Sequentially Estimating the Structural Equation by Power Transformation

JAEDO CHOI

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

Email: jaedohi@umich.edu

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

Email: moonr@usc.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

November, 2021.

Abstract

This study provides an econometric methodology to test a linear structural relationship among economic variables. For this, we propose the so-called distance-difference (DD) test and show that it has omnibus power against arbitrary nonlinear structural relationships. If the DD-test rejects the linear model hypothesis, a sequential testing procedure assisted by the DD-test 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 and moment selection criteria. Finally, we provide an empirical illustration by investigating the relationship between the value-added and its production factors using firm-level data from the United States. By our methodology, we affirm that the production function has exhibited a factor-biased technological change instead of Hicks-neutral technology presumed by Cobb-Douglas production function.

Key Words: GMM estimation; Model linearity testing; Model specification testing; Gaussian stochastic process; Sequential testing procedure; Factor-biased technological change.

Subject Classification: C12, C13, C26, C52, O14.

Acknowledgements: The Co-Editor, Patrik Guggenberger, and two anonymous referees provided very helpful comments for which we are most grateful. The authors are also benefited from discussions 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).

* Corresponding author.

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 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 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 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 a linear structural model hypothesis using a distance-difference (DD) test as in Baek, Cho, and Phillips (2015). We then derive the null limit distribution of the DD-test and show that it has omnibus power against a nonlinear structural model.

Cho and Phillips (2018) further developed a sequential testing procedure using the QLR-test to consistently estimate a nonlinear reduced-form equation. In this study, we apply the sequential testing procedure to the DD-test as in Cho and Phillips (2018) and show that the unknown polynomial structural model can be consistently estimated using the current 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 approximate structural equation.

We also compare our testing procedures with some widely used ones in the literature. We first consider Horowitz's (2006) and Breunig's (2015) specifications tests in addition to Sargan (1958, 1988) and Hansen's (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 four tests and find that they can complement each other. In particular, our simulation suggests that the DD-sequential testing procedure selects the correct model more often than the J-sequential testing when the sample size is relatively small. Second, we investigate Andrews's (1999) procedure 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) with the sequential testing procedure using simulations, finding that the sequential testing and MSCs can supplement each other.

In the semi/nonparametric literature, studies such as Hong and White (1995), Ai and Chen (2003), Newey and Powell (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 is fully parametric. In addition, the DD-test can be used as a diagnostic test before applying the semi/nonparameteric methodologies. If the null model is not rejected by the DD-test, no need would arise to estimate the structural model using semi/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 testing the linearity condition by formally introducing the DD-test. The null limit distribution and power properties of the test are also examined. The same section also examines the asymptotic size of the DD-test and the influence of weak instrumental variables to the DD-test. Section 3 extends the linear structure testing to polynomial structures using the same test. Furthermore, the sequential testing is applied to estimate the polynomial structural equation. Section 4 reports Monte Carlo simulations and compares the DD-test with other methodologies, while Section 5 presents an empirical illustration. Finally, Section 6 concludes the paper. Mathematical proofs and other supplementary information are presented in Appendix.

Before moving to the next section, we introduce some useful mathematical notations. For functions $f(\cdot)$ and $j = 1, 2, \ldots$, let $(d^j/d^j x)f(\bar{x})$ 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, where n denotes the sample size throughout the paper.

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 a 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, viz., $m(\cdot)$, 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, and Balassa (1964) and Samuelson's (1964) linear structural model between the ratio of purchasing power parity to exchange rate and the per capita income differentials. As another example, a simple log-linear production function is linear with respect to log production factors. If it is subject to input bias problem, the model turns out to contain an endogenous variable. We detail this in Section 5 for our empirical illustration.

We are motivated to test the null hypothesis from the possibility that the linear model is arbitrarily misspecified. The following examples specifically illustrate our motivation. As our first 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 terms of the log-linear production function model subject to input bias problem, prior literature attempts to explain the recent large distributional consequences of factor shares by attributing them to the change of production technology (*e.g.*, Krusell et al., 2000; Antràs, 2004; Karabarbounis and Neiman, 2014; Piketty, 2014; Acemoglu and Restrepo, 2018; Raval, 2019; Oberfield and Raval, 2021). Nevertheless, if the production technology exhibits a log-linear production function, factor shares always remain constant, so that the typically assumed Cobb-Douglas technology must be misspecified. 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.

In this study, we aim to provide a test methodology that consistently detects arbitrary nonlinearity rather well and overcomes the challenges associated with sieve series estimation, and thus leads to a simple and straightforward testing procedure. Specifically, we first extend the approach of Bierens (1990) and Baek, Cho, and Phillips (2015), who estimate $m(\cdot)$ when X_t is exogenous.

To this end, we first 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 and test the linearity hypothesis. That is, we compare Sargan (1958, 1988) and Hansen's (1982) J-tests implied by $\mathcal{M}$ and the linear model, and reject the linearity hypothesis if the difference between the two J-tests is sufficiently large. We formally define our test below; this is the DD-test.

The DD-test based on $\mathcal{M}$ has the following useful properties over other methodologies in the literature. First, the DD-test is consistent for general nonlinearity, because the power transform X_t^{γ} in $\mathcal{M}$ is a sieve basis. Note that a number of studies in prior literature have tested the linearity condition, and most of them rely on the semi/nonparametric method. For example, Chen and Pouzo (2015) estimate $m(\cdot)$ 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 show that $m(\cdot)$ can be consistently estimated by letting the number of sieve series to increase as n increases, further introducing a methodology to test the correct model assumption consistently.[1] 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 Liao, 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 to compare the GMM distances measured by the linear model and $\mathcal{M}$ in parallel with the LR-test. Note that the degree of sieve basis γ is estimated to obtain the optimum sieve that best improves the DD-test, instead of including the maximum number of sieve series

[1]For X_t as an exogenous variable, Hong and White (1995) estimate $m(\cdot)$ using a sieve series estimation method, to provide an omnibus specification testing.

limited by n. Second, we compute the DD-test 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 can play the role of diagnostic test before estimating $m(\cdot)$ using other methodology. Note that Ai and Chen (2003), Newey and Powell (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 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 et al. (2015) test for the linear model misspecification using the model similar to $\mathcal{M}$, as mentioned above. Their methodology can be easily applied even when testing correctly specified structural models by treating the employed instrumental variables as conditioning variables. Second, Sargan (1958, 1988) and Hansen's (1982) J-test typically tests the structurally correct model assumption as well as the validity of instrumental variables. Thus, the J-test 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 presumes valid instrumental variables and focuses on testing for structural model misspecification with omnibus power against general nonlinear functions. Third, Horowitz (2006), Breunig (2015), and Zhu (2020) also provide tests with similar goals to the DD-test, but high level of computational complexity is required for their applications in addition to satisfying their model conditions. When implementing Horowitz's (2006) test, the empirical researcher has to first estimate the eigenvalues of the covariance kernel characterizing the Gaussian process associated with the null limit distribution of his test. Further, it requires that the number of instrumental variables is identical to the number of explanatory variables. Breunig's (2015) test is also defined by first approximating the associated test basis based on L_2 norm, so that the empirical researcher has to select a reference measure and basis functions before applying his test. Likewise, Zhu's (2020) test is defined by a reference measure selected by the researcher. In contrast, the DD-test is designed to be straightforward for empirical applications. Despite this simplicity, the DD-test exhibits comparable powers to the other tests as our simulations in Section 4 demonstrate.

2.2 Testing Environment and Assumptions

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

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

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, viz., $\mathbf{Z}_t \in \mathbb{R}^p$ with $p > k+2$. For notational simplicity, we also divide the parameter vector $\boldsymbol{\delta}_*$ into $(\xi_{1*}, \boldsymbol{\eta}_*')'$ such that $\mathbf{E}_t'\boldsymbol{\delta}_* = \xi_{1*}X_t + \mathbf{D}_t'\boldsymbol{\eta}_*$.

We next 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:

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

and 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 $\mathbf{Y} := (Y_1, \ldots, Y_n)'$; $\mathbf{X}(\gamma) := (X_1^{\gamma}, \ldots, X_n^{\gamma})'$; $\mathbf{V}_t := (1, \mathbf{E}_t')'$; $\mathbf{V} := [\mathbf{V}_1', \ldots, \mathbf{V}_n']'$; $\mathbf{Z} := [\mathbf{Z}_1', \ldots, \mathbf{Z}_n']'$; $\boldsymbol{\varsigma} := (\xi_0, \boldsymbol{\delta}')'$; and $\mathbf{M}_n$ is a weighting matrix. That is, the GMM estimator is obtained by minimizing the quadratic distance function: $\widehat{\boldsymbol{\omega}}_n := \arg\min_{\boldsymbol{\omega}\in\boldsymbol{\Omega}} d_n(\boldsymbol{\omega})$. We also let $\widetilde{\boldsymbol{\omega}}_n := \arg\min_{\boldsymbol{\omega}\in\boldsymbol{\Omega}} d_n(\boldsymbol{\omega})$ 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. Note that $\mathcal{M}$ could be misspecified under a general nonlinear structure between Y_t and X_t. As Hall and Inoue (2003) have pointed out, in such a case, the power function in $\mathcal{M}$ 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, $\mathcal{M}$ is correctly specified for the linear structure between Y_t and X_t. We therefore impose the following hypothesis:

$$\mathcal{H}_0 : \mathbb{E}[m_t(\xi_0, \boldsymbol{\delta}, \beta_*, \gamma_*)\mathbf{Z}_t] = \mathbf{0} \;\text{ for some } \boldsymbol{\delta} \;\text{ and }\; \xi_0, \quad \text{for } \;\beta_* = 0 \;\text{ or }\; \gamma_* = 0 \;\text{ or }\; 1.$$

We may further partition $\mathcal{H}_0$ into the following sub-conditions:

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

to generate a linear structure between Y_t and X_t and thus hypothesize the researcher's interest. The negation of $\mathcal{H}_0$ is an alternative hypothesis: $\mathcal{H}_1 : \beta_* \neq 0$, $\gamma_* \neq 0$ and $\gamma_* \neq 1$. For simplicity, we let $\boldsymbol{\Omega}_0 := \{\boldsymbol{\omega} \in \boldsymbol{\Omega} : \beta = 0,\ \gamma = 0,\ \text{or } \gamma = 1\}$ and $\boldsymbol{\Omega}_1 := \boldsymbol{\Omega} \setminus \boldsymbol{\Omega}_0$ be 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's (1977, 1987) identification arises under $\mathcal{H}_{0,1}$. That is, γ_* is identified only when $\beta_* \neq 0$, and by this, a standard test does not follow a standard chi-squared distribution under $\mathcal{H}_{0,1}$. In general, a Gaussian process is involved with the null limit distribution when Davies's (1977, 1987) identification problem arises. Similarly, if $\gamma_* = 0$, only $\xi_{0*} + \beta_*$ is identified, implying that ξ_{0*} and β_* are not separately identified, and Davies's

(1977, 1987) identification problem arises in a different manner under $\mathcal{H}_{0,2}$ from $\mathcal{H}_{0,1}$. Likewise, Davies's (1977, 1987) identification problem arises under $\mathcal{H}_{0,3}$, implying that neither ξ_{1*} nor β_* is separately identified. Thus, we find three composite Davies's (1977, 1987) identification problems with $\mathcal{H}_0$. A proper testing methodology should tackle all these separate identification problems using a single test. We call this the trifold identification problem following Baek, Cho, and Phillips (2015). By the trifold identification problem, it becomes challenging to test the linear structure via Wald's (1943) test principle. As Baek, Cho, and Phillips (2015) pointed out, when a multifold identification problem is associated with the null hypothesis, the standard Wald-test can have a unbounded null limit distribution, because a parameter value 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:

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

Note that $\mathcal{M}$ approximates the unknown functional form of $m(\cdot)$ by the power transform, and the DD-test exploits this approximation to gain the test statistic marginal power; this is exactly the same motivation as that of the QLR-test. The DD- and QLR-tests are defined similarly, but have different features. 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.

Before examining the asymptotic behaviors of the DD-test under the different hypotheses, we formalize the above DGP and model conditions along with others as collected in the following assumption:

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

(ii) 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 $\{U_t, \mathbf{Z}_t, \mathbf{E}_t, U_{t-1}, \mathbf{Z}_{t-1}, \mathbf{E}_{t-1}, \ldots\}$;

(iii) (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 $\mathbf{D}_t$;

(iv) (a) $\mathrm{var}(n^{-1/2}\mathbf{Z}'\mathbf{U})$ converges to $\boldsymbol{\Sigma}$ as $n \to \infty$;

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

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

(ii) for the measurable functions $m(\cdot)$ and $(\xi_{0*}, \boldsymbol{\delta}_*')' \in \mathbb{R}^{2+k}$, $Y_t = \xi_{0*} + \mathbf{E}_t'\boldsymbol{\delta}_* + m(X_t) + U_t$; 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 $\mathbf{V}_t := (1, \mathbf{E}_t')'$. □

Assumption 3. *An SSE sequence $\{M_t\}$ exists such that*

(i) $\mathbb{E}[M_t^4] < \infty$ and $\sup_{\gamma\in\boldsymbol{\Gamma}} |X_t^\gamma| \leq M_t$; and

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

Assumption 4. *(i) For all $\epsilon > 0$, $\mathbb{E}[\mathbf{G}_t(\cdot)\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t\mathbf{G}_t(\cdot)']$ is PD uniformly on $\boldsymbol{\Gamma}(\epsilon)$, where $\mathbf{G}_t(\gamma) := (X_t^\gamma, \mathbf{V}_t')'$ and $\boldsymbol{\Gamma}^c(\epsilon) := \{\gamma \in \boldsymbol{\Gamma} : |\gamma| \geq \epsilon \ \text{ or } \ |\gamma - 1| \geq \epsilon\}$;*

(ii) $\mathbb{E}[\mathbf{G}_{t,0}\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t\mathbf{G}_{t,0}']$ is PD, where $\mathbf{G}_{t,0} := (L_t, \mathbf{V}_t')'$; and

(iii) $\mathbb{E}[\mathbf{G}_{t,1}\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t\mathbf{G}_{t,1}']$ is PD, where $\mathbf{G}_{t,1} := (X_tL_t, \mathbf{V}_t')'$. □

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. In addition, Assumption 4 lets $\widehat{\boldsymbol{\omega}}_n$ be asymptotically non-degenerate even under $\mathcal{H}_0$.

(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), $\text{var}(n^{-1/2}\mathbf{Z}'\mathbf{U})$ 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.

(e) Although $\mathcal{M}$ supposes a fixed form of model, the model condition can be flexibly modified without difficulty. As an example, if the intercept term is not needed for model construction, we can simply remove it from $\mathcal{M}$, and the DD-test can be applied in parallel to the presence of the intercept term. For such a case, the rank condition needs to be modified to $p > k + 1$.

(f) The interior parameter condition for β_* in Assumption 2 removes the asymptotic chance for $\widehat{\boldsymbol{\omega}}_n$ to exist as a set. If zero is on the boundary of the parameter space, $\widehat{\beta}_n$ can be asymptotically zero with non-negligible probability under the null. For such a case, any $\gamma \in \boldsymbol{\Gamma}$ becomes $\widehat{\gamma}_n$ as it is a placeholder, so that $\widehat{\boldsymbol{\omega}}_n$ can exist as a set. Instead, we impose the zero interior parameter space condition to make $\widehat{\beta}_n$ be zero with probability converging to zero and remove the asymptotic probability for $\widehat{\boldsymbol{\omega}}_n$ to exists as a set. Similarly, we do not impose the parameter space condition for γ so that $\widehat{\gamma}_n$ is 0 or 1 with probability converging to zero by the same rationale.

(g) Although we here focus on testing the structural linearity, our methodology can be used to test the linearity hypothesis of the exogenous variable. Instead of βX_t^γ, we may modify $\mathcal{M}$ by introducing the power transformation of the exogenous variable and test its linearity hypothesis by the methodology described in Section 2.3. □

2.3 Testing Structural Linearity

We now examine how the trifold identification problem is associated with the null limit distribution. For this purpose, we first define three tests denoted below as $\mathcal{D}_n^{(\beta=0)}(\epsilon)$, $\mathcal{D}_n^{(\gamma=0)}$, and $\mathcal{D}_n^{(\gamma=1)}$ that test the three sub-conditions $\mathcal{H}_{0,1}$, $\mathcal{H}_{0,2}$, and $\mathcal{H}_{0,3}$, respectively; and we next derive their limit approximations under their respective sub-condition. We finally show how the null approximations are interrelated with each other, that we exploit to obtain the limit distribution of $\mathcal{D}_n$ under $\mathcal{H}_0$.

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. That is, we obtain $\min_\gamma \min_\beta \min_\varsigma d_n(\boldsymbol{\omega})$. If we let $\mathbf{Q}_1 := \ddot{\mathbf{Z}}(\mathbf{I}-\ddot{\mathbf{Z}}'\mathbf{V}(\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V})^{-1}\mathbf{V}'\ddot{\mathbf{Z}})\ddot{\mathbf{Z}}'$, $\ddot{\mathbf{Z}} := \mathbf{Z}\mathbf{M}_n^{1/2}$, and $\mathbf{U} := (U_1, \ldots, U_n)'$, then it follows that

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

where $\mathcal{D}_n^{(\beta=0)}(\epsilon)$ denotes the DD-test testing $\mathcal{H}_{0,1}$. Here, the γ space is modified from $\boldsymbol{\Gamma}$ to $\boldsymbol{\Gamma}^c(\epsilon)$ and 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 limit distribution under $\mathcal{H}_0$.

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

Lemma 1. *Given Assumptions 1, 2, 3, 4, and $\mathcal{H}_{0,1}$, for each $\epsilon > 0$, we have $\mathcal{D}_n^{(\beta=0)}(\epsilon) \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}^c(\epsilon)} \mathcal{Z}_1^2(\gamma)$, where for each $\epsilon > 0$, $\{\mathcal{Z}_1(\gamma) : \gamma \in \boldsymbol{\Gamma}^c(\epsilon)\}$ is a zero mean Gaussian process such that for each pair (γ, γ'), $\mathbb{E}[\mathcal{Z}_1(\gamma)\mathcal{Z}_1(\gamma')] = \rho_1(\gamma,\gamma') := \kappa_1(\gamma,\gamma')/\{\sigma_1^2(\gamma)\sigma_1^2(\gamma')\}^{1/2}$, $\kappa_1(\gamma,\gamma') := \mathbb{E}[X_t^\gamma \widetilde{\mathbf{Z}}_t']\mathbf{J}_1\widetilde{\boldsymbol{\Sigma}}\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma'}]$, $\sigma_1^2(\gamma) := \mathbb{E}[X_t^\gamma \widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^\gamma]$, $\widetilde{\mathbf{Z}}_t := \mathbf{M}_0^{1/2}\mathbf{Z}_t$, $\widetilde{\boldsymbol{\Sigma}} := \mathbf{M}_0^{1/2}\boldsymbol{\Sigma}\mathbf{M}_0^{1/2}$, and $\mathbf{J}_1 := \mathbf{I} - \mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}_t'](\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t']\mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}_t'])^{-1}\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t']$.* □

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, $\widetilde{\mathcal{Z}}_1(\gamma) := \boldsymbol{\pi}_1(\gamma)'\boldsymbol{\mathcal{U}}$, where $\boldsymbol{\pi}_1(\gamma) := \mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^\gamma]/\sigma_1^2(\gamma)^{1/2}$ and $\boldsymbol{\mathcal{U}} \sim N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}})$, 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 $\kappa_1(\gamma,\gamma') = \mathbb{E}[X_t^\gamma \widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma'}]$, 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 rank condition in Assumption 2(*iii*) should be linked to $\mathbf{J}_1$. If $\mathbb{E}[\mathbf{V}_t\mathbf{Z}_t']$ had been a square matrix with $k = p$,

$\mathbf{J}_1 = \mathbf{0}$, so that we cannot test the hypothesis by the DD-test statistic. The condition on the relationship between p and k in Assumption 1(i), *i.e.*, $p > k + 2$, implies that the DD-test statistic is applicable only to overidentified models. $\square$

We next examine the limit distribution of $\mathcal{D}_n$ 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 any other parameter, to obtain

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

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

by applying a second-order Taylor expansion, where $\mathbf{C}_0 := [L_1, \ldots, L_n]'$, and $\mathcal{D}_n^{(\gamma=0;\beta)}$ (resp. $\mathcal{D}_n^{(\gamma=0;\xi_0)}$) denotes the DD-test designed to test $\mathcal{H}_{0,2}$ by treating β_* (resp. ξ_{0*}) as an unidentified parameter. Here, the distance functions $d_n(\cdot;\beta)$ and $d_n(\cdot;\xi_0)$ in (4) and (5) are interpreted as functions of γ when optimizing them with respect to γ while keeping β and ξ_0 as nuisance parameters, respectively. In addition, 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 $\mathcal{D}_n$ irrespective of the optimization order. We let $\mathcal{D}_n^{(\gamma=0)}$ denote the DD-test designed to test $\mathcal{H}_{0,2}$ and contain the null limit distribution in the following lemma:

Lemma 2. *Given Assumptions 1, 2, 3, 4, and* $\mathcal{H}_{0,2}$, $\mathcal{D}_n^{(\gamma=0)} = \{\mathbf{C}_0'\mathbf{Q}_1\mathbf{U}\}^2/\{n\mathbf{C}_0'\mathbf{Q}_1\mathbf{C}_0\} + o_{\mathbb{P}}(1) \overset{A}{\sim} \mathcal{Z}_0^2$, *where* $\mathcal{Z}_0 \overset{A}{\sim} N(0, \kappa_0^2)$ *and* $\kappa_0^2 := \mathbb{E}[L_t\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\widetilde{\mathbf{\Sigma}}\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_tL_t]/\mathbb{E}[L_t\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_tL_t]$. $\square$

Remarks.

