SEQUENTIALLY TESTING POLYNOMIAL MODEL HYPOTHESES USING POWER TRANSFORMS OF REGRESSORS

By

Jin Seo Cho and Peter C. B. Phillips

July 2016

COWLES FOUNDATION DISCUSSION PAPER NO. 2060

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Sequentially Testing Polynomial Model Hypotheses using Power Transforms of Regressors

JIN SEO CHO

School of Economics

Yonsei University, 50 Yonsei-ro

Seodaemun-gu, Seoul, 120-749, Korea

PETER C.B. PHILLIPS

Yale University, University of Auckland

Singapore Management University &

University of Southampton

First version: May, 2013. This version: July, 2016

Abstract

We provide a methodology for testing a polynomial model hypothesis by extending the approach and results of Baek, Cho, and Phillips (2015; BCP) that tests for neglected nonlinearity using power transforms of regressors against arbitrary nonlinearity. We examine and generalize the BCP quasi-likelihood ratio test dealing with the multifold identification problem that arises under the null of the polynomial model. The approach leads to convenient asymptotic theory for inference, has omnibus power against general nonlinear alternatives, and allows estimation of an unknown polynomial degree in a model by way of sequential testing, a technique that is useful in the application of sieve approximations. Simulations show good performance in the sequential test procedure in identifying and estimating unknown polynomial order. The approach, which can be used empirically to test for misspecification, is applied to a Mincer (1958, 1974) equation using data from Card (1995). The results confirm that Mincer's log earnings equation is easily shown to be misspecified by including nonlinear effects of experience and schooling on earnings, with some flexibility required in the respective polynomial degrees.

Keywords: QLR test; Asymptotic null distribution; Misspecification; Mincer equation; Nonlinearity; Polynomial model; Power Gaussian process; Sequential testing.

JEL Classification: C12, C18, C46, C52.

Acknowledgements: The authors benefited from discussions with Seungmoon Choi, Chirok Han, Tae-Hwan Kim, Dong Jin Lee, Jong Hwa Lee, Hyungsik Roger Moon, and conference participants at SETA and NZESG (University of Waikato, 2016). Phillips acknowledges support from the NSF under Grant No. SES 12-58258.

1 Introduction

Polynomial models are popularly used in empirical work to address departures from linearity. When linear model assumptions are violated in the data or suspected of violation, polynomial specifications are often introducted to detect and cope with unknown forms of neglected nonlinearity. Quadratic, cubic, quartic, and even higher degree polynomial models are flexible, easy to estimate using least squares, and may be justified in terms of sieve approximation techniques in the context of general nonparametric formulations of nonlinearity.

Nevertheless, the validity of a polynomial model is often verified in only a limited fashion. For any pre-specified polynomial model, its given degree may be insufficient to detect nonlinearity in the data or it may be redundantly too high. Test statistics that are available in the literature do not tell the researcher the degree of nonlinearity to be included in the model without iterative testing when they reject the specified polynomial model.

The present paper makes a twofold contribution. First, we provide a methodology for testing a polynomial model hypothesis and detecting whether there is further neglected nonlinearity in the model. The approach adopted extends recent work of Baek, Cho, and Phillips (2015, BCP henceforth) for testing arbitrary nonlinearity using power transforms of regressors. The methodology is a convenient way of delivering an omnibus test for neglected nonlinearity by simple augmented regression. Second, we exploit the flexible feature of power transforms by estimating polynomial degree in a manner that assists in specifying a parsimonious polynomial model. For this purpose, we sequentially test the polynomial model hypothesis by increasing the polynomial degree and controlling the overall type-I error in the sequential testing procedure. The approach has a natural application in sieve nonparametric estimation for determining the dimension of a suitable sieve space.

Power transforms of regressors have been popular in the literature since Tukey's (1957, 1977) suggestion of the power transform as a mechanism to link the log linear model to the linear model. Box and Cox (1964) further developed the theory, leading to the so-called Box-Cox transform which elegantly corroborates Tukey's (1957, 1977) ladder formula showing the log transform as a limit form as the power exponent converges to zero. BCP used an augmented form of the Box-Cox transform in constructing a quasi-likelihod ratio (QLR) test for neglected nonlinearity. Power transforms are also popular in time series modeling, where Ding, Granger, and Engle (1993) and Duan (1997), for example, introduced the asymmetric power GARCH and augmented GARCH models by applying power transform methods. In developing nonlinear regression asymptotics, Wu (1981) and Phillips (2007) examined power transforms of time trends and showed that estimating such models involves asymptotic collinearities which lead to complications in implementation and limit theory as reviewed briefly below.

The approach pursued here extends the linear null model framework of BCP to a more general polynomial class, develops omnibus tests for further neglected nonlinearity by examining the effect of the power transform on prediction errors, and provides a statistical algorithm for estimating the degree of a polynomial model by sequentially testing the polynomial model. While in principle this approach may seem straightforward, it has not been attempted in the prior literature using power transform methods mainly because of the multiple identification problem that arises when testing the polynomial model assumption. Cho and Ishida (2012) and BCP showed that testing the linear model assumption by the power transform method introduces a trifold identification problem (bifold in the case of a location model). If the null model is an m-th degree polynomial model, identification is aggravated by the fact that there are now $m+2$ different ways to identify the model, leading to what we call a *multifold identification problem*. To the best of our knowledge, this multifold identification problem has never been addressed in the literature.

The goal of the present paper is to tackle this problem and provide a methodology for empirically testing a null polynomial model and identifying polynomial degree by means of sequential testing. Specifically, we consider two time-series models in parallel to BCP. The first case involves strictly stationary data and the quasi-likelihood ratio (QLR) test statistic of the null polynomial model here is shown to have a limit distribution in terms of a functional of a Gaussian process induced by the presence of multifold identification under the null, and we also show that the QLR test statistic possesses omnibus power under the alternative. That is, it consistently rejects the null polynomial model under an arbitrary alternative hypothesis. As we demonstrate below, the covariance kernel of this Gaussian process is dependent upon both the data generating process (DGP) and the model assumptions, so that the null limit critical values are case-dependent. Next, we examine the polynomial time-trend stationary model. Although the QLR test statistic in this case still converges weakly to a functional of a Gaussian process due to multifold identification, the covariance kernel is regular in the sense that if the prediction error is a martingale difference sequence (MDS), the null limit distribution is invariant to the conditional variance of the prediction error and to the degree of the null polynomial model. This invariance has the convenient implication that asymptotic critical values can be tabulated and these are provided by simulating a certain exponential Gaussian process (as in Cho and White, 2010). For these two time series contexts, we provide a sequential testing methodology that yields a consistent estimator of the polynomial degree by iterative hypothesis testing without resorting to data snooping. The methodology relies on suitable control of the overall test significance level to ensure a slow passage to zero as the sample size tends to infinity.

This estimation and inferential methodology has numerous applications in applied work. For example, the classic Mincer (1958, 1974) equation predicts individual log earnings as the sum of a linear function of schooling years and a quadratic function of years of potential experience. This equation has long been influential in empirical studies of human capital and similar second degree polynomials involving variables such as age and age squared are ubiquitous in empirical work in attempts to capture nonlinear effects in econometric modeling. These empirical models are also used as a primary motivation for the use of sieve approximations in nonparametric econometrics.

The second degree polynomial model of the Mincer equation provides a natural platform to apply the testing methodology developed in the current study. Accordingly, we apply the QLR test statistic to the Mincer equation and test the empirical adequacy of its form for explaining log earnings, using the national longitudinal survey data from Card (1995). Revisiting this application and testing the specification using the methods developed here, we conclude that the Mincer equation fails to capture the nonlinearity of earnings with respect to years of experience if the model is extended to include other explanatory variables.

The paper is organized as follows. Section 2 derives the null limit distribution of the QLR test statistic for the strictly stationary case. This section examines asymptotic power and develops a sequential testing algorithm for detecting polynomial degree in practical applications. Section 3 extends the analysis to the polynomial time-trend case. Section 4 reports simulations to assess finite sample performance and the adequacy of the sequential testing algorithm. Section 5 provides an empirical application of the methodology to a Mincer earnings equation. Concluding remarks are given in Section 6. Proofs are in the Appendix. For notational simplicity we use $(d^j/d^jx)f(0)$ to denote $(d^j/dx^j)f(x)|_{x=0}$ for some function f and positive integer j. Other notation is standard.

2 Sequential QLR Testing for Nonlinearity with Stationary Data

This section assumes stationary data and develops the QLR machinery for testing neglected nonlinearity and sequential testing to determine polynomial degree.

2.1 Model Formulation

We suppose that the researcher specifies a model $\mathcal{M}_m$ to characterize the systematic component $\mathbb{E}[y_t|z_t]$ of a scalar endogenous variable y_t given a set of covariates $z_t := (x_t(m)', d_t')' := (1, x_t, \ldots, x_t^m, d_t')'$ that involve m-th degree polynomial components of some process x_t. The model $\mathcal{M}_m$ is formulated as

$$\mathcal{M}_m := \{\mathbb{E}[y_t|z_t] = \mu_t(\cdot) : \Omega \mapsto \mathbb{R} \text{ with } \mu_t(\alpha, \eta, \beta, \gamma) := \ x_t(m)'\alpha + d_t'\eta + \beta x_t^{\gamma}\}, \tag{1}$$

in which the power transform component x_t^{γ} is introduced to allow for possible additional nonlinearity in $\mathbb{E}[y_t|z_t]$ beyond conventional polynomial effects. In (1), the variables $(y_t, x_t, d_t')' \in \mathbb{R}^{2+k}$ ($k \in \mathbb{N}$) are assumed to be strictly stationary and ergodic, x_t is strictly nonnegative with probability 1, and the parameter space for $\omega := (\alpha', \eta', \beta, \gamma)' := (\alpha_0, \ldots, \alpha_m, \eta', \beta, \gamma)'$ is $\Omega \subset \mathbb{R}^{3+m+k}$. It is further assumed that the signal matrix $Z'Z = \sum_{t=1}^{n} z_t z_t'$ is nonsingular, where $Z = [z_1, ..., z_n]'$ is the observation matrix and n is the sample size. This model extends the framework of BCP where it is assumed that the base model is linear and $m = 1$. The model $\mathcal{M}_m$ is motivated by the concern that an m-th degree polynomial model may not be flexible enough to detect any remaining nonlinearity in $\mathbb{E}[y_t|z_t]$. This model is specifically formulated to facilitate testing the following hypothesis:

$$\mathcal{H}_{0,m} : \exists(\alpha_*', \eta_*')', \ \ \mathbb{E}[y_t|x_t, d_t] = x_t(m)'\alpha_* + d_t'\eta_* \text{ with probability } 1, \tag{2}$$

so that the m-th degree polynomial model becomes the null model whereas $\mathcal{M}_m$ is treated as the alternative.

Many irregular issues of identification are entailed by transition from $\mathcal{M}_m$ to the null model. In particular, the null model can be separately generated from $\mathcal{M}_m$ by imposing a number of restrictions, each of which bears its own model identification signature (*c.f.*, Davies, 1977, 1987). Thus, if the parameter space of γ, denoted by Γ, contains the elements $\{0, 1, \ldots, m\}$, there are $(m + 2)$ different ways to obtain the null model from $\mathcal{M}_m$. First, for each $j = 1, 2, \ldots, m + 1$, if $\gamma_* = j - 1$, the coefficient of x_t^{j-1} becomes $(\alpha_{(j-1)*} + \beta_*)$, thereby leading to the null model. Nevertheless, $\alpha_{(j-1)*}$ and β_* are not separately identified although their sum is identified. Second, the null model is obtained by letting $\beta_* = 0$, but γ_* is itself not identified, leading to a further identification problem. As a result, there are $(m + 2)$ different ways to obtain the null model from $\mathcal{M}_m$, and, accordingly, $(m + 2)$ different identification problems. We may separately state these in terms of the explicit sub-hypotheses

$$\mathcal{H}_{0,m}^{(1)} : \gamma_* = 0; \quad \ldots \quad \mathcal{H}_{0,m}^{(m+1)} : \gamma_* = m; \quad \text{and} \quad \mathcal{H}_{0,m}^{(m+2)} : \beta_* = 0.$$

In the following subsections, we examine the limit distribution of the QLR test statistic defined as

$$QLR_n := n\left(1 - \frac{\widehat{\sigma}_{n,A}^2}{\widehat{\sigma}_{n,0}^2}\right)$$

under each null hypothesis. Here, $\widehat{\sigma}^2_{n,A}$ and $\widehat{\sigma}^2_{n,0}$ are the means of the squared residuals obtained respectively from the model $\mathcal{M}_m$ and the null model hypothesis. The quasi-likelihood (QL) function is

$$L_n(\alpha,\eta,\beta,\gamma) := -\sum_{t=1}^{n}(y_t - \ x_t(m)'\alpha - d_t'\eta - \beta x_t^{\gamma})^2,$$

so that $\widehat{\sigma}^2_{n,A} := -n^{-1}\max_{\alpha,\eta,\beta,\gamma} L_n(\alpha,\eta,\beta,\gamma)$ and $\widehat{\sigma}^2_{n,0} := -n^{-1}\max_{\alpha,\eta,\beta,\gamma} L_n(\alpha,\eta,0,\gamma)$, where γ in the latter is simply a placeholder whose value is irrelevant under the null. For notational simplicity, we also let $c := j-1$ from now. Therefore, c runs from 0 to m given that $j = 1,2,\ldots,m+1$. As we demonstrate below, the QLR test possesses omnibus power for detecting neglected nonlinearity.

The following conditions that are assumed throughout this section to fix ideas and develop an asymptotic theory of inference.

Assumption 1. *(i) $(y_t, x_t, d_t')' \in \mathbb{R}^{2+k}$ ($k \in \mathbb{N}$) is a strictly stationary and absolutely regular process with mixing coefficient β_ℓ such that for some $r > 1$, $\sum_{\ell=1}^{\infty} \ell^{1/(r-1)}\beta_\ell < \infty$, $\mathbb{E}[|y_t|] < \infty$, and x_t is nonnegative with probability 1;*

(ii) The model for $\mathbb{E}[y_t|x_t,d_t]$ is specified as $\mathcal{M}_m := \{\mu_t(\cdot) : \Omega \mapsto \mathbb{R} : \mu_t(\alpha,\eta,\beta,\gamma) := \ x_t(m)'\alpha + d_t'\eta + \beta x_t^{\gamma}\}$, where Ω is the parameter space of $\omega := (\alpha',\eta',\beta,\gamma)'$, $z_t := (\ x_t(m)', d_t')'$, and n is the sample size;

(iii) $\Omega = (\prod_{i=0}^{m}\bar{A}_i) \times H \times \bar{B} \times \Gamma$ such that H, $\bar{B}$, and Γ are convex and compact parameter spaces in $\mathbb{R}^k$, $\mathbb{R}$, and $\mathbb{R}$, respectively, with $0,1,\cdots,m$ being interior elements of Γ, and for $i = 0,1,\cdots,m$, $\bar{A}_i$ is also a convex and compact parameter space in $\mathbb{R}$; and

(iv) $Z'Z = \sum_{t=1}^{n} z_t z_t'$ is nonsingular with probability 1. $\square$

Assumption 2. *(i) For each $\epsilon > 0$, $A(\gamma) := \mathbb{E}[G_t(\gamma)G_t(\gamma)']$ and $B(\gamma) := \mathbb{E}[u_t^2 G_t(\gamma)G_t(\gamma)']$ are positive definite uniformly on $\Gamma(\epsilon) := \{\gamma \in \Gamma : \gamma \notin \cup_{i=0}^{m}(i-\epsilon, i+\epsilon)\}$, where $G_t(\gamma) := [z_t',\ x_t(m)'\log(x_t), x_t^{\gamma}]'$, and $u_t := y_t - \mathbb{E}[y_t|z_t]$;*

(ii) $\{u_t, \mathcal{F}_t\}$ is an MDS, where $\mathcal{F}_t$ is the smallest σ-field generated by $\{z_{t+1}, u_t, z_t, u_{t-1}, \cdots\}$;

(iii) There is a strictly stationary and ergodic sequence $\{m_t, s_t\}$ such that for $i = 1,2,\cdots,k$, $|d_{t,i}| \leq m_t$, $\mathbb{E}[m_t^{4\rho}] < \infty$, $\mathbb{E}[s_t^8] < \infty$, where $d_{t,i}$ is the i-th row element of d_t, and

(iii.a) $|u_t| \leq m_t$, $|x_t^m| \leq s_t$, and $|\log(x_t)| \leq s_t$;

(iii.b) $|x_t^m| \leq m_t$, $|u_t| \leq s_t$, and $|\log(x_t)| \leq s_t$; or

(iii.c) $|\log(x_t)| \leq m_t$, $|u_t| \leq s_t$, and $|x_t^m| \leq s_t$;

(iv) $\sup_{\gamma\in\Gamma}|x_t^{\gamma}| \leq m_t$ and $\sup_{\gamma\in\Gamma}|x_t^{\gamma}\log(x_t)| \leq m_t$; and

(v) $r = \rho$. $\square$

In the above notation, it would be more precise to write z_t as $z_t(m)$, which accords more closely to the definition $z_t := (x_t(m)', d_t')'$ in terms of $x_t(m)$. However, we suppress the argument m for notational simplicity and it will be implicit in what follows until we examine sequential testing. The majorization and moment conditions given in Assumption 2(*iii*) are alternates and do not imply one another. It transpires that if at least one of these conditions separately holds, then the desired results given below follow. Further details regarding these conditions are provided when claims relevant to the conditions are stated and discussed below.

2.2 Limit Distribution of the QLR Test Statistic under $\mathcal{H}_{0,m}^{(j)} : \gamma_* = c$ with $c = 0, 1, \ldots, m$

We first examine the limit behavior of the QLR test statistic under $\mathcal{H}_{0,m}^{(j)} : \gamma_* = c$, where $c = 0, 1