\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)]$; this implies that the J- and DD test statistics supplement each other. If the J-test statistic rejects the null but the DD test statistic does not, the rejection is highly related to $\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] \neq \mathbf{0}$.

(c) From Theorem 2(*i*), the DD test statistic has a consistent power even when the power transform misspecifies the functional form of $m(\cdot)$. If the power transformation correctly specifies the functional form of $m(\cdot)$, the power obtained consistently is trivial. That is, if for some $\gamma_* \in \mathbf{\Gamma} \setminus \{0, 1\}$, $m(X_t) = \beta_* X_t^{\gamma_*}$, then $n^{-1}\mathcal{D}_{n,1} = \beta_*^2 \mathbb{E}[X_t^{\gamma_*}\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma_*}] + o_{\mathbb{P}}(1)$; this is strictly positive at the limit, implying that $\mathcal{D}_{n,1}$ has nontrivial asymptotic power.

(d) For an intuitive proof of Theorem 2(*i*), assume that $\mathbf{M}_0 = \mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}$. Now, we have $\mu_1^2(\gamma) = \mathrm{corr}[Q_t, U_t(\gamma)]^2 \mathrm{var}[Q_t]$, where $Q_t := \widetilde{\mathbf{Z}}_t'\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)]$ and $U_t(\gamma) := \widetilde{\mathbf{Z}}_t'\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]$, and so the DD test statistic does not have an asymptotic power if $\mathrm{corr}[Q_t, U_t(\cdot)]^2 \equiv 0$. Nevertheless, $\mathrm{corr}[Q_t, U_t(\cdot)]^2$ cannot be zero, because $\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}] = \mathbf{0}$ implies that $\mathbb{E}[\widetilde{\mathbf{Z}}_t | X_t] = \mathbf{0}$, since $\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}] = \nabla_{\boldsymbol{\tau}}\mathbb{E}[\exp(\gamma \log (X_t) + \boldsymbol{\tau}'\widetilde{\mathbf{Z}}_t)]|_{\boldsymbol{\tau}=\mathbf{0}}$ and $\mathbb{E}[\exp(\gamma\log(X_t) + \boldsymbol{\tau}'\widetilde{\mathbf{Z}}_t)]$ is a moment generating function of $(\log(X_t), \widetilde{\mathbf{Z}}_t')'$. Here, if $\mathbb{E}[\widetilde{\mathbf{Z}}_t | X_t] = \mathbf{0}$, then $\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] = \mathbf{0}$ by the law of iterated expectation, since it is contradictory to the condition that $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] \neq \mathbf{0}$, leading to a nonzero correlation coefficient between Q_t and $U_t(\cdot)$.

(e) From Theorem 2(*ii*), the DD test statistic has a nontrivial power against a local alternative converging to zero at the rate of $n^{-1/2}$. Note that the asymptotic local power can be gained by shifting the locality parameter of $\mathcal{Z}_1(\cdot)$ by $\nu_1(\cdot)$, which is different from 0, as implied by Theorem 2(*i*). □

3 Extension to Testing the Polynomial Model Hypothesis

3.1 Motivation and Model

In this section, we extend the linear structure testing condition to testing a polynomial structure. Here, a sequential testing procedure can be used to estimate a nonlinear polynomial structure consistently.

The main aim of sequential testing is to estimate a parsimonious structural model. Note that semi- and nonparametric sieve estimations exploit as many sieve bases as the sample size allows and lead to possibly unnecessary estimation errors for the estimator. A sequential testing procedure is a machinery process to avoid unnecessary estimation error.

We believe that the empirical researcher would approximate the unknown functional form of $m(\cdot)$ using the polynomial model specified as

$$\mathcal{M}_q := \left\{ m_{t,q}(\boldsymbol{\omega}^{(q)}) := Y_t - \mathbf{X}_{t,q}'\boldsymbol{\xi}^{(q)} - \mathbf{D}_t'\boldsymbol{\eta} - \beta X_t^{\gamma} : \boldsymbol{\omega}^{(q)} \in \boldsymbol{\Omega}^{(q)} \subset \mathbb{R}^{k+q+3} \right\},$$

where $\boldsymbol{\omega}^{(q)} := (\boldsymbol{\xi}^{(q)\prime}, \boldsymbol{\eta}', \beta, \gamma)'$, $\mathbf{X}_{t,q} := (1, X_t, X_t^2, \ldots, X_t^q)'$, $\boldsymbol{\xi}^{(q)} := (\xi_0, \xi_1, \ldots, \xi_q)'$, and k and $q \in \mathbb{N}$. As earlier, we assume that for some $\boldsymbol{\omega}_*^{(q)} \in \boldsymbol{\Omega}^{(q)}$, $Y_t = \mathbf{X}_{t,q}'\boldsymbol{\xi}_*^{(q)} + \mathbf{D}_t'\boldsymbol{\eta}_* + m(X_t) + U_t$ such that $\mathbb{E}[\mathbf{Z}_t U_t] = \mathbf{0}$, and X_t and $\mathbf{D}_t$ are endogenous and exogenous variables respectively. Note that this structure generalizes the linear structure in Section 2. If $q = 1$, then $\mathcal{M}_q$ is identical to $\mathcal{M}_1$, whereas the structural equation is possibly nonlinear for $q > 1$.

The motivation of $\mathcal{M}_q$ comes from estimating a reduced-form equation through sieve approximation. Each polynomial term forms a sieve basis, with the unknown reduced-form equation well known to be approximated arbitrarily well through a polynomial function by increasing its degree. Another standard method is to estimate the unknown sieve estimation degree using information criteria (*e.g.*, Chen and Liu, 2014). Cho and Phillips (2018) apply the sequential testing procedure based on QLR test statistic to the polynomial model and find that it can consistently estimate the nonlinear reduced-form equation.

We apply the sequential testing procedure in Cho and Phillips (2018) to the nonlinear structure using the DD test statistic. Since the structural form of $m(\cdot)$ is unknown, the sieve estimation motivates to approximate $m(\cdot)$ using a higher-degree polynomial function. If the DD test statistic does not reject the high-degree polynomial model, the sequential testing procedure would take it as $m(\cdot)$ or its close approximation, enabling the researcher to develop an economic theory consistent with the empirical estimate obtained using the sequential testing procedure.

Another motivation to use sequential testing stems from the moment selection criterion (MSC) developed by Andrews (1999). We discuss this motivation by relating the sequential testing procedure to specifically the Bayesian-type MSC among others, which is defined as $BC_{n,q} := \bar{\mathcal{J}}_{n,q} - (m - k - q - 1)\log(n)/n$, where $\bar{\mathcal{J}}_{n,q} := n^{-1}\mathcal{J}_{n,q}$ and $\mathcal{J}_{n,q}$ is the J-test statistic designed to test the q^{th}-degree polynomial structural equation such that $q = 1, 2, \ldots, \bar{q} < \infty$. The MSC selects the polynomial model with the smallest value of $BC_{n,q}$ for $q = 1, 2, \ldots, \bar{q}$. If $q_* < \bar{q}$, Andrews (1999) shows that the Bayesian-type MSC asymptotically selects the q_*^{th}-degree polynomial model. The same result can be obtained via

$$\Delta BC_{n,q} := BC_{n,q+1} - BC_{n,q} = \bar{\mathcal{J}}_{n,q+1} - \bar{\mathcal{J}}_{n,q} + \frac{1}{n}\log(n)$$

under some regularity conditions. If $q \geq q_*$, $\text{plim}_{n\to\infty}\Delta BC_{n,q} = 0$, because the probability limits of $\bar{\mathcal{J}}_{n,q+1}$ and $\bar{\mathcal{J}}_{n,q}$ are identical since the q^{th}-degree polynomial model is nested in a higher-degree polynomial model. Thus, if $\lim_{n\to\infty}\mathbb{P}(\Delta BC_{n,q}<0)=1$ for every $q<q_*$, then q_* must be the smallest q among the qs, such that $\text{plim}_{n\to\infty}\Delta BC_{n,q}$ is zero. From this feature, we can consistently estimate q_* by sequentially testing whether $\text{plim}_{n\to\infty}\Delta BC_{n,q}$ is less than or equal to 0 from $q=1$ to $q=\bar{q}$ until we cannot reject the hypothesis that $\text{plim}_{n\to\infty}\Delta BC_{n,q}=0$.

We design our sequential testing procedure to ensure the undergoing supposition. The procedure using $\Delta BC_{n,q}$ would work properly if $\lim_{n\to\infty}\mathbb{P}(\Delta BC_{n,q}<0)<1$ holds for every $q<q_*$. Otherwise, the procedure would fail to estimate q_* consistently. We thus avoid this fallacy by replacing $\Delta BC_{n,q}$ with the DD test statistic. The DD test statistic has omnibus power, implying that for every $q<q_*$, $\lim_{n\to\infty}\mathbb{P}(\mathcal{D}_{n,q}<0)<1$, where $\mathcal{D}_{n,q}$ is the DD test statistic testing the q-th degree polynomial hypothesis, as formally defined below. Therefore, the fallacy probability becomes negligible as n increases.

We discuss this approach more specifically below. For this, we first examine the q^{th}-degree polynomial model testing and then apply the sequential testing procedure to estimate the polynomial structure.

3.2 Inference Using the DD Test Statistic

We assume that the empirical researcher is testing whether the q^{th}-degree polynomial model is adequate or not for the nonlinear structure by letting the null model be the q^{th}-degree polynomial function.

The testing procedure using $\mathcal{M}_q$ is similar to that using $\mathcal{M}_1$. Note that $\mathcal{M}_q$ can be transformed into the q^{th}-degree polynomial model in $q+2$ different ways, as with the linear model:

$$\mathcal{H}'_{0,1}:\beta_*=0,\quad \mathcal{H}'_{0,2}:\gamma_*=0,\quad \cdots,\quad \mathcal{H}_{0,q+1}:\gamma_*=q-1,\quad \text{or}\quad \mathcal{H}'_{0,q+2}:\gamma_*=q.$$

Since any of these hypotheses would generate the q^{th}-degree polynomial model, we treat them as the sub-hypotheses of $\mathcal{H}'_0:=\cup_{s=1}^{q+2}\mathcal{H}'_{0,s}$, which is now the null hypothesis of this section. Each sub-hypothesis has its own identification problem: γ_* is not identified under $\mathcal{H}'_{0,1}$; for $s=0,1,\ldots,q$, β_* and $\xi_{s,*}$ are not separately identified under $\mathcal{H}'_{0,s+2}$. This forms a multifold identification problem that generalizes the trifold identification problem in Section 2.3.

We then use the DD test statistic to overcome the multifold identification problem. For this, we define the DD test statistic as

$$\mathcal{D}_{n,q}:=n^{-1}\{d_n(\widetilde{\boldsymbol{\omega}}_n^{(q)})-d_n(\widehat{\boldsymbol{\omega}}_n^{(q)})\},$$

where $\widetilde{\boldsymbol{\omega}}_n^{(q)}:=\arg\min_{\boldsymbol{\omega}^{(q)}\in\boldsymbol{\Omega}^{(q)}}d_n(\boldsymbol{\omega}^{(q)})$, subject to $\beta=0$, $\widehat{\boldsymbol{\omega}}_n^{(q)}:=\arg\min_{\boldsymbol{\omega}^{(q)}\in\boldsymbol{\Omega}^{(q)}}d_n(\boldsymbol{\omega}^{(q)})$, and

$$d_n(\boldsymbol{\omega}^{(q)}):=(\mathbf{Y}-\beta\mathbf{X}(\gamma)-\mathbf{V}_q\boldsymbol{\varsigma}^{(q)})\mathbf{Z}\mathbf{M}_n\mathbf{Z}'(\mathbf{Y}-\beta\mathbf{X}(\gamma)-\mathbf{V}_q\boldsymbol{\varsigma}^{(q)}).$$

Here, we assume that $\mathbf{V}_q := [\mathbf{V}_{1,q}', \ldots, \mathbf{V}_{n,q}']'$, $\mathbf{V}_{t,q} := (1, \mathbf{W}_{t,q}')' := (1, \mathbf{X}_{t,q}', \mathbf{D}_t')'$, and $\boldsymbol{\varsigma}^{(q)} := (\boldsymbol{\xi}^{(q)'}, \boldsymbol{\eta}')'$, so that $\boldsymbol{\omega}^{(q)} = (\boldsymbol{\varsigma}^{(q)'}, \beta, \gamma)'$. Note that if $q = 1$, $\mathbf{V}_q$ and $\mathbf{V}_{t,q}$ would be identical to $\mathbf{V}$ and $\mathbf{V}_t$ in Section 2.3, respectively.

We can obtain the null limit distribution of the DD test statistic as in the linear model case. For this, we extend the earlier model and moment conditions, to have the following assumption:

Assumption 4. *(i) The structural relationship between Y_t and $\mathbf{W}_t$ is specified as $\mathcal{M}_q := \{m_{t,q}(\boldsymbol{\omega}^{(q)}) := Y_t - \mathbf{X}_{t,q}'\boldsymbol{\xi}^{(q)} - \mathbf{D}_t'\boldsymbol{\eta} - \beta X_t^{\gamma} : \boldsymbol{\omega}^{(q)} \in \boldsymbol{\Omega}^{(q)} \subset \mathbb{R}^{k+q+3}\}$, where $\boldsymbol{\Omega}^{(q)} := \boldsymbol{\Xi}^{(q)} \times \boldsymbol{\Delta} \times \mathbf{B} \times \boldsymbol{\Gamma}^{(q)}$ such that $\boldsymbol{\Xi}^{(q)}$, $\boldsymbol{\Delta}$, $\mathbf{B}$, and $\boldsymbol{\Gamma}^{(q)} := [\underline{\gamma}, \overline{\gamma}]$ are convex and compact in $\mathbb{R}^q$, $\mathbb{R}^{k+1}$, $\mathbb{R}$, and $\mathbb{R}$, respectively; 0, 1, $\ldots$, and q are interior elements of $\boldsymbol{\Gamma}^{(q)}$;*

(ii) for the measurable functions $m(\cdot)$ and $(\xi_{0*}, \boldsymbol{\delta}_*^{(q)\prime})' \in \mathbb{R}^{1+k+q}$, $Y_t = \xi_{0*} + \mathbf{W}_{t,q}'\boldsymbol{\delta}_*^{(q)} + m(X_t) + U_t$, where $\mathbf{W}_{t,q} := (1, \mathbf{X}_{t,q}', \mathbf{D}_t')'$ and $\mathbf{X}_{t,q} := (1, X_t, X_t^2, \ldots, X_t^q)'$;

(iii) $\mathbb{E}[\mathbf{V}_{t,q}\mathbf{Z}_t']$ and $\mathbf{V}_q'\mathbf{Z}$ have full row ranks uniformly in n, where $\mathbf{V}_{t,q} = (1, \mathbf{W}_{t,q}')'$ and $\mathbf{V}_q := [\mathbf{V}_{1,q}', \ldots, \mathbf{V}_{n,q}']'$;

(iv) an SSE sequence $\{M_t\}$ exists such that $\mathbb{E}[M_t^2] < \infty$ and $\sup_{\gamma \in \boldsymbol{\Gamma}^{(q)}} |X_t^{\gamma}| \leq M_t$; and

(v) $\mathbb{E}[X_t^{4q}] < \infty$ and $\mathbb{E}[L_t^4] < \infty$. □

Remarks.

(a) The parameter space condition in Assumption 2 is modified to include $0, 1, \ldots, q$ as interior elements of $\boldsymbol{\Gamma}^{(q)}$.

(b) Note that if $q = 1$, Assumption 4 would imply Assumptions 2 and 3. □

Under the above conditions, we can obtain the properties of the DD test statistic as in linearity testing. For this, we follow the approach of the linear model case as follows. Assume that

$$\mathcal{D}_{n,q}^{(\beta=0)} := -\inf_{\gamma \in \boldsymbol{\Gamma}^{(q)}(\epsilon)} \inf_{\beta} n^{-1}\{d_n(\beta; \gamma) - d_n(0; \gamma)\} = \sup_{\gamma \in \boldsymbol{\Gamma}^{(q)}} \frac{1}{n} \frac{\{\mathbf{X}(\gamma)'\mathbf{Q}_q\mathbf{U}\}^2}{\mathbf{X}(\gamma)'\mathbf{Q}_q\mathbf{X}(\gamma)},$$

to obtain the null limit distribution of the DD test statistic under $\mathcal{H}_{0,1}'$, where $d_n(\beta; \gamma) := \min_{\boldsymbol{\varsigma}^{(q)}} d_n(\boldsymbol{\omega}^{(q)})$ and $\mathbf{Q}_q := \ddot{\mathbf{Z}}\{\mathbf{I} - \ddot{\mathbf{Z}}'\mathbf{V}_q(\mathbf{V}_q'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}_q')^{-1}\mathbf{V}_q'\ddot{\mathbf{Z}}\}\ddot{\mathbf{Z}}'$. Next, as in the linear case, for each $s = 0, 1, \ldots, q$, assume that

$$\mathcal{D}_{n,q}^{(\gamma=s)} := \max[\mathcal{D}_{n,q}^{(\gamma=s;\xi_s)}, \mathcal{D}_{n,q}^{(\gamma=s;\beta)}],$$

to obtain the null limit distribution of the DD test statistic under $\mathcal{H}_{0,2+s}$, where

$$\mathcal{D}_{n,q}^{(\gamma=s;\beta)} := -\inf_{\beta}\inf_{\gamma} n^{-1}\{d_n(\gamma; \beta) - d_n(1; \beta)\} \quad \text{and} \quad \mathcal{D}_{n,q}^{(\gamma=s;\xi_s)} := -\inf_{\xi_s}\inf_{\gamma} n^{-1}\{d_n(\gamma; \xi_s) - d_n(1; \xi_s)\}$$

with $d_n(\gamma;\beta) := \min_{\boldsymbol{\varsigma}^{(q)}} d_n(\boldsymbol{\omega}^{(q)})$, $d_n(\gamma;\xi_s) := \min_{\boldsymbol{\xi}_{-s}^{(q)},\boldsymbol{\eta},\beta} d_n(\boldsymbol{\omega}^{(q)})$, and $\boldsymbol{\xi}_{-s}^{(q)} := (\xi_0, \ldots, \xi_{s-1}, \xi_{s+1}, \ldots, \xi_q)'$. We obtain all these statistics by optimizing the GMM distance function with regard to the unidentified parameters under each sub-null hypothesis $\mathcal{H}'_{0,2+s}$ in the final stage and the null limit approximation of the DD test statistic as their maximum, as in the linear model case. That is, if we assume that

$$\mathcal{D}_{n,q} := \max[\mathcal{D}_{n,q}^{(\beta=0)}, \mathcal{D}_{n,q}^{(\gamma=0)}, \mathcal{D}_{n,q}^{(\gamma=1)}, \cdots, \mathcal{D}_{n,q}^{(\gamma=q)}],$$

then the DD test statistic would be asymptotically equivalent to $\mathcal{D}_{n,q}$ under $\mathcal{H}'_0$ by analogy.

We can also obtain the omnibus power property of the DD test statistic as in the linear model case. For the desired properties, we assume that for the measurable function $m(\cdot)$, $Y_t = \mathbf{X}'_{t,q}\boldsymbol{\xi}_*^{(q)} + \mathbf{D}'_t\boldsymbol{\eta}_* + m(X_t) + U_t$ such that $\mathbb{E}[U_t\mathbf{Z}_t] = \mathbf{0}$, with possibly no (β, γ), such that $m(X_t) = \beta X_t^{\gamma}$ with probability 1. Given this assumption, we can obtain $\mathrm{plim}_{n\to\infty} n^{-2}\{d_n(\widetilde{\boldsymbol{\omega}}_n^{(q)}) - d_n(\widehat{\boldsymbol{\varsigma}}_n^{(q)}(\gamma), \gamma)\} = \sup_{\gamma\in\boldsymbol{\Gamma}^{(q)}} \mu_q^2(\gamma)$ by applying the ergodic theorem, where $\widehat{\boldsymbol{\varsigma}}_n^{(q)}(\gamma) := \arg\min_{\boldsymbol{\varsigma}} d_n(\boldsymbol{\varsigma}, \gamma)$, $\boldsymbol{\varsigma}^{(q)} := (\boldsymbol{\xi}^{(q)\prime}, \boldsymbol{\eta}')'$, and for each $\gamma \in \boldsymbol{\Gamma}^{(q)}$,

$$\mu_q(\gamma) := \frac{\mathbb{E}[m(X_t)\widetilde{\mathbf{Z}}'_t]\mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]}{\{\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}'_t]\mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]\}^{1/2}}.$$

Here, $\mathbf{J}_q := \mathbf{I} - \mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}'_{t,q}](\mathbb{E}[\mathbf{V}_{t,q}\widetilde{\mathbf{Z}}'_t]\ \mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}'_{t,q}])^{-1}\mathbb{E}[\mathbf{V}_{t,q}\widetilde{\mathbf{Z}}'_t]$. From this, if $\sup_{\gamma\in\boldsymbol{\Gamma}^{(q)}} \mu_q^2(\gamma) > 0$, the DD test statistic will have consistent power.

We collect these null and alternative limit properties, to obtain the following corollary:

Corollary 1. *Given Assumption 1 and 4,*

(i) $\mathcal{D}_{n,q} \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}^{(q)}} \mathcal{Z}_q^2(\gamma)$ *under* $\mathcal{H}'_0$*, where for each* $\gamma \in \boldsymbol{\Gamma}^{(q)}(\epsilon)$, $\mathcal{Z}_q(\gamma) \sim N(0, \rho_q(\gamma,\gamma))$*, and for each pair* (γ,γ'), $\mathbb{E}[\mathcal{Z}_q(\gamma)\mathcal{Z}_q(\gamma')] = \rho_q(\gamma,\gamma') := \kappa_q(\gamma,\gamma')/\{\sigma_q^2(\gamma)\sigma_q^2(\gamma')\}^{1/2}$, $\kappa_q(\gamma,\gamma') := \mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}'_t]\mathbf{J}_q \widetilde{\mathbf{\Sigma}}\mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma'}]$*, and* $\sigma_q^2(\gamma) := \mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}'_t]\mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]$*;*

(ii) if $\mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] \neq \mathbf{0}$*, and possibly there is no* (β,γ) *such that* $m(X_t) = \beta X_t^{\gamma}$ *with probability 1, then for some* $\widetilde{\gamma} \in \boldsymbol{\Gamma}^{(q)}$, $n^{-1}\mathcal{D}_{n,q} = \mu_q^2(\widetilde{\gamma}) + o_{\mathbb{P}}(1)$ *such that* $\mu_q^2(\widetilde{\gamma}) > 0$*; and*

(iii) if for a measurable function $s(\cdot)$, $m(X_t) = n^{-1/2}s(X_t)$ *with probability 1,* $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t s(X_t)] \neq \mathbf{0}$*, and possibly there is no* (β,γ) *such that* $s(X_t) = \beta X_t^{\gamma}$ *with probability 1, then* $\mathcal{D}_{n,q} \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}^{(q)}}\{\mathcal{Z}_q(\gamma) + \nu_q(\gamma)\}^2$*, where* $\nu_q(\cdot) := \mathbb{E}[X_t^{(\cdot)}\widetilde{\mathbf{Z}}_t]\mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t s(X_t)]/\sigma_1(\cdot)$. □

Remarks.

(a) Corollary 1 generalizes the consequences in Theorems 1 and 2 to the polynomial model case; we can obtain its proof by iterating the proofs of Theorems 1 and 2. We summarize the proof as follows:

first, for each $\epsilon > 0$, it follows that $\mathcal{D}_{n,q}^{(\beta=0)}(\epsilon) \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}^{(q)}(\epsilon)} \mathcal{Z}_q^2(\gamma)$ by extending Lemma 1, where $\mathbf{\Gamma}^{(q)}(\epsilon) := \mathbf{\Gamma}^{(q)} \setminus \cup_{j=0}^{q}(j-\epsilon, j+\epsilon)$; second, for each $s = 0, 1, \ldots, q$, it follows that $\mathcal{D}_{n,q}^{(\gamma=s)} = \{\mathbf{C}_s'\mathbf{Q}_q\mathbf{U}\}^2/\{n\mathbf{C}_s'\mathbf{Q}_q\mathbf{C}_s\} + o_{\mathbb{P}}(1)$ under $\mathcal{H}_{0,s+2}' : \gamma_* = s$; finally, if we assume that $N_{n,q}(\gamma) := \{\mathbf{X}(\gamma)'\mathbf{Q}_q\mathbf{U}\}^2$ and $D_{n,q}(\gamma) := n\mathbf{X}(\gamma)'\mathbf{Q}_q\mathbf{X}(\gamma)$, then for each $s = 0, 1, 2, \ldots, q$,

$$\text{plim}_{\gamma\to s}\frac{N_{n,q}(\gamma)}{D_{n,q}(\gamma)} = \frac{1}{n}\frac{\{\mathbf{C}_s'\mathbf{Q}_q\mathbf{U}\}^2}{\mathbf{C}_s'\mathbf{Q}_q\mathbf{C}_s} = \mathcal{D}_{n,q}^{(\gamma=s)} + o_{\mathbb{P}}(1);$$

this implies that the GMM distance obtained under $\mathcal{H}_{0,1}'$ becomes larger than those obtained under $\mathcal{H}_{0,s}'$ with $s = 2, 3, \ldots, q+2$. Thus, $\mathcal{D}_{n,q} = \mathcal{D}_{n,q}^{(\beta=0)} + o_{\mathbb{P}}(1)$ under $\mathcal{H}_0'$, as in the linear model case. Since this proof slightly generalizes that already shown for the linearity condition, we do not repeat the essentially same proof in the appendix.

(b) Note that the covariance kernel of $\mathcal{Z}_q(\cdot)$ is different from that of $\rho_1(\cdot,\cdot)$ in Lemma 1. This depends on both the model and DGP conditions. For the same DGP, different polynomial models provide different covariance kernels. Likewise, for the same model, different DGPs provide different covariance kernels. Furthermore, the null limit distribution of the DD test statistic depends on $\mathbf{\Gamma}^{(q)}$. We obtain different null limit distributions with different $\mathbf{\Gamma}^{(q)}$.

(c) Despite the different properties of $\mathcal{Z}_q(\cdot)$ and $\mathcal{Z}_1(\cdot)$, the asymptotic critical values can be obtained similarly to $\mathcal{Z}_1(\cdot)$. Under mild regularity conditions, we can estimate $\boldsymbol{\pi}_q(\cdot) := \mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}]/\sigma_q^2(\gamma)^{1/2}$ consistently by its sample analog, assuming that $\widetilde{\mathcal{Z}}_q(\cdot) := \boldsymbol{\pi}_q(\cdot)\mathbf{G}$ simulates $\sup_{\gamma\in\mathbf{\Gamma}^{(q)}} \widetilde{\mathcal{Z}}_q^2(\gamma)$, where $\mathbf{G} \sim N(\mathbf{0}, \widetilde{\mathbf{\Sigma}})$ as before.

(d) Corollaries 1(*ii* and *iii*) extend the properties of Theorem 2 under the fixed and local alternative hypotheses, respectively. $\square$

3.3 Sequentially Estimating Correct Polynomial Model

Corollary 1 provides a basis for a system of sequential testing using polynomial models. By applying the sequential testing procedure to Corollary 1, we can estimate the unknown degree of the polynomial model consistently. For this, we assume that $\bar{q}$ is the maximum degree of the polynomial models considered, and $I(\bar{q}) := \{1, 2, \ldots, \bar{q}\}$ is a set of model indices, so that $\bar{q}$ number of models are considered here in total. We also assume that $\mathbf{\Gamma}^{(\bar{q})}$ includes the elements of $I(\bar{q})$ as interior elements and $\mathbf{\Gamma}^{(\bar{q})}$ is identical to $\mathbf{\Gamma}^{(q)}$ in $\mathcal{M}_q$ for each $q \in I(\bar{q})$. We further assume that q_* is the minimum degree polynomial model correctly specified. Note that if the q^{th}-degree polynomial model is correctly specified, every polynomial model with a degree higher than q is also correctly specified. The goal of the sequential testing procedure is to estimate q_* to derive the most parsimonious and correctly specified model. If $q_* \notin I(\bar{q})$, every model is misspecified.

Our sequential testing procedure is performed in the following order:

- **Step 1**: We compute $\mathcal{D}_{n,1}$ using $\mathcal{M}_1$ and compare it with the critical value $cv_1(\alpha_n)$ in Corollary 1 at the level of α_n. If the $\mathcal{D}_{n,1}$ is less than or equal to $cv_1(\alpha_n)$, we stop this sequential testing procedure and conclude that the structural relationship is linear. Otherwise, we move to the next step.

- **Step 2**: For $q = 2, 3, \ldots, \bar{q}$, compute $\mathcal{D}_{n,q}$ and iterate the same testing procedure using the critical value $cv_q(\alpha_n)$ implied by the same level of significance α_n as in Step 1 and the null limit distribution in Corollary 1. If there is any $q \in I(\bar{q})$ such that $\mathcal{D}_{n,q}$ is less than or equal to $cv_q(\alpha_n)$, we let the degree estimator be $\widehat{q}_n := \min\{q \in I(\bar{q}) : \mathcal{D}_{n,q} \leq cv_q(\alpha_n)\}$.

- **Step 3**: If there is no $q \in I(\bar{q})$ such that $\mathcal{D}_{n,q}$ is less than or equal to $cv_q(\alpha_n)$, we conclude that $\mathcal{M}(\bar{q}) := \{\mathcal{M}_q : q \in I(\bar{q})\}$ is not adequate to capture the structural nonlinearity between Y_t and X_t.

Here, the level of significance α_n is set to depend on the sample size. The degree estimation error due to the sequential testing procedure would not vanish if it were fixed at a certain level. Therefore, we allow it to converge to zero gradually as the sample size increases. Thus, the degree estimation error vanishes as n increases (*e.g.*, Cho and Phillips, 2018). Theorem 3 discusses how to choose α_n in order to estimate q_* consistently:

Theorem 3. *Given that for each $q \in I(\bar{q})$, Assumptions 1 and 4 hold with $\mathbf{\Gamma}^{(q)}$ being $\mathbf{\Gamma}^{(\bar{q})}$,*

(i) if for each $\alpha \in (0, 1)$, $\alpha_n = \alpha$ and $q_* \in I(\bar{q})$, then for each $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|\widehat{q}_n - q_*| > \epsilon) = \alpha$;

(ii) if for each $q \in I(\bar{q})$, (a) $\mathbb{P}(\sup_{\gamma\in\mathbf{\Gamma}^{(\bar{q})}} \mathcal{Z}_q(\gamma) \geq a_q) \leq 1/2$ for some a_q, (b) $\lim_{n\to\infty} \alpha_n = 0$, and (c) $\lim_{n\to\infty} \log(\alpha_n)/n = 0$, then for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|\widehat{q}_n - q_*| > \epsilon) = 0$. □

Remarks.

(a) From Theorem 3(*i*), if α_n does not converge to zero as n tends to infinity, the degree estimator does not vanish to zero. Theorem 3(*ii*) provides the conditions for α_n to converge to zero so that the degree estimation error converges to zero. Note that the possibility of estimating a degree less than q_* gets smaller as n increases, because the DD test statistic has omnibus power for $q < q_*$.

(b) The regularity conditions in Theorem 3(*ii*) are weaker than those in theorem 2 of Cho and Phillips (2018), because they presume a locally stationary Gaussian process with covariance structure dominated by that of the standardized $\mathcal{Z}_q(\cdot)$. However, Theorem 3(*ii*) does not assume such a Gaussian process.

(c) Since the proof of Theorem 3(*i*) is straightforward from Corollary 1, we do not include it in the appendix. Theorem 3(*ii*) can be proved by applying Borel's theorem on the upper probability bound of an extreme Gaussian stochastic process (*e.g.*, Piterbarg, 1996, p. 13). $\square$

4 Simulation

In this section, we use simulation to examine the DD test statistic and compare its numerical performance with those of the J-test statistic and MSC.

We consider a simple DGP structure for our simulation. First, we assume that $(D_t, G_t, \mathbf{W}_t')' \sim$ IID $N(\mathbf{0}_4, \mathbf{I}_4)$, where $\mathbf{W}_t \in \mathbb{R}^2$ and $G_t \in \mathbb{R}$. Second, we assume that $U_t := G_t$ and $X_t := \exp(\frac{1}{2}(\mathbf{W}_t'\boldsymbol{\iota}_2 + G_t))$. Note that X_t is positively valued with probability 1. Third, we consider

$$Y_t := 1 + D_t + X_t + X_t^2 + U_t,$$

so that Y_t is quadratically associated with X_t, and X_t and U_t are correlated through G_t. Finally, we propose two plans for the instrumental variables. We first let $\mathbf{Z}_t$ be $(1, D_t, \mathbf{W}_t', W_{1,t}^2, W_{2,t}^2, W_{1,t}^3)' \in \mathbb{R}^7$, where $W_{j,t}$ denotes the j^{th}-row element of $\mathbf{W}_t$, and then assume that $\mathbf{Z}_t := (1, D_t, \mathbf{W}_t', W_{1,t}^2, W_{2,t}^2, W_{1,t}^3, W_{2,t}^3)' \in \mathbb{R}^8$, to estimate the unknown parameters using GMM estimation. We call the two instrumental variable sets as Sets A and B respectively, with the different sets employed to examine how the DD test statistic responds to the different instrumental variables.

The main goal of an empirical researcher is to estimate the structural relationship between Y_t and $(1, X_t, D_t)$. We assume that the researcher specifies the following set of models without having an idea of the true structure between Y_t and X_t: for each $q \in I(3) := \{1, 2, 3\}$,

$$\mathcal{M}_q' := \left\{ m_{t,q}(\boldsymbol{\omega}^{(q)}) := Y_t - \xi_0 - X_t\xi_1 - \ldots - X_t^q\xi_q - D_t\eta - \beta X_t^\gamma : \boldsymbol{\omega}^{(q)} \in \boldsymbol{\Omega}^{(q)} \subset \mathbb{R}^{4+q} \right\},$$

where $\boldsymbol{\omega}^{(q)} := (\xi_0, \ldots, \xi_q, \eta, \beta, \gamma)'$, and $\boldsymbol{\Omega}^{(q)}$ is the parameter space of $\boldsymbol{\omega}^{(q)}$. In particular, we assume that the parameter space of γ is $\boldsymbol{\Gamma} = [-0.25, 3.50]$, so that the third-degree polynomial model is nested in $\mathcal{M}_q'$ for every q. The other parameter spaces are not restricted. Given this model assumption, we also assume that the researcher estimates the unknown parameters using the GMM estimation method with weighting matrix $\mathbf{M}_n = (n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$. Since U_t is independent of $\mathbf{Z}_t$, this weighting matrix can reduce the size distortion of the J-test statistic without resorting to the bootstrap methodology (*e.g.*, Hall and Horowitz, 1996). Note that $\mathcal{M}_2'$ and $\mathcal{M}_3'$ are correctly specified models, where $\mathcal{M}_2'$ is the most parsimonious model among correctly specified ones. The empirical researcher's main goal is therefore achieved when estimating $q_* = 2$ consistently.

Given the DGP and model assumptions, we perform our simulations in the following order.

(a) We compute the DD test statistic for each model and test whether the structural model is correctly specified. Here, we estimate the covariance matrix $\widetilde{\mathbf{\Sigma}}$ in Corollary 1 by $\widetilde{\sigma}_{q,n}^2(n^{-1}\mathbf{Z}'\mathbf{Z})$, where $\widetilde{\sigma}_{q,n}^2 := n^{-1}\widetilde{\mathbf{U}}^{(q)\prime}\widetilde{\mathbf{U}}^{(q)}$ and $\widetilde{\mathbf{U}}^{(q)} := \mathbf{Y} - \mathbf{V}_q\widetilde{\boldsymbol{\varsigma}}_n^{(q)}$. Note that $\widetilde{\boldsymbol{\omega}}_n^{(q)} \equiv (\widetilde{\boldsymbol{\varsigma}}_n^{(q)\prime}, 0, \gamma)'$. We fix the significant level at 10%, 5%, and 1%, and examine how the DD test statistic size distortion evolves as the sample size increases from 100 to 1,000 by 100 observations.

(b) We also perform sequential testing using the J-test statistic. We compute the J-test statistics using the following models. For each $q \in I(3)$,

$$\mathcal{M}_q'' := \left\{ m_{t,q}(\bar{\boldsymbol{\omega}}^{(q)}) := Y_t - \xi_0 - X_t\xi_1 - \ldots - X_t^q\xi_q - D_t\eta : \bar{\boldsymbol{\omega}}^{(q)} \in \bar{\boldsymbol{\Omega}}^{(q)} \subset \mathbb{R}^{2+q} \right\},$$

where $\bar{\boldsymbol{\omega}}^{(q)} := (\xi_0, \xi_1, \ldots, \xi_q, \eta)$ and $\bar{\boldsymbol{\Omega}}^{(q)}$ is the parameter space of $\bar{\boldsymbol{\omega}}^{(q)}$. Note that $\mathcal{M}_q''$ is the q^{th}-degree polynomial model with respect to X_t without having its power transform, obviating the identification problem. Assume that $\mathcal{J}_{n,q}$ denotes the J-test statistic that can test the q^{th}-degree polynomial model.

(c) We allow the significance level to depend on the sample size so that the estimation error degree becomes zero as the sample size increases. To apply Theorem 3(*ii*), we first assume that $\alpha_n = 1/n^{1/2}$, $1/n^{3/4}$, and $1/n$, and examine how the estimation error is formed using simulation. Note that the significance levels converge to zero by all the level plans, with the convergence rate of $\alpha_n = 1/n$ faster than the others. In addition to the DD test statistic, we apply the J-test statistic to the sequential testing procedure. We call these sequential testing procedures the DD- and J-sequential testing procedures, respectively.

(d) Finally, we apply the MSCs in Andrews (1999). We examine three MSCs, the Akaike-type model MSC, Bayesian-type MSC, and Hannan-Quinn-type MSC; specifically, they are

$$\text{AIC-MSC} := \bar{\mathcal{J}}_{n,q} - 2(m-q-2)/n, \qquad \text{Bayesian-MSC} := \bar{\mathcal{J}}_{n,q} - \log(n)(m-q-2)/n,$$

$$\text{Hannan-Quinn-MSC} := \bar{\mathcal{J}}_{n,q} - \kappa\log(\log(n))(m-q-2)/n,$$

respectively, where $\bar{\mathcal{J}}_{n,q} := n^{-1}\mathcal{J}_{n,q}$, m is the number of instrumental variables; we assume that κ is 2.01 following Andrews (1999). The model that performs best is the one with the smallest MSC. $\square$

We iteratively perform this four-step simulation for the data sets formed by Sets A and B and report the simulation results in Tables 1 and 2. Table 1 presents the results obtained through the first two steps, and Table 2 presents the results obtained through the next two steps. Specifically, they report the precision rates of the methods. For example, if $\widehat{q}_{n,i}$ denotes the degree estimated by the i^{th}- simulation, the precision rate is computed using $r^{-1}\sum_{i=1}^{r}\mathbf{I}(\widehat{q}_{n,i}=q_*)\times 100$, where r is the number of simulations repeated and $\mathbf{I}(\cdot)$ is the indicator function.

The simulation results reported in Table 1 are summarized as follows:

(a) When the significance level is fixed irrespective of n, the degrees estimated using the DD- and J-sequential testing procedures yield the results as desired by Theorem 3(i): if the significance level is fixed at α, the estimated precision rate would converge to $(1-\alpha)\times 100$ for $q=2$ as n increases. For example, when $\alpha=0.10$, the DD- and J-sequential testing procedures give precision rates close to 90% for $q=2$.

(b) When the significance level is relatively high, the J-sequential testing procedure gives more precise estimates than the procedure using the DD test statistic. If the significance level is low, the DD-sequential testing procedure would yield more precise estimates.

(c) The J-sequential testing procedure yields a wider distribution for $\widehat{q}_n$ than the procedure using the DD test statistic. Note that for every sample size n, $\widehat{q}_n=1$ or 4 is more often observed than $\widehat{q}_n=3$ for the J-sequential testing procedure unless $\widehat{q}_n=2$. However, $\widehat{q}_n=3$ is more often observed than $\widehat{q}_n=1$ or 4 for the DD-sequential testing procedure, implying that the estimated degree distribution is more concentrated around $q_*=2$ for the DD-sequential testing procedure. □

Table 2 reports the precision rate of each estimation method by letting $\alpha_n=n^{-1/2}$, $n^{-3/4}$, and n^{-1}. Table 2 is summarized as follows:

(a) As n increases, the estimation errors decrease when applying the DD- and J-sequential testing procedures. Note that the difference between $\widehat{q}_n$ and q_* also decreases under both methods. Furthermore, for any level plan, smaller estimation errors are observed for the data sets with larger n, and so the degree estimation errors based on $\alpha_n=n^{-1}$ are smaller than the others.

(b) As n increases, the estimation errors when using MSCs also decrease, with the Bayesian MSC estimating q_* more efficiently than the other two MSCs.

(c) Overall, the DD-sequential testing procedure performs better than the MSCs and the J-sequential testing procedure. For each sample size, we indicate the method performing best in boldface font. Note that the DD-sequential testing procedure overall dominates the other methods.

(d) As regards the DD-sequential testing procedure, no level plan uniformly dominates the other plans. If n is small, $\alpha_n = n^{-1/2}$ performs better than the others; if n is moderately large, $\alpha_n = n^{-3/4}$ becomes dominant; and if n is sufficiently large, $\alpha_n = n^{-1}$ becomes dominant over the other plans. This implies that the level plan itself needs to be carefully selected so as to depend on n.

(e) For a small n, the DD-sequential testing procedure performs better than the J-sequential testing procedure, but this is not true for every level plan and sample size. For example, for $\alpha_n = n^{-1/2}$, if n increases, the J-sequential testing procedure will perform better than the procedure using the DD test statistic. In other words, the estimation error when using the J-sequential testing procedure will become zero more quickly than when using the DD test statistic procedure. This also implies that the degree estimation error when using the J-sequential testing procedure is better controlled than when using the DD-sequential testing procedure for large n. However, for $\alpha_n = n^{-3/4}$ or n^{-1}, this dominance relationship is reversed and observed for all ns, and so the estimation error when using the DD-sequential testing procedure shows more precise rates for all ns.

(f) As regards the MSC, all the MSCs are always dominated by the DD-sequential testing procedure with $\alpha_n = n^{-3/4}$. □

These simulations prove that we can efficiently estimate the most parsimonious and correctly specified polynomial structures using the DD-sequential testing procedure.

5 Empirical Application

In the literature on human capital, identifying the return to education has been a popular research topic. Card (1995) examines the monetary return to education using college proximity as an instrument variable, constructed from the Young Men Cohort of the National Longitudinal Survey (NLS). Specifically, his empirical analysis shows that individual educational variations are created by proximity to college, and uses this variation to estimate the structural earnings equations, to find the instrumental variable estimates about 50∼60% higher than those measured by ordinary least squares (OLS) estimation. This distinct result is empirically important from the viewpoint of Griliches (1977) and Willis (1986), who pointed out the possibility of significantly large omitted variable bias when estimating the returns of education due to individual unobservable abilities.

Note that all of the models estimated by Card (1995) are linear, and the estimates might suffer from model misspecification bias. Although the instrumental variables used are orthogonal to the unobservable ability variables, the estimates can suffer from the endogeneity arising from the omitted variable bias when the education effect is nonlinear. The power direction of the J-test statistic differs from that of the DD test statistic and the linear structure assumption may not be valid for their data sets.

We therefore conduct model misspecification testing separately using the DD test statistic. We design this investigation to affirm or refute the original inference of Card (1995) and further use it as an empirical illustration of the DD-sequential testing procedure.

The structural model posited by Card (1995) has a standard model structure to enable GMM estimation. The basic linear model is given as

$$\log(wage_t) = \alpha + \beta_1 educ_t + \mathbf{X}_t' \boldsymbol{\eta} + U_t, \tag{10}$$

where $\log(wage_t)$ is the log wage, $educ_t$ is the schooling years, $\mathbf{X}_t$ is the set of observable controls, and U_t is the structural error. Because $educ_t$ and U_t can be correlated, the OLS estimation may be biased. To remedy the endogeneity problem, Card (1995) employed the proximity variables to four-year private or public colleges as instrumental variables for schooling years, as mentioned above. He further lets the set of controls to include family structure, parental education, residential location dummies, region and urban/rural indicators, age, and race dummies. Here, the family structure provides the dummy variables indicating whether the individual lived with a single mother or both parents at age 14. For parental education level, the dummy variable is constructed to indicate whether both parents had under 12 years of schooling.

We extend Card's (1995) structural models to polynomial structural models. In addition to his original instrumental variables, we include other instrumental variables since multiple instrumental variables are required for DD-sequential testing. We generate the other instrumental variables by multiplying the proximity variables to four-year private or public college or to family structure. That is, we let the instrumental variables be the two original proximity variables and their interactions, with the dummy variable indicating whether the individual was from a single-mother family. These new instrumental variables arise because the geographical variations in college proximity have heterogeneous effects over individuals with different family characteristics. The dependent variable is the log wage in 1976 or 1978.

In our data analysis, we first verify whether the employed instrumental variables are weak. Following Andrews, Stock, and Sun (2019), we compute the effective first-stage F-test statistics, as reported in Table 3. Note that they are close to 10, in accordance with Staiger and Stock's (1997) rule of thumb. From this, we conclude that the current estimations do not significantly suffer from the weak instrumental variable problem.

We next apply the DD- and J-sequential testing procedures to the linear model and data adopting the

MSC approach. The testing results are reported in Table 3. For the DD-sequential testing procedure, we assume that $\mathbf{\Gamma} = [0.5, 3.5]$, and transform the model in (10) into $\mathcal{M}_q$ with $q \in \{1, 2\}$, so that the DD test statistics can be computed for each q. Next, using the null limit distribution in Corollary 1, we compute the DD test statistics p-values and report them in parentheses. For this, we obtain the null weak limit in Corollary 1 through 500 independent simulation experiments. The figures in parentheses denoting the percentiles for simulated weak limits are greater than the respective DD test statistics figures. We further estimate the J-test statistics p-values through a similar simulation. The desired null limit distributions are obtained by estimating the weighting matrix using White's (1980) heteroskedasticity consistent covariance matrix estimator, and the estimated p-values are reported in parentheses below the J-test statistics. Finally, we apply Andrew's (1999) Bayesian MSC to the same data and models, using the optimal weighting matrix recommended by Andrews (1999). We report the degrees selected by these three methods in boldface.

Both the DD- and J- sequential testing procedures estimate the same polynomial degree model. That is, the log wage with respect to education years is structurally linear for both the 1976 and 1978 data sets. Thus, the linear models specified by Card (1995) are correct, as further affirmed by the Bayesian MSC. The linear models with respect to education minimize the Bayesian MSC values for both data sets.

We finally report the model coefficients estimated using the OLS and TSLS methods in Table 4. The TSLS estimations are obtained using the instrumental variables employed for the DD sequential testing procedure. Notwithstanding the different instrumental variables, the estimates obtained are more or less similar to what Card (1995) reported, although the estimates in Table 4 are slightly higher than those given by Card (1995). Therefore, we draw the same conclusion as Card (1995): the returns to education are distinctively higher than those measured by OLS estimation.

6 Concluding Remarks

In this study, we provide an econometric method to estimate a correct structural model. For this, we proceed in three steps. First, we provide the DD test statistic and show how it has omnibus power against an arbitrary nonlinear structure. We also derive the null and local alternative limit distributions of the DD test statistic. Second, we approximate the nonlinear structural equation using a polynomial function if the linear model is rejected, and provide a sequential testing procedure to consistently estimate the degree of polynomial function. This procedure can consistently estimate the polynomial function when it is finite, with the significance level converging to zero as the sample size tends toward infinity. These properties and their performance relative to the J-sequential testing procedure and MSCs are also compared through simulation. Third, we investigate the structural relationship between the log wage and education years using Card's (1995) National Longitudinal Survey data. Our methodology shows that his linear model is correctly specified and leads to his inference results.

Appendix: Proofs

Before proving the main claims of this study, we provide some preliminary lemmas to facilitate the proofs. For notational simplicity, we assume that $\mathbf{F} := \mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}$ and $\ddot{\mathbf{P}} := \ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}$.

Lemma A1. *Given Assumptions 1, 2, and 3,*

(i) $\mathbf{Z}'\mathbf{U} = O_{\mathbb{P}}(\sqrt{n})$,

(ii) $\mathbf{V}'\mathbf{V} = O_{\mathbb{P}}(n)$, $\mathbf{C}_0\mathbf{Z} = O_{\mathbb{P}}(n)$, $\mathbf{V}'\mathbf{Z} = O_{\mathbb{P}}(n)$, $\mathbf{Z}'\mathbf{Z} = O_{\mathbb{P}}(n)$, *and* $\mathbf{K}_1'\mathbf{Z} = O_{\mathbb{P}}(n)$, *where for* $j = 1, 2, \ldots$, $\mathbf{K}_j := [\mathbf{L}_j \vdots \mathbf{0}_{n\times k}]$ *and* $\mathbf{L}_j := [L_1^j, \ldots, L_n^j]'$*;*

(iii) $\mathbf{L}_2\mathbf{Z} = O_{\mathbb{P}}(n)$, *and* $\mathbf{K}_2\mathbf{Z} = O_{\mathbb{P}}(n)$*;*

(iv) $\mathbf{Z}'\mathbf{U} = o_{\mathbb{P}}(n)$. □

Lemma A2. *For* $j = 1, 2, \ldots$, *let* $d_n^{(j)}(0;\xi_0) := (\partial^j/\partial\gamma^j)d_n(\gamma;\xi_0)|_{\gamma=0}$. *Given Assumptions 1, 2, 3, and* $\mathcal{H}_{0,2}$,

(i) $d_n^{(1)}(0;\xi_0) = -2(\xi_{0*} - \xi_0)\mathbf{C}_0'\mathbf{Q}_1\mathbf{U} + 2\mathbf{U}'\mathbf{K}_1\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} + \mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}(\ddot{\mathbf{P}}'\mathbf{K}_1 + \mathbf{K}_1'\ddot{\mathbf{P}})\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U}$*;*

(ii) $d_n^{(1)}(0;\xi_0) = -2(\xi_{0*} - \xi_0)\mathbf{C}_0'\mathbf{Q}_1\mathbf{U} + O_{\mathbb{P}}(n)$*; and*

(iii) $d_n^{(2)}(0;\xi_0) = 2(\xi_{0*} - \xi_0)^2\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0 + o_{\mathbb{P}}(n^2)$. □

Lemma A3. *Given Assumptions 1, 2, 3, and* $\mathcal{H}_{0,2}$,

(i) $\mathcal{D}_{n,1}^{(\gamma=0;\beta)} = \{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2/\{n\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0\} + o_{\mathbb{P}}(1)$*; and*

(ii) $\mathcal{D}_{n,1}^{(\gamma=0;\beta)} = O_{\mathbb{P}}(1)$. □

Lemma A4. *Given Assumptions 1, 2, 3, and* $\mathcal{H}_{0,2}$,

(i) $\mathcal{D}_{n,1}^{(\gamma=0;\xi_0)} = \{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2/\{n\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0\} + o_{\mathbb{P}}(1)$*;*

(ii) $\mathcal{D}_{n,1}^{(\gamma=0;\xi_0)} = O_{\mathbb{P}}(1)$. □

Lemma A5. *Given Assumptions 1, 2, and 3,*

(i) $\mathbf{Z}'\mathbf{U} = O_{\mathbb{P}}(\sqrt{n})$*;*

(ii) $\mathbf{V}'\mathbf{V} = O_{\mathbb{P}}(n)$, $\mathbf{C}_1\mathbf{Z} = O_{\mathbb{P}}(n)$, $\mathbf{V}'\mathbf{Z} = O_{\mathbb{P}}(n)$, $\mathbf{Z}'\mathbf{Z} = O_{\mathbb{P}}(n)$, *and* $\overline{\mathbf{K}}_1'\mathbf{Z} = O_{\mathbb{P}}(n)$, *where for* $j = 1, 2, \ldots$, $\overline{\mathbf{K}}_j := [\mathbf{0}_{n\times 1} \vdots \mathbf{C}_j \vdots \mathbf{0}_{n\times k}]$*;*

(iii) $\mathbf{C}_2\mathbf{Z} = O_{\mathbb{P}}(n)$, *and* $\overline{\mathbf{K}}_2\mathbf{Z} = O_{\mathbb{P}}(n)$*;*

(iv) $\mathbf{Z}'\mathbf{U} = o_{\mathbb{P}}(n)$. □

Lemma A6. *For $j = 1, 2, \ldots$, $d_n^{(j)}(1;\xi_1) := (\partial^j/\partial\gamma^j) d_n(\gamma;\xi_1)|_{\gamma=1}$. Given Assumptions 1, 2, 3, and $\mathcal{H}_{0,3}$,*

(i) $d_n^{(1)}(1;\xi_1) = -2(\xi_{1*} - \xi_1)\mathbf{C}_1'\mathbf{Q}_1\mathbf{U} - 2\mathbf{U}'\overline{\mathbf{K}}_1\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} + \mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}(\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1\ddot{\mathbf{P}})\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U}$*;*

(ii) $d_n^{(1)}(1;\xi_1) = -2(\xi_{1*} - \xi_1)\mathbf{C}_1'\mathbf{Q}_1\mathbf{U} + O_{\mathbb{P}}(n)$*; and*

(iii) $d_n^{(2)}(1;\xi_1) = 2(\xi_{1*} - \xi_1)\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1 + o_{\mathbb{P}}(n^2)$. □

Lemma A7. *Given Assumptions 1, 2, 3, and $\mathcal{H}_{0,3}$,*

(i) $\mathcal{D}_{n,1}^{(\gamma=1;\beta)} = \{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2/\{n\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1\} + o_{\mathbb{P}}(1)$*; and*

(ii) $\mathcal{D}_{n,1}^{(\gamma=1;\beta)} = O_{\mathbb{P}}(1)$. □

Lemma A8. *Given Assumptions 1, 2, 3, and $\mathcal{H}_{0,3}$,*

(i) $\mathcal{D}_{n,1}^{(\gamma=1;\xi_1)} = \{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2/\{n\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1\} + o_{\mathbb{P}}(1)$*; and*

(ii) $\mathcal{D}_{n,1}^{(\gamma=1;\xi_1)} = O_{\mathbb{P}}(1)$. □

Proof of Lemma A1: (*i*) $\mathbf{Z}'\mathbf{U} = [\sum_t Z_{tj}U_t]$. Since $\mathbb{E}[Z_{tj}^2 U_t^2] < \mathbb{E}[Z_{tj}^4]^{1/2}\mathbb{E}[U_t^4]^{1/2}$ by the Cauchy Schwarz inequality, $\mathbb{E}[Z_{tj}^4] < \infty$, and $\mathbb{E}[U_t^4] < \infty$ hold by the Assumption 3, we can apply the CLT and obtain the desired result.

(*ii*) By the definition of $\mathbf{K}_1$, if $\mathbf{C}_0'\mathbf{Z} = O_{\mathbb{P}}(n)$, $\mathbf{K}_1'\mathbf{Z} = O_{\mathbb{P}}(n)$. We assu me that $\mathbf{R}$ is a generic notation for $\mathbf{V}$, $\mathbf{C}_0$, and $\mathbf{Z}$. As $\mathbf{R}'\mathbf{Z} = [\sum R_{tj}Z_{ti}]$, the result follows by ergodicity if $\mathbb{E}[|R_{tj}Z_{ti}|] < \infty$, which holds by the Cauchy-Schwarz inequality and the fact that $\mathbb{E}[Z_{ti}^2] < \infty$, $\mathbb{E}[V_{tj}^2] < \infty$, and $\mathbb{E}[\log^2(X_t)] < \infty$ by Assumption 3.

(*iii*) Similarly, by the definition of $\mathbf{K}_2$, if $\mathbf{L}_2'\mathbf{Z} = O_{\mathbb{P}}(n)$, $\mathbf{K}_2'\mathbf{Z} = O_{\mathbb{P}}(n)$. As $\mathbb{E}[\log^4(X_t)] < \infty$ and $\mathbb{E}[Z_{ti}^2] < \infty$, the result similarly follows from ergodicity and the Cauchy-Schwarz inequality.

(*iv*) This simply follows from the fact that $\{\mathbf{Z}_t U_t\}$ is a mixingale sequence by the Assumption 1 and applying LLN. ■

Proof of Lemma A2: (*i*) We can obtain the first-order derivative with respect to γ as follows:

$$
\begin{aligned}
d_n^{(1)}(0;\xi_0) = &- 2\mathbf{P}(\xi_0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)[\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)]^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{P}(\xi_0) \\
&- \mathbf{P}(\xi_0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)(d/d\gamma)[\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)]^{-1}\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{P}(\xi_0).
\end{aligned}
$$

Note that

$$
(d/d\gamma)[\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)]^{-1} = -\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1 + \mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}, \tag{1}
$$

and that $\mathbf{P}(\xi_0) = \mathbf{Y} - \xi_0\boldsymbol{\iota} = \mathbf{V}[\xi_{0*} - \xi_0, \boldsymbol{\delta}_*']' + \mathbf{U} = \mathbf{V}\boldsymbol{\kappa}(\xi_0) + \mathbf{U}$ by assuming that $\boldsymbol{\kappa}(\xi_0) := [\xi_{0*} - \xi_0, \boldsymbol{\delta}_*']'$.

For notational simplicity, we suppress ξ_0 in $\boldsymbol{\kappa}(\xi_0)$. From $\mathbf{H}(0) = \mathbf{V}$ and $\ddot{\mathbf{P}} := \ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}$, it follows that

$$d_n^{(1)}(0;\xi_0) = -\underbrace{2(\mathbf{V}\boldsymbol{\kappa} + \mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'(\mathbf{V}\boldsymbol{\kappa} + \mathbf{U})}_{(A)} + \underbrace{(\mathbf{V}\boldsymbol{\kappa} + \mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1 + \mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\ddot{\mathbf{P}}'(\mathbf{V}\boldsymbol{\kappa} + \mathbf{U})}_{(B)}.$$

We now examine each RHS component. The first component (A) can be expressed as a sum of following four components:

(a) $-2\boldsymbol{\kappa}'\mathbf{V}\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{P}}\boldsymbol{\kappa} = -2\boldsymbol{\kappa}'\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}\boldsymbol{\kappa}$;

(b) $-2\boldsymbol{\kappa}'\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U}$;

(c) $-2\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{P}}\boldsymbol{\kappa}$; and

(d) $-2\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U}$.

Next, the second component (B) can also be expressed as the sum of four other components:

(a) $\boldsymbol{\kappa}'(\ddot{\mathbf{P}}'\mathbf{K}_1 + \mathbf{K}_1'\ddot{\mathbf{P}})\boldsymbol{\kappa} = 2\boldsymbol{\kappa}'\mathbf{K}_1'\ddot{\mathbf{P}}\boldsymbol{\kappa}$;

(b) $\boldsymbol{\kappa}'\ddot{\mathbf{P}}'\mathbf{K}_1\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} + \mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{P}}\boldsymbol{\kappa} = 2\boldsymbol{\kappa}'\ddot{\mathbf{P}}'\mathbf{K}_1\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U}$;

(c) $\boldsymbol{\kappa}'\mathbf{K}_1'\ddot{\mathbf{P}}\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} + \mathbf{U}'\ddot{\mathbf{P}}'\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{K}_1\boldsymbol{\kappa} = 2\boldsymbol{\kappa}'\mathbf{K}_1'\ddot{\mathbf{P}}\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U}$;

(d) $\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1 + \mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U}$.

By adding and organizing all these terms according to their order of convergence, we obtain the following:

(a) $-2\boldsymbol{\kappa}'\mathbf{K}_1'\ddot{\mathbf{P}}\boldsymbol{\kappa} + 2\boldsymbol{\kappa}'\mathbf{K}_1'\ddot{\mathbf{P}}\boldsymbol{\kappa} = 0$;

(b, c) $-2\boldsymbol{\kappa}'\{\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}' + \mathbf{K}_1'\ddot{\mathbf{P}}\mathbf{F}^{-1}\ddot{\mathbf{P}}'\}\mathbf{U} = -2(\xi_{0*} - \xi_0)\mathbf{C}_0\mathbf{Q}_1\mathbf{U}$; and

(d) $\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1 + \mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} - 2\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U}$.

Hence, the first-order derivative can be obtained as

$$d_n^{(1)}(0;\xi_0) = -2(\xi_{0*} - \xi_0)\mathbf{C}_0'\mathbf{Q}_1\mathbf{U} - 2\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U} + \mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1 + \mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U}.$$

(*ii*) Given the result in (*i*), by applying the result of Lemma A1, we obtain

$$\begin{aligned} d_n^{(1)}(0;\xi_0) &= -2(\xi_{0*} - \xi_0)\underbrace{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}}_{O_{\mathbb{P}}(n^{3/2})} - 2\underbrace{\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U}}_{O_{\mathbb{P}}(n)} + \underbrace{\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1 + \mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U}}_{O_{\mathbb{P}}(n)} \\ &= -2(\xi_{0*} - \xi_0)\mathbf{C}_0'\mathbf{Q}_1\mathbf{U} + O_{\mathbb{P}}(n). \end{aligned}$$

(*iii*) The second-order derivative is obtained as

$$
\begin{aligned}
d_n^{(2)}(0;\xi_0) = & -2\mathbf{P}(\xi_0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{K}_1[\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)]^{-1}\mathbf{K}_1'\mathbf{P}(\xi_0) \\
& -4\mathbf{P}(\xi_0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)(d/d\gamma)[\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)]^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{P}(\xi_0) \\
& -2\mathbf{P}(\xi_0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)[\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)]^{-1}\mathbf{K}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{P}(\xi_0) \\
& -\mathbf{P}(\xi_0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)(d^2/d\gamma^2)[\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)]^{-1}\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{P}(\xi_0),
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{d^2}{d\gamma^2}[\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)]^{-1} = & 2\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1} \\
& -\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_2+\mathbf{K}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}+2\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{K}_1]\mathbf{F}^{-1},
\end{aligned}
$$

and (1) shows the specific form of $(d/d\gamma)[\mathbf{H}(0)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{H}(0)]^{-1}$. Using these results, we arrange the terms to obtain

$$
\begin{aligned}
d_n^{(2)}(0;\xi_0) = & -2(\mathbf{V}\boldsymbol{\kappa}+\mathbf{U})'\{\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{K}_1\mathbf{F}^{-1}\ddot{\mathbf{P}}'+\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{K}_1\mathbf{F}^{-1}\mathbf{K}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\}(\mathbf{V}\boldsymbol{\kappa}+\mathbf{U}) \\
& +4(\mathbf{V}\boldsymbol{\kappa}+\mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'(\mathbf{V}\boldsymbol{\kappa}+\mathbf{U}) \\
& -2(\mathbf{V}\boldsymbol{\kappa}+\mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{M}^{-1}\ddot{\mathbf{P}}'(\mathbf{V}\boldsymbol{\kappa}+\mathbf{U}) \\
& -(\mathbf{V}\boldsymbol{\kappa}+\mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}[2\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{K}_1+\ddot{\mathbf{P}}'\mathbf{K}_2+\mathbf{K}_2'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\ddot{\mathbf{P}}'(\mathbf{V}\boldsymbol{\kappa}+\mathbf{U}).
\end{aligned}
$$

By organizing each term according to their order of convergence and applying Lemma A1, because $\mathbb{E}[\mathbf{Z}_tU_t]=0$, we can obtain

- $-2\boldsymbol{\kappa}'\{\ddot{\mathbf{P}}'\mathbf{K}_1\mathbf{F}^{-1}\mathbf{K}_1+\mathbf{K}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\}\mathbf{V}\boldsymbol{\kappa}+4\boldsymbol{\kappa}'[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{P}}\boldsymbol{\kappa}-2\boldsymbol{\kappa}'[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\boldsymbol{\kappa}+2\boldsymbol{\kappa}'[2\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{K}_1+\mathbf{K}_2'\ddot{\mathbf{P}}+\ddot{\mathbf{P}}'\mathbf{K}_2]\boldsymbol{\kappa} = 2(\boldsymbol{\kappa}'\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{K}_1\boldsymbol{\kappa}-2\boldsymbol{\kappa}'\mathbf{K}_1'\ddot{\mathbf{P}}\mathbf{M}^{-1}\ddot{\mathbf{P}}'\boldsymbol{\kappa}) = 2(\xi_{0*}-\xi_0)^2\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0 = O_{\mathbb{P}}(n^2)$.

- $-4\boldsymbol{\kappa}'\ddot{\mathbf{P}}'\mathbf{K}_1\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U}+4\boldsymbol{\kappa}'[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U}-4\boldsymbol{\kappa}'[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U}-2\boldsymbol{\kappa}'\mathbf{K}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U}-2\boldsymbol{\kappa}'\ddot{\mathbf{P}}'\mathbf{K}_2\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} = -2(\xi_{0*}-\xi_0)[\mathbf{L}_2'\mathbf{Q}_1\mathbf{U}-2\mathbf{C}_0'\mathbf{Q}_1\mathbf{K}_1\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U}+2\mathbf{C}_0'\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_1'\mathbf{Q}_1\mathbf{U}] = o_{\mathbb{P}}(n^2)$.

- $-2\mathbf{U}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{K}_1\mathbf{F}^{-1}\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U}-2\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\mathbf{K}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U}+4\mathbf{U}'\ddot{\mathbf{P}}(\ddot{\mathbf{P}}'\mathbf{V})^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\mathbf{K}_1'\mathbf{Z}\,\mathbf{M}_n\mathbf{Z}'\mathbf{U}+2\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\{[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\mathbf{K}_1+\mathbf{K}_1'\ddot{\mathbf{P}}]-\mathbf{K}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{K}_1-\ddot{\mathbf{P}}'\mathbf{K}_2\}\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} = o_{\mathbb{P}}(n^2)$.

Therefore, by adding all these terms, we can have $d_n^{(2)}(0;\xi_0) = 2(\xi_{0*}-\xi_0)^2\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0 + o_{\mathbb{P}}(n^2)$. ■

Proof of Lemma A3: (*i*) By applying a second-order Taylor expansion to $d_n(\gamma;\beta)$ and optimizing with respect to γ, we have

$$\inf_{\gamma}\{d_n(\gamma;\beta) - d_n(0;\beta)\} = -\frac{\{d_n^{(1)}(0;\beta)\}^2}{2d_n^{(2)}(0;\beta)} + o_{\mathbb{P}}(1).$$

Given this, we note that $d_n^{(1)}(0;\beta) := (d/d\gamma)d_n(0;\beta) = 2\beta\mathbf{C}_0'\mathbf{Q}_1\mathbf{U} = O_{\mathbb{P}}(n^{3/2})$ and $L_n^{(2)}(0;\beta) := (d^2/d\gamma^2)L_n(0;\beta) = \beta^2\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0 - \beta\mathbf{L}_2'\mathbf{Q}_1\mathbf{U} = O_{\mathbb{P}}(n^2)$. From this, it follows that

$$\begin{aligned}\mathcal{D}_{n,1}^{(\gamma=0;\beta)} &= -\inf_{\gamma\in\mathbf{\Gamma}} n^{-1}\{d_n(\gamma;\beta) - d_n(0;\beta)\} \\ &= \frac{\{n^{-3/2}\beta\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2}{n^{-2}(\beta^2\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0 - \beta\mathbf{L}_2'\mathbf{Q}_1\mathbf{U})} + o_{\mathbb{P}}(1) = \frac{\{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2}{n\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0 y} + o_{\mathbb{P}}(1),\end{aligned}$$

because $\mathbf{L}_2'\mathbf{Q}_1\mathbf{U} = o_{\mathbb{P}}(n^2)$, as shown in (*ii*).

(*ii*) We separate the proof into three parts. First, we note that $\mathbf{C}_0'\mathbf{Q}_1\mathbf{U} = \mathbf{C}_0'\ddot{\mathbf{Z}}(\mathbf{I} - \ddot{\mathbf{Z}}'\mathbf{V}(\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V})^{-1}\mathbf{V}'\ddot{\mathbf{Z}})\ddot{\mathbf{Z}}'\mathbf{U}$. Lemmas A1(*i*, *ii*) and Assumption 3 imply that $\mathbf{C}_0'\mathbf{Q}_1\mathbf{U} = O_{\mathbb{P}}(n^{3/2})$. Similarly, Lemmas A1(*ii*) and Assumption 3 imply that $\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0 = O_{\mathbb{P}}(n^2)$. Further, Lemmas A1(*ii*, *iii*, and *iv*) and Assumption 3 imply that $\mathbf{L}_2'\mathbf{Q}_1\mathbf{U} = o_{\mathbb{P}}(n^2)$. By combining all these results, we obtain the desired result. ■

Proof of Lemma A4: (*i*) By applying a second-order Taylor expansion to $d_n(\cdot;\xi_0)$ and optimizing with respect to γ, we have

$$\inf_{\gamma\in\mathbf{\Gamma}}\{d_n(\gamma;\xi_0) - d_n(0;\xi_0)\} = -\frac{\{d_n^{(1)}(0;\xi_0)\}^2}{2d_n^{(2)}(0;\xi_0)} + o_{\mathbb{P}}(n) = -\frac{\{2(\xi_{0*}-\xi_0)\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2}{4(\xi_{0*}-\xi_0)^2\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0} + o_{\mathbb{P}}(n).$$

Therefore,

$$\mathcal{D}_{n,1}^{(\gamma=0;\xi_0)} = -\inf_{\gamma} n^{-1}\{d_n(\gamma;\xi_0) - d_n(0;\xi_0)\} = \frac{\{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2}{n\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0} + o_{\mathbb{P}}(1).$$

(*ii*) The desired result follows from Lemmas A3 and A4(*i*). ■

Proof of Lemma A5: The proof of this lemma is similar to that of Lemma A1. ■

Proof of Lemma A6: (*i*) We can obtain the first-order derivative with respect to γ as follows:

$$\begin{aligned}d_n^{(1)}(1;\xi_1) = &-2\widetilde{\mathbf{P}}(\xi_1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)[\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)]^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{P}}(\xi_1) \\ &- \widetilde{\mathbf{P}}(\xi_1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)(d/d\gamma)[\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)]^{-1}\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{P}}(\xi_1).\end{aligned}$$

Note that

$$(d/d\gamma)[\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)]^{-1} = -\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}, \tag{2}$$

and that $\widetilde{\mathbf{P}}(\xi_1) = \mathbf{Y} - \xi_1\mathbf{X} = \mathbf{V}[\xi_{0*}, \xi_{1*} - \xi_1, \boldsymbol{\eta}_*']' + \mathbf{U} = \mathbf{V}\boldsymbol{\zeta}(\xi_1) + \mathbf{U}$ by assuming that $\boldsymbol{\zeta}(\xi_1) := [\xi_0, \xi_{1*} - \xi_1, \boldsymbol{\eta}_*']'$. For notational simplicity, we further suppress ξ_1 of $\boldsymbol{\zeta}(\xi_1)$. From this, it follows that since $\widetilde{\mathbf{H}}(1) = \mathbf{V}$,

$$d_n^{(1)}(1;\xi_1) = -2(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U}) + (\mathbf{V}\boldsymbol{\zeta} + \mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\ddot{\mathbf{P}}'(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U}).$$

Note that this is the same as $d_n^{(1)}(0;\xi_0)$ in Lemma A2(*i*) when we replace $\boldsymbol{\zeta}$, $\mathbf{C}_1$, and $\overline{\mathbf{K}}_1$ with $\boldsymbol{\kappa}$, $\mathbf{C}_0$, and $\mathbf{K}_1$, respectively.

(*ii*) From (*i*) and Lemmas A2 and A5, we can infer that $d_n^{(1)}(1;\xi_1) = -2(\xi_{1*} - \xi_1)\mathbf{C}_1'\mathbf{Q}_1\mathbf{U} + O_{\mathbb{P}}(n)$.

(*iii*) The second-order derivative is

$$\begin{aligned}
d_n^{(2)}(1;\xi_1) = &- 2\widetilde{\mathbf{P}}(\xi_1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1[\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)]^{-1}\overline{\mathbf{K}}_1'\widetilde{\mathbf{P}}(\xi_1)\\
&- 4\widetilde{\mathbf{P}}(\xi_1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)(d/d\gamma)[\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)]^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{P}}(\xi_1)\\
&- 2\widetilde{\mathbf{P}}(\xi_1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)[\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)]^{-1}\overline{\mathbf{K}}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{P}}(\xi_1)\\
&- \widetilde{\mathbf{P}}(\xi_1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)(d^2/d\gamma^2)[\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)]^{-1}\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{P}}(\xi_1),
\end{aligned}$$

where

$$\begin{aligned}
\frac{d^2}{d\gamma^2}[\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)]^{-1} = &-\mathbf{F}^{-1}[\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_2 + \overline{\mathbf{K}}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V} + 2\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1]\mathbf{F}^{-1}\\
&+ 2\mathbf{F}^{-1}[\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}]\mathbf{F}^{-1}[\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}]\mathbf{F}^{-1},
\end{aligned}$$

and (2) shows the specific form of $(d/d\gamma)[\widetilde{\mathbf{H}}(1)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\widetilde{\mathbf{H}}(1)]^{-1}$. By using these results and arranging the terms, we obtain

$$\begin{aligned}
d_n^{(2)}(1;\xi_1) = &- 2(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U})'\{\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1\mathbf{F}^{-1}\ddot{\mathbf{P}}' + \ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1\mathbf{F}^{-1}\overline{\mathbf{K}}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\}(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U})\\
&+ 4(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U})\\
&- 2(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U})\\
&- (\mathbf{V}\boldsymbol{\zeta} + \mathbf{U})'\ddot{\mathbf{P}}\mathbf{F}^{-1}[2\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1 + \ddot{\mathbf{P}}'\overline{\mathbf{K}}_2 + \overline{\mathbf{K}}_2'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'(\mathbf{V}\boldsymbol{\zeta} + \mathbf{U}).
\end{aligned}$$

If we reorganize the terms according to their order of convergence by applying Lemma A5 and because $\mathbb{E}[\mathbf{Z}_t U_t] = 0$, we obtain

- $-2\boldsymbol{\zeta}'\{\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1\mathbf{F}^{-1}\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\}\mathbf{V}\boldsymbol{\zeta} + 4\boldsymbol{\zeta}'[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{P}}\boldsymbol{\zeta} - 2\boldsymbol{\zeta}'[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\,\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\boldsymbol{\zeta} + 2\boldsymbol{\zeta}'[2\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_2'\ddot{\mathbf{P}} + \ddot{\mathbf{P}}'\overline{\mathbf{K}}_2]\boldsymbol{\zeta} = 2(\boldsymbol{\zeta}'\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1\boldsymbol{\zeta} - 2\boldsymbol{\zeta}'\overline{\mathbf{K}}_1'\ddot{\mathbf{P}}\mathbf{M}^{-1}\ddot{\mathbf{P}}'\boldsymbol{\zeta}) = 2(\xi_{1*} - \xi_1)^2\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1 = O_{\mathbb{P}}(n^2)$.

- $-4\boldsymbol{\zeta}'\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1\mathbf{F}^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U} + 4\boldsymbol{\zeta}'[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U} - 4\boldsymbol{\zeta}'[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} - 2\boldsymbol{\zeta}'\overline{\mathbf{K}}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U} - 2\boldsymbol{\zeta}'\ddot{\mathbf{P}}'\overline{\mathbf{K}}_2\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} = -2(\xi_{1*} - \xi_1)[\mathbf{C}_2'\mathbf{Q}_1\mathbf{U} - 2\mathbf{C}_1'\mathbf{Q}_1\overline{\mathbf{K}}_1\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} + 2\mathbf{C}_1'\ddot{\mathbf{P}}\mathbf{F}^{-1}\overline{\mathbf{K}}_1'\mathbf{Q}_1\mathbf{U}] = o_{\mathbb{P}}(n^2)$.

- $-2\mathbf{U}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1\mathbf{F}^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U} - 2\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\overline{\mathbf{K}}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U} + 4\mathbf{U}'\ddot{\mathbf{P}}(\ddot{\mathbf{P}}'\mathbf{V})^{-1}[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{U} + 2\mathbf{U}'\ddot{\mathbf{P}}\mathbf{F}^{-1}\{[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}] - \overline{\mathbf{K}}_1'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\overline{\mathbf{K}}_1 - \ddot{\mathbf{P}}'\overline{\mathbf{K}}_2\}\mathbf{F}^{-1}\ddot{\mathbf{P}}'\mathbf{U} = o_{\mathbb{P}}(n^2)$.

Therefore, we combine all these terms and obtain $d_n^{(2)}(1;\xi_1) = 2(\xi_{1*} - \xi_1)^2\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1 + o_{\mathbb{P}}(n^2)$. ■

Proof of Lemma A7: (*i*) By applying a second-order Taylor expansion to $d_n(\gamma;\beta)$ and optimizing with respect to γ, we have

$$\inf_{\gamma\in\boldsymbol{\Gamma}}\{d_n(\gamma;\beta) - d_n(1;\beta)\} = -\frac{\{d_n^{(1)}(1;\beta)\}^2}{2d_n^{(2)}(1;\beta)} + o_{\mathbb{P}}(n) = -\frac{\{\beta\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{\beta^2\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1 - \beta\mathbf{C}_2'\mathbf{Q}_1\mathbf{U}} + o_{\mathbb{P}}(n),$$

where $d_n^{(1)}(1;\beta) := (d/d\gamma)d_n(1;\beta) = -2\beta\mathbf{C}_1'\mathbf{Q}_1\mathbf{U} = O_{\mathbb{P}}(n^{3/2})$ and $d_n^{(2)}(1;\beta) := (d^2/d\gamma^2)d_n(1;\beta) = -\beta^2\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1 + \beta\mathbf{C}_2'\mathbf{Q}_1\mathbf{U} = O_{\mathbb{P}}(n^2)$. In (*ii*), we show that $\mathbf{C}_2'\mathbf{Q}_1\mathbf{U} = o_{\mathbb{P}}(n)$, so that

$$\begin{aligned}\mathcal{D}_{n,1}^{(\gamma=1;\beta)} &= -\inf_{\gamma\in\boldsymbol{\Gamma}} n^{-1}\{d_n(\gamma;\beta) - d_n(1;\beta)\}\\ &= \frac{\{n^{-3/2}\beta\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{n^{-2}(\beta^2\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1 - \beta\mathbf{C}_2'\mathbf{Q}_1\mathbf{U})} + o_{\mathbb{P}}(1) = \frac{\{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{n\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1} + o_{\mathbb{P}}(1),\end{aligned}$$

as desired.

(*ii*) We proceed with the proof in three components. First, $\mathbf{C}_1'\mathbf{Q}_1\mathbf{U} = \mathbf{C}_1'\ddot{\mathbf{Z}}(\mathbf{I} - \ddot{\mathbf{Z}}'\mathbf{V}(\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V})^{-1}\mathbf{V}'\ddot{\mathbf{Z}})\ddot{\mathbf{Z}}'\mathbf{U}$. Lemmas A5(*i*, *ii*) and Assumption 3 imply that $\mathbf{C}_1'\mathbf{Q}_1\mathbf{U} = O_{\mathbb{P}}(n^{3/2})$. Similarly, Lemma A5(*ii*) and Assumption 3 imply that $\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1 = O_{\mathbb{P}}(n^2)$. Furthermore, Lemmas A5(*ii*, *iii*, and *iv*) and Assumption 3 imply that $\mathbf{C}_2'\mathbf{Q}_1\mathbf{U} = o_{\mathbb{P}}(n^2)$. By combining all these results, we obtain $\mathcal{D}_{n,1}^{(\gamma=1;\beta)} = O_{\mathbb{P}}(1)$. ■

Proof of Lemma A8: (*i*) By applying a second-order Taylor expansion to $d_n(\gamma;\xi_1)$ and optimizing with respect to γ, we have

$$\inf_{\gamma\in\mathbf{\Gamma}}\{d_n(\gamma;\xi_1)-d_n(1;\xi_1)\} = -\frac{\{d_n^{(1)}(1;\xi_1)\}^2}{2d_n^{(2)}(1;\xi_1)} + o_{\mathbb{P}}(n) = -\frac{\{2(\xi_{1*}-\xi_1)\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{4(\xi_{1*}-\xi_1)^2\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1} + o_{\mathbb{P}}(n)$$

by Lemmas A6(*ii* and *iii*). Therefore, it follows that

$$\begin{aligned}\mathcal{D}_{n,1}^{(\gamma=1;\xi_1)} &= -\inf_{\gamma\in\mathbf{\Gamma}} n^{-1}\{d_n(\gamma;\xi_1)-d_n(1;\xi_1)\}\\ &= \frac{\{n^{-3/2}(\xi_{1*}-\xi_1)\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{n^{-2}(\xi_{1*}-\xi_1)^2\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1} + o_{\mathbb{P}}(1) = \frac{\{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{n\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1} + o_{\mathbb{P}}(1),\end{aligned}$$

as desired.

(*ii*) The desired result follows from Lemmas A7 and A8(*i*). ■

We now prove the main claims of this study.

Proof of Lemma 1: We can apply the ULLN to each row of $\{n^{-1/2}\sum_{t=1}^n X_t^\gamma \mathbf{Z}_t\}$, so that for each $j = 1, 2, \ldots, m$, we have

$$\sup_{\gamma\in\mathbf{\Gamma}}\left|n^{-1}\sum_{t=1}^n X_t^\gamma Z_{t,j} - \mathbb{E}[X_t^\gamma Z_{t,j}]\right| \xrightarrow{\mathbb{P}} 0, \tag{3}$$

where $Z_{t,j}$ is the j^{th}-row element of $\mathbf{Z}_t$. This result mainly follows from theorem 3(a) of Andrews (1992). In particular, Assumption 2 implies that $\mathbf{\Gamma}$ is totally bounded; for each j, $\mathbb{E}[|X_t^\gamma Z_{t,j}|] \leq \mathbb{E}[M_t^2] < \infty$ by Assumption 3, so that for each $\gamma \in \mathbf{\Gamma}$, the ergodic theorem holds for $n^{-1}\sum_{t=1}^n X_t^\gamma Z_{t,j}$; and finally, $X_t^{(\cdot)} Z_{t,j}$ is Lipschitz continuous because for each j,

$$|X_t^\gamma Z_{t,j} - X_t^{\gamma'} Z_{t,j}| \leq \sup_{\gamma\in\mathbf{\Gamma}}|X_t^\gamma L_t|\cdot|Z_{t,j}|\cdot|\gamma-\gamma'| \leq M_t^2|\gamma-\gamma'|, \tag{4}$$

where $M_t^2 = O_{\mathbb{P}}(1)$. These three conditions are the assumptions required for theorem 3(a) of Andrews (1992) to prove the ULLN. This also implies that $\mathbb{E}[X_t^{(\cdot)}\mathbf{V}_t]$ is continuous on $\mathbf{\Gamma}$. Note that

$$\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U} = \mathbf{X}(\gamma)'\ddot{\mathbf{Z}}[\mathbf{I} - \ddot{\mathbf{Z}}'\mathbf{V}\mathbf{F}^{-1}\mathbf{V}'\ddot{\mathbf{Z}}]\ddot{\mathbf{Z}}'\mathbf{U},$$

to obtain $\sup_{\gamma\in\mathbf{\Gamma}}|n^{-3/2}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U} - n^{-1/2}\mathbb{E}[X_t^\gamma\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\widetilde{\mathbf{Z}}'\mathbf{U}| = o_{\mathbb{P}}(1)$, because $\mathbf{M}_n \xrightarrow{\mathbb{P}} \mathbf{M}_0$ and $n^{-1}\sum_{t=1}^n \mathbf{Z}_t\mathbf{V}_t' \xrightarrow{\mathbb{P}} \mathbb{E}[\mathbf{Z}_t\mathbf{V}_t']$ by ergodicity, where $\widetilde{\mathbf{Z}} := \mathbf{M}_0^{1/2}\mathbf{Z}$. Furthermore, we can apply the CLT to $n^{-1/2}\mathbf{Z}'\mathbf{U}$, so that $n^{-1/2}\mathbf{Z}'\mathbf{U} \overset{A}{\sim} N(0,\mathbf{\Sigma})$, implying that $n^{-1/2}\mathbf{X}(\cdot)'\mathbf{Q}_1\mathbf{U} \Rightarrow \mathcal{G}(\cdot)$, where $\mathcal{G}(\cdot)$ is a Gaussian stochastic process whose covariance kernel is identical to $\kappa(\cdot,\cdot)$.

Second, we apply the ULLN to $n^{-2}\mathbf{X}(\cdot)'\mathbf{Q}_1\mathbf{X}(\cdot)$. We separate our proof into two parts. We first show that $\sup_{\gamma\in\mathbf{\Gamma}}|n^{-2}\mathbf{X}(\gamma)\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{X}-\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t'X_t^{\gamma}]|=o_{\mathbb{P}}(1)$, and then show that $\sup_{\gamma\in\mathbf{\Gamma}}|n^{-2}\mathbf{X}(\gamma)'\mathbf{Z}\mathbf{G}_n\mathbf{Z}'\mathbf{X}(\gamma)-\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t']\mathbf{G}_0\mathbb{E}[\mathbf{Z}_tX_t^{\gamma}]|=o_{\mathbb{P}}(1)$, where $\mathbf{G}_n:=\mathbf{M}_n\mathbf{Z}'\mathbf{V}\mathbf{F}^{-1}\mathbf{V}'\mathbf{Z}\mathbf{M}_n$ and $\mathbf{G}_0:=\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t\mathbf{V}_t'](\mathbb{E}[\mathbf{V}_t\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t\mathbf{V}_t'])^{-1}\mathbb{E}[\mathbf{V}_t\mathbf{Z}_t']\mathbf{M}_0$.

For the first part, we note the following triangle inequality:

$$\begin{aligned}\sup_{\gamma\in\mathbf{\Gamma}}|n^{-2}\mathbf{X}(\gamma)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{X}(\gamma)-\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t']\mathbb{E}[\widetilde{\mathbf{Z}}_tX_t^{\gamma}]| &\leq \sup_{\gamma\in\mathbf{\Gamma}}|(n^{-1}\mathbf{X}(\gamma)'\mathbf{Z}-\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t])\mathbf{M}_nn^{-1}\mathbf{Z}'\mathbf{X}(\gamma)|\\ &+\sup_{\gamma\in\mathbf{\Gamma}}|\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t](\mathbf{M}_n-\mathbf{M}_0)n^{-1}\mathbf{Z}'\mathbf{X}(\gamma)|+\sup_{\gamma\in\mathbf{\Gamma}}|\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t](n^{-1}\widetilde{\mathbf{Z}}'\mathbf{X}(\gamma)-\mathbb{E}[\widetilde{\mathbf{Z}}_tX_t^{\gamma}])|.\end{aligned}$$

$\sup_{\gamma\in\mathbf{\Gamma}}|(n^{-1}\mathbf{X}(\gamma)'\mathbf{Z})-\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t']|=o_{\mathbb{P}}(1)$ by (3), and $|\mathbf{M}_n-\mathbf{M}_0|=o_{\mathbb{P}}(1)$ by Assumption 1. Moreover, $\sup_{\gamma\in\mathbf{\Gamma}}|n^{-1}\mathbf{X}(\gamma)'\mathbf{Z}|=O_{\mathbb{P}}(1)$, by Assumption 3, ensuring that $\sup_{\gamma\in\mathbf{\Gamma}}|\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t']|=O(1)$. Thus, $\sup_{\gamma\in\mathbf{\Gamma}}|n^{-2}\mathbf{X}(\gamma)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{X}(\gamma)-\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t']\mathbb{E}[\widetilde{\mathbf{Z}}_tX_t^{\gamma}]|=o_{\mathbb{P}}(1)$.

For the second part, note that

$$\begin{aligned}\sup_{\gamma\in\mathbf{\Gamma}}|n^{-2}\mathbf{X}(\gamma)'\mathbf{Z}\mathbf{G}_n\mathbf{Z}'\mathbf{X}(\gamma)-\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t']\mathbf{G}_0\mathbb{E}[\mathbf{Z}_tX_t^{\gamma}]| &\leq \sup_{\gamma\in\mathbf{\Gamma}}|(n^{-1}\mathbf{X}(\gamma)'\mathbf{Z}-\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t])\mathbf{G}_nn^{-1}\mathbf{Z}'\mathbf{X}(\gamma)|\\ &+\sup_{\gamma\in\mathbf{\Gamma}}|\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t](\mathbf{G}_n-\mathbf{G}_0)n^{-1}\mathbf{Z}'\mathbf{X}(\gamma)|+\sup_{\gamma\in\mathbf{\Gamma}}|\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t]\mathbf{G}_0(n^{-1}\mathbf{Z}'\mathbf{X}(\gamma)-\mathbb{E}[\mathbf{Z}_tX_t^{\gamma}])|.\end{aligned}$$

Here, $\mathbf{G}_n=\mathbf{G}_0+o_{\mathbb{P}}(1)$, because $|\mathbf{M}_n-\mathbf{M}_0|=o_{\mathbb{P}}(1)$ and $n^{-1}\mathbf{Z}'\mathbf{V}=\mathbb{E}[\mathbf{Z}_t\mathbf{V}_t']+o_{\mathbb{P}}(1)$ by Assumptions 1, 3, and the ergodicity. Therefore, $\sup_{\gamma\in\mathbf{\Gamma}}|n^{-2}\mathbf{X}(\gamma)'\mathbf{Z}\mathbf{G}_n\mathbf{Z}'\mathbf{X}(\gamma)-\mathbb{E}[X_t^{\gamma}\mathbf{Z}_t']\mathbf{G}_0\mathbb{E}[\mathbf{Z}_tX_t^{\gamma}]|=o_{\mathbb{P}}(1)$, as for the first part.

From these two parts, it follows that $\sup_{\gamma\in\mathbf{\Gamma}}|n^{-2}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)-\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_tX_t^{\gamma}]|=o_{\mathbb{P}}(1)$, by noting that $\mathbf{M}_0^{1/2}\mathbf{J}_1\mathbf{M}_0^{1/2}=\mathbf{M}_0-\mathbf{G}_0$, and the desired result follows from the definition of $\sigma_1^2(\cdot)$. ■

Proof of Lemma 2: The desired result follows from Lemmas A3 and A4. Specifically, we apply the martingale CLT and continuous mapping theorem to derive the asymptotic null distribution of $\mathcal{Z}_0$. ■

Proof of Lemma 3: The desired result follows from Lemmas A7 and A8. Specifically, we apply the martingale CLT and continuous mapping theorem to derive the asymptotic null distribution of $\mathcal{Z}_1$. ■

Proof of Lemma 4: (*i*) Letting γ to converge to zero,

$$\text{plim}_{\gamma\to0}N_n^{(2)}(\gamma)=\text{plim}_{\gamma\to0}2\{(d/d\gamma)\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2+2\{\mathbf{X}(\gamma)'\mathbf{Q}_1(d/d\gamma)\mathbf{X}(\gamma)\}=2\{\mathbf{C}_0\mathbf{Q}_1\mathbf{U}\}^2,$$

because $\text{plim}_{\gamma\to 0}(d/d\gamma)\mathbf{X}(\gamma) = \mathbf{C}_0$ and $\text{plim}_{\gamma\to 0}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U} = \boldsymbol{\iota}'\mathbf{Q}_1\mathbf{U} = 0$. Furthermore,

$$\begin{aligned}\text{plim}_{\gamma\to 0}D_n^{(2)}(\gamma) =&\text{plim}_{\gamma\to 0}2n\{(d^2/d\gamma^2)\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)\}^2\\ &+\text{plim}_{\gamma\to 0}2n\{(d/d\gamma)\mathbf{X}(\gamma)'\mathbf{Q}_1(d/d\gamma)\mathbf{X}(\gamma)\} = 2n\mathbf{C}_0\mathbf{Q}_1\mathbf{C}_0,\end{aligned}$$

because $\text{plim}_{\gamma\to 0}(d/d\gamma)\mathbf{X}(\gamma) = \mathbf{C}_0$ and $\text{plim}_{\gamma\to 0}(d^2/d\gamma^2)\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U} = \mathbf{L}_2'\mathbf{Q}_1\boldsymbol{\iota} = 0$.

We now let γ to converge to 1.

$$\text{plim}_{\gamma\to 1}N_n^{(2)}(\gamma) = \text{plim}_{\gamma\to 1}2\{(d/d\gamma)\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2 + 2\{\mathbf{X}(\gamma)'\mathbf{Q}_1(d/d\gamma)\mathbf{X}(\gamma)\} = 2\{\mathbf{C}_1\mathbf{Q}_1\mathbf{U}\}^2,$$

because $\text{plim}_{\gamma\to 1}(d/d\gamma)\mathbf{X}(\gamma) = \mathbf{C}_1$ and $\text{plim}_{\gamma\to 1}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U} = \mathbf{X}'\mathbf{Q}_1\mathbf{U} = 0$. Furthermore,

$$\begin{aligned}\text{plim}_{\gamma\to 1}D_n^{(2)}(\gamma) =&\text{plim}_{\gamma\to 1}2n\{(d^2/d\gamma^2)\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)\}^2\\ &+\text{plim}_{\gamma\to 1}2n\{(d/d\gamma)\mathbf{X}(\gamma)'\mathbf{Q}_1(d/d\gamma)\mathbf{X}(\gamma)\} = 2n\mathbf{C}_1\mathbf{Q}_1\mathbf{C}_1,\end{aligned}$$

because $\text{plim}_{\gamma\to 1}(d/d\gamma)\mathbf{X}(\gamma) = \mathbf{C}_1$ and $\text{plim}_{\gamma\to 0}(d^2/d\gamma^2)\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U} = \mathbf{C}_2'\mathbf{Q}_1\mathbf{X} = 0$. ■

Proof of Theorem 1: From Lemma 4, we have

$$\sup_{\gamma\in\boldsymbol{\Gamma}}\frac{1}{n}\frac{\{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)} \geq \max\left[\frac{1}{n}\frac{\{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0}, \frac{1}{n}\frac{\{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1}\right].$$

Thus, the desired result follows from Lemmas 1, 2, and 3. ■

Proof of Theorem 2: (*i*) For notational simplicity, for each $\gamma \in \boldsymbol{\Gamma}$, we assume that $\mathbf{g}(\gamma) := \mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]$ and $\mathbf{h} := \mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)]$. Note that from (9), it follows that

$$d_0 - d(\gamma) = \left\{\frac{\mathbf{h}'\mathbf{g}(\gamma)}{\sqrt{\mathbf{h}'\mathbf{h}}\sqrt{\mathbf{g}(\gamma)'\mathbf{g}(\gamma)}}\right\}^2(\mathbf{h}'\mathbf{h}),$$

so that $d_0 - d(\cdot) \geq 0$. Therefore, if $\sup_{\gamma\in\boldsymbol{\Gamma}}(d_0 - d(\gamma)) = 0$, it implies that $c(\cdot) := \langle \mathbf{h}, \mathbf{g}(\cdot)\rangle \equiv 0$.

We prove the given claim by contradiction. Now, assume that $c(\cdot) \equiv 0$ on $\boldsymbol{\Gamma}$. From the condition that $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] \neq \mathbf{0}$, it follows that $\mathbf{h} \neq \mathbf{0}$, and so $\mathbf{g}(\cdot) \equiv \mathbf{0}$ from the assumption that $c(\cdot) \equiv 0$ and $\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}] \equiv \mathbf{0}$. If we assume that $M(\cdot,\cdot)$ denotes the moment generating function of $(\log(X_t), \widetilde{\mathbf{Z}}_t')'$, so that $M(\gamma, \boldsymbol{\tau}) := \mathbb{E}[\exp(\gamma\log(X_t) + \boldsymbol{\tau}'\widetilde{\mathbf{Z}}_t)]$, then for each γ, $\mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t] = \nabla_{\boldsymbol{\tau}}M(\gamma,\boldsymbol{\tau})|_{\boldsymbol{\tau}=\mathbf{0}}$, so that $\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}] \equiv \mathbf{0}$ implies that $\mathbb{E}[\widetilde{\mathbf{Z}}_t|\log(X_t)] = \mathbf{0}$ with probability 1 by applying theorem 1 of Bierens (1982) to the moment generating function. Note that $\log(\cdot)$ is a one-to-one mapping from $\mathbb{R}^+$ to $\mathbb{R}$, so that it is a measure preserving transformation. This implies that $\mathbb{E}[\widetilde{\mathbf{Z}}_t|X_t] = \mathbf{0}$ with probability 1. We may multiply $m(X_t)$ to each side and apply the law of iterated expectation: $\mathbb{E}[m(X_t)\mathbb{E}[\widetilde{\mathbf{Z}}_t|X_t]] = \mathbb{E}[m(X_t)\widetilde{\mathbf{Z}}_t] = \mathbf{0}$. Note that this is a contradiction to the condition that $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] = \mathbf{0}$. Therefore, for some $\widetilde{\gamma}$, $c(\widetilde{\gamma}) \neq 0$, and this implies that $d_0 - d(\widetilde{\gamma}) > 0$.

(*ii*) Because $d_n(\beta,\gamma) = (\mathbf{Y} - \beta\mathbf{X}(\gamma))'\mathbf{Q}_1(\mathbf{Y} - \beta\mathbf{X}(\gamma))$ and $\mathbf{Y} = \mathbf{V}\boldsymbol{\varsigma}_* + n^{-1/2}\mathbf{s} + \mathbf{U}$, where $\mathbf{s} := (s(X_1),\ldots,s(X_n))'$, we have

$$\mathcal{D}_{n,1} = \sup_{\gamma\in\mathbf{\Gamma}} \frac{\{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{Y}\}^2}{n\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)} = \sup_{\gamma\in\mathbf{\Gamma}} \frac{\{n^{-2}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{s} + n^{-3/2}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U})\}^2}{n^{-2}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)}.$$

From Lemma 1, we have $n^{-3/2}\mathbf{X}(\cdot)'\mathbf{Q}_1\mathbf{U} \Rightarrow \mathcal{G}(\cdot)$ and $\sup_{\gamma\in\mathbf{\Gamma}} |n^{-2}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma) - \sigma_1^2(\gamma)| \overset{\mathbb{P}}{\to} 0$, where $\sigma_1^2(\gamma) := \mathbb{E}[X_t^\gamma\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t s(X_t)]$. Note that $n^{-2}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{s} = n^{-2}\mathbf{X}(\gamma)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{s} - n^{-2}\mathbf{X}(\gamma)'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}\mathbf{F}^{-1}\mathbf{V}'\ddot{\mathbf{Z}}$ $\ddot{\mathbf{Z}}'\mathbf{s}$. In the proof of Lemmas 1 and A1, we saw that $\sup_{\gamma\in\mathbf{\Gamma}} |n^{-1}\mathbf{X}(\gamma)'\,\mathbf{Z} - \mathbb{E}[X_t^\gamma\mathbf{Z}_t]| \overset{\mathbb{P}}{\to} \mathbf{0}$ and $n^{-1}\mathbf{V}'\mathbf{Z} \overset{\mathbb{P}}{\to}$ $\mathbb{E}[\mathbf{V}_t\mathbf{Z}_t']$. Furthermore, if we apply the ergodic theorem, $n^{-1}\mathbf{Z}'\mathbf{s} \overset{\mathbb{P}}{\to} \mathbb{E}[\mathbf{Z}_t s(X_t)]$ by the moment condition that $\mathbb{E}[s^2(X_t)] < \infty$. Thus, we have $\sup_{\gamma\in\mathbf{\Gamma}} |n^{-2}\mathbf{Z}(\gamma)'\mathbf{Q}_1\mathbf{s} - \mathbb{E}[X_t^\gamma\widetilde{\mathbf{Z}}_t]\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t s(X_t)]| \overset{\mathbb{P}}{\to} \mathbf{0}$. Therefore, it follows that

$$\mathcal{D}_{n,1} \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}} \frac{\{\mathbb{E}[X_t^\gamma\widetilde{\mathbf{Z}}_t]\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t s(X_t)] + \mathcal{G}(\gamma)\}^2}{\sigma_1^2(\gamma)} = \sup_{\gamma\in\mathbf{\Gamma}}\{\nu_1(\gamma) + \mathcal{Z}_1(\gamma)\}^2$$

by the definitions of $\nu_1(\cdot) := \mathbb{E}[X_t^{(\cdot)}\widetilde{\mathbf{Z}}_t]\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t s(X_t)]/\sigma_1(\cdot)$ and $\mathcal{Z}_1(\cdot) := \mathcal{G}(\cdot)/\sigma(\cdot)$. This completes the proof. ■

Proof of Theorem 3: (*i*) This is obvious from Corollary 1.

(*ii*) For the given claim, note that $\lim_{n\to\infty}\mathbb{P}(\widehat{q}_n > q_*) = \lim_{n\to\infty}\alpha_n = 0$ by the given condition. Furthermore, for any $q < q_*$, if $cv_q(\alpha_n) = o(n)$, then $\lim_{n\to\infty}\mathbb{P}(\mathcal{D}_{n,q} > cv_q(\alpha_n)) = 1$, implying that the desired result follows if $cv_q(\alpha_n)) = o(n)$. We show this as follows.

First, note that $\sup_{\gamma\in\mathbf{\Gamma}(\bar{q})}\mathcal{Z}_q^2(\gamma) \leq \sup_{\gamma\in\mathbf{\Gamma}(\bar{q})}\max^2[0,\mathcal{Z}_q(\gamma)] + \sup_{\gamma\in\mathbf{\Gamma}(\bar{q})}\min^2[0,\mathcal{Z}_q(\gamma)]$. This implies that for any $u > 0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{\gamma\in\mathbf{\Gamma}(\bar{q})}\mathcal{Z}_q^2(\gamma) \geq u^2\right) &\leq \mathbb{P}\left(\sup_{\gamma\in\mathbf{\Gamma}(\bar{q})}\mathcal{Z}_q(\gamma) \geq \frac{u}{\sqrt{2}}\right) + \mathbb{P}\left(\inf_{\gamma\in\mathbf{\Gamma}(\bar{q})}\mathcal{Z}_q(\gamma) \leq -\frac{u}{\sqrt{2}}\right)\\ &= 2\mathbb{P}\left(\sup_{\gamma\in\mathbf{\Gamma}(\bar{q})}\mathcal{Z}_q(\gamma) \geq \frac{u}{\sqrt{2}}\right)\end{aligned}$$

from the inequality in the proof of theorem 2 of Cho and Phillips (2018). We further note that Borel's inequality (*e.g.*, Piterbarg, 1996, p. 13) implies that

$$\mathbb{P}\left(\sup_{\gamma\in\mathbf{\Gamma}^{(\bar{q})}} \mathcal{Z}_q(\gamma) \geq \frac{u}{\sqrt{2}}\right) \leq 2\Psi\left(\frac{u/\sqrt{2}-a_q}{\sigma_q}\right),$$

and so it follows that

$$\mathbb{P}\left(\sup_{\gamma\in\mathbf{\Gamma}^{(\bar{q})}} \mathcal{Z}_q^2(\gamma) \geq u^2\right) \leq 4\Psi\left(\frac{u/\sqrt{2}-a_q}{\sigma_q}\right) \leq 2\exp\left(-\frac{u^2-2\sqrt{2}ua_q+a_q^2}{4\sigma_q^2}\right)$$

from the fact that $\Psi(\cdot) \leq \frac{1}{2}\exp(-(\cdot)^2/2)$. We now let the left-hand side of this inequality and u^2 to be α_n and $cv_q(\alpha_n)$, respectively. Then, it follows that

$$-\frac{\log(\alpha_n)}{n} \geq \frac{1}{n}\left(\frac{a_q^2}{4\sigma_q^2}-\log(2)\right)+\frac{1}{4\sigma_q^2}\left(\frac{cv_q(\alpha_n)}{n}\right)-\frac{a_q}{\sqrt{2}\sigma_q^2}\left(\frac{cv_q(\alpha_n)}{n^2}\right)^{1/2}.$$

Note that $n^{-1}(a_q^2/(4\sigma_q^2)-\log(2)) \to 0$, and the sum of the last terms is greater than zero, provided that $cv_q^{1/2}(\alpha_n) > 2\sqrt{2}a_q$ and is achieved as $\alpha_n \to 0$. Furthermore, the given condition implies that $-\log(\alpha_n)/n \to 0$, so that

$$\frac{1}{4\sigma_q^2}\left(\frac{cv_q(\alpha_n)}{n}\right)-\frac{a_q}{\sqrt{2}\sigma_q^2}\left(\frac{cv_q(\alpha_n)}{n^2}\right)^{1/2}=o(1).$$

Therefore, it follows that $cv_q(\alpha_n)=o(n)$, as desired. ■

References

AI, C. AND CHEN, X. (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

ANDREWS, D. W. K. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation,“ *Economietrica*, 62, 543–564.

ANDREWS, D. W. K., STOCK, J., AND SUN, L. (2019): "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

ANGRIST, J. AND KRUEGER, A. (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 106, 979–1014.

AKAIKE, H. (1973): "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second Internation Symposium on Information Theory*. Eds. B.N. Petrov and F. Csake. Budapest: Akademiai Kiado, 267–281.

BALASSA, B. (1964): "The Purchasing Power Parity Doctrine: A Reappraisal," *Journal of Political Economy*, 72, 584–596.

BAEK, Y.I., CHO, J.S., AND PHILLIPS, C.B. (2015): "Testing Linearity Using Power Transforms of Regressors," *Journal of Econometrics*, 187, 376–384.

BIERENS, H. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

BIERENS, H. (1990): "A Consistent Conditional Moment Test of Functional Forms," *Econometrica*, 58, 1443–1458.

CARD, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. Eds. L.N. Christofides, E.K. Grant, and R. Swidinsky. Toronto: University of Toronto Press, 1995.

CARD, D. AND KRUEGER, A. (1992): "Does School Quality Matter? Return to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, 100, 1–40.

CHEN, X. AND LIU, Z. (2014): "Sieve M Inference on Irregular Parameters," *Journal of Econometrics*, 182, 70–86.

CHEN, X. AND POUZO, D. (2015): "Sieve Wald and QLR Inferences Semi/Nonparametric Conditional Moment Models," *Econometrica*, 83, 1013–1079.

CHO, J. S., CHEONG, T. U., AND WHITE, H. (2011): "Experience with the Weighted Bootstrap in Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Economic Theory and Econometrics*, 22:2, 60–91

CHO, J.S. AND ISHIDA, I. (2012): "Testing for the Effects of Omitted Power Transformations," *Economic Letters*, 117, 287–290.

CHO, J.S. AND PHILLIPS, P.C.B. (2018): "Sequentially Testing Polynomial Model Hypothesis Using the Power Transform of Regressors," *Journal of Applied Econometrics*, 33, 141–159.

DAVIES, R. (1977): "The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics*, 114, 361–394.

DAVIES, R. (1987): "Hypothesis Testing When a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 74, 33–43.

GRILICHES, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45, 1–22.

HALL, A. AND INOUE, A. (2003): "Inference When a Nuisance Parameter is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

HALL, P. AND HOROWITZ, J. (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64, 891–916.

HANNAN, E. AND QUINN, B. (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society*, Series B, 41, 190–195.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, B. (1996): "Inference When a Nuisance Parameter is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

HONG, Y. AND WHITE, H. (1995): "Consistent Specification Testing Via Nonparametric Series Regression," *Econometrica*, 63, 1133–1159.

MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281–302.

MINCER, J. (1997): "Changes in Wage Inequality, 1970–1990," *Research in Labor Economics*, 16, 1–18.

NEWEY, W.K. (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229–256.

NEWEY, W.K. AND POWELL, J.L. (2003): "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71, 1557–1569.

SAMUELSON, P.A. (1964): "Theoretical Notes on Trade Problems," *Review of Economics and Statistics*, 46, 145–164.

SARGAN, J. D. (1958): “The Estimation of Economic Relationships Using Instrumental Variables,” *Econometrica*, 26, 393–415.

SARGAN, J. D. (1988): *Lectures on Advanced Econometric Theory*. Oxford: Basil Blackwell.

SCHWARZ, G. E. (1978): “Estimating the dimension of a model,” *Annals of Statistics*, 6, 461–464.

SCOTT, D. J. (1973): “Central Limit Theorems for Martingales and for Processes with Stationary Increments Using a Skorokhod Representation Approach,” *Advances in Applied Probability*, 5, 119–137.

STAIGER, D. AND STOCK, J. (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65, 557–586.

STOCK, J. AND YOGO, M. (1997): “Testing for Weak Instruments in Linear IV Regression,” in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Eds. D. W. K. Andrews and J. H. Stock, Chapter 5, pp. 80–108. Cambridge: Cambridge University Press.

PITERBARG, V. (1996): *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. Translations of Mathematical Monographs, 148. Providence: American Mathematical Society.

WALD, A. (1943): “Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large,” *Transactions of the American Mathematical Society*, 54, 426–482.

WHITE, H. (1980): “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Tests for Heteroskedasticity,” *Econometrica*, 48, 817–838.

WILLIS, R. (1986): “Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions,” *Handbook of Labor Economics*. Eds. O. Ashenfelter and R. Layard, Vol. 1, Chapter 10, 525–602. Amsterdam: Elseviwr Science Publishers BV.

Sets	α	Test Stat.	$q \setminus n$	100	200	300	400	500	600	700	800	900	1,000
A	10%	$\mathcal{D}_n$	1	19.23	7.20	3.37	0.90	0.77	0.13	0.27	0.13	0.10	0.03
			2^*	67.70	80.23	84.33	85.70	87.20	87.07	86.70	87.43	86.87	87.10
			3	11.07	10.33	10.17	11.03	9.83	10.60	11.17	10.10	10.70	10.13
			≥ 4	2.00	2.23	2.13	2.37	2.20	2.20	1.87	2.33	2.33	2.73
		$\mathcal{J}_n$	1	32.53	15.63	7.33	2.97	1.60	0.63	0.40	0.47	0.30	0.10
			2^*	59.40	76.47	84.10	87.87	89.53	89.70	89.70	90.43	90.17	90.27
			3	4.87	4.27	4.57	5.03	4.53	5.37	5.10	4.23	5.13	4.77
			≥ 4	3.20	3.63	4.00	4.13	4.33	4.30	4.80	4.87	4.40	4.87
	5%	$\mathcal{D}_n$	1	25.30	10.13	4.77	1.60	0.97	0.33	0.33	0.20	0.17	0.03
			2^*	67.57	84.00	88.57	91.00	93.43	92.97	92.63	93.17	92.93	93.20
			3	6.60	5.10	5.90	6.60	5.13	5.67	6.27	5.60	6.00	5.80
			≥ 4	0.53	0.77	0.77	0.80	0.47	1.03	0.77	1.03	0.90	0.97
		$\mathcal{J}_n$	1	38.33	20.03	11.03	5.03	2.57	1.40	0.97	0.97	0.50	0.23
			2^*	57.93	76.33	84.53	89.70	93.43	94.20	94.37	94.53	94.97	95.10
			3	2.47	2.27	2.23	3.20	2.07	2.40	2.67	2.37	2.70	2.50
			≥ 4	1.27	1.37	2.20	2.07	1.93	2.00	2.00	2.13	1.83	2.17
	1%	$\mathcal{D}_n$	1	37.30	19.13	9.50	4.27	2.37	1.20	0.80	0.87	0.47	0.20
			2^*	61.03	79.20	88.97	94.00	96.50	97.13	97.73	97.77	97.93	98.27
			3	1.60	1.67	1.50	1.63	1.07	1.53	1.40	1.20	1.40	1.47
			≥ 4	0.07	0.00	0.03	0.10	0.07	0.13	0.07	0.17	0.20	0.07
		$\mathcal{J}_n$	1	50.97	31.63	19.70	10.77	6.07	4.07	2.23	1.80	1.20	0.63
			2^*	48.23	67.47	79.43	88.27	93.07	94.87	96.87	97.27	97.80	98.30
			3	0.53	0.63	0.47	0.60	0.57	0.70	0.57	0.53	0.70	0.70
			≥ 4	0.27	0.27	0.40	0.37	0.30	0.37	0.33	0.40	0.30	0.37
B	10%	$\mathcal{D}_n$	1	19.10	7.60	3.90	1.20	0.87	0.27	0.30	0.30	0.17	0.03
			2^*	66.67	79.17	82.97	84.80	86.03	86.00	85.97	86.63	87.27	87.97
			3	11.07	10.93	10.20	10.77	10.23	10.73	11.10	10.00	10.13	9.57
			≥ 4	2.87	2.30	2.93	3.23	2.87	3.00	2.63	3.07	2.43	2.43
		$\mathcal{J}_n$	1	32.80	16.70	8.50	3.93	2.00	1.00	0.67	0.70	0.40	0.13
			2^*	59.03	75.87	82.87	85.77	88.83	89.70	89.80	90.00	90.27	90.20
			3	4.13	3.80	3.87	4.50	3.47	4.63	4.03	3.87	4.23	4.33
			≥ 4	4.03	3.63	4.77	5.80	5.70	4.67	5.50	5.43	5.10	5.33
	5%	$\mathcal{D}_n$	1	25.30	10.97	5.20	1.93	1.20	0.47	0.40	0.33	0.27	0.10
			2^*	67.40	82.63	87.40	90.40	92.53	91.70	92.27	92.27	93.10	93.03
			3	6.20	5.47	6.00	6.50	5.33	6.50	6.17	6.47	5.87	6.00
			≥ 4	1.10	0.93	1.40	1.17	0.93	1.33	1.17	0.93	0.77	0.87
		$\mathcal{J}_n$	1	37.83	21.43	11.93	5.77	3.20	1.90	1.33	1.00	0.70	0.27
			2^*	58.07	74.93	83.77	89.07	92.53	93.53	94.20	94.73	95.00	95.10
			3	2.27	1.73	1.87	2.87	1.97	2.13	2.10	1.87	2.33	2.13
			≥ 4	1.83	1.90	2.43	2.30	2.30	2.43	2.37	2.40	1.97	2.50
	1%	$\mathcal{D}_n$	1	35.03	18.50	9.63	4.50	2.40	1.57	1.00	0.83	0.53	0.20
			2^*	62.93	79.63	88.70	93.43	96.27	96.63	97.30	97.57	97.73	97.87
			3	1.83	1.67	1.43	1.97	1.20	1.67	1.47	1.50	1.63	1.90
			≥ 4	0.20	0.20	0.23	0.10	0.13	0.13	0.23	0.10	0.10	0.03
		$\mathcal{J}_n$	1	47.67	31.00	20.03	11.63	6.77	4.37	2.50	2.10	1.40	0.73
			2^*	51.47	68.33	79.23	87.50	92.20	94.53	96.93	97.17	97.67	98.07
			3	0.43	0.33	0.27	0.47	0.73	0.60	0.30	0.20	0.60	0.70
			≥ 4	0.43	0.33	0.47	0.40	0.30	0.50	0.27	0.53	0.33	0.50

Table 1: ESTIMATED POLYNOMIAL DEGREE BY THE DD- AND J-TEST STATISTICS (IN PERCENTAGE). Number of Replications: 3,000. This table shows the estimated polynomial degrees by sequentially applying the DD- and J-test statistics when the significance levels are fixed irrespective of sample size. The true polynomial equation degree is 2, as indicated by the asterisks (*). DGP: $Y_t = 1 + D_t + X_t + X_t^2 + U_t$, $U_t = G_t$, $X_t := \exp(\frac{1}{2}(\mathbf{W}_t'\boldsymbol{\iota}_2 + G_t))$, and $(D_t, G_t, \mathbf{W}_t')' \sim$ IID $N(\mathbf{0}_4, \mathbf{I}_4)$. Model: $\mathcal{M}_q' := \{m_{t,q}(\boldsymbol{\omega}^{(q)}) := Y_t - \xi_0 - X_t\xi_1 - \ldots - X_t^q\xi_q - D_t\eta - \beta X_t^\gamma : \boldsymbol{\omega}^{(q)} \in \boldsymbol{\Omega}^{(q)}\}$ with $q \in I(3)$, $\boldsymbol{\omega}^{(q)} := (\xi_0, \ldots, \xi_q, \eta, \beta, \gamma)'$, and $\Gamma := [-0.25, 3.5]$. Sets A and B assume that $\mathbf{Z}_t := (1, D_t, \mathbf{W}_t, W_{1,t}^2, W_{t,2}^2, W_{t,1}^3)'$ and $\mathbf{Z}_t := (1, D_t, \mathbf{W}_t, W_{1,t}^2, W_{t,2}^2, W_{t,1}^3, W_{t,2}^3)'$, respectively.

Sets	Methods	Test Stat. $\backslash$ n	100	200	300	400	500	600	700	800	900	1,000
A	Seq. Est.	$\mathcal{D}_n$	**67.70**	**83.07**	87.93	91.00	93.83	93.93	94.40	95.67	95.00	95.47
	with	$\mathcal{J}_n$	59.40	76.63	84.93	89.70	93.70	94.63	95.23	95.83	96.53	96.63
	$\alpha_n = n^{-1/2}$	(Hypo. Rate)	(90.00)	(92.92)	(94.22)	(95.00)	(95.52)	(95.91)	(96.22)	(96.46)	(96.66)	(96.83)
	Seq. Est.	$\mathcal{D}_n$	67.23	82.17	**89.40**	**94.00**	**96.57**	**97.40**	**97.83**	**98.30**	98.43	98.90
	with	$\mathcal{J}_n$	55.77	70.97	80.87	88.73	93.03	94.67	96.77	97.20	97.87	98.63
	$\alpha_n = n^{-3/4}$	(Hypo. Rate)	(96.83)	(98.11)	(98.61)	(98.88)	(99.05)	(99.17)	(99.26)	(99.33)	(99.39)	(99.43)
	Seq. Est.	$\mathcal{D}_n$	61.03	76.50	85.70	92.32	94.87	96.30	97.90	98.20	**98.63**	**99.17**
	with	$\mathcal{J}_n$	48.23	63.60	73.70	83.40	89.00	91.50	95.07	95.67	97.17	98.20
	$\alpha_n = n^{-1}$	(Hypo. Rate)	(99.00)	(99.50)	(99.66)	(99.75)	(99.80)	(99.83)	(99.85)	(99.87)	(99.88)	(99.90)
	Akaike-MSC		64.03	74.70	76.90	78.13	77.53	78.20	77.83	78.83	78.03	79.43
	Bayesian-MSC		65.50	81.20	89.27	93.20	95.67	96.37	97.13	97.40	97.47	97.93
	Hannan-Quinn-MSC		67.10	81.80	86.53	88.80	90.63	90.80	90.70	92.43	91.40	92.63
B	Seq. Est.	$\mathcal{D}_n$	66.67	**81.70**	86.70	90.40	93.20	92.63	94.00	94.47	94.57	95.17
	with	$\mathcal{J}_n$	59.03	75.87	83.67	89.07	92.83	93.87	95.07	95.77	96.03	96.57
	$\alpha_n = n^{-1/2}$	(hypo. rate)	(90.00)	(92.92)	(94.22)	(95.00)	(95.52)	(95.91)	(96.22)	(96.46)	(96.66)	(96.83)
	Seq. Est.	$\mathcal{D}_n$	67.20	81.57	**89.23**	**93.37**	**96.17**	**96.93**	**97.53**	**97.93**	98.17	98.87
	with	$\mathcal{J}_n$	56.77	71.00	80.20	87.87	92.00	94.23	96.73	97.23	97.80	98.37
	$\alpha_n = n^{-3/4}$	(hypo. rate)	(96.83)	(98.11)	(98.61)	(98.88)	(99.05)	(99.17)	(99.26)	(99.33)	(99.39)	(99.43)
	Seq. Est.	$\mathcal{D}_n$	62.93	77.57	85.87	92.10	94.57	96.17	97.80	98.03	**98.50**	**99.17**
	with	$\mathcal{J}_n$	51.47	65.17	74.97	83.40	89.03	91.63	95.00	95.67	97.33	98.13
	$\alpha_n = n^{-1}$	(hypo. rate)	(99.00)	(99.50)	(99.66)	(99.75)	(99.80)	(99.83)	(99.85)	(99.87)	(99.88)	(99.90)
	Akaike-MSC		63.50	73.33	76.57	77.10	77.73	77.23	77.07	78.83	78.67	78.93
	Bayesian-MSC		66.20	80.63	88.90	92.57	95.60	95.73	97.17	97.43	97.20	97.33
	Hannan-Quinn-MSC		**67.47**	80.70	85.20	88.47	90.37	90.10	90.47	91.13	91.77	91.73

Table 2: PRECISION RATES OF THE DD- AND J-SEQUENTIAL TESTING PROCEDURES AND MSCS (IN PERCENTAGE). Number of Replications: 3,000. This table shows the correctly estimated polynomial degree percentages obtained using the DD- and J-sequential testing procedures and MSCs. The figures in parentheses denote $(1 - \alpha_n) \times 100$, where $\alpha_n = 1/n^{1/2}$, $1/n^{3/4}$, and $1/n$, with the best performing result for each sample size indicated in boldface; this is fixed irrespective of sample size. The true polynomial equation degree is 2, as indicated by asterisks (*). DGP: $Y_t = 1 + D_t + X_t + X_t^2 + U_t$, $U_t = G_t$, $X_t := \exp(\frac{1}{2}(\mathbf{W}_t'\boldsymbol{\iota}_2 + G_t))$, and $(D_t, G_t, \mathbf{W}_t')' \sim$ IID $N(\mathbf{0}_4, \mathbf{I}_4)$. Model: $\mathcal{M}_q' := \{m_{t,q}(\boldsymbol{\omega}^{(q)}) := Y_t - \xi_0 - X_t\xi_1 - \ldots - X_t^q\xi_q - D_t\eta - \beta X_t^\gamma : \boldsymbol{\omega}^{(q)} \in \boldsymbol{\Omega}^{(q)}\}$ with $q \in I(3)$, $\boldsymbol{\omega}^{(q)} := (\xi_0, \ldots, \xi_q, \eta, \beta, \gamma)'$, and $\Gamma := [-0.25, 3.5]$. Sets A and B assume that $\mathbf{Z}_t := (1, D_t, \mathbf{W}_t, W_{1,t}^2, W_{t,2}^2, W_{t,1}^3)'$ and $\mathbf{Z}_t := (1, D_t, \mathbf{W}_t, W_{1,t}^2, W_{t,2}^2, W_{t,1}^3, W_{t,2}^3)'$, respectively.

Dependent Variable	Degree (q) \ Test	$\mathcal{J}_{n,q}$	$\mathcal{D}_{n,q}$	MSC	Effective F-statistic
log(wage) in 1976	1	**0.4733**	**0.4680**	$-\mathbf{0.0069}$	9.173
		(0.370)	(0.134)	-	(10%)
	2	0.0055	0.0001	-0.0053	
		(0.976)	(0.998)	-	
log(wage) in 1978	1	**0.6810**	**0.4529**	$-\mathbf{0.0083}$	8.917
		(0.198)	(0.112)	-	(10%)
	2	0.3480	0.2249	-0.0060	
		(0.225)	(0.806)	-	

Table 3: APPLICATION OF THE DD- AND J-SEQUENTIAL TESTING PROCEDURES AND MSCS TO CARDS (1995) DATA SET. This table shows the J- and DD-sequential testing procedure to Card's (1995) data set. The model controls are the location dummy variables, family characteristics, experience, and the squared of experience. The parameter space for γ is $[0.5, 3.5]$ for all specifications. The figures in parentheses denote the test statistics p-values. Conditional heteroskedasticity is assumed to estimate $\mathbb{E}[U_t^2\mathbf{Z}_t\mathbf{Z}_t']$; we obtained the critical values of the test statistics through simulation based on the estimated covariance matrix. The simulation was replicated 500 times. The polynomial degree selected by each method is denoted in boldface font.

Dependent Variable	log(wage) in 1976		log(wage) in 1978	
Estimation	OLS (1)	TSLS (2)	OLS (3)	TSLS (4)
Education	0.074^{***}	0.144^{***}	0.070^{***}	0.151^{***}
	(0.004)	(0.035)	(0.004)	(0.036)
Experience	0.086^{***}	0.111^{***}	0.070^{***}	0.099^{***}
	(0.007)	(0.014)	(0.007)	(0.015)
Experience-Squared	-0.236^{***}	-0.230^{***}	-0.215^{***}	-0.211^{***}
/100	(0.032)	(0.035)	(0.035)	(0.039)
Black	-0.189^{***}	-0.144^{***}	-0.196^{***}	-0.149^{***}
Indicator	(0.019)	(0.029)	(0.020)	(0.029)
Live in South	-0.141^{***}	-0.102^{**}	-0.097^{**}	-0.042
	(0.033)	(0.039)	(0.035)	(0.045)
Live in SMSA	0.161^{***}	0.126^{***}	0.170^{***}	0.128^{***}
	(0.015)	(0.023)	(0.017)	(0.026)
Single Mother	0.011	0.003	0.002	0.009
Family Indicator	(0.029)	(0.032)	(0.031)	(0.034)
Both Parents' Education Below	–0.012	0.078	–0.022	0.087
High School	(0.020)	(0.049)	(0.022)	(0.053)
Regional dummies	Yes	Yes	Yes	Yes
n	3,010	3,010	2,438	2,438
Adj R^2	0.29	0.20	0.29	0.17

Table 4: ESTIMATION RESULTS OF THE ORIGINAL CARD'S (1995) MODEL BY OLS AND TSLS METHODS. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Robust standard errors are provided in parentheses. The dependent variables are the log hourly wages in 1976 and 1978 as shown in Columns (1)∼(2) and (3)∼(4), respectively. In Columns (2) and (4), we employ the instrumental variables constructed using the proximity variable of living near four-year public or private college and their interaction with the dummy variable indicating the single-mother family structure.