(a) The null limit distribution in Lemma 2 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 $\mathbf{M}_0 = \mathbf{\Sigma}^{-1}$, then $\kappa_0^2 = 1$, and so $\mathcal{D}_n^{(\gamma=0)} \overset{A}{\sim} \mathcal{X}_1^2$ 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 below examine their joint distribution along with the weak limits under $\mathcal{H}_{0,3}$. $\square$

Next, we examine the limit distribution of $\mathcal{D}_n$ under $\mathcal{H}_{0,3} : \gamma_* = 1$. The process is parallel 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

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

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

by applying a second-order Taylor approximation to $d_n(\cdot)$, where for $j = 1, 2, ...,$ we let $\mathbf{C}_j := [X_t^j L_1, \ldots, X_n^j L_n]$. Here, $\mathcal{D}_n^{(\gamma=1;\beta)}$ (resp. $\mathcal{D}_n^{(\gamma=1;\xi_1)}$) denotes the DD-test 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, and interpreting $d_n(\cdot;\beta)$ and $d_n(\cdot;\xi_1)$ in (6) and (7) as functions of γ when optimizing them with respect to γ while keeping β and ξ_1 as nuisance parameters, respectively. As earlier, the RHS parameters of (6) and (7) are asymptotically free of β and ξ_1, respectively, under our regularity conditions, so that maximization with respect to β and ξ_0 in (6) and (7) respectively becomes innocuous in obtaining the null limit distribution. Furthermore, the null approximation in (6) is identical to that in (7), implying that the limit distribution under $\mathcal{H}_{0,3}$ is uniquely obtained irrespective of the optimization order. We let $\mathcal{D}_n^{(\gamma=1)}$ denote the DD-test designed to test $\mathcal{H}_{0,3}$ and contain its null limit distribution in the following lemma:

Lemma 3. *Given Assumptions 1, 2, 3, 4, and $\mathcal{H}_{0,3}$, $\mathcal{D}_n^{(\gamma=1)} = \{\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}\}^2/\{n\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1\} + o_{\mathbb{P}}(1) \overset{A}{\sim} \mathcal{Z}_1^2$, 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 limit distribution of $\mathcal{D}_n$ 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 the 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 $\mathrm{plim}_{\gamma\to 0}N_n(\gamma) = 0$ and $\mathrm{plim}_{\gamma\to 0}D_n(\gamma) = 0$, 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) $\mathrm{plim}_{\gamma\to 0}N_n^{(1)}(\gamma) = 0$ *and* $\mathrm{plim}_{\gamma\to 0}D_n^{(1)}(\gamma) = 0$*;*

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

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

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

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

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

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

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

Therefore, the biggest GMM distance is obtained under $\mathcal{H}_{0,1}$ without restricting $\mathbf{\Gamma}$ by ϵ. This implies that the limit distribution of the DD-test 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, 4, and $\mathcal{H}_0$, $\mathcal{D}_n \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}} \mathcal{Z}_1^2(\gamma)$, where $\mathcal{Z}_1(0)$ and $\mathcal{Z}_1(1)$ are defined as the limits of $\mathcal{D}_n^{(\gamma=0)}$ and $\mathcal{D}_n^{(\gamma=1)}$, respectively.* □

If γ is restricted to either 0 or 1 in $\mathcal{M}$, it becomes a linear model, so that the DD-test becomes zero, implying that the null limit distribution of the DD-test without the restriction cannot be achieved by letting $\gamma = 0$ or 1, mainly because the DD-test has to be always greater than zero. We therefore simply let $\mathcal{Z}_1^2(0)$ and $\mathcal{Z}_1^2(1)$ be the weak limits obtained from the right sides of (8) so that the null limit distribution of the DD-test is not affected by this selection and $\mathcal{Z}_1^2(\cdot)$ becomes continuous on $\mathbf{\Gamma}$, enabling us to apply the maximal principle for the existence of $\sup_{\gamma\in\mathbf{\Gamma}} \mathcal{Z}_1^2(\gamma)$.

The null limit distribution of the DD-test can be obtained through simulation. If $\widehat{\boldsymbol{\pi}}_{n,1}(\cdot)$ and $\widehat{\mathbf{\Sigma}}_n$ consistently estimate $\boldsymbol{\pi}_1(\cdot)$ and $\widetilde{\mathbf{\Sigma}}$, respectively, the limit distribution of $\sup_{\gamma\in\mathbf{\Gamma}}(\widehat{\boldsymbol{\pi}}_{n,1}(\gamma)'\widehat{\boldsymbol{\mathcal{U}}})^2$ would estimate the null limit distribution of $\mathcal{D}_n$, provided that $\widehat{\boldsymbol{\mathcal{U}}} \sim N(\mathbf{0}, \widehat{\mathbf{\Sigma}}_n)$. Therefore, the empirical researcher can apply Hansen's (1996) weighted bootstrap and obtain the asymptotic critical values as detailed in Section 4 (see also Cho et al., 2011). Here, when computing $\sup_{\gamma\in\mathbf{\Gamma}}(\widehat{\boldsymbol{\pi}}_{n,1}(\gamma)'\widehat{\boldsymbol{\mathcal{U}}})^2$, it would be more straightforward to apply the grid search method than applying an optimization algorithm, because the dimension of γ is one.

2.4 Testing for Structural Nonlinearity

The DD-test 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.

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, that we denote as $d_0 := \mathrm{plim}_{n\to\infty} n^{-2} d_n(\widetilde{\omega}_n)$, and obtain it by the ergodic theorem:

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

Here, if ς_0 is the argument for d_0, then $\varsigma_0 = \varsigma_* + (\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t']\mathbb{E}[\widetilde{\mathbf{Z}}_t\mathbf{V}_t'])^{-1}\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t']\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)]$, 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 $d(\gamma) := \min_{\boldsymbol{\varsigma},\beta} \operatorname{plim}_{n\to\infty} n^{-2} d_n(\boldsymbol{\varsigma}, \beta, \gamma)$,

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

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

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

by some tedious algebra. 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 $\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{Z}}_t]$, so that the RHS of (9) is obtained by the L'Hôpital rule for $\gamma = 0$ or 1. Therefore,

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

Note that the two limits are still strictly positive.

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

$$\mu_1^2(\cdot) := \rho^2(\mathbf{h}, \mathbf{g}(\cdot)) \cdot (\mathbf{h}'\mathbf{h}) := \frac{\{\mathbf{h}'\mathbf{g}(\cdot)\}^2}{\{\mathbf{h}'\mathbf{h}\} \cdot \{\mathbf{g}(\cdot)'\mathbf{g}(\cdot)\}} \cdot (\mathbf{h}'\mathbf{h})$$

and $\mathbf{h} := \mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)]$ and $\mathbf{g}(\gamma) := \mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]$. Indeed, $\sup_{\gamma\in\boldsymbol{\Gamma}} \mu_1^2(\gamma)$ 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, 3, and 4,*

(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 $\widetilde{\gamma} \in \boldsymbol{\Gamma}$, $d(\widetilde{\gamma}) \in (0, d_0)$ and $n^{-1}\mathcal{D}_n = d_0 - d(\widetilde{\gamma}) + o_{\mathbb{P}}(1)$; 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 $\mathcal{D}_n \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}}\{\mathcal{Z}_1(\gamma) + \nu_1(\gamma)\}^2$, 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 a 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. Note that the relationship between p and k and the rank condition in Assumptions 1(*i*) and 2(*iii*), respectively imply $\mathbf{J}_1 \neq \mathbf{0}$ as remarked below Lemma 1, from which $\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] \neq \mathbf{0}$ is virtually imposed by the nonzero condition in Theorem 2(*i*).

(b) The DD-test draws its consistent power from a factor different from that for the J-test, that 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 draws its power from the uncentered correlation between $\mathbf{h}$ and $\mathbf{g}(\cdot)$; this implies that the J- and DD-tests supplement each other. If the J-test rejects the null but the DD-test 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 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 $\gamma_* \in \mathbf{\Gamma}\backslash\{0,1\}$, $m(X_t) = \beta_* X_t^{\gamma_*}$, then $n^{-1}\mathcal{D}_n = \beta_*^2 \mathbb{E}[X_t^{\gamma_*}\widetilde{\mathbf{Z}}_t']\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma_*}] + o_{\mathbb{P}}(1)$; this is strictly positive at the limit, implying that $\mathcal{D}_n$ has nontrivial asymptotic power.

(d) For an intuitive proof of Theorem 2(*i*), note that the DD-test does not have an asymptotic power if and only if $\rho(\mathbf{g}(\cdot), \mathbf{h}) \equiv 0$, but this condition is contradictory to the provided condition. Suppose that for each γ, $\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}] = \mathbf{0}$, and this implies that $\mathbb{E}[\widetilde{\mathbf{Z}}_t | X_t] = \mathbf{0}$ from the fact that $\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}] = \nabla_{\boldsymbol{\tau}} \mathbb{E}[\exp(\gamma \log\ (X_t) + \boldsymbol{\tau}'\widetilde{\mathbf{Z}}_t)]|_{\boldsymbol{\tau}=\mathbf{0}}$ and $\mathbb{E}[\exp(\gamma \log(X_t) + \boldsymbol{\tau}'\widetilde{\mathbf{Z}}_t)]$ is a moment generating function of $(\log(X_t), \widetilde{\mathbf{Z}}_t')'$, so that if $\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}] = \mathbf{0}$, then $\mathbb{E}[\widetilde{\mathbf{Z}}_t | \log(X_t)] = \mathbf{0}$. Here, $\log(\cdot)$ is a measure-preserving transformation, so that $\mathbb{E}[\widetilde{\mathbf{Z}}_t | \log(X_t)] = \mathbb{E}[\widetilde{\mathbf{Z}}_t | X_t]$. Therefore, $\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] = \mathbf{0}$ is implied by the law of iterated expectation. Note that it is contradictory to the requirement that $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] \neq \mathbf{0}$, leading to a nonzero correlation coefficient between $\mathbf{h}$ and $\mathbf{g}(\gamma)$ for some γ.

(e) The intuition of the power property in Theorem 2(*i*) should be straightforward. The DD-test may be comparable to the specification test in Bierens (1990) that first computes an infinite number of moment conditions and next chooses the worst moment condition to gain the power. Differently from this, the DD-test first computes the infinite number of moments given by $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)]$ and $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}]$ and next gains the power by choosing γ of $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]$ that best approximates $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)]$ in terms of the uncentered correlation coefficient.

(f) From Theorem 2(*ii*), the DD-test 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 $\nu_1(\cdot)$, that is different from 0, as implied by Theorem 2(*i*). □

2.5 Asymptotic Uniform Inference

The null limit distribution in Theorem 1 can be used to implement the DD-test uniformly on the parameter space. In this section, we examine this feature by investigating whether the asymptotic size defined as $\limsup_{n\to\infty} \sup_{\boldsymbol{\omega}_* \in \boldsymbol{\Omega}} \mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_n(\boldsymbol{\alpha}))$ is less than or equal to $\boldsymbol{\alpha}$ as examined by Andrews and Cheng (2015) and Elliott, Müller, and Watson (2015) in their frameworks.

Testing linearity by the DD-test can be applied to testing different hypotheses. Note that Theorem 1 implies that the null limit distribution is determined by that testing $\mathcal{H}_{0,1} : \beta_* = 0$. Due to this aspect, we focus on the hypothesis $\mathcal{H}_0' : \beta_* = \beta_0$ for the asymptotic size, where β_0 is not necessarily equal to zero. We first note that the DD-test can be written as

$$\mathcal{D}_n = \sup_{\gamma\in\mathbf{\Gamma}} n^{-1}\{d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n) - d_n(\widehat{\xi}_{0,n}(\gamma), \widehat{\boldsymbol{\delta}}_n(\gamma), \widehat{\beta}_n(\gamma), \gamma)\},$$

where $(\widehat{\xi}_{0,n}(\gamma), \widehat{\boldsymbol{\delta}}_n(\gamma), \widehat{\beta}_n(\gamma)) := \arg\inf_{\xi_0,\boldsymbol{\delta},\beta} d_n(\boldsymbol{\omega})$ and $(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \widetilde{\gamma}_n) := \arg\inf_{\xi_0,\boldsymbol{\delta},\gamma} d_n(\boldsymbol{\omega})$ subject to $\beta_* = \beta_0$. Note that $\widetilde{\gamma}_n$ is a placehold if $\beta_0 = 0$, but it has to be estimated, otherwise.

The null limit distribution of the DD-test is obtained under some mild regularity conditions. For this derivation, we further suppose the following conditions along with the earlier ones:

Assumption 5. *(i)* $\{\boldsymbol{\omega} \in \mathbf{\Omega} : \mathbb{E}[m_t(\boldsymbol{\omega})\mathbf{Z}_t] = \mathbf{0}\}$ *has a unique element as an interior element of* $\mathbf{\Omega}$*, provided that* $\beta_* \neq 0$ *or* $\gamma_* \neq 0, 1$*;*

(ii) For all $\epsilon > 0$, $\mathbb{E}[\mathbf{G}_{t,2}(\cdot)\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t\mathbf{G}_{t,2}(\cdot)']$ *is PD uniformly on* $\mathbf{\Gamma}(\epsilon)$*, where* $\mathbf{G}_{t,2}(\gamma) := (X_t^\gamma, X_t^\gamma L_t, \mathbf{V}_t')'$. □

By Assumption 5(*i*), the model is identified unless $\mathcal{H}_0$ is supposed.

The derivation of the null limit distribution is facilitated by exploiting the structure of the DD-test. We first note that

$$d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n) = \inf_{\gamma\in\mathbf{\Gamma}}[\mathbf{U} - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))]'\mathbf{Q}_1[\mathbf{U} - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))] \tag{10}$$

under $\mathcal{H}_0'$ and

$$d_n(\widehat{\xi}_{0,n}(\gamma), \widehat{\boldsymbol{\delta}}_n(\gamma), \widehat{\beta}_n(\gamma), \gamma) = [\mathbf{U} - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))]'\mathbf{P}(\gamma)[\mathbf{U} - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))], \tag{11}$$

where $\mathbf{P}(\gamma) := \mathbf{Q}_1 - \mathbf{Q}_1\mathbf{X}(\gamma)(\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma))^{-1}\mathbf{X}(\gamma)'\mathbf{Q}_1$. The right sides on (10) and (11) are obtained by first concentrating the GMM distance with respect to $(\xi_0, \boldsymbol{\delta})$ and $(\xi_0, \boldsymbol{\delta}, \beta)$, respectively. If we further let $\mathbf{H}(\gamma) := \mathbf{J}_1 - \mathbf{g}(\gamma)(\mathbf{g}(\gamma)'\mathbf{g}(\gamma))^{-1}\mathbf{g}(\gamma)'$ and $\mathbf{d}(\gamma) := \mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^\gamma \log(X_t)]$ and suppose that $n^{-1/2}\mathbf{U}'\ddot{\mathbf{Z}} \Rightarrow \boldsymbol{\mathcal{U}} \sim N(\mathbf{0}, \widetilde{\mathbf{\Sigma}})$ by applying Assumption 1, it now follows that

$$\begin{aligned}\mathcal{D}_n &\Rightarrow \inf_s[\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)]'\mathbf{J}_1[\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)] - \inf_s[\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)]'\mathbf{H}(\gamma_*)[\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)]\\ &= \mathcal{Z}_2^2(\gamma_*) := \frac{\{\mathbf{g}(\gamma_*)'\mathbf{K}(\gamma_*)\boldsymbol{\mathcal{U}}\}^2}{\mathbf{g}(\gamma_*)'\mathbf{K}(\gamma_*)\mathbf{g}(\gamma_*)}\end{aligned}$$

under $\mathcal{H}_0'$, provided that $\beta_0 \neq 0$, for $\gamma_* \neq 0, 1$. Here, in order to obtain the weak limit, we used the fact that $\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*) = \mathbf{D}(\gamma_*)(\gamma - \gamma_*) + o_{\mathbb{P}}((\gamma - \gamma_*))$ and $\mathbf{J}_1$ is idempotent, where we let $\mathbf{K}(\gamma_*) := \mathbf{I} - \mathbf{d}(\gamma_*)(\mathbf{d}(\gamma_*)'\mathbf{d}(\gamma_*))^{-1}\mathbf{d}(\gamma_*)'$, $\mathbf{D}(\gamma_*) := [D_1(\gamma_*) \ldots, D_n(\gamma_*)]'$ with $D_t(\gamma_*) := X_t^{\gamma_*} L_t$, and s captures the asymptotic distance measured by $\sqrt{n}(\gamma - \gamma_*)$. Note that the null limit distribution is free of β_*. In contrast, if $\beta_0 = 0$ and/or $\gamma_* = 0, 1$,

$$\mathcal{D}_n = n^{-1}\mathbf{U}'\mathbf{Q}_1\mathbf{U} - \inf_{\gamma\in\mathbf{\Gamma}} n^{-1}\mathbf{U}'\mathbf{P}(\gamma)\mathbf{U} = \sup_{\gamma\in\mathbf{\Gamma}} \frac{1}{n}\frac{\{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)} \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}} \frac{\{\mathbf{g}(\gamma)'\boldsymbol{\mathcal{U}}\}^2}{\mathbf{g}(\gamma)'\mathbf{g}(\gamma)}$$

under $\mathcal{H}_0'$. By letting $\boldsymbol{\pi}_1(\cdot) := \{\mathbf{g}(\cdot)'\mathbf{g}(\cdot)\}^{-1/2}\mathbf{g}(\cdot)$, note that the consequent limit becomes the same weak limit as given in Theorem 1. We contain these null limit distributions in the following theorem.

Theorem 3. *Given Assumptions 1, 2, 3, 4, 5, and* $\mathcal{H}_0': \beta_* = \beta_0$,

(i) *if* $\beta_0 = 0$ *and/or* $\gamma_* = 0$, *1,* $\mathcal{D}_n \Rightarrow \sup_{\gamma \in \mathbf{\Gamma}} \mathcal{Z}_1^2(\gamma)$*; and*

(ii) *if* $\beta_0 \neq 0$ *and* $\gamma_* \neq 0$, 1, $\mathcal{D}_n \Rightarrow \mathcal{Z}_2^2(\gamma_*)$ *such that* $\mathcal{Z}_2(\gamma_*) \sim N(0, \kappa_2^2(\gamma_*))$ *and* $\kappa_2^2(\gamma_*) := \mathbf{g}(\gamma_*)'\mathbf{K}(\gamma_*)\widetilde{\mathbf{\Sigma}}\mathbf{K}(\gamma_*)$ $\mathbf{g}(\gamma_*)/\mathbf{g}(\gamma_*)'\mathbf{K}(\gamma_*)\mathbf{g}(\gamma_*)$. □

As a remark to Theorem 3, the null limit distribution in Theorem 3(*ii*) is also be derived by using the following null approximation:

$$\mathcal{D}_n = \frac{1}{n}\frac{\{\mathbf{X}(\gamma_*)'(\mathbf{Q}_1 - \mathbf{Q}_1\mathbf{D}(\gamma_*)(\mathbf{D}(\gamma_*)'\mathbf{Q}_1\mathbf{D}(\gamma_*))^{-1}\mathbf{D}(\gamma_*)'\mathbf{Q}_1)\mathbf{U}\}^2}{\mathbf{X}(\gamma_*)'(\mathbf{Q}_1 - \mathbf{Q}_1\mathbf{D}(\gamma_*)(\mathbf{D}(\gamma_*)'\mathbf{Q}_1\mathbf{D}(\gamma_*))^{-1}\mathbf{D}(\gamma_*)'\mathbf{Q}_1)\mathbf{X}(\gamma_*)} + o_{\mathbb{P}}(1). \tag{12}$$

If we apply the CLT or ergodic theorem to the relevant quantities on the RHS of (12), the same null limit distribution is also obtained as given in Theorem 3(*ii*).

The null limit distributions in Theorem 3 are obtained by fixing the unknown parameter value, but we can exploit them to examine the asymptotic size. Note that the GMM distances given by (10) and (11) are influenced by only β_* and γ_*, so that we can focus on its parameter space for the asymptotic uniform inference. We separate the parameter space into two subsets by noting that the null limit distribution depends on whether the model is identified or not. That is, for every $\epsilon > 0$, we let $\mathbf{B}(\epsilon) := \{\beta \in \mathbf{B} : |\beta| < \epsilon\}$ and $\mathbf{B}^c(\epsilon) := \mathbf{B} \setminus \mathbf{B}(\epsilon)$. Likewise, we let $\mathbf{\Gamma}(\epsilon) := \{\gamma \in \mathbf{\Gamma} : |\gamma| < \epsilon \text{ or } |\gamma - 1| < \epsilon\}$ and $\mathbf{\Gamma}^c(\epsilon) := \mathbf{\Gamma} \setminus \mathbf{\Gamma}(\epsilon)$. Note that $\mathbf{\Upsilon}_0 := \lim_{\epsilon \downarrow 0} \mathbf{B}(\epsilon) \times \mathbf{\Gamma}(\epsilon)$ collects the null parameter values for (β_*, γ_*) when the null model is not identified, so that the null limit distribution in Theorem 3(*i*) is obtained for $(\beta_*, \gamma_*) \in \mathbf{\Upsilon}_0$. On the other hand, if $(\beta_*, \gamma_*) \in \mathbf{\Upsilon}^c(\epsilon) := \mathbf{B}^c(\epsilon) \times \mathbf{\Gamma}^c(\epsilon)$, the null limit distribution in Theorem 3(*ii*) applies, so that we can handle the identified model case by Assumption 5(*ii*). We now provide the results on the asymptotic size in the following theorem:

Theorem 4. *Given Assumptions 1, 2, 3, 4, 5, and* $\mathcal{H}_0': \beta_* = \beta_0$,

(i) $\limsup_{n\to\infty} \sup_{(\beta_*,\gamma_*)\in\mathbf{\Upsilon}_0} \mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_1(\boldsymbol{\alpha})) = \boldsymbol{\alpha}$, *where* $cv_1(\boldsymbol{\alpha}) := \inf\{x \in \mathbb{R}^+ : F_1(x) \geq 1 - \boldsymbol{\alpha}\}$ *and for every* $x \geq 0$, $F_1(x) := \mathbb{P}(\sup_{\gamma\in\mathbf{\Gamma}} \mathcal{Z}_1^2(\gamma) \leq x)$;

(ii) *for any* $\epsilon > 0$, $\limsup_{n\to\infty} \sup_{(\beta_*,\gamma_*)\in\mathbf{\Upsilon}^c(\epsilon)} \mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_2(\boldsymbol{\alpha})) = \boldsymbol{\alpha}$, *where* $cv_2(\boldsymbol{\alpha}) := \inf\{x \in \mathbb{R}^+ : F_2(x) \geq 1 - \boldsymbol{\alpha}\}$ *and for every* $x \geq 0$, $F_2(x) := \mathbb{P}(\mathcal{Z}_2^2(\gamma_*) \leq x)$*; and*

(iii) $\limsup_{n\to\infty} \sup_{(\beta_*,\gamma_*)\in\mathbf{B}\times\mathbf{\Gamma}} \mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv(\boldsymbol{\alpha})) \leq \boldsymbol{\alpha}$, *where* $cv(\boldsymbol{\alpha}) := cv_1(\boldsymbol{\alpha})$ *if* $(\beta_*, \gamma_*) \in \mathbf{\Upsilon}_0$*; and* $cv(\boldsymbol{\alpha}) := cv_2(\boldsymbol{\alpha})$*, otherwise.* □

Remarks.

(a) Theorem 4(*i*) directly follows from Theorem 3(*i*) as all parameter values in $\mathbf{\Upsilon}_0$ characterize the linear model and Theorem 4(*i*) obtains the null limit distribution under the linear model assumption.

(b) We prove Theorem 4(*ii*) in Appendix by showing that Theorem 3(*ii*) holds uniformly on $\mathbf{\Gamma}^c(\epsilon)$ for any $\epsilon > 0$, and for this, we derive the null weak limit of the DD-test indexed by β_* and γ_* as a functional of a Gaussian process defined on the parameter space, so that for each $\boldsymbol{\omega}_*$, it follows that $\mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_2(\boldsymbol{\alpha})) = \boldsymbol{\alpha}$.

(c) Despite the asymptotic result in Theorem 4(*ii*), it does not necessarily imply that the empirical rejection rate of the DD-test is equally performing irrespective of the value of (β_*, γ_*). In Section 4.2, we demonstrate by simulation that for finite n, the empirical rejection rate of the DD-test can be different from $\boldsymbol{\alpha}$, depending on whether β_* is close to 0 and/or γ_* is close to 0 or 1.

(d) Theorem 4(*iii*) is obtained by combining the results in Theorems 4(*i* and *ii*). □

2.6 Weak Instrumental Variables

Many empirical studies often estimate the model using both strong and weak instrumental variables (*e.g.*, Angrist and Keueger, 1991). In this section, we examine the influence of weak instrumental variables to the DD-test. As it turns out, the null limit distribution of the DD-test is virtually determined by strong instrumental variables.

For this examination, we slightly generalize the earlier assumption. We first partition the instrumental variable $\mathbf{Z}_t$ into $\mathbf{S}_t \in \mathbb{R}^{p_s}$ and $\mathbf{W}_t \in \mathbb{R}^{p_w}$ such that $\mathbf{Z}_t \equiv (\mathbf{S}_t', \mathbf{W}_t')'$ and $p \equiv p_s + p_w$. Here, $\mathbf{S}_t$ and $\mathbf{W}_t$ denote strong and weak instrumental variables, respectively. The earlier discussion in Section 2.3 assumes that $p_w = 0$, so that $\mathbf{Z}_t = \mathbf{S}_t$. Due to the presence of the weak instrumental variable $\mathbf{W}_t$, it is not valid to suppose Assumption 2(*iii*) any longer. We therefore modify it into the following assumption:

Assumption 6. *(i)* $\mathbf{Z}_t \equiv (\mathbf{S}_t', \mathbf{W}_t')'$*, where* $\mathbf{S}_t \in \mathbb{R}^{p_s}$ *and* $\mathbf{W}_t \in \mathbb{R}^{p_w}$*;*

(ii) $\mathbb{E}[\mathbf{V}_t\mathbf{S}_t']$ *and* $\sum_{t=1}^n \mathbf{V}_t\mathbf{S}_t'$ *have full column ranks uniformly in* n*, respectively; and*

(iii) $\mathbf{W}_t = n^{-1/2}\boldsymbol{\mu}_w + \mathbf{W}_{0t}$ *such that* $\sum_{t=1}^n \mathbf{W}_{0t}\mathbf{V}_t = O_{\mathbb{P}}(\sqrt{n})$. □

Note that Assumption 6(*iii*) implies that $\mathbb{E}[\mathbf{V}_t\mathbf{W}_t'] = n^{-1/2}\mathbb{E}[\mathbf{V}_t]\boldsymbol{\mu}_w'$, so that the influence of $\mathbf{W}_t$ to $\mathbf{V}_t$ reduces to zero as n tends to infinity at the rate of $n^{-1/2}$, by which we desire to capture the feature of weak instrumental variable.

Given this, we derive the null limit distribution in parallel to the null limit distribution in Theorem 1. Note that the finite sample analog of the DD-test is given as

$$\mathcal{D}_n = \sup_{\gamma \in \boldsymbol{\Gamma}} \frac{1}{n} \frac{\{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)}$$

under $\mathcal{H}_0$. Each component on the right side has the following limit behavior:

(i) $n^{-1}\mathbf{X}(\cdot)'\ddot{\mathbf{Z}} = n^{-1}[\mathbf{X}(\cdot)'\ddot{\mathbf{S}}, \mathbf{X}(\cdot)'\ddot{\mathbf{W}}] = [\mathbb{E}[X_t^{(\cdot)}\widetilde{\mathbf{S}}_t], \mathbf{0}_{1\times p_w}'] + o_{\mathbb{P}}(1)$ uniformly on $\boldsymbol{\Gamma}$, where $\ddot{\mathbf{S}}$ and $\widetilde{\mathbf{S}}_t$ are such that $[\ddot{\mathbf{S}}, \ddot{\mathbf{W}}] = [\mathbf{S}, \mathbf{W}]\mathbf{M}_n^{1/2}$ and $[\widetilde{\mathbf{S}}_t', \widetilde{\mathbf{W}}_t']' := \mathbf{M}_0^{1/2}[\mathbf{S}_t', \mathbf{W}_t']'$;

(ii) $n^{-1/2}\mathbf{U}'\ddot{\mathbf{Z}} = n^{-1/2}[\mathbf{U}'\ddot{\mathbf{S}}, \mathbf{U}'\ddot{\mathbf{W}}] \Rightarrow [\boldsymbol{\mathcal{U}}_s', \boldsymbol{\mathcal{U}}_w'] =: \boldsymbol{\mathcal{U}}'$, where $\boldsymbol{\mathcal{U}}_s$ and $\boldsymbol{\mathcal{U}}_w$ denote the weak limits of $n^{-1/2}\mathbf{U}'\ddot{\mathbf{Z}}$ driven by the strong and weak instrumental variables, respectively; and

(iii) $n^{-1}\mathbf{V}'\ddot{\mathbf{Z}} = n^{-1}[\mathbf{V}'\ddot{\mathbf{S}}, \mathbf{V}'\ddot{\mathbf{W}}] = [\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t'], \mathbf{0}_{(2+k)\times p_w}] + o_{\mathbb{P}}(1)$.

As proving these is elementary, we do not separately derive these limits in Appendix.

The null limit distribution can be obtained by combining the separately obtained limits according to the sample analog of the DD-test. We contain it in the following theorem:

Theorem 5. *Given Assumptions 1, 2(i, ii), 3, 4, 6, and $\mathcal{H}_0$, if $p_s > 2+k$, $\mathcal{D}_n \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}} \mathcal{Z}_s^2(\gamma)$, where $\mathcal{Z}_s(\cdot) := \boldsymbol{\pi}_s(\cdot)'\boldsymbol{\mathcal{U}}_s$, $\boldsymbol{\pi}_s(\cdot) := \mathbf{J}_s\mathbb{E}[\mathbf{X}_t^{(\cdot)}\widetilde{\mathbf{S}}_t]/\{\mathbb{E}[\mathbf{X}_t^{(\cdot)}\widetilde{\mathbf{S}}_t']\mathbf{J}_s\mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{X}_t^{(\cdot)}]\}^{1/2}$ with $\mathbf{J}_s := \mathbf{I}_{p_s} - \mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{V}_t'](\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t']\mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{V}_t'])^{-1}\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t']$.* □

Remarks.

(a) Theorem 5 implies that the null limit distribution of the DD-test is virtually determined by the strong instrumental variables. The weak instrumental variables may affect the null limit distribution through $\mathbf{M}_n$ if the empirical researcher selects the weighting matrix to influence $\mathbf{M}_0$ through the weak instrumental variables. Otherwise, the null limit distribution is identical to that given in Theorem 1.

(b) Note that $\mathbf{J}_s$ is not well defined if $\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t']\mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{V}_t']$ is singular. Given that $\mathbb{E}[\mathbf{V}_t\mathbf{S}_t']$ is a full-column rank matrix, Theorem 5 avoids having this singular matrix problem by supposing that $p_s > 2 + k$. From this aspect, the DD-test can still be successfully exploited if the number of strong instrumental variables is greater than the number of explanatory variables. □

3 Extension to Testing the Polynomial Model Hypothesis

3.1 Motivation and Model

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

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

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 $\boldsymbol{\omega}_*^{(q)} \in \mathbf{\Omega}^{(q)}$, $Y_t = \mathbf{X}_{t,q}'\boldsymbol{\xi}_*^{(q)} + \mathbf{D}_t'\boldsymbol{\eta}_* + m(X_t) + U_t$ such that $\mathbb{E}[\mathbf{Z}_tU_t] = \mathbf{0}$, 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 $\mathcal{M}$, whereas the structural equation is possibly nonlinear for $q > 1$.

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/nonparametric sieve estimation exploits as many sieve bases as the sample size allows and leads to possibly unnecessary estimation errors for the estimator. A sequential testing procedure is a machinery process to avoid unnecessary estimation error.

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 Liao, 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. 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 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 MSC developed by Andrews (1999). We discuss this motivation by focusing on the Bayesian-type MSC among others and relating it to the sequential testing procedure. The Bayesian-type MSC is defined as $BC_{n,q} := \bar{\mathcal{J}}_{n,q} - (p - k - q - 1)\log(n)/n$, 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 rephrased in terms of

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

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 has omnibus power, implying that for every $q < q_*$, $\lim_{n\to\infty}\mathbb{P}(\mathcal{D}_{n,q} < 0) < 1$, where we let $\mathcal{D}_{n,q}$ be the DD-test statistic testing the q-th degree polynomial hypothesis, as formally defined below. Therefore, the fallacy probability becomes negligible as n increases.

The sequential testing procedure also tackles the data snooping bias that arises from testing multiple hypotheses. As pointed out by Hosoya (1989) and Cho and Phillips (2018) among others, a sequential testing procedure needs to test from the smallest hypothesis to bigger ones. Otherwise, it fails to control type-I error consistently, leading to a data snooping bias. We therefore start from testing a linear model hypothesis and increase the polynomial degree to the second degree in case we reject the linear model. In this manner, we continue increasing the degree one by one until we fail to reject the null model, enabling us to test the polynomial model hypothesis in an inclusive manner, and by this, we can eliminate a data snooping bias asymptotically. In addition, we also design our testing procedure to estimate q_* consistently by letting the level of significance depend on n. If a sequential testing procedure is combined with a constant level of significance, $\boldsymbol{\alpha}$ say, it estimates q_* inconsistently with an asymptotic probability for the estimated polynomial degree to differ from q_*

being equal to $\boldsymbol{\alpha}$. We therefore let the level of significance level converge to zero as n increases, that lets the estimated degree converge to q_* in probability. We discuss this 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

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 $\mathcal{M}$. 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 $\mathcal{H}''_0 := \cup_{s=1}^{q+2}\mathcal{H}''_{0,s}$, that 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 to overcome the multifold identification problem. For this, we define the DD-test 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 $\mathbf{V}_q := [\mathbf{V}'_{1,q}, \ldots, \mathbf{V}'_{n,q}]'$, $\mathbf{V}_{t,q} := (1, \mathbf{E}'_{t,q})' := (1, \mathbf{X}'_{t,q}, \mathbf{D}'_t)'$, and $\boldsymbol{\varsigma}^{(q)} := (\boldsymbol{\xi}^{(q)\prime}, \boldsymbol{\eta}')'$, so that $\boldsymbol{\omega}^{(q)} = (\boldsymbol{\varsigma}^{(q)\prime}, \beta, \gamma)'$. 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, so that $\mathcal{D}_{n,1} = \mathcal{D}_n$.

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

Assumption 7. *(i) The structure between Y_t and $\mathbf{E}_t$ 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, ..., and q are interior elements of $\boldsymbol{\Gamma}^{(q)}$;*

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

(iv) an SSE sequence $\{M_t\}$ exists such that $\mathbb{E}[M_t^4] < \infty$ and $\sup_{\gamma\in\boldsymbol{\Gamma}^{(q)}} |X_t^\gamma| \leq M_t$;

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

(vi) $\mathbb{E}[\mathbf{G}_t(\cdot)\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t\mathbf{G}_t(\cdot)']$ is PD uniformly on $\boldsymbol{\Gamma}$, where $\mathbf{G}_t(\gamma) := (X_t^\gamma, \mathbf{V}_{t,q}')'$; and

(vii) for $j = 1, 2, \ldots, q$, $\mathbb{E}[\mathbf{G}_{t,j}\mathbf{Z}_t']\mathbf{M}_0\mathbb{E}[\mathbf{Z}_t\mathbf{G}_{t,j}']$ is PD, where $\mathbf{G}_{t,j} := (X_t^j \log(X_t), \mathbf{V}_{t,q}')'$. □

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 7 would imply Assumptions 2, 3, and 4. □

Under the above conditions, we can obtain the properties of the DD-test as for the linearity testing. For this, we follow the approach of the linear model case. Let

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

to obtain the null limit distribution of the DD-test 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 for the linear model case, for each $s = 0, 1, \ldots, q$, let

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

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

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

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. The null limit approximation of the DD-test is obtained as their maximum, as for the linear model case. That is, if we let

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

then the DD-test $\mathcal{D}_{n,q}$ would be asymptotically equivalent to $\widetilde{\mathcal{D}}_{n,q}$ and $\widetilde{\mathcal{D}}_{n,q} = \mathcal{D}_{n,q}^{(\beta=0)} + o_{\mathbb{P}}(1)$ under $\mathcal{H}_0''$ by analogy, so that $\mathcal{D}_{n,q} = \mathcal{D}_{n,q}^{(\beta=0)} + o_{\mathbb{P}}(1)$ under $\mathcal{H}_0''$.

The omnibus power of the DD-test is also obtained as for 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, it follows that $\text{plim}_{n\to\infty} n^{-2}\{d_n(\widetilde{\boldsymbol{\omega}}_n^{(q)}) - d_n(\widehat{\boldsymbol{\varsigma}}_n^{(q)}(\gamma), \gamma)\} = \sup_{\gamma\in\boldsymbol{\Gamma}^{(q)}} \mu_q^2(\gamma)$ 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

$\gamma \in \mathbf{\Gamma}^{(q)}$,

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

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 $\sup_{\gamma\in\mathbf{\Gamma}^{(q)}} \mu_q^2(\gamma) > 0$, the DD-test would have a consistent power.

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

Corollary 1. *Given Assumption 1 and 7,*

(i) $\mathcal{D}_{n,q} \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}^{(q)}} \mathcal{Z}_q^2(\gamma)$ *under* $\mathcal{H}_0''$*, where* $\{\mathcal{Z}_q(\gamma) : \gamma \in \mathbf{\Gamma}^{(q)}\}$ *is a zero mean Gaussian process such that for each pair* (γ,γ'), $\mathbb{E}[\mathcal{Z}_q(\gamma)\mathcal{Z}_q(\gamma')] = \rho_q(\gamma,\gamma') := \kappa_q(\gamma,\gamma')/\{\sigma_q^2(\gamma)\sigma_q^2(\gamma')\}^{1/2}$, $\kappa_q(\gamma,\gamma') := \mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t']\mathbf{J}_q$ $\widetilde{\mathbf{\Sigma}}\mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma'}]$, $\sigma_q^2(\gamma) := \mathbb{E}[X_t^{\gamma}\widetilde{\mathbf{Z}}_t']\mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{\gamma}]$, *and for each* $j = 0,1,\ldots,q$, $\mathcal{Z}_q^2(j)$ *is the limit of* $\mathcal{D}_{n,q}^{(\gamma=j)}$*;*

(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* $\widetilde{\gamma} \in \mathbf{\Gamma}^{(q)}$, $n^{-1}\mathcal{D}_{n,q} = \mu_q^2(\widetilde{\gamma}) + o_{\mathbb{P}}(1)$ *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* $\mathcal{D}_{n,q} \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}^{(q)}}\{\mathcal{Z}_q(\gamma) + \nu_q(\gamma)\}^2$*, 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 prove them by iterating the proofs of Theorems 1 and 2. We summarize the proof as follows: first, for each $\epsilon > 0$, it follows that $\mathcal{D}_{n,q}^{(\beta=0)}(\epsilon) \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}^{(q,c)}(\epsilon)} \mathcal{Z}_q^2(\gamma)$ under $\mathcal{H}_{0,1}''$ by extending Lemma 1, where $\mathbf{\Gamma}^{(q,c)}(\epsilon) := \mathbf{\Gamma}^{(q)} \setminus \cup_{j=0}^{q}(j-\epsilon, j+\epsilon)$; 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 $s = 0,1,2,\ldots,q$,

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

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 for 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 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 depends on $\mathbf{\Gamma}^{(q)}$. We obtain different null limit distributions with different $\mathbf{\Gamma}^{(q)}$.

(c) Despite the different properties between $\mathcal{Z}_q(\cdot)$ 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 $\boldsymbol{\pi}_q(\cdot) := \mathbf{J}_q\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}]/\sigma_q^2(\cdot)^{1/2}$ consistently by its sample analog, letting $\widetilde{\mathcal{Z}}_q(\cdot) := \boldsymbol{\pi}_q(\cdot)\boldsymbol{\mathcal{U}}$ and simulating $\sup_{\gamma\in\mathbf{\Gamma}^{(q)}} \widetilde{\mathcal{Z}}_q^2(\gamma)$, where $\boldsymbol{\mathcal{U}} \sim N(\mathbf{0},\widetilde{\mathbf{\Sigma}})$ as before.

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

3.3 Sequentially Estimating Correct Polynomial Model

Corollary 1 provides a system basis for 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 $\mathcal{D}_n$ using $\mathcal{M}$ and compare it with the critical value $cv_1(\boldsymbol{\alpha}_n)$ in Corollary 1 at the level of $\boldsymbol{\alpha}_n$. If the $\mathcal{D}_n$ is less than or equal to $cv_1(\boldsymbol{\alpha}_n)$, 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 $cv_q(\boldsymbol{\alpha}_n)$ implied by the same level of significance $\boldsymbol{\alpha}_n$ as in given 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 $cv_q(\boldsymbol{\alpha}_n)$, we let the degree estimator be $\widehat{q}_n := \min\{q \in I(\bar{q}) : \mathcal{D}_{n,q} \leq cv_q(\boldsymbol{\alpha}_n)\}$.
- **Step 3**: If there is no $q \in I(\bar{q})$ such that $\mathcal{D}_{n,q}$ is less than or equal to $cv_q(\boldsymbol{\alpha}_n)$, 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 $\boldsymbol{\alpha}_n$ is set to depend on n. 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 n increases. Thus, the degree estimation error vanishes as n increases (*e.g.*, Cho and Phillips, 2018). Theorem 6 discusses how to choose $\boldsymbol{\alpha}_n$ in order to estimate q_* consistently:

Theorem 6. *Given that for each $q \in I(\bar{q})$, Assumptions 1 and 7 hold with $\mathbf{\Gamma}^{(q)}$ being $\mathbf{\Gamma}^{(\bar{q})}$,*

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

(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) $\lim_{n\to\infty} \boldsymbol{\alpha}_n = 0$, and (c) $\lim_{n\to\infty} \log(\boldsymbol{\alpha}_n)/n = 0$, then for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|\widehat{q}_n - q_*| > \epsilon) = 0$. □

Remarks.

(a) From Theorem 6(*i*), if $\boldsymbol{\alpha}_n$ does not converge to zero as n tends to infinity, the degree estimator does not vanish to zero. Theorem 6(*ii*) provides the conditions for $\boldsymbol{\alpha}_n$ 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 of the omnibus power of the DD-test for $q < q_*$.

(b) The regularity conditions in Theorem 6(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 6(ii) does not assume such a Gaussian process.

(c) Since the proof of Theorem 6(i) is straightforward from Corollary 1, we do not include it in Appendix. Theorem 6(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). □

4 Simulation

In this section, we simulate the DD-test and compare its performance with other tests and MSCs. We first compare their empirical sizes and powers and next their capability of estimating the unknown degree of the polynomial model. Finally, we examine the asymptotic size of the DD-test.

4.1 Empirical Size and Power

We suppose two simple DGPs and estimate the unknown parameters by GMM, to examine the empirical size properties of the DD-test. We proceed in the following steps:

- Step 1: We first generate data as follows:

$$Y_t = \beta_{1*} X_t + \beta_{2*} m(X_t) + U_t,$$

where $X_t := \sum_{j=1}^{4} Z_{tj} + U_t^2 \mathbf{1}(|U_t| \leq b_d)$ such that $U_t \sim$ IID $N(0,1)$, $Z_{t1} \sim$ IID $U(0,1)$, Z_{t2} and $Z_{t3} \sim$ IID Beta$(5,5)$, and $Z_{t1} \sim$ IID Beta$(5,3)$. Here, $\mathbf{1}(\cdot)$ is the indicator function, and U_t is bounded between $[-b_d, b_d]$ when defining X_t. The unspecified β_{1*}, β_{2*}, $m(\cdot)$, and b_d are going to be given below to characterize the null and alternative hypotheses. We denote this as DGP A. As our next DGP, we let $X_t := \sum_{j=1}^{4} Z_{tj} + U_t^2$ such that $U_t \sim$ IID $N(0,1)$, $Z_{t1} \sim$ IID Half-$N(0,1)$, $Z_{t2} \sim$ IID Beta$(5,5)$, and $Z_{t3} \sim$ IID Beta$(5,3)$, and $Z_{t3} \sim$ IID $\mathcal{X}_1^2$. We denote this as DGP B. For both DGPs, each of $Z_{t1}, \ldots, Z_{t4}$, and U_t is independently drawn. Note that X_t is always positively valued and correlated with U_t, whereas U_t is not correlated with Z_{t1}, …, Z_{t4}.

- Step 2: We estimate the unknown parameters by GMM by letting

$$\mathcal{M}_1^o := \left\{ m_t(\boldsymbol{\omega}) := Y_t - X_t \xi - \beta X_t^{\gamma} : \boldsymbol{\omega} \in \boldsymbol{\Omega} \subset \mathbb{R}^3 \right\},$$

so that $\mathbf{V}_t := X_t$. Here, we let $\gamma \in \boldsymbol{\Gamma} := [-0.25, 2.25]$, and this lets a quadratic model in X_t be nested in the $\mathcal{M}_1^o$. The other are parameters are not restricted. We also let $\mathbf{Z}_t := (Z_{t1}, \ldots, Z_{t4})'$ and $\mathbf{M}_n := (n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$. Here, we do not contain unity in $\mathbf{V}_t$ and $\mathbf{Z}_t$, because the PD matrix condition in Assumption 2(*iv*) does not hold by this. Here, we obtain the empirical size and power of the DD-test by applying Hansen's (1996) weighted bootstrap. Specifically, after estimating $\boldsymbol{\omega}_*$ by GMM, we let $\widehat{U}_t := m_t(\widehat{\boldsymbol{\omega}}_n)$ and compute

$$\widehat{\mathcal{G}}_b^2 := \sup_{\gamma \in \mathbf{\Gamma}} \left(\frac{1}{\sqrt{n}} \widehat{\boldsymbol{\pi}}_n(\gamma)' \sum_{t=1}^{n} \mathbf{Z}_t \widehat{U}_t G_t^{(b)} \right)^2$$

for $b = 1, 2, \ldots, B$, where for each b, $G_t^{(b)}$ is independently drawn from $N(0,1)$. Here, $\widehat{\boldsymbol{\pi}}_n(\cdot)$ is the sample analog of $\boldsymbol{\pi}_1(\cdot)$ defined in the remark below Theorem 1, and $n^{-1/2} \sum_{t=1}^{n} \mathbf{Z}_t \widehat{U}_t G_t^{(b)}$ corresponds to $\widehat{\boldsymbol{\mathcal{U}}}$. Instead of $\widehat{U}_t$, we can also use the residual generated by $\widetilde{\boldsymbol{\omega}}_n$. We finally compute the empirical p-value by $\widehat{p}_n := B^{-1} \sum_{b=1}^{B} \mathbf{1}(\widehat{\mathcal{G}}_b^2 > \mathcal{D}_n)$ and reject the linearity hypothesis of $\widehat{p}_n < \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is the level of significance.

- Step 3: We also apply other tests in the literature for comparison purpose. First, we apply Horowitz's (2006) test that requires that explanatory and instrumental variables have to be constrained on the unit interval. Furthermore, the number of instrumental variables has to be the same as that of the explanatory variables. So, the DGP and model conditions of our simulated data need to be modified accordingly. We redefine the instrumental variable by scaling down the sum of instrumental variables by the maximum value, viz., $\widetilde{X}_t := X_t / \max[X_1, \ldots, X_n]$ and $\widetilde{Z}_t := \sum_{j=1}^{4} Z_{tj} / \max[\sum_{j=1}^{4} Z_{1j}, \ldots, \sum_{j=1}^{4} Z_{nj}]$. Using them, we estimate the null model by GMM and apply his test. Following his recommendation, we estimate 25 largest eigenvalues of the covariance matrix estimator, to obtain the null limit distribution of his test. We let $\mathcal{H}_n$ denote his test. Second, we apply Breunig's (2015) test. Out of his two tests used for the simulation, we employ the second test.[2] We obtain the null limit distribution of his test by estimating 200 largest eigenvalues of the covariance matrix estimator following him, letting $\mathcal{B}_n$ denote his test. Finally, we apply J-test as defined by Sargan (1958, 1988) and Hansen (1982). Note that J-test can also be used to test correct model specification because the null hypothesis of the J-test does not hold unless the model is correctly specified. We denote it as $\mathcal{J}_n$. $\square$

We now report the size properties of the tests. For this purpose, we let $\boldsymbol{\beta}_* = (1, 0)'$ in DGP A so that a linear structural relationship holds between Y_t and X_t. We let $b_d = 1$. The simulation results are reported in the first panel of Table 1. The total number of experiments and B are $5,000$ and 500, respectively. The simulation results are summarized as follows:

(a) For every n of consideration, the DD-test exhibits empirical rejection rates more or less similar to the nominal significance levels, affirming Theorem 1. This aspect also implies that the DD-test controls type-I errors efficiently.

(b) Horowitz's, Breunig's, and Sargan's tests also control type-I errors efficiently. Although $\mathcal{B}_n$ suffers from size distortion for high levels of significance, it is not substantial. $\square$

We next examine the empirical size under DGP B. We let $\boldsymbol{\beta}_* = (1, 0)'$ and contain the simulation results in the second panel of Table 1. The simulation results are summarized as follows:

(a) The DD-test exhibits empirical rejection rates more or less similar to the nominal significance levels. When n is

[2] His second test outperforms the first test under our simulation environment.

small, the empirical rejection rates are slightly different from the nominal levels but the size distortion disappears soon as n increases.

(b) Horowitz's, Breunig's, and Sargan's tests control type-I errors efficiently. □

We next examine the empirical power properties. For this purpose, we first generate data according to the following two plans:

- DGP A′: $\boldsymbol{\beta}_* = (1, -0.4)'$, $m(X_t) = X_t^2$, and $b_d = 1$;
- DGP A″: $\boldsymbol{\beta}_* = (1, -0.4)'$, $m(X_t) = X_t^2$, and $b_d = 3$.

The explanatory variables and instrumental variables are generated according to DGP A. Note that $\mathcal{M}_1^o$ is correctly specified for both DGPs A′ and A″. The simulation results are reported in the first two panels of Table 2. They are obtained by letting the total number of experiments and B be $3,000$ and 500, respectively. The simulation results are summarized as follows:

(a) The DD-test shows consistent power, that is the same for the other tests.

(b) The DD-test has a comparable power against the other tests. When $b_d = 1$ (DGP A′), $\mathcal{H}_n$ is the most powerful and $\mathcal{J}_n$ is least favored, but this relationship is modified if b_d increases to 3 (DGP A″). That is, $\mathcal{D}_n$ becomes most powerful, but $\mathcal{J}_n$ is still least favored. This aspect implies that the power ranking among the tests depends on the size of b_d. □

We next examine the empirical power under DGP B under the same simulation environment as for the previous one. For this, we generated data according to following two DGPs:

- DGP B′: $\boldsymbol{\beta}_* = (1, 1)'$ and $m(X_t) = \tanh(-X_t/2)$;
- DGP B″: $\boldsymbol{\beta}_* = (1, 2)'$ and $m(X_t) = 2|\sin(-X_t/5)|$.

The explanatory variables and instrumental variables are generated according to DGP B. Note that $\mathcal{M}_1^o$ is misspecified for both DGPs B′ and B″ contrary to the earlier simulation. The simulation results are reported in the final two panels of Table 2. We summarize the simulation results as follows:

(a) The DD-test shows consistent power, that is the same for the other tests. This is consistent with Theorem 2.

(b) The DD-test is most powerful, and $\mathcal{B}_n$ is least powerful for both DGPs B′ and B″. □

In addition to the power examinations in Table 2, we conducted power simulations using data obtained by different DGPs and models. We describe our experiences as follows: First, when $\mathcal{M}_1^o$ is misspecified, if $\boldsymbol{\Gamma}$ is selected to be too narrow, the overall power of the DD-test is not so great as that with a bigger $\boldsymbol{\Gamma}$. Unless the DD-test suffers from size distortion, it is recommended to select a moderately wide interval for $\boldsymbol{\Gamma}$. Second, for different DGPs, we could observe power rankings different from that in Table 2. Under different DGP conditions, any of the four test statistics could be most powerful. It is challenging to rank the powers of the tests under a generic DGP condition. Third, the regularity conditions of the four tests are different. For example, the DD-test requires X_t to be positive, and further the rank condition needs to be satisfied. In contrast, Horowitz's (2006) test requires the number of the explanatory variables is identical to that of the instrumental variables and further that they are defined on the unit interval. Similarly, Breunig's (2015) test performs depending on the supports of the explanatory variables and instrumental variables. These aspects say that they can supplement each other.

4.2 Application to the Sequential Testing Procedure

We next examine the DD-test and its application to the sequential testing procedure by simulation and compare it with J-test and MSCs.

We conduct simulations according to the DGP and model conditions as given in the following plan:

- Step 1: We assume that $(D_t, U_t)' \sim$ IID $N(\mathbf{0}, \mathbf{I}_2)$ and generate the following 11 instrumental variables: $Z_{t1} \sim$ IID $U(0,1)$, Z_{t2} and $Z_{t3} \sim$ IID $\mathcal{X}_1^2$, Z_{t4} and $Z_{t5} \sim$ IID Rayleigh(1), Z_{t6} and $Z_{7t} \sim$ IID Half-$N(0,1)$, Z_{t8} and $Z_{t9} \sim$ IID Beta$(5,3)$, and Z_{t10} and $Z_{t11} \sim$ IID Beta$(5,5)$. Each of D_t, U_t, Z_{t1}, $\ldots$, Z_{t11} is independently distributed, and all of Z_{t1}, $\ldots$, Z_{t11} are positively valued. We further let $X_t := \sum_{j=1}^{11} Z_{tj} + U_t^2$, so that X_t is also positively valued with probability 1, and X_t and U_t are correlated, but U_t is not correlated with $Z_{t1}, \ldots, Z_{t10}$, and Z_{t11}.
- Step 2: We consider the following structural equation:

$$Y_t := \beta_{1*} D_t + \beta_{2*} X_t + \beta_{3*} X_t^2 + U_t,$$

  so that if we let $\boldsymbol{\beta}_* = (1, 1, 0.005)'$, Y_t is quadratically associated with X_t.
- Step 3: We estimate $\boldsymbol{\beta}_*$ by GMM. For this, we let our models be specified as follows: for each $q \in I(3) := \{1, 2, 3\}$,

$$\mathcal{M}_q' := \left\{ m_{t,q}(\boldsymbol{\omega}^{(q)}) := Y_t - D_t\eta - X_t\xi_1 - \ldots - X_t^q \xi_q - \beta X_t^\gamma : \boldsymbol{\omega}^{(q)} \in \boldsymbol{\Omega}^{(q)} \subset \mathbb{R}^{3+q} \right\},$$

  where $\boldsymbol{\omega}^{(q)} := (\xi_1, \ldots, \xi_q, \eta, \beta, \gamma)'$, and $\boldsymbol{\Omega}^{(q)}$ is the parameter space of $\boldsymbol{\omega}^{(q)}$. In particular, we assume that the parameter space of γ is $\boldsymbol{\Gamma} = [0.50, 3.50]$, so that the third-degree polynomial model is nested in $\mathcal{M}_q'$ for $q = 2, 3$. The other parameter spaces are not restricted. Given this model assumption, we also let $\mathbf{Z}_t := (D_t, Z_{t1}, Z_{t2}, \ldots, Z_{t11})'$, $\mathbf{M}_n = (n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$, and for each $q \in I(3)$, $\mathbf{V}_{t,q} := (D_t, X_t, \ldots, X_t^q)$. The GMM estimator is obtained by minimizing the GMM distance for each $q \in I(3)$. □

Note that $\mathcal{M}_2'$ and $\mathcal{M}_3'$ are correctly specified models, and $\mathcal{M}_2'$ is the most parsimonious model. Therefore, the main goal of the sequential testing procedure is achieved when $q_* = 2$ is consistently estimated.

Given the DGP and model conditions, we perform our simulations in the following three steps:

- Step 1: Using the DD-test, we test whether the structural model is correctly specified. Here, we apply Hansen's (1996) weighted bootstrap as in Section 4.1. The bootstrap iteration B is 300, and we fix the significant level at 10%, 5%, and 1%. We also apply J-test.
- Step 2: By letting the significance level decrease as n increases, we can apply Theorem 6(*ii*). Specifically, we let $\boldsymbol{\alpha}_n = 1/n^{1/2}$, $1/n^{3/4}$, and $1/n$ and examine how the degree estimation error is formed. Note that the significance levels converge to zero by these level plans but the convergence rate of $\boldsymbol{\alpha}_n = 1/n$ is faster than the others. We

also apply the sequential testing procedure to the J-test. We call them the DD- and J-sequential testing procedures, respectively.

- Step 3: 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

$$\text{AIC-MSC} := \bar{\mathcal{J}}_{n,q} - 2(p-q-1)/n, \qquad \text{Bayesian-MSC} := \bar{\mathcal{J}}_{n,q} - \log(n)(p-q-1)/n,$$

$$\text{Hannan-Quinn-MSC} := \bar{\mathcal{J}}_{n,q} - \kappa \log(\log(n))(p-q-1)/n,$$

respectively, where $\bar{\mathcal{J}}_{n,q} := n^{-1}\mathcal{J}_{n,q}$; and we let κ be 2.01 following Andrews (1999). The model performing best is the one with the smallest MSC. □

We iteratively perform this three-step simulations and report the simulation results in Tables 3 and 4. Table 3 presents the results obtained through Step 1, and Table 4 reports the results obtained by Steps 2 and 3. Specifically, they report the precision rate of each method. For example, if $\widehat{q}_{n,r}$ denotes the degree estimated by the r^{th}- simulation, the precision rate is computed by $R^{-1}\sum_{r=1}^{R}\mathbf{I}(\widehat{q}_{n,r} = q_*) \times 100$, where R is the total number of experiments. We let R be 3,000.

The simulation results in Table 3 are summarized as follows:

(a) When the significance level is fixed, the degree estimator obtained by the DD-sequential testing procedure yields the results predicted by Theorem 6(i). If the significance level is fixed at $\boldsymbol{\alpha}$, the estimated precision rate converges to $(1-\boldsymbol{\alpha}) \times 100$ for $q = 2$ as n increases.

(b) Similar results are obtained for the J-sequential testing procedure. Nevertheless, we also note that the J-sequential testing procedure is asymptotically more conservative than the DD-sequential testing procedure. For example, when $n = 4,500$ and $\boldsymbol{\alpha}$ = 10%, the J-sequential testing procedure produces more precise estimation results than the DD-sequential testing procedure. It is mainly because the J-test is more conservative than the DD-test. In other words, it is more difficult to control type-I error. □

We now examine the sequential testing procedures obtained by letting the significance levels depend on n. Table 4 reports the simulation results of each estimation method when $\boldsymbol{\alpha}_n = n^{-1/2}$, $n^{-3/4}$, and n^{-1}. We summarize the simulation results as follows:

(a) As n increases, the estimation errors decrease by applying the DD-sequential testing procedure. Furthermore, for any significance level, smaller estimation errors are observed for the data sets with larger n, so that the degree estimation errors based on $\boldsymbol{\alpha}_n = n^{-1}$ are smaller than the others.

(b) Likewise, the J-sequential testing procedure also estimates q_* consistently. Nevertheless, we note that the DD-sequential testing procedure better controls the precision rate. Here, the hypothetical rate defined as $(1-\boldsymbol{\alpha}_n) \times 100$ denotes the precision rate desired by each sequential testing procedure. Nevertheless, note that the precision rates produced by the DD-sequential testing procedure are closer to the hypothetical rates than the J-sequential testing procedure.

(c) As n increases, the estimation errors of using MSCs also decrease. The Bayesian-type MSC estimates q_* more efficiently than the other two MSCs.

(d) For a small n, the DD-sequential testing procedure performs better than the J-sequential testing procedure, but this is not true for every $\boldsymbol{\alpha}$ and n. For example, for $\boldsymbol{\alpha}_n = n^{-1/2}$, if n increases, the J-sequential testing procedure estimates q_* more precisely than the DD-test. It is mainly because the J-test is more conservative than the DD-test. However, for $\boldsymbol{\alpha}_n = n^{-1}$, this dominance relationship is reversed as n increases. The DD-test is better controlled, so that the estimation error from the DD-sequential testing procedure shows more precise rates than the J-sequential testing procedure. □

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

In addition to the reported simulations, we also conducted different simulations using different DGP and model conditions, producing different simulation results depending on the choice of significance level. Nonetheless, the overall appeal of the DD-test remains still effective as in Tables 3 and 4.

4.3 Asymptotic Uniform Inference

In this section, we examine the asymptotic size of the DD-test by simulation. Specifically, we provide simulation evidence that

$$\lim_{n\to\infty} \sup_{\boldsymbol{\omega}_*} \mathbb{P}_{\boldsymbol{\omega}_*}[\mathcal{D}_n > cv_n(\boldsymbol{\alpha})] = \lim_{n\to\infty} \inf_{\boldsymbol{\omega}_*} \mathbb{P}_{\boldsymbol{\omega}_*}[\mathcal{D}_n > cv_n(\boldsymbol{\alpha})] = \boldsymbol{\alpha}$$

under $\mathcal{H}_0' : \beta_* = \beta_0$ against $\mathcal{H}_1' : \beta_* \neq \beta_0$, when the model $\mathcal{M}$ in Section 2.2 is specified.

We proceed with our simulations in the following steps:

- Step 1: We generate data according to the following DGP condition:

$$Y_t = \xi_* X_t + \beta_* X_t^{\gamma_*} + U_t,$$

  where $X_t := \prod_{j=1}^{12} Z_{tj} \sum_{j=1}^{12} Z_{tj} + U_t^2$ such that Z_{t1} and $Z_{t2} \sim$ IID $U[0,1]$, Z_{t3} and $Z_{t4} \sim$ IID Beta$[5,3]$, Z_{t5} and $Z_{t6} \sim$ IID Beta$[5,5]$, Z_{t7} and $Z_{t8} \sim$ IID $\mathcal{X}_1^2$, and $Z_{t9}, \ldots, Z_{t12} \sim$ Half-$N(0,1)$. Each of $Z_{t1}, \ldots, Z_{t12}$, and U_t is independently drawn, and all of $Z_{t1}, \ldots, Z_{t12}$ are positively valued.

- Step 2: Given the DGP condition, we let our model be defined as follows:

$$\mathcal{M}'' := \{m_{t,q}(\boldsymbol{\omega} := Y_t - X_t\xi - \beta X_t^{\gamma} : \boldsymbol{\omega} \in \boldsymbol{\Omega}\}$$

  with $\boldsymbol{\omega} := (\xi, \beta, \gamma)'$, and $\boldsymbol{\Gamma} := [0.50, 3.50]$. We also let $\mathbf{Z}_t$ and $\mathbf{M}_n$ be $(Z_{t1}, \ldots, Z_{t12})$ and $(n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$, respectively. Here, the parameter values are specified as $\xi_* = 1$, $\beta_* \in \{-0.75, -0.5, 0.25, 0.00, 0.25, 0.50, 0.75\}$ and $\gamma_* \in \{0.50, 0.75, 1.00, 1.25, 1.50\}$.

- Step 3: We compute the empirical rejection rates of the DD-test under $\mathcal{H}_0' : \beta_* = \beta_0$ using the above data. For this computation, we separately consider the models with and without the identification problem. If $\beta_* = 0.00$ or $\gamma_* = 1.00$, the model $\mathcal{M}''$ is not identified. We therefore test the linear model hypothesis by using the null limit distribution in Theorem 1 and by applying Hansen's (1996) weighted bootstrap described in Section 4.1. That is, we obtain the asymptotic critical value by $cv_n(\boldsymbol{\alpha}) := \inf\{x \geq 0 : \widehat{F}_B(x) \geq 1 - \boldsymbol{\alpha}\}$, where $\widehat{F}_B(\cdot)$ is the empirical distribution of $\{\widehat{\mathcal{G}}_1^2, \ldots, \widehat{\mathcal{G}}_B^2\}$. On the other hand, for $\beta_* \neq 0$ and $\gamma_* \neq 1$, the model $\mathcal{M}''$ is now identified. Thus, we approximate the DD-test as (12) and we next apply Hansen's (1996) weighted bootstrap similarly to testing the linear model hypothesis. Specifically, for $b = 1, \ldots, B$, we first let

$$\widehat{\mathcal{G}}_b^2 = \frac{1}{n} \frac{\{\mathbf{X}(\widehat{\gamma}_n)'(\mathbf{Q}_1 - \mathbf{Q}_1\mathbf{D}(\widehat{\gamma}_n)(\mathbf{D}(\widehat{\gamma}_n)'\mathbf{Q}_1\mathbf{D}(\widehat{\gamma}_n))^{-1}\mathbf{D}(\widehat{\gamma}_n)'\mathbf{Q}_1)\ddot{\mathbf{U}}_b\}^2}{\mathbf{X}(\widehat{\gamma}_n)'(\mathbf{Q}_1 - \mathbf{Q}_1\mathbf{D}(\widehat{\gamma}_n)(\mathbf{D}(\widehat{\gamma}_n)'\mathbf{Q}_1\mathbf{D})^{-1}\mathbf{D}(\widehat{\gamma}_n)'\mathbf{Q}_1)\mathbf{X}(\widehat{\gamma}_n)}$$

by following (12), where $\ddot{\mathbf{U}}_b := [\widetilde{U}_1 G_1^{(b)}, \ldots, \widetilde{U}_n G_n^{(b)}]'$, $\widetilde{U}_t := m(\widetilde{\boldsymbol{\omega}}_n)$, and $G_t^{(b)}$ is independently drawn from $N(0,1)$ with respect to t and b. From this, we obtain the asymptotic critical value as for the linearity testing case. □

Table 5 reports the simulation results that are obtained from the data with each combination of ξ_*, β_* and γ_*. The first and second panels are obtained by letting $n = 500$ and $5,000$, respectively. The level of significance is $\boldsymbol{\alpha} = 5\%$. The simulation results are summarized as follows:

(a) For $n = 500$, it is not quite clear that the empirical rejection rate is close to 5% uniformly on the parameter space of consideration. It is evident that the empirical rejection rate is close to 5% when $\beta_* = 0.00$ or $\gamma_* = 1.00$; or when (β_*, γ_*) is quite different from $(0.00, 1.00)$. On the contrary, if β_* or γ_* is close to 0.00 or 1.00, respectively, the empirical rejection rate is quite different from 5%. For example, if $(\beta_*, \gamma_*) = (0.75, 1.25)$, the empirical rejection rate is obtained as 4.67, whereas if $(\beta_*, \gamma_*) = (0.25, 1.25)$, the empirical rejection rate is obtained as 1.53. This aspect implies that the DD-test can have a finite sample size distortion when the parameters are close to $\boldsymbol{\Upsilon}_0$.

(b) For $n = 5,000$, the size distortion of the DD-test substantially reduces. Most empirical rejection rates are close to 5%, and this feature is observable even for the parameters close to the linear model. For example, if $(\beta_*, \gamma_*) = (0.25, 1.25)$, the empirical rejection rate is obtained as 5.10, that is quite different from $n = 500$. The only exceptional cases are $(\beta_*, \gamma_*) = (-0.25, 0.75)$ and $(0.25, 0.75)$ as their empirical rejection rates are 2.13 and 2.00, respectively, but their respective empirical rejection rates are 3.40 and 4.40 when $n = 10,000$, implying that the finite sample size distortion further reduces as n further increases. For the other significance levels 1% and 10%, we could obtain similar results.

(c) For finite n, the DD-test can be usefully exploited if the researcher wishes a conservative test. It is mainly because of the unidentified model feature and the fact that the DD-test is constructed by the LR-test principle. The intuition is straightforward. If we concentrate $d_n(\boldsymbol{\omega})$ with respect to $(\xi_0, \boldsymbol{\delta})$ to obtain the concentrated GMM estimator $(\widehat{\xi}_{0,n}(\beta,\gamma), \widehat{\boldsymbol{\delta}}_n(\beta,\gamma))$ and draw $d_n(\beta,\gamma) := d_n(\widehat{\xi}_{0,n}(\beta,\gamma), \widehat{\boldsymbol{\delta}}_n(\beta,\gamma), \beta, \gamma)$ as a function of (β,γ), it becomes a very flat function on the space of (β,γ) under the null of linearity, that is a consequence of the multifold identification problem. This fact is still effective even if (β_*, γ_*) is close to $\mathbf{\Upsilon}_0$ without belonging to $\mathbf{\Upsilon}_0$. That is, $d_n(\cdot,\cdot)$ is still quite a flat function, although it is minimized at (β_*, γ_*) at the limit. This flat function implies that $d_n(\cdot,\cdot)$ is poorly approximated by a quadratic function for finite n, so that when testing $\mathcal{H}_0' : \beta_* = \beta_0 (\neq 0)$, the DD-test statistic measuring the distance of two GMM distances is likely to be smaller than the asymptotic critical value obtained by approximating $d_n(\cdot,\cdot)$ through Taylor's expansion, so that $\mathbb{P}_{(\beta_*,\gamma_*)}(\mathcal{D}_n > cv_n(\boldsymbol{\alpha})) \leq \boldsymbol{\alpha}$, as revealed by the current simulation. Note that for $n = 500$, the empirical rejection rates of the DD-test are less than 5% for most (β_*, γ_*)'s around $(0.00, 1.00)$ but get to close 5% from below as n increases to $5,000$. That is, if n is finite, the type-I error can be controlled at a level less than or equal to $\boldsymbol{\alpha}$ around $\mathbf{\Upsilon}_0$. This is certainly an advantage of using the current methodology. In contrast, if a simple polynomial model is instead specified without the power transform to test linear versus quadratic models say, the unidentified model feature cannot be exploited any longer to test the coefficient of the quadratic term, implying that if the coefficient is close to zero, the finite sample type-I error can be quite different from $\boldsymbol{\alpha}$ in an unexpected way as Leeb and Pötscher (2005) illustrate using a simple linear model example. □

These results provide simulation evidence that Hansen's (1996) weighted bootstrap is useful for the DD-test to become a valid testing procedure uniformly on the assumed parameter space.

5 Production Function Estimation Using Firm-Level Data

Recently, the literature has witnessed large distributional consequences of shares across different factors for production. For example, there was a large rise in wage inequality between skilled and unskilled workers and also a decline in labor shares over capital shares. Karabarbounis and Neiman (2014) and Piketty (2014) empirically examine the decline of the labor share; Krusell et al. (2000) and Acemoglu and Restrepo (2018) report wage inequality between skilled and unskilled workers.

The distributional consequence of factor shares is attributed to the factor-biased technological change. The studies mentioned above argue that the recent technological changes have favored some factors over others, ensuing the recently discovered large distributional consequence across factors. For example, Krusell et al. (2000) attribute the increase in wage inequality between skilled and unskilled workers to the skill bias technological change.

Behind the factor-biased technological change, the key assumption lies in the fact that the production function is log-nonlinear in factors, that cannot be related to the typically assumed Cobb-Douglas production technology. Cobb-Douglas function implicitly assumes that any technological change is Hicks-neutral instead of being factor-biased.[3] Therefore, any technological change under Cobb-Douglas technology leads to a proportional increase in the output obtained from any combination of inputs, so that the technological change cannot be related to the distributional consequence. Meanwhile,

[3] Production function $F(L, K, A)$ is said to exhibit labor-augmenting (resp. capital-augmenting) technology if $\frac{\partial}{\partial A}\left(\frac{\partial F(L,K,A)/\partial L}{\partial F(L,K,A)/\partial K}\right) >$ (resp. $<$) 0 (*e.g.*, Acemoglu, 2008), where L, K, and A are labor, capital and technology shock, respectively. The log-linearity property of Cobb-Douglas function implies that $\frac{\partial}{\partial A}\left(\frac{\partial F(L,K,A)/\partial L}{\partial F(L,K,A)/\partial K}\right) = 0$.

the factor-biased technological change affects the effective unit of one factor disproportionately relative to other factors, from which the large distributional consequence follows across factors. Therefore, it is important to affirm whether the production technology is well approximated by Cobb-Douglas function or not in evaluating the prior studies in terms of the factor-biased technological change.

The DD-test can be usefully exploited for this purpose. Note that the DD-test can straightforwardly test the log-linearity of Cobb-Douglas production function. If Cobb-Douglas production technology cannot be rejected, the recent studies need to be carefully revisited as they may incorrectly attribute the recent rise in wage inequality and decline in labor shares to the factor-biased technological change. Otherwise, we cannot reject the factor-biased technological change as a potential explanation of the recently discovered distributional conse quence of factor shares.

We consider the following Cobb-Douglas production function with labor and capital:

$$\log(Y_t) = \beta_{l*} \log(L_t) + \beta_{k*} \log(K_t) + \log(A_t), \tag{13}$$

where Y_t is the output of firm t measured by value-added, L_t is the labor input, K_t is the capital stock, and A_t is the productivity shock. We test the output is produced according to the log-linear technology in labor and capital. Our null and alternative hypotheses are stated as follows:

$\mathcal{H}_0^\dagger$: Production function is log-linear in factors, viz., Cobb-Douglas. vs. $\mathcal{H}_1^\dagger$: Production function is not log-linear.

Note that (13) suffers from the fundamental endogeneity issue, that is well known since Griliches and Mairesse (1995). The endogeneity problem arises because L_t and K_t are endogenously selected by firm t based on A_t that is unobservable to the empirical researcher, although it is observable to the firm. The literature calls this the input bias problem. As detailed below more, it is standard in prior literature to treat K_t as a dynamic input, because it is pre-determined in the current period by a pre-existing investment plan that makes estimating the coefficient of log capital free from the input bias problem (*e.g.*, Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015).

There are prior empirical studies for the same purpose, and our approach using the DD-test is different from them. Most prior studies directly estimate a prespecified production function that treats Cobb-Douglas technology as a special case. For example, Antràs (2004), Raval (2019) and Oberfield and Raval (2021) draw empirical economic implications by estimating the constant elasticity substitution (CES) production technology. Note that CES production function is log-nonlinear and nests Cobb-Douglas production technology as a special case by letting the elasticity of substitution converge to zero, extending the model scope assumed by Cobb-Douglas function. Nevertheless, it is also possible that the assumed CES model is still misspecified, letting the misspecification play a certain role in testing Cobb-Douglas production technology. The DD-test approach is different from the prior empirical studies. Note that the DD-test can detect any log-nonlinearity without imposing a specific structure to the production function because of its omnibus power against arbitrary nonlinearity. We further apply the sequential testing procedure based upon the DD-test and estimate the production function supported by empirical data. We also attempt to draw economic implications from this empirical analysis.

We specifically apply the DD-test to the control function approach to estimate the production function (*e.g.*, Olley and Pakes, 1996; Levinsohn and Petrin, 2003). For this purpose, we impose more structural assumptions on the production function in (13) to estimate the following production function: for each firm t,

$$\log(Y_t) = \beta_{l*} \log(L_t) + \beta_{k*} \log(K_t) + U_t \ \text{ such that } \ U_t := W_t + V_t. \tag{14}$$

The error term U_t has two components. First, we let V_t be an IID error term to which the firm does not respond by supposing that it captures measurement or specification error. Next, W_t is a firm-specific time varying productivity shock that is observable to the firm but unavailable information to the empirical researcher. We let W_t introduce the estimation bias by supposing that the firm chooses its static inputs labor L_t after observing W_t, that in turn makes $\log(L_t)$ be correlated with the error term. We assume that W_t follows the first-order Markov process, viz., $W_t = \mathbb{E}[W_t|W_t^{(-1)}] + \varepsilon_t$, where ε_t is an innovation in the current period and the superscript '(-1)' is used to denote the first-lagged W_t. For example, for $j = 1, 2, \ldots$, $W_t^{(-j)}$ denotes the j-lagged W_t with respect to time index. We also suppose that K_t is a dynamic input that is adjusted with one-period lag by noting that the investment made in the previous period increases the capital stock in the current period. We also assume that ε_t is realized after the firm first makes its investment decision in the previous period. From this supposition, $\log(K_t)$ is uncorrelated with ε_t, whereas it is correlated with W_t mainly because the firm makes the investment decision in the previous period based on its anticipation on W_t conditional on $W_t^{(-1)}$, viz., $\mathbb{E}[W_t|W_t^{(-1)}]$.

Using the control function approach, the input bias problem can be resolved. Following Levinsohn and Petrin (2003), we assume that log material input $\log(M_t)$ is a proxy variable for W_t, implying that the material input can be written as $\log(M_t) = m(W_t, \log(K_t))$ such that $m(\cdot)$ is strictly increasing with respect to W_t for each value of $\log(K_t)$. Given the strict monotonicity of $m(\cdot)$ with respect to W_t, it is not difficult to show that for some function $g(\cdot)$, $W_t = g(\log(M_t), \log(K_t))$, so that W_t can be written as a function of the observables $\log(M_t)$ and $\log(K_t)$. As W_t follows the first-order Markov process, we can express $\mathbb{E}[W_t|W_t^{(-1)}]$ as follows: for some function $f(\cdot)$,

$$\mathbb{E}[W_t|W_t^{(-1)}] = f(W_t^{(-1)}) = f(g(\log(M_t^{(-1)}), \log(K_t^{(-1)}))) =: h(\log(M_t^{(-1)}), \log(K_t^{(-1)})), \tag{15}$$

where $M_t^{(-1)}$ and $K_t^{(-1)}$ denote the first-lagged M_t and K_t, respectively. By substituting (15) into (14), we can rewrite the production function as $\log(Y_t) = \beta_{l*} \log(L_t) + \beta_{k*} \log(K_t) + h(\log(M_t^{(-1)}), \log(K_t^{(-1)})) + \varepsilon_t + V_t$. Here, we can nonparametrically control $h(\log(M_t^{(-1)}), \log(K_t^{(-1)}))$ using $(\log(M_t^{(-1)}), \log(K_t^{(-1)}))$. For example, Wooldridge (2009) approximates $h(\log(M_t^{(-1)}), \log(K_t^{(-1)}))$ by polynomials of $\log(M_t^{(-1)})$ and $\log(k_t^{(-1)})$. As another example, Dhyne et al. (2017) uses the first-order approximation of $h(\log(M_t^{(-1)}), \log(K_t^{(-1)}))$ viz., $h(\log(M_t^{(-1)}), \log(K_t^{(-1)})) \approx \gamma_{m*} \log(M_t^{(-1)}) + \gamma_{k*} \log(K_t^{(-1)})$. By substituting this approximate into the production function, we derive the following production function:

$$\log(Y_t) = \beta_{l*}\log(L_t) + \beta_{k*}\log(K_t) + \gamma_{m*}\log(M_t^{(-1)}) + \gamma_{k*}\log(K_t^{(-1)}) + \varepsilon_t + V_t, \tag{16}$$

that we now regard as our regression model. Once we condition out W_t using the lagged proxy variable and the lagged capital stock, $\log(K_t)$ becomes uncorrelated with the error term because K_t is a dynamic input that cannot be adjusted contemporaneously to the innovation ε_t, whereas $\log(L_t)$ is a static input that can be flexibly adjusted by the firm after it observes ε_t, letting the estimation of β_{l*} be subject to the input bias problem. Therefore, $\log(Y_t)$, $\log(L_t)$, $[\log(K_t), \log(M_t^{(-1)}), \log(K_t^{(-1)})]'$, and $\varepsilon_t + V_t$ correspond to Y_t, X_t, $\mathbf{D}_t$, and U_t, respectively in terms of the notations in Section 2.2.

We overcome the input bias problem by employing the GMM estimator with the weighting matrix assuming conditional homoskedastic error on the instrumental variable, so that the GMM estimator becomes equivalent to the two-stage least squares estimator. Note that the model structure provides a set of valid instrumental variables. Specifically, for each $j = 1, 2, \ldots$, $\mathbb{E}[\log(L_t^{(-j)})(\varepsilon_t + V_t)] = 0$, so that a large set of instrumental variables can be constructed by flexibly employing the lagged labor as valid instrumental variables. Using these instrumental variables, we can apply the DD-test for a desired empirical inference.

We detail the data structure for empirical application. Compustat data are used for the estimation that cover 2,140 public firms in the United States in the year of 2019. The variables of value-added Y_t, employment L_t, material input M_t, and capital stock K_t are constructed by following İmrohoroğlu and Tüzel (2014) without missing observations. In Appendix, we provide more detailed information on the data construction. Furthermore, the observations in the data set trivially satisfy the positive endogenous variable condition because the firm-level data set covers the firms with more than a single employee. For the instrumental variables, we specifically let them be 3 lagged log labors or their squares as follows:

$$\mathbf{Z}_t := \left[\log(L_t^{(2016)}), \log(L_t^{(2017)}), \log(L_t^{(2018)}), \log^2(L_t^{(2016)}), \log^2(L_t^{(2017)}), \log^2(L_t^{(2018)})\right]',$$

where for example, $L_t^{(2016)}$ denotes the employment of the t-th firm in the year of 2016. We below test whether the selected instrumental variables are strong enough to apply the DD-test using Kleinbergen and Papp's (2006) and Stock and Yogo's (2005) F-tests.

Table 6 reports the OLS and GMM estimates of (13). We present the OLS estimates results in columns (1)-(2), and the GMM estimates results in columns (3)-(4). In columns (1) and (3), we report the estimation results of Cobb-Douglas production function by OLS and GMM, respectively. Column (4) assumes that Cobb-Douglas production technology is misspecified with respect to the endogenous variable $\log(L_t)$ and remedies the misspecification by adding its square term on the right side. We summarize the estimation results as follows:

(a) The OLS estimates of $\log(L_t)$ are slightly bigger than the corresponding GMM estimates. The direction of the biases is consistent with the model assumption that labor can be adjusted contemporaneously to the innovation ε_t.

Because the firm with positive ε_t uses more labor inputs, it leads to the upward bias of the OLS estimate.

(b) At the bottom panel of Table 6, Kleinbergen and Papp's (2006) F-test is reported. The test values are bigger than the rule-of-thumb value 10 for the models in column (3) and (4) (cf., Staiger and Stock, 1997), implying that the instrumental variables do not suffer from the weak instrumental variable problem. When Cragg and Donald's (1993) F-test is applied to the models in column (3) and (4), they are obtained as 11,320 and 10,618, respectively. These values are sufficiently bigger than the critical values of 5% level of significance in Stock and Yogo (2005) that are 19.28 and 15.72, respectively. This reaffirms that the instrumental variables are not weak.

(c) At the bottom panel of Table 6, the DD-test is provided and it rejects the hypothesis of Cobb-Douglas production technology given in column (3), whereas it does not reject for the model in column (4). This implies that adding the power transform of $\log(L_t)$ on the right side does not reduce the GMM distance significantly.

(d) The J-test at the bottom panel also rejects the model in columns (3) but does not, for the model in column (4), that is consistent with the DD-test. Although the J-test in column (3) does not say why the orthogonality condition violates, the DD-test ascribes the violation to the model misspecification.

(e) The negative coefficient of the quadratic log labor term in column (4) implies that the labor-augmenting technological change leads to the decline in the labor share. The negative coefficient indicates that an increase in the labor-augmenting technology decreases the marginal revenue product of labor (MRPL) relative to the marginal revenue product of capital (MRPK), as our estimate implies that $\text{MRPL}_t/\text{MRPK}_t = (0.80 - 2 \times 0.01 \times \log(A_t^L L_t))$, where A_t^L is the labor-augmenting productivity shock. This makes firms substitute more toward capital and in turn decreases the ratio of labor expenditure to the value-added, so that the labor-augmenting technological change with our preferred functional form explains the recent decline of labor shares. $\square$

From this empirical analysis applying the DD- and J-tests sequentially, we essentially conclude that Cobb-Douglas production technology misspecifies the firm-level production function in the United States, that is remedied by adding $\log^2(L_t)$ to Cobb-Douglas production technology. Furthermore, the estimated production technology is consistent with the recently discovered decline of labor shares from the empirical literature.

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 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. 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 by simulation. Third, we provide an empirical illustration by investigating the relationship between the value-added and its production factors using firm-level data from the United States. Using the DD-test, we affirm that the production function has exhibited factor-biased technological changes instead of Hicks-neutral technology presumed by Cobb-Douglas production function.

A Appendix

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof of Lemma A1: (*i*) $\mathbf{Z}'\mathbf{U} = [\sum_t Z_{tj}U_t]$. Since $\mathbb{E}[Z_{tj}^2U_t^2] < \mathbb{E}[Z_{tj}^4]^{1/2}\mathbb{E}[U_t^4]^{1/2}$ 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 assume 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 $\mathbb{E}[|R_{tj}Z_{ti}|] < \infty$, that 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 $\{\mathbf{Z}_tU_t\}$ 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:

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

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

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

where

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

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

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

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

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

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,

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

(*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 A1, 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

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

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

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

If we reorganize the terms according to their order of convergence by applying Lemmas A1 and A5 and the fact $\mathbb{E}[\mathbf{Z}_tU_t] = 0$, we obtain

- $-2\zeta'\{\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1\mathbf{F}^{-1}\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_2'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\}\mathbf{V}\boldsymbol{\zeta} + 4\zeta'[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}\overline{\mathbf{K}}_1'\ddot{\mathbf{P}}\boldsymbol{\zeta} - 2\zeta'[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\mathbf{F}^{-1}[\ddot{\mathbf{P}}'\overline{\mathbf{K}}_1 + \overline{\mathbf{K}}_1'\ddot{\mathbf{P}}]\boldsymbol{\zeta} +$

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

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

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

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

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

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 A1(*i*), 3, and A5(*i*) imply that $\mathbf{C}_1'\mathbf{Q}_1\mathbf{U}=O_{\mathbb{P}}(n^{3/2})$. Similarly, Lemma A5(*i*) and Assumption 3 imply that $\mathbf{C}_1'\mathbf{Q}_1\mathbf{C}_1=O_{\mathbb{P}}(n^2)$. Furthermore, Lemmas A1(*iv*), A5(*i* and *ii*) 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 $\mathcal{D}_n^{(\gamma=1;\beta)}=O_{\mathbb{P}}(1)$. ■

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

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

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

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

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 uniform law of large numbers (ULLN) to each row of $\{n^{-1/2}\sum_{t=1}^n X_t^\gamma\mathbf{Z}_t\}$, so that for each j, we have

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

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, $\mathbb{E}[|X_t^{\gamma}Z_{t,j}|]\le\mathbb{E}[M_t^2]<\infty$ 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,

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

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 $\mathbf{M}_n\overset{\mathbb{P}}{\to}\mathbf{M}_0$ and $n^{-1}\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{V}_t'\overset{\mathbb{P}}{\to}\mathbb{E}[\mathbf{Z}_t\mathbf{V}_t']$ 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:

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

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

For the second part, note that

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

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,

$$\operatorname*{plim}_{\gamma\to 0} N_n^{(2)}(\gamma) = \operatorname*{plim}_{\gamma\to 0} 2\{(d/d\gamma)\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2 + 2\{\mathbf{X}(\gamma)'\mathbf{Q}_1(d/d\gamma)\mathbf{X}(\gamma)\} = 2\{\mathbf{C}_0\mathbf{Q}_1\mathbf{U}\}^2,$$

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

$$\operatorname*{plim}_{\gamma\to 0} D_n^{(2)}(\gamma) = \operatorname*{plim}_{\gamma\to 0} 2n\{(d^2/d\gamma^2)\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)\}^2 + \operatorname*{plim}_{\gamma\to 0} 2n\{(d/d\gamma)\mathbf{X}(\gamma)'\mathbf{Q}_1(d/d\gamma)\mathbf{X}(\gamma)\} = 2n\mathbf{C}_0\mathbf{Q}_1\mathbf{C}_0,$$

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

We now let γ to converge to 1.

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

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

$$\operatorname*{plim}_{\gamma\to 1} D_n^{(2)}(\gamma) = \operatorname*{plim}_{\gamma\to 1} 2n\{(d^2/d\gamma^2)\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)\}^2 + \operatorname*{plim}_{\gamma\to 1} 2n\{(d/d\gamma)\mathbf{X}(\gamma)'\mathbf{Q}_1(d/d\gamma)\mathbf{X}(\gamma)\} = 2n\mathbf{C}_1\mathbf{Q}_1\mathbf{C}_1,$$

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

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 $\mathbf{\Gamma}$. 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 let $M(\cdot,\cdot)$ denote the moment generating function of $(\log(X_t), \widetilde{\mathbf{Z}}_t')'$, viz., $M(\gamma, \boldsymbol{\tau}) := \mathbb{E}[\exp(\gamma \log(X_t) + \boldsymbol{\tau}'\widetilde{\mathbf{Z}}_t)]$, 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 now 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 $\mathbf{J}_1\mathbb{E}[\widetilde{\mathbf{Z}}_t m(X_t)] \neq \mathbf{0}$. 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

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

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

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

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*) The proof is the same as that of Theorem 1.

(*ii*) We first note from (10) that

$$\begin{aligned}
&n^{-1}d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n)\\
&\quad = \inf_{\gamma\in\mathbf{\Gamma}} [n^{-1/2}\mathbf{U}'\ddot{\mathbf{Z}} - \beta_*\sqrt{n}(\gamma - \gamma_*)n^{-1}\mathbf{D}(\gamma_*)'\ddot{\mathbf{Z}}]\mathbf{J}_1[n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*\sqrt{n}(\gamma - \gamma_*)n^{-1}\ddot{\mathbf{Z}}'\mathbf{D}(\gamma_*)] + o_{\mathbb{P}}(1)\\
&\quad \Rightarrow \inf_s [\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)]'\mathbf{J}_1[\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)]
\end{aligned}$$

by noting that $\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*) = \mathbf{D}(\gamma_*)(\gamma - \gamma_*) + o_{\mathbb{P}}((\gamma - \gamma_*))$; $\mathbf{Q}_1 = \ddot{\mathbf{Z}}(\mathbf{J}_1 + o_{\mathbb{P}}(1))\ddot{\mathbf{Z}}'$; $\mathbf{J}_1$ is idempotent; and

$n^{-1}\mathbf{D}(\cdot)'\ddot{\mathbf{Z}} \overset{\mathbb{P}}{\to} \mathbb{E}[X_t^{(\cdot)} \log(X_t)\widetilde{\mathbf{Z}}_t]$ uniformly on $\mathbf{\Gamma}$ by the ULLN. Likewise, we note from (11) that

$$\begin{aligned}
&\inf_{\gamma\in\mathbf{\Gamma}} n^{-1} d_n(\widehat{\xi}_{0,n}(\gamma), \widehat{\boldsymbol{\delta}}_n(\gamma), \widehat{\beta}_n(\gamma), \gamma) \\
&\quad = \inf_{\gamma\in\mathbf{\Gamma}} [n^{-1/2}\mathbf{U}'\ddot{\mathbf{Z}} - \beta_*\sqrt{n}(\gamma-\gamma_*)n^{-1}\mathbf{D}(\gamma_*)'\ddot{\mathbf{Z}}]\mathbf{H}(\gamma)[n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*\sqrt{n}(\gamma-\gamma_*)n^{-1}\ddot{\mathbf{Z}}'\mathbf{D}(\gamma_*)] + o_{\mathbb{P}}(1) \\
&\quad = \inf_{\gamma\in\mathbf{\Gamma}} [n^{-1/2}\mathbf{U}'\ddot{\mathbf{Z}} - \beta_*\sqrt{n}(\gamma-\gamma_*)n^{-1}\mathbf{D}(\gamma_*)'\ddot{\mathbf{Z}}]\mathbf{H}(\gamma_*)[n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*\sqrt{n}(\gamma-\gamma_*)n^{-1}\ddot{\mathbf{Z}}'\mathbf{D}(\gamma_*)] + o_{\mathbb{P}}(1) \\
&\quad \Rightarrow \inf_s [\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)]'\mathbf{H}(\gamma_*)[\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)]
\end{aligned}$$

by further noting that $\mathbf{P}(\gamma_*) = \ddot{\mathbf{Z}}(\mathbf{H}(\gamma_*) + o_{\mathbb{P}}(1))\ddot{\mathbf{Z}}'$. Here, the second equality holds by noting that

$$\begin{aligned}
d_n(\beta,\gamma) &:= d_n(\widehat{\xi}_{0,n}(\beta,\gamma), \widehat{\boldsymbol{\delta}}_n(\beta,\gamma), \beta, \gamma) \\
&= [\mathbf{U} - (\beta-\beta_*)\mathbf{X}(\gamma) - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))]'\mathbf{Q}_1[\mathbf{U} - (\beta-\beta_*)\mathbf{X}(\gamma) - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))] \\
&= [\mathbf{U} - (\beta-\beta_*)\mathbf{X}(\gamma) - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))]'\mathbf{Q}_1[\mathbf{U} - (\beta-\beta_*)\mathbf{X}(\gamma) - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))] + o_{\mathbb{P}}(n^{-1}),
\end{aligned}$$

where we let $(\widehat{\xi}_{0,n}(\beta,\gamma), \widehat{\boldsymbol{\delta}}_n(\beta,\gamma)) := d_n(\xi_0, \boldsymbol{\delta}, \beta, \gamma)$. Here, we further use that $\mathbf{X}(\gamma) = \mathbf{X}(\gamma_*) + \mathbf{D}(\gamma_*)(\gamma-\gamma_*) + o_{\mathbb{P}}((\gamma-\gamma_*))$ to obtain that

$$d_n(\beta,\gamma) = [\mathbf{U} - (\beta-\beta_*)\mathbf{X}(\gamma_*) - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))]'\mathbf{Q}_1[\mathbf{U} - (\beta-\beta_*)\mathbf{X}(\gamma_*) - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))] + o_{\mathbb{P}}(n^{-1}),$$

so that if we optimize this with respect to β, it follows that $\widehat{\beta}_n(\gamma) = \beta_* + (\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma))^{-1}\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}$ and

$$\begin{aligned}
&d_n(\widehat{\beta}_n(\gamma), \gamma) \\
&\quad = [\mathbf{U} - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))]'[\mathbf{Q}_1 - \mathbf{Q}_1\mathbf{X}(\gamma_*)(\mathbf{X}(\gamma_*)'\mathbf{Q}_1\mathbf{X}(\gamma_*))^{-1}\mathbf{X}(\gamma_*)'\mathbf{Q}_1][\mathbf{U} - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))] + o_{\mathbb{P}}(n^{-1}) \\
&\quad = [\mathbf{U} - \beta_*(\gamma-\gamma_*)\mathbf{D}(\gamma_*)]'\ddot{\mathbf{Z}}\mathbf{H}(\gamma_*)\ddot{\mathbf{Z}}'[\mathbf{U} - \beta_*(\gamma-\gamma_*)\mathbf{D}(\gamma_*)] + o_{\mathbb{P}}(n^{-1}).
\end{aligned}$$

We now note that $d_n(\widehat{\xi}_{0,n}(\gamma), \widehat{\boldsymbol{\delta}}_n(\gamma), \widehat{\beta}_n(\gamma), \gamma) = d_n(\widehat{\beta}_n(\gamma), \gamma) + o_{\mathbb{P}}(n^{-1})$, leading to the second equality.

Therefore, it now follows that

$$\begin{aligned}
\mathcal{D}_n &= n^{-1} d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n) - \inf_{\gamma\in\mathbf{\Gamma}} n^{-1} d_n(\widehat{\xi}_{0,n}(\gamma), \widehat{\boldsymbol{\delta}}_n(\gamma), \widehat{\beta}_n(\gamma), \gamma) \\
&\Rightarrow \inf_s [\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)]'\mathbf{J}_1[\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)] - \inf_s [\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)]'\mathbf{H}(\gamma_*)[\boldsymbol{\mathcal{U}} - \beta_* s\mathbf{d}(\gamma_*)] = \frac{\{\mathbf{g}(\gamma_*)'\mathbf{K}(\gamma_*)\boldsymbol{\mathcal{U}}\}^2}{\mathbf{g}(\gamma_*)'\mathbf{K}(\gamma_*)\mathbf{g}(\gamma_*)},
\end{aligned}$$

whose distribution is the same as that of $\mathcal{Z}_2^2(\gamma_*)$ as desired. ■

Proof of Theorem 4 (*i*) For each $(\beta_*, \gamma_*) \in \mathbf{\Upsilon}_0$, we can rewrite $\xi_{0*} + \mathbf{E}_t'\boldsymbol{\delta}_* + \beta_* X_t^{\gamma_*}$ as a linear model of $(1, X_t, \mathbf{D}_t')'$, so that $\mathcal{D}_n \Rightarrow \sup_{\gamma\in\mathbf{\Gamma}} \mathcal{Z}_1^2(\gamma)$ uniformly on $\mathbf{\Upsilon}_0$, implying that $\limsup_{n\to\infty} \sup_{(\beta_*,\gamma_*)\in\mathbf{\Upsilon}} \mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_1(\boldsymbol{\alpha})) = \boldsymbol{\alpha}$.

(ii) Before proving the statement, we note that if we let $\boldsymbol{\nu}_n(\cdot) := n^{1/2}\{n^{-1}\mathbf{X}(\cdot)'\ddot{\mathbf{Z}}\mathbf{J}_1 - \mathbf{g}(\cdot)\}$,

$$n^{-1/2}\mathbf{J}_1\ddot{\mathbf{Z}}'(\mathbf{X}(\cdot) - \mathbf{X}(\circ)) = \sqrt{n}\{\mathbf{g}(\cdot) - \mathbf{g}(\circ)\} + \boldsymbol{\nu}_n(\cdot) - \boldsymbol{\nu}_n(\circ) = \mathbf{d}(\circ)\triangle_n + \triangle_n^2 O(n^{-1/2}) + \boldsymbol{\nu}_n(\cdot) - \boldsymbol{\nu}_n(\circ), \quad (5)$$

where $\triangle_n$ translates $\sqrt{n}(\cdot - \circ)$, and the second equality follows from the fact that $\mathbf{g}(\cdot) - \mathbf{g}(\circ) = \mathbf{d}(\circ)(\cdot - \circ) + O(1)(\cdot - \circ)^2$ and that for each $j = 1, 2, \ldots, p$, $\mathbb{E}[Z_{t,j}\log^2(X_t)X_t^{(\cdot)}] < \infty$ uniformly on $\mathbf{\Gamma}$ by Assumptions 1(iii) and 3, leading to the uniformly bounded second-order derivative of $\mathbf{g}(\cdot)$ on $\mathbf{\Gamma}^c(\epsilon)$. Furthermore, $\boldsymbol{\nu}_n(\cdot)$ converges to a multivariate Gaussian process by applying a FCLT. Therefore, if we let $d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n; \beta_*, \gamma_*) := d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n)$ to indicate the dependence on the unknown parameters of $d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n)$,

$$d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \gamma; \beta_*, \gamma_*) = [\mathbf{U} - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))]'\mathbf{Q}_1[\mathbf{U} - \beta_*(\mathbf{X}(\gamma) - \mathbf{X}(\gamma_*))]$$

under $\mathcal{H}_0'$. We now note that $n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} = O_{\mathbb{P}}(1)$, $\mathbf{Q}_1 = \ddot{\mathbf{Z}}(\mathbf{J}_1 + o_{\mathbb{P}}(1))\ddot{\mathbf{Z}}'$, and (3), so that under $\mathcal{H}_0'$,

$$\begin{aligned} n^{-2}d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \cdot; \beta_*, \circ) &= [n^{-1}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_* n^{-1}\ddot{\mathbf{Z}}'(\mathbf{X}(\cdot) - \mathbf{X}(\circ))]'\mathbf{Q}_1[n^{-1}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_* n^{-1}\ddot{\mathbf{Z}}'(\mathbf{X}(\cdot) - \mathbf{X}(\circ))] \\ &= \beta_*^2[\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}] - \mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\circ)}]]'\mathbf{J}_1[\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}] - \mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\circ)}]] + o_{\mathbb{P}}(1) \end{aligned}$$

uniformly on $\mathbf{\Gamma}^c(\epsilon)$. Therefore, the right side is $o_{\mathbb{P}}(1)$ if $(\cdot) = (\circ)$ uniformly on $\mathbf{\Gamma}^c(\epsilon)$, implying that if we let $\widetilde{\gamma}_n(\gamma_*) := \arg\inf_{\gamma\in\mathbf{\Gamma}^c(\epsilon)} d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \gamma; \beta_*, \gamma_*)$, then $\widetilde{\gamma}_n(\circ) - \circ = o_{\mathbb{P}}(1)$ uniformly on $\mathbf{\Gamma}^c(\epsilon)$, so that $\|\boldsymbol{\nu}_n(\widetilde{\gamma}_n(\circ)) - \boldsymbol{\nu}_n(\circ)\| = o_{\mathbb{P}}(1)$. Next, we note that if we let $\mathbf{U}_n(\cdot, \circ) := n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} + \boldsymbol{\nu}_n(\cdot) - \boldsymbol{\nu}_n(\circ)$, then $\mathbf{U}_n(\widetilde{\gamma}_n(\circ), \circ) = n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} + \boldsymbol{\nu}_n(\widetilde{\gamma}_n(\circ)) - \boldsymbol{\nu}_n(\circ) = n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} + o_{\mathbb{P}}(1)$

$$\begin{aligned} &n^{-1}d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n(\circ); \beta_*, \circ) \\ &\quad = [n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_* n^{-1/2}\ddot{\mathbf{Z}}'(\mathbf{X}(\widetilde{\gamma}_n(\circ)) - \mathbf{X}(\circ))]'\mathbf{Q}_1[n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_* n^{-1/2}\ddot{\mathbf{Z}}'(\mathbf{X}(\widetilde{\gamma}_n(\circ)) - \mathbf{X}(\circ))] \\ &\quad = [\mathbf{U}_n(\widetilde{\gamma}_n(\circ), \circ) - \beta_*\{\mathbf{d}(\circ)\widetilde{\triangle}_n + \widetilde{\triangle}_n^2 O(n^{-1/2})\}]'\mathbf{J}_1[\mathbf{U}_n(\widetilde{\gamma}_n(\circ), \circ) - \beta_*\{\mathbf{d}(\circ)\widetilde{\triangle}_n + \widetilde{\triangle}_n^2 O(n^{-1/2})\}] + o_{\mathbb{P}}(1) \\ &\quad = [n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*\{\mathbf{d}(\circ)\widetilde{\triangle}_n + \widetilde{\triangle}_n^2 O(n^{-1/2})\}]'\mathbf{J}_1[n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*\{\mathbf{d}(\circ)\widetilde{\triangle}_n + \widetilde{\triangle}_n^2 O(n^{-1/2})\}] + o_{\mathbb{P}}(1), \end{aligned}$$

where $\widetilde{\Delta}_n := \sqrt{n}(\widetilde{\gamma}_n(\circ) - \circ)$ and $\widetilde{\gamma}_n(\circ)$ satisfies the following asymptotic first-order condition:

$$[n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*\{\mathbf{d}(\circ)\widetilde{\triangle}_n + \widetilde{\triangle}_n^2 O(n^{-1/2})\}]'\mathbf{J}_1[-\beta_*\{\mathbf{d}(\circ) + \tilde{\triangle}_n O(n^{-1/2})\}] = o_{\mathbb{P}}(1).$$

If we solve for $\widetilde{\triangle}_n$ from this condition, it follows that $\widetilde{\triangle}_n = \beta_*^{-1}(\mathbf{d}(\circ)'\mathbf{J}_1\mathbf{d}(\circ))^{-1}(\mathbf{d}(\circ)'\mathbf{J}_1 n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U}) + o_{\mathbb{P}}(1) \Rightarrow \beta_*^{-1}(\mathbf{d}(\circ)'\mathbf{J}_1\mathbf{d}(\circ))^{-1}(\mathbf{d}(\circ)'\mathbf{J}_1\boldsymbol{\mathcal{U}})$. That is, $\widetilde{\triangle}_n$ is $O_{\mathbb{P}}(1)$ uniformly on $\mathbf{\Gamma}^c(\epsilon)$, so that

$$n^{-1}d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n(\circ); \beta_*, \circ) \Rightarrow \boldsymbol{\mathcal{U}}'\mathbf{K}(\circ)\boldsymbol{\mathcal{U}}. \quad (6)$$

This null weak limit is free of β_*, although it has different null limit distributions for different γ_*'ys.

Next, we let $d_n(\beta,\gamma;\beta_*,\gamma_*) := d_n(\beta,\gamma)$ to indicate the dependence of the unknown parameters and note that

$$\begin{aligned} n^{-2}d_n(\beta,\gamma;\beta_*,\gamma_*) =&[n^{-1}\ddot{\mathbf{Z}}'\mathbf{U} - (\beta-\beta_*)n^{-1}\ddot{\mathbf{Z}}'\mathbf{X}(\gamma) - \beta_*(n^{-1}\ddot{\mathbf{Z}}'\mathbf{X}(\gamma) - n^{-1}\ddot{\mathbf{Z}}'\mathbf{X}(\gamma_*))]' \\ &\times (\mathbf{J}_1 + o_{\mathbb{P}}(1))[n^{-1}\ddot{\mathbf{Z}}'\mathbf{U} - (\beta-\beta_*)n^{-1}\ddot{\mathbf{Z}}'\mathbf{X}(\gamma) - \beta_*(n^{-1}\ddot{\mathbf{Z}}'\mathbf{X}(\gamma) - n^{-1}\ddot{\mathbf{Z}}'\mathbf{X}(\gamma_*))], \end{aligned}$$

so that if we note that $n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} = O_{\mathbb{P}}(1)$, $\mathbf{Q}_1 = \ddot{\mathbf{Z}}(\mathbf{J}_1 + o_{\mathbb{P}}(1))\ddot{\mathbf{Z}}'$, and (3),

$$\begin{aligned} n^{-2}d_n(\diamond,\cdot;\star,\circ) = &[(\diamond-\star)\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}] + (\star)(\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}] - \mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\circ)}])]' \\ &\times \mathbf{J}_1[(\diamond-\star)\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}] + (\star)(\mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\cdot)}] - \mathbb{E}[\widetilde{\mathbf{Z}}_t X_t^{(\circ)}])] + o_{\mathbb{P}}(1) \end{aligned}$$

uniformly on $\boldsymbol{\Upsilon}^c(\epsilon)$. Therefore, the right side is $o_{\mathbb{P}}(1)$ if $(\diamond,\cdot) = (\star,\circ)$ uniformly on $\boldsymbol{\Upsilon}^c(\epsilon)$, implying that if we let $(\widehat{\beta}_n(\beta_*,\gamma_*), \widehat{\gamma}_n(\beta_*,\gamma_*)) := \arg\inf_{(\beta,\gamma)\in\boldsymbol{\Upsilon}^c(\epsilon)} d_n(\beta,\gamma;\beta_*,\gamma_*)$, then $\|(\widehat{\beta}_n(\star,\circ),\widehat{\gamma}_n(\star,\circ)) - (\star,\circ)\| = o_{\mathbb{P}}(1)$ uniformly on $\boldsymbol{\Upsilon}^c(\epsilon)$, so that $\|\boldsymbol{\nu}_n(\widetilde{\gamma}_n(\star,\circ)) - \boldsymbol{\nu}_n(\circ)\| = o_{\mathbb{P}}(1)$.

We also let $d_n(\widehat{\xi}_{0,n}(\gamma), \widehat{\boldsymbol{\delta}}_n(\gamma), \widehat{\beta}_n(\gamma), \gamma; \beta_*,\gamma_*) = d_n(\widehat{\xi}_{0,n}(\gamma), \widehat{\boldsymbol{\delta}}_n(\gamma), \widehat{\beta}_n(\gamma), \gamma)$ to emphasize its dependence on the unknown parameter. We further let $d_n(\cdot;\beta_*,\circ) := d_n(\widehat{\xi}_{0,n}(\cdot), \widehat{\boldsymbol{\delta}}_n(\cdot), \widehat{\beta}_n(\cdot), \cdot;\beta_*,\circ)$ and note that it follows from (11) that

$$d_n(\cdot;\beta_*,\circ) = [\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*(\ddot{\mathbf{Z}}'\mathbf{X}(\cdot) - \ddot{\mathbf{Z}}'\mathbf{X}(\circ))]'(\mathbf{H}(\cdot) + o_{\mathbb{P}}(1))[\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*(\ddot{\mathbf{Z}}'\mathbf{X}(\cdot) - \ddot{\mathbf{Z}}'\mathbf{X}(\circ))]$$

by noting that $\mathbb{P}(\cdot) = \ddot{\mathbf{Z}}'(\mathbf{H}(\cdot) + o_{\mathbb{P}}(1))\ddot{\mathbf{Z}}$. If we now use (5) and let $\widehat{\gamma}_n(\circ) := \widehat{\gamma}_n(\beta_*,\circ)$ for simplicity,

$$\begin{aligned} &n^{-1}d_n(\widehat{\gamma}_n(\circ);\beta_*,\circ) \\ &=[\mathbf{U}_n(\widehat{\gamma}_n(\circ),\circ) - \beta_*\{\mathbf{d}(\circ)\widehat{\triangle}_n + \widehat{\triangle}_n^2 O(n^{-1/2})\}]'\mathbf{H}(\widehat{\gamma}_n(\circ))[\mathbf{U}_n(\widehat{\gamma}_n(\circ),\circ) - \beta_*\{\mathbf{d}(\circ)\widehat{\triangle}_n + \widehat{\triangle}_n^2 O(n^{-1/2})\}] + o_{\mathbb{P}}(1) \\ &=[n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*\{\mathbf{d}(\circ)\widehat{\triangle}_n + \widehat{\triangle}_n^2 O(n^{-1/2})\}]'\mathbf{H}(\widehat{\gamma}_n(\circ))[n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*\{\mathbf{d}(\circ)\widehat{\triangle}_n + \widehat{\triangle}_n^2 O(n^{-1/2})\}] + o_{\mathbb{P}}(1), \end{aligned}$$

where $\widehat{\triangle}_n := \sqrt{n}(\widehat{\gamma}_n(\circ) - \circ)$ and $\widehat{\gamma}_n(\circ)$ asymptotically satisfies the first-order condition:

$$[n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U} - \beta_*\{\mathbf{d}(\circ)\widehat{\triangle}_n + \widehat{\triangle}_n^2 O(n^{-1/2})\}]'\mathbf{H}(\widehat{\gamma}_n(\circ))[-\beta_*\{\mathbf{d}(\circ) + \widehat{\triangle}_n O(n^{-1/2})\}] = o_{\mathbb{P}}(1).$$

If we solve for $\widehat{\triangle}_n$ from this condition, it follows that $\widehat{\triangle}_n = \beta_*^{-1}(\mathbf{d}(\circ)'\mathbf{H}(\widehat{\gamma}(\circ))\mathbf{d}(\circ))^{-1}(\mathbf{d}(\circ)'\mathbf{H}(\widehat{\gamma}(\circ))n^{-1/2}\ddot{\mathbf{Z}}'\mathbf{U}) + o_{\mathbb{P}}(1) \Rightarrow \beta_*^{-1}(\mathbf{d}(\circ)'\mathbf{H}(\circ)\mathbf{d}(\circ))^{-1}(\mathbf{d}(\circ)'\mathbf{H}(\circ)\boldsymbol{\mathcal{U}})$. That is, $\widehat{\triangle}_n$ is $O_{\mathbb{P}}(1)$ uniformly on $\boldsymbol{\Gamma}^c(\epsilon)$. Here, $|\widehat{\gamma}_n(\circ) - \circ| = o_{\mathbb{P}}(1)$ uniformly on $\boldsymbol{\Gamma}^c(\epsilon)$, implying that $\mathbf{H}(\widehat{\gamma}_n(\circ)) = \mathbf{H}(\circ) + o_{\mathbb{P}}(1)$ uniformly on $\boldsymbol{\Gamma}^c(\epsilon)$. Therefore,

$$n^{-1}d_n(\widehat{\gamma}_n(\circ);\beta_*,\circ) \Rightarrow \boldsymbol{\mathcal{U}}'(\mathbf{H}(\circ) - \mathbf{H}(\circ)\mathbf{d}(\circ)(\mathbf{d}(\circ)'\mathbf{H}(\circ)\mathbf{d}(\circ))^{-1}\mathbf{d}(\circ)'\mathbf{H}(\circ))\boldsymbol{\mathcal{U}} \tag{7}$$

using the fact that $\mathbf{H}(\circ)$ is idempotent uniformly on $\boldsymbol{\Gamma}^c(\epsilon)$. As for $n^{-1}d_n(\widetilde{\xi}_{0,n}, \widetilde{\boldsymbol{\delta}}_n, \beta_0, \widetilde{\gamma}_n(\circ);\beta_*,\circ)$, the null weak limit of $n^{-1}d_n(\widehat{\gamma}_n(\circ);\beta_*,\circ)$ is free of β_*, although it has different null limit distributions for different γ_*'s.

Finally, we now obtain the null limit distribution of $\mathcal{D}_n$ by emphasizing the dependence of $\mathcal{D}_n$ on (β_*,γ_*) via $\mathcal{D}_n(\beta_*,\gamma_*)$. If we now combine (6) with (7),

$$\mathcal{D}_n(\cdot,\circ) = n^{-1}d_n(\widetilde{\xi}_{0,n},\widetilde{\boldsymbol{\delta}}_n,\beta_0,\widetilde{\gamma}_n(\circ);\cdot,\circ) - n^{-1}d_n(\widehat{\gamma}_n(\circ);\cdot,\circ) \Rightarrow \mathcal{Z}_2^2(\circ)$$

under $\mathcal{H}_0'$ by the fact that the null weak limits of $n^{-1}d_n(\widetilde{\xi}_{0,n},\widetilde{\boldsymbol{\delta}}_n,\beta_0,\widetilde{\gamma}_n(\circ);\beta_*,\circ)$ and $n^{-1}d_n(\widehat{\gamma}_n(\circ);\beta_*,\circ)$ are free of β_*. This implies the desired result.

(*iii*) We note that

$$\sup_{(\beta_*,\gamma_*)\in\mathbf{B}\times\boldsymbol{\Gamma}}\mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv(\boldsymbol{\alpha})) = \max\left[\sup_{(\beta_*,\gamma_*)\in\boldsymbol{\Upsilon}_0}\mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_1(\boldsymbol{\alpha})), \lim_{\epsilon\downarrow 0}\sup_{(\beta_*,\gamma_*)\in\boldsymbol{\Upsilon}^c(\epsilon)}\mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_2(\boldsymbol{\alpha}))\right],$$

and

$$\begin{aligned}&\limsup_{n\to\infty}\max\left[\sup_{(\beta_*,\gamma_*)\in\boldsymbol{\Upsilon}_0}\mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_1(\boldsymbol{\alpha})), \lim_{\epsilon\downarrow 0}\sup_{(\beta_*,\gamma_*)\in\boldsymbol{\Upsilon}^c(\epsilon)}\mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_2(\boldsymbol{\alpha}))\right]\\ &\quad\le \max\left[\limsup_{n\to\infty}\sup_{(\beta_*,\gamma_*)\in\boldsymbol{\Upsilon}_0}\mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_1(\boldsymbol{\alpha})), \limsup_{n\to\infty}\lim_{\epsilon\downarrow 0}\sup_{(\beta_*,\gamma_*)\in\boldsymbol{\Upsilon}^c(\epsilon)}\mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv_2(\boldsymbol{\alpha}))\right] \le \boldsymbol{\alpha},\end{aligned}$$

where the last inequality follows from (*i*) and (*ii*). Therefore, it follows that $\limsup_{n\to\infty}\sup_{(\beta_*,\gamma_*)\in\mathbf{B}\times\boldsymbol{\Gamma}}\mathbb{P}_{\boldsymbol{\omega}_*}(\mathcal{D}_n > cv(\boldsymbol{\alpha})) \le \boldsymbol{\alpha}$. ■

Proof of Theorem 5: We first note that

$$\begin{aligned}\frac{1}{n^{3/2}}\mathbf{X}(\cdot)'\mathbf{Q}_1\mathbf{U} &= \frac{1}{n}\mathbf{X}(\cdot)'\ddot{\mathbf{Z}}\left(\mathbf{I}-\frac{1}{n}\ddot{\mathbf{Z}}'\mathbf{V}\left(\frac{1}{n^2}\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}\right)^{-1}\frac{1}{n}\mathbf{V}'\ddot{\mathbf{Z}}\right)\frac{1}{\sqrt{n}}\ddot{\mathbf{Z}}'\mathbf{U}\\ &\Rightarrow [\mathbb{E}[X_t^{(\cdot)}\widetilde{\mathbf{S}}_t'],\mathbf{0}'y]\left(\mathbf{I}-[\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t'],\mathbf{0}']'\left(\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t']\mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{V}_t']\right)^{-1}[\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t'],\mathbf{0}']\right)\boldsymbol{\mathcal{U}} = \mathbb{E}[\mathbf{X}_t^{(\cdot)}\widetilde{\mathbf{S}}_t']\mathbf{J}_s\boldsymbol{\mathcal{U}}_s\end{aligned}$$

uniformly on $\boldsymbol{\Gamma}$. Here, note that

$$\frac{1}{n^2}\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V} = \left(\frac{1}{n}\mathbf{V}'\ddot{\mathbf{S}}\right)\left(\frac{1}{n}\ddot{\mathbf{S}}'\mathbf{V}\right)+\left(\frac{1}{n}\mathbf{V}'\ddot{\mathbf{W}}\right)\left(\frac{1}{n}\ddot{\mathbf{W}}'\mathbf{V}\right) = \mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t']\mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{V}_t'] + o_{\mathbb{P}}(1)$$

by applying Assumption 6(*iii*) and the ergodic theorem, and $(\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t']\mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{V}_t'])^{-1}$ is well defined because of Assumption

6(*ii*). Next, we note that

$$
\begin{aligned}
\frac{1}{n^2}\mathbf{X}(\cdot)'\mathbf{Q}_1\mathbf{X}(\cdot) &= \frac{1}{n}\mathbf{X}(\cdot)'\ddot{\mathbf{Z}}\left(\mathbf{I}-\frac{1}{n}\ddot{\mathbf{Z}}'\mathbf{V}\left(\frac{1}{n^2}\mathbf{V}'\ddot{\mathbf{Z}}\ddot{\mathbf{Z}}'\mathbf{V}\right)^{-1}\frac{1}{n}\mathbf{V}'\ddot{\mathbf{Z}}\right)\frac{1}{n}\ddot{\mathbf{Z}}'\mathbf{X}(\cdot) \\
&= [\mathbb{E}[X_t^{(\cdot)}\widetilde{\mathbf{S}}_t'],\mathbf{0}']\left(\mathbf{I}-[\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t'],\mathbf{0}']'\left(\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t']\mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{V}_t']\right)^{-1}[\mathbb{E}[\mathbf{V}_t\widetilde{\mathbf{S}}_t'],\mathbf{0}']\right)[\mathbb{E}[X_t^{(\cdot)}\widetilde{\mathbf{S}}_t'],\mathbf{0}']'+o_{\mathbb{P}}(1) \\
&= \mathbb{E}[\mathbf{X}_t^{(\cdot)}\widetilde{\mathbf{S}}_t']\mathbf{J}_s\mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{X}_t^{(\cdot)}]+o_{\mathbb{P}}(1)
\end{aligned}
$$

uniformly on $\boldsymbol{\Gamma}$. Therefore, it now follows that

$$
\mathcal{D}_n = \sup_{\gamma\in\boldsymbol{\Gamma}}\frac{1}{n}\frac{\{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{U}\}^2}{\mathbf{X}(\gamma)'\mathbf{Q}_1\mathbf{X}(\gamma)} \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}}\left(\frac{\mathbb{E}[\mathbf{X}_t^{(\gamma)}\widetilde{\mathbf{S}}_t']\mathbf{J}_s\boldsymbol{\mathcal{U}}_s}{\{\mathbb{E}[\mathbf{X}_t^{(\gamma)}\widetilde{\mathbf{S}}_t']\mathbf{J}_s\mathbb{E}[\widetilde{\mathbf{S}}_t\mathbf{X}_t^{(\gamma)}]\}^{1/2}}\right)^2
$$

under $\mathcal{H}_0$. Here, if we apply the definition of $\boldsymbol{\pi}_s(\cdot)$ to the right side, it is equivalent to $\sup_{\gamma\in\boldsymbol{\Gamma}}(\boldsymbol{\pi}_s(\gamma)'\boldsymbol{\mathcal{U}}_s)^2$, and this completes the proof. ■

Proof of Theorem 6: (*i*) This is obvious from Corollary 1.

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

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

$$
\mathbb{P}\left(\sup_{\gamma\in\boldsymbol{\Gamma}^{(\bar{q})}}\mathcal{Z}_q^2(\gamma)\ge u^2\right) \le \mathbb{P}\left(\sup_{\gamma\in\boldsymbol{\Gamma}^{(\bar{q})}}\mathcal{Z}_q(\gamma)\ge\frac{u}{\sqrt{2}}\right)+\mathbb{P}\left(\inf_{\gamma\in\boldsymbol{\Gamma}^{(\bar{q})}}\mathcal{Z}_q(\gamma)\le-\frac{u}{\sqrt{2}}\right) = 2\mathbb{P}\left(\sup_{\gamma\in\boldsymbol{\Gamma}^{(\bar{q})}}\mathcal{Z}_q(\gamma)\ge\frac{u}{\sqrt{2}}\right)
$$

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

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

and so it follows that

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

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

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

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 $cv_q^{1/2}(\boldsymbol{\alpha}_n) > 2\sqrt{2}a_q$ and is achieved as $\boldsymbol{\alpha}_n \to 0$. Furthermore, the given condition implies that $-\log(\boldsymbol{\alpha}_n)/n \to 0$, so that

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

Therefore, it follows that $cv_q(\boldsymbol{\alpha}_n) = o(n)$, as desired. ■

A.2 Data Construction

Using Compustat fundamental annual, we construct firm-level value-added, capital stock, employment, and material inputs following the data-cleaning procedure conducted by İmrohoroğlu and Tüzel (2014). We supplement Compustat with the Gross Domestic Product (GDP) price deflator, the investment price deflator, and the national wage index from the Social Security Administration. See the online Appendix of İmrohoroğlu and Tüzel (2014) for more details. When we construct the data set, we use their code that is downloadable from

`https://sites.google.com/usc.edu/selale-tuzel/home?authuser=2`.

The data set is constructed by the following procedure:

(a) We exclude financial firms (SIC $\in [6000, 6999]$) and regulated firms (SIC $\in [4900, 4999]$).

(b) We keep only observations with positive values on sales (SALE), total assets (AT), number of employees (EMP), gross property, plant, and equipment (PPEGT), depreciation (DP), accumulated depreciation (DPACT), and capital expenditures (CAPX).

(c) We compute the material input (M) by total expenditure (TE) minus labor expenditure (LE), where TE is obtained as SALE minus operating income before depreciation and amortization (OIBDP), and LE is computed by EMP multiplied by the national wage index from the Social Security Administration.

(d) We compute the value-added by SALE minus M. Both SALE and M are deflated by the GDP price deflator.

(e) We use EMP as the labor input.

(f) Finally, we deflate PPEGT using the investment price deflator and use this as the capital stock. When deflating PPEGT, we use the deflator corresponding to the average age of capital of each year. □

References

ACEMOGLU, D. (2008): *Introduction to Modern Economic Growth*, Princeton University Press.

ACEMOGLU, D. AND P. RESTREPO (2018): "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 108, 1488–1542.

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451.

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

AKAIKE, H. (1973): "Information Theory and an Extension of the Maximum Likelihood Principle," in *2nd International Symposium on Information Theory, 1973*, ed. by B. Petrov and F. Csake, Budapest: Akademiai Kiado, 268–281.

ANDREWS, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

——— (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–563.

ANDREWS, D. W. K. AND X. CHENG (2015): "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2353–2211.

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

ANTRÀS, P. (2004): "Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution," *The BE Journal of Macroeconomics*, 4, 1–36.

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

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

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

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

BREUNIG, C. (2015): "Goodness-of-Fit Tests based on Series Estimators in Nonparametric Instrumental Regression," *Journal of Econometrics*, 184, 328–346.

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

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

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

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

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

CRAGG, J. G. AND S. G. DONALD (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 33, 222–240.

DAVIES, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64, 247–254.

——— (1987): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 74, 33–43.

DHYNE, E., A. PETRIN, V. SMEETS, AND F. WARZYNSKI (2017): "Multi Product Firms, Import Competition, and the Evolution of Firm-Product Technical Efficiencies," NBER Working Paper 23637, NBER.

ELLIOTT, G., U. MÜLLER, AND M. WATSON (2015): "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811.

GRILICHES, Z. AND J. MAIRESSE (1995): "Production Functions: The Search for Identification," NBER Working Paper 5067, NBER.

HALL, A. R. AND A. INOUE (2003): "The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics*, 114, 361–394.

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

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

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

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

HOROWITZ, J. L. (2006): "Testing a Parametric Model against a Nonparametric Alternative with Identification through Instrumental Variables," *Econometrica*, 74, 521–538.

HOSOYA, Y. (1989): "Hierarchical Statistical Models and a Generalized Likelihood Ratio Test," *Journal of the Royal Statistical Society: Series B (Methodological)*, 51, 435–447.

İMROHOROĞLU, A. AND E. TÜZEL (2014): "Firm-level Productivity, Risk, and Return," *Management Science*, 60, 2073–2090.

KARABARBOUNIS, L. AND B. NEIMAN (2014): "The Global Decline of the Labor Share," *Quarterly Journal of Economics*, 129, 61–103.

KLEINBERGEN, F. AND R. PAPP (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

KRUSELL, P., L. E. OHANIAN, J. V. RÍOS-RULL, AND G. L. VIOLANTE (2000): "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 68, 1029–1053.

LEEB, H. AND B. PÖTSCHER (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21, 21–59.

LEVINSOHN, J. AND A. PETRIN (2003): "Estimating Production Functions using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–341.

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

——— (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 J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

OBERFIELD, E. AND D. RAVAL (2021): "Micro Data and Macro Technology," *Econometrica*, 89, 703–732.

OLLEY, G. S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

PIKETTY, T. (2014): *Capital in the 21st Century*, Harvard University Press.

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

RAVAL, D. R. (2019): "The Micro Elasticity of Substitution and Non-Neutral Technology," *RAND Journal of Economics*, 50, 142–167.

SAMUELSON, P. A. (1964): "Theoretical Notes on Trade Problems," *The Review of Economics and Statistics*, 46, 145–154.

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

——— (1988): *Lectures on Advanced Econometric Theory*, Oxford: Basil Blackwell.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *The 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 J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STOCK, J. H. AND M. YOGO (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock, Cambridge University Press, 80–108.

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.

WOOLDRIDGE, J. M. (2009): "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 104, 112–114.

ZHU, Y. (2020): "Inference in Nonparametric/Semiparametric Moment Equality Models with Shape Restrictions," *Quantitative Economics*, 11, 609–636.

DGP	Test Stat.	$\boldsymbol{\alpha} \setminus n$	100	200	300	400	500
A	$\mathcal{D}_n$	1%	0.52	1.02	1.12	1.08	0.98
		5%	3.54	4.20	5.08	4.74	4.96
		10%	8.64	9.14	9.86	10.08	10.12
	$\mathcal{H}_n$	1%	0.72	1.14	1.30	1.06	1.26
		5%	4.80	5.50	5.18	5.34	5.72
		10%	10.06	10.26	10.52	10.36	10.44
	$\mathcal{B}_n$	1%	0.90	1.28	1.22	1.22	1.08
		5%	4.90	5.38	5.38	5.40	5.08
		10%	9.34	10.30	10.06	10.46	11.12
	$\mathcal{J}_n$	1%	0.72	0.88	0.90	0.96	1.08
		5%	4.80	5.04	4.76	4.58	4.62
		10%	9.62	9.94	9.68	9.66	9.60
B	$\mathcal{D}_n$	1%	2.14	1.96	1.90	1.62	1.52
		5%	7.50	6.64	7.04	5.94	6.26
		10%	12.90	11.76	11.82	11.42	10.76
	$\mathcal{H}_n$	1%	1.06	1.06	1.12	1.20	1.36
		5%	5.10	4.90	4.78	5.30	5.08
		10%	10.30	10.26	9.72	9.88	10.58
	$\mathcal{B}_n$	1%	1.42	1.18	1.42	1.24	1.04
		5%	5.76	5.60	5.70	5.32	5.36
		10%	10.54	10.94	10.92	10.74	10.44
	$\mathcal{J}_n$	1%	0.92	0.80	0.92	1.04	0.96
		5%	5.06	4.88	5.28	4.58	4.42
		10%	9.90	9.84	9.82	9.48	9.28

Table 1: EMPIRICAL REJECTION RATES UNDER THE NULL (IN PERCENTAGE). Number of Replications: 5,000. This table shows the empirical rejection rates of the DD-and the other test statistics under the null hypothesis. DGP A: $Y_t = X_t + U_t$ and $X_t := \sum_{j=1}^{4} Z_{tj} + U_t^2 \cdot \mathbf{1}(|U_t| \leq 1)$ such that $U_t \sim$ IID $N(0,1)$, $Z_{t1} \sim$ IID $U(0,1)$, Z_{t2} and $Z_{t3} \sim$ IID Beta$(5,5)$, and $Z_{t1} \sim$ IID Beta$(5,3)$; and DGP B: $Y_t = X_t + U_t$ and $X_t := \sum_{j=1}^{4} Z_{tj} + U_t^2$ such that $U_t \sim$ IID $N(0,1)$, $Z_{t1} \sim$ IID Half-$N(0,1)$, $Z_{t2} \sim$ IID Beta$(5,5)$, and $Z_{t3} \sim$ IID Beta$(5,3)$, and $Z_{t3} \sim$ IID $\mathcal{X}_1^2$. Each of $Z_{t1}, \ldots, Z_{t4}$, and U_t is independently distributed. Model: $\mathcal{M}_1^o := \{m_t(\boldsymbol{\omega}) := Y_t - X_t\xi - \beta X_t^{\gamma} : \boldsymbol{\omega} \in \boldsymbol{\Omega} \subset \mathbb{R}^3\}$ with $\boldsymbol{\Gamma} := [-0.25, 2.25]$. Estimation: GMM estimation with $\mathbf{V}_t := X_t$, $\mathbf{Z}_t := (Z_{t1}, \ldots, Z_{t4})'$, and $\mathbf{M}_n := (n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$ for $\mathcal{D}_n$ and $\mathcal{H}_n$; and GMM estimation with $\mathbf{V}_t := \widetilde{X}_t$, $\mathbf{Z}_t := \widetilde{Z}_t$, and $\mathbf{M}_n := (n^{-1}\sum_{t=1}^{n} \widetilde{Z}_t^2)^{-1}$ for $\mathcal{H}_n$, where $\widetilde{X}_t := X_t / \max[X_1, \ldots, X_n]$ and $\widetilde{Z}_t := \sum_{j=1}^{4} Z_{tj} / \max[\sum_{j=1}^{4} Z_{1j}, \ldots, \sum_{j=1}^{4} Z_{nj}]$. The weighted bootstrap is applied to $\mathcal{D}_n$, and the bootstrap number is 300.

DGP	Test Stat.	$\boldsymbol{\alpha} \setminus n$	100	200	300	400	500
A$'$	$\mathcal{D}_n$	1%	26.63	52.20	70.07	83.53	91.40
		5%	46.47	72.70	84.97	93.60	97.23
		10%	57.73	81.77	89.57	96.83	98.63
	$\mathcal{H}_n$	1%	69.97	97.37	99.87	100.0	100.0
		5%	87.73	99.47	100.0	100.0	100.0
		10%	93.73	99.87	100.0	100.0	100.0
	$\mathcal{B}_n$	1%	25.07	59.40	81.37	92.93	98.40
		5%	49.17	80.93	93.33	97.87	99.87
		10%	62.00	88.90	96.43	99.27	99.97
	$\mathcal{J}_n$	1%	7.72	22.44	40.26	57.58	72.3
		5%	22.30	46.12	64.72	78.74	88.12
		10%	34.20	59.60	75.56	86.50	93.68
A$''$	$\mathcal{D}_n$	1%	76.83	78.17	79.93	82.90	84.20
		5%	83.87	85.37	86.50	88.60	89.83
		10%	87.67	88.30	90.03	91.13	91.43
	$\mathcal{H}_n$	1%	12.27	24.37	35.47	45.27	55.50
		5%	28.80	44.30	57.90	67.27	75.83
		10%	40.77	56.13	68.47	77.43	83.50
	$\mathcal{B}_n$	1%	1.73	4.40	6.57	10.10	14.17
		5%	8.90	15.63	20.90	27.80	33.33
		10%	17.20	26.60	32.80	40.97	46.47
	$\mathcal{J}_n$	1%	1.32	2.34	3.28	4.16	6.12
		5%	7.98	10.00	11.78	14.80	19.10
		10%	14.22	17.92	20.02	25.10	30.52
B$'$	$\mathcal{D}_n$	1%	50.57	82.03	94.43	98.73	99.77
		5%	70.23	93.70	98.53	99.83	99.97
		10%	79.87	96.07	99.40	99.93	100.00
	$\mathcal{H}_n$	1%	38.17	79.07	94.43	98.97	99.83
		5%	65.93	92.97	98.90	99.87	100.00
		10%	78.37	96.37	99.80	99.97	100.00
	$\mathcal{B}_n$	1%	1.43	3.00	2.97	3.23	4.10
		5%	6.10	8.87	8.70	10.33	12.43
		10%	11.97	14.57	15.70	17.50	19.27
	$\mathcal{J}_n$	1%	24.20	62.13	84.90	94.77	98.57
		5%	47.80	80.27	95.00	98.83	99.80
		10%	60.70	88.30	97.53	99.50	99.90
B$''$	$\mathcal{D}_n$	1%	41.73	65.93	83.27	89.43	95.93
		5%	57.87	80.77	91.87	95.87	98.80
		10%	65.90	87.40	95.47	97.73	99.20
	$\mathcal{H}_n$	1%	14.20	35.50	58.73	74.50	87.90
		5%	35.43	65.60	83.77	91.80	97.40
		10%	50.97	79.50	91.70	96.33	98.73
	$\mathcal{B}_n$	1%	1.17	1.33	1.77	2.10	1.87
		5%	5.87	5.73	6.17	7.00	6.83
		10%	10.90	11.23	12.07	12.40	12.40
	$\mathcal{J}_n$	1%	21.30	46.17	66.40	80.90	89.13
		5%	38.33	65.37	82.47	90.27	95.87
		10%	49.17	74.80	88.50	93.73	97.90

Table 2: EMPIRICAL REJECTION RATES UNDER THE ALTERNATIVES (IN PERCENTAGE). Number of Replications: 3,000. This table shows the empirical rejection rates of the DD-and the other test statistics under the alternatives. DGP A$'$: $Y_t = X_t - 0.4X_t^2 + U_t$ with $b_d = 1$; DGP A$''$: $Y_t = X_t - 0.4X_t^2 + U_t$ with $b_d = 3$, where $X_t := \sum_{j=1}^4 Z_{tj} + U_t^2 \cdot \mathbf{1}(|U_t| \leq b_d)$ such that $U_t \sim$ IID $N(0,1)$, $Z_{t1} \sim$ IID $U(0,1)$, Z_{t2} and $Z_{t3} \sim$ IID Beta$(5,5)$, and $Z_{t4} \sim$ IID Beta$(5,3)$; DGP B$'$: $Y_t = X_t + \tanh(-X_t/2) + U_t$; DGP B$''$: $Y_t = X_t + 2|\sin(-X_t/5)| + U_t$, where $X_t := \sum_{j=1}^4 Z_{tj} + U_t^2$ such that $U_t \sim$ IID $N(0,1)$, $Z_{t1} \sim$ IID Half-$N(0,1)$, $Z_{t2} \sim$ IID Beta$(5,5)$, and $Z_{t3} \sim$ IID Beta$(5,3)$, and $Z_{t3} \sim$ IID $\mathcal{X}_1^2$. Each of $Z_{t1}, \ldots, Z_{t4}$, and U_t is independently distributed. Model: $\mathcal{M}_1^o := \{m_t(\boldsymbol{\omega}) := Y_t - X_t\xi - \beta X_t^{\gamma} : \boldsymbol{\omega} \in \boldsymbol{\Omega} \subset \mathbb{R}^3\}$ with $\boldsymbol{\Gamma} := [-0.25, 2.25]$. Estimation: GMM estimation with $\mathbf{V}_t := X_t$, $\mathbf{Z}_t := (Z_{t1}, \ldots, Z_{t4})'$, and $\mathbf{M}_n := (n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$ for $\mathcal{D}_n$ and $\mathcal{H}_n$; and GMM estimation with $\mathbf{V}_t := \widetilde{X}_t$, $\mathbf{Z}_t := \widetilde{Z}_t$, and $\mathbf{M}_n := (n^{-1}\sum_{t=1}^n \widetilde{Z}_t^2)^{-1}$ for $\mathcal{H}_n$, where $\widetilde{X}_t := X_t / \max[X_1, \ldots, X_n]$ and $\widetilde{Z}_t := \sum_{j=1}^4 Z_{tj} / \max[\sum_{j=1}^4 Z_{1j}, \ldots, \sum_{j=1}^4 Z_{nj}]$. The weighted bootstrap is applied to $\mathcal{D}_n$, and the bootstrap number is 300.

$\boldsymbol{\alpha}$	Test Stat.	$q \setminus n$	100	500	1,000	1,500	2,000	2,500	3,000	3,500	4,000	4,500
10%	$\mathcal{D}_n$	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		2^*	89.83	93.07	92.80	93.13	91.77	91.50	91.60	89.90	90.53	91.40
		3	9.27	6.60	6.83	6.57	8.03	8.20	8.13	9.80	9.10	8.43
		≥ 4	0.90	0.33	0.37	0.30	0.20	0.30	0.27	0.30	0.37	0.17
	$\mathcal{J}_n$	1	86.00	51.30	16.60	3.83	1.03	0.13	0.00	0.00	0.00	0.00
		2^*	7.90	41.80	76.97	90.67	92.60	93.10	93.23	93.47	93.17	93.33
		3	2.47	1.80	1.70	1.33	1.57	1.77	1.87	1.60	1.33	1.90
		≥ 4	3.63	5.10	4.73	4.17	4.80	5.00	4.90	4.93	5.50	4.77
5%	$\mathcal{D}_n$	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		2^*	95.23	97.57	96.43	97.00	96.77	96.53	96.50	95.53	96.07	95.90
		3	4.47	2.43	3.57	2.97	3.20	3.43	3.47	4.43	3.93	4.00
		≥ 4	0.30	0.00	0.00	0.03	0.03	0.03	0.03	0.03	0.00	0.10
	$\mathcal{J}_n$	1	93.23	62.90	25.93	7.73	2.20	0.40	0.03	0.00	0.00	0.00
		2^*	4.10	33.90	71.27	89.60	94.87	96.10	97.13	96.93	97.07	97.00
		3	1.17	1.07	0.73	0.83	0.77	1.07	0.70	0.63	0.83	0.90
		≥ 4	1.50	2.13	2.07	1.83	2.17	2.43	2.13	2.43	2.10	2.10
1%	$\mathcal{D}_n$	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		2^*	99.10	99.83	99.70	99.53	99.47	99.40	99.40	99.20	99.20	99.40
		3	0.90	0.17	0.30	0.47	0.53	0.60	0.60	0.80	0.80	0.60
		≥ 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$\mathcal{J}_n$	1	98.30	81.37	47.07	19.13	6.50	1.70	0.37	0.10	0.03	0.00
		2^*	1.30	18.20	52.40	80.33	93.13	97.50	99.03	99.27	99.37	99.47
		3	0.13	0.13	0.17	0.33	0.20	0.30	0.20	0.30	0.27	0.20
		≥ 4	0.27	0.30	0.37	0.20	0.17	0.50	0.40	0.33	0.33	0.33

Table 3: 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-tests when the significance level ($\boldsymbol{\alpha}$) is fixed. The true polynomial equation degree is 2, as indicated by the asterisks (*). DGP: $Y_t = D_t + X_t + 0.005X_t^2 + U_t$, $X_t := \sum_{j=1}^{11} Z_{tj} + U_t^2$, $(D_t, U_t)'y \sim$ IID $N(\mathbf{0}, \mathbf{I}_2)$, $Z_{t1} \sim$ IID $U(0,1)$, Z_{t2} and $Z_{t3} \sim$ IID $\mathcal{X}_1^2$, Z_{t4} and $Z_{t5} \sim$ IID Rayleigh(1), Z_{t6} and $Z_{7t} \sim$ IID Half $N(0,1)$, Z_{t8} and $Z_{t9} \sim$ IID Beta$(5,3)$, and Z_{t10} and $Z_{t11} \sim$ IID Beta$(5,5)$. Each of D_t, U_t, Z_{t1}, …, Z_{t11} is independently distributed. Model: $\mathcal{M}_q' := \{m_{t,q}(\boldsymbol{\omega}^{(q)}) := Y_t - D_t\eta - X_t\xi_1 - \ldots - X_t^q\xi_q - \beta X_t^\gamma : \boldsymbol{\omega}^{(q)} \in \boldsymbol{\Omega}^{(q)}\}$ with $q \in I(3)$, $\boldsymbol{\omega}^{(q)} := (\xi_1, \ldots, \xi_q, \eta, \beta, \gamma)'$, and $\boldsymbol{\Gamma} := [0.50, 3.50]$. Estimation: GMM estimation by letting $\mathbf{Z}_t := (D_t, Z_{t1}, \ldots, Z_{t11})'$ and $\mathbf{V}_{t,q} := (D_t, X_t, \ldots, X_t^q)'$.

Methods	Test Stat.	$q \setminus n$	100	500	1,000	1,500	2,000	2,500	3,000	3,500	4,000	4,500
Seq. Est. with $\boldsymbol{\alpha}_n = n^{-1/2}$	$\mathcal{D}_n$	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		2^*	89.83	97.87	97.97	98.60	98.93	98.57	98.73	98.50	98.60	98.90
		3	9.27	2.13	2.03	1.40	1.07	1.43	1.27	1.50	1.40	1.10
		≥ 4	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$\mathcal{J}_n$	1	86.00	64.53	31.67	11.83	3.87	1.07	0.17	0.07	0.00	0.00
		2^*	7.90	32.70	66.77	86.87	94.93	97.33	98.80	98.67	99.00	99.10
		3	2.47	0.80	0.47	0.43	0.40	0.67	0.30	0.63	0.33	0.27
		≥ 4	3.63	1.97	1.10	0.87	0.80	0.93	0.73	0.63	0.67	0.63
	(Hypo. Rate)		90.00	95.53	96.84	97.42	97.76	98.00	98.17	98.31	98.42	98.51
Seq. Est. with $\boldsymbol{\alpha}_n = n^{-3/4}$	$\mathcal{D}_n$	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		2^*	96.90	99.83	99.83	99.60	99.73	99.67	99.87	99.70	99.73	99.83
		3	3.03	0.17	0.17	0.40	0.27	0.33	0.13	0.30	0.27	0.17
		≥ 4	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$\mathcal{J}_n$	1	95.23	81.90	55.03	27.37	11.93	3.77	1.50	0.33	0.13	0.03
		2^*	3.23	17.77	44.70	72.43	87.87	96.03	98.43	99.60	99.80	99.97
		3	0.70	0.10	0.07	0.07	0.07	0.10	0.03	0.00	0.00	0.00
		≥ 4	0.83	0.23	0.20	0.13	0.13	0.10	0.03	0.07	0.07	0.00
	(Hypo. Rate)		96.84	99.05	99.44	99.59	99.67	99.72	99.75	99.78	99.80	99.82
Seq. Est. with $\boldsymbol{\alpha}_n = n^{-1}$	$\mathcal{D}_n$	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		2^*	99.10	99.87	99.87	99.80	99.87	99.67	99.87	99.70	99.73	99.83
		3	0.90	0.13	0.13	0.20	0.13	0.33	0.13	0.30	0.27	0.17
		≥ 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$\mathcal{J}_n$	1	98.30	92.30	73.70	44.87	25.00	10.30	3.63	1.30	0.40	0.20
		2^*	1.30	7.60	26.23	55.07	74.97	89.70	96.37	98.70	99.60	99.80
		3	0.13	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
		≥ 4	0.27	0.10	0.07	0.07	0.00	0.00	0.00	0.00	0.00	0.00
	(Hypo. Rate)		99.00	99.80	99.90	99.93	99.95	99.96	99.97	99.97	99.98	99.98
MSC	Akaike	1	49.80	7.36	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		2	32.94	72.50	80.52	83.10	83.00	83.08	83.28	81.48	82.04	83.50
		3	17.26	20.14	19.06	16.90	17.00	16.92	16.72	18.52	17.96	16.50
	Hannan-Quinn	1	78.98	23.58	2.82	0.30	0.00	0.00	0.00	0.00	0.00	0.00
		2	17.56	71.12	92.16	95.94	95.98	95.78	96.44	95.84	96.22	96.42
		3	3.46	5.30	5.02	3.76	4.02	4.22	3.56	4.16	3.78	3.58
	Bayesian	1	95.86	57.60	15.88	3.44	0.60	0.10	0.00	0.00	0.00	0.00
		2	4.02	42.16	83.82	96.16	99.10	99.30	99.78	99.64	99.74	99.76
		3	0.12	0.24	0.30	0.40	0.30	0.60	0.22	0.36	0.26	0.24

Table 4: PRECISION RATES OF THE SEQUENTIAL TESTING PROCEDURES AND MSCS (IN PERCENTAGE). Number of Replications: 3,000. This table shows the estimated polynomial degrees by sequentially applying the DD- and J-tests when the significance level depends on the sample size: $\boldsymbol{\alpha}_n = n^{-1/2}$, $n^{-3/4}$, or n^{-1}. The true polynomial equation degree is 2, as indicated by the asterisks (*). DGP: $Y_t = D_t + X_t + 0.005X_t^2 + U_t$, $X_t := \sum_{j=1}^{11} Z_{tj} + U_t^2$, $(D_t, U_t)' \sim$ IID $N(\mathbf{0}, \mathbf{I}_2)$, $Z_{t1} \sim$ IID $U(0,1)$, Z_{t2} and $Z_{t3} \sim$ IID $\mathcal{X}_1^2$, Z_{t4} and $Z_{t5} \sim$ IID Rayleigh(1), Z_{t6} and $Z_{7t} \sim$ IID Half $N(0,1)$, Z_{t8} and $Z_{t9} \sim$ IID Beta$(5,3)$, and Z_{t10} and $Z_{t11} \sim$ IID Beta$(5,5)$. Each of D_t, U_t, Z_{t1}, ..., Z_{t11} is independently distributed. Model: $\mathcal{M}_q' := \{m_{t,q}(\boldsymbol{\omega}^{(q)}) := Y_t - D_t\eta - X_t\xi_1 - \ldots - X_t^q\xi_q - \beta X_t^\gamma : \boldsymbol{\omega}^{(q)} \in \boldsymbol{\Omega}^{(q)}\}$ with $q \in I(3)$, $\boldsymbol{\omega}^{(q)} := (\xi_1, \ldots, \xi_q, \eta, \beta, \gamma)'$, and $\boldsymbol{\Gamma} := [0.50, 3.50]$. Estimation: GMM estimation by letting $\mathbf{Z}_t := (D_t, Z_{t1}, \ldots, Z_{t11})'$ and $\mathbf{V}_{t,q} := (D_t, X_t, \ldots, X_t^q)'$.

	$n = 500$						
$\gamma_* \setminus \beta_*$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
0.50	6.13	4.57	1.57	4.63	1.83	4.17	5.27
0.75	3.43	2.53	0.80	4.63	0.93	2.13	3.57
1.00	4.73	4.90	4.10	4.63	4.63	4.63	4.63
1.25	4.17	3.13	1.77	4.63	1.53	3.80	4.67
1.50	5.23	6.00	4.53	4.63	4.60	5.37	4.77
	$n = 5,000$						
$\gamma_* \setminus \beta_*$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
0.50	6.33	6.30	5.20	4.73	4.90	5.77	5.77
0.75	5.83	5.43	2.13	4.73	2.00	6.00	6.27
1.00	4.00	4.53	4.17	4.73	4.07	4.33	4.00
1.25	5.73	5.93	5.00	4.73	5.10	5.33	6.47
1.50	5.03	4.77	5.47	4.73	5.13	5.30	5.30

Table 5: EMPIRICAL REJECTION RATES (IN PERCENT). This table shows the empirical rejection rates of the DD-test statistic for $n = 500$ and $n = 5,000$. The level of significance $\boldsymbol{\alpha}$ is 5%. When $\beta_* = 0.00$ or $\gamma_* = 1.00$, we apply the critical values obtained by applying Hansen's (1996) weighted bootstrap to Theorem 1. For the other cases, the critical values are obtained by applying the weighted bootstrap to the null approximation given in Theorem 4. When $n = 10,000$, the empirical rejection rates of the DD-test are modified to 3.40 and 4.40 for $(\beta_*, \gamma_*) = (-0.25, 0.75)$ and $(0.25, 0.75)$, respectively. The number of experiments is 3,000, and the bootstrap iteration is 300.

	OLS Estimation		GMM Estimation	
	(1)	(2)	(3)	(4)
$\log(L_t)$	0.61^{***}	0.84^{***}	0.60^{***}	0.80^{***}
	(0.02)	(0.06)	(0.02)	(0.06)
$\log(K_t)$	0.56^{***}	0.55^{***}	0.56^{***}	0.55^{***}
	(0.09)	(0.09)	(0.09)	(0.10)
$\log^2(K_t)$				
$\log^2(L_t)$		-0.01^{***}		-0.01^{***}
		(0.00)		(0.00)
$\log(K_t^{(2018)})$	-0.32^{***}	-0.31^{***}	-0.32^{***}	-0.31^{***}
	(0.09)	(0.09)	(0.09)	(0.09)
$\log(M_t^{(2018)})$	0.19^{***}	0.19^{***}	0.20^{***}	0.20^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
KP-$\mathcal{F}_n$			13,638	9,928
SY-$\mathcal{F}_n$			$11{,}320^{**}$	$10{,}618^{**}$
$\mathcal{D}_n$			4.92	0.04
p-value of $\mathcal{D}_n$			(0.00)	(0.94)
$\mathcal{J}_n$			19.72	2.31
p-value of $\mathcal{J}_n$			(0.00)	(0.68)
n	2,140	2,140	2,140	2,140

Table 6: THE OLS AND GMM ESTIMATES OF PRODUCTION FUNCTION. This table reports the OLS and GMM estimates of (13). The dependent variable is the log of value-added. In columns (1)-(2), the OLS estimates are reported, and in columns (3)-(4), the GMM estimates are reported. KP-$\mathcal{F}_n$ and SY-$\mathcal{F}_n$ denote Kleibergen and Papp's (2006) and Stock and Yogo's (2005) F-tests, respectively, and n denotes the total number of the firms available in the year of 2019. Robust standard errors are given in the parentheses below the estimates, and *, **, and *** denote that $p < 0.1$, $p < 0.05$, and $p < 0.01$, respectively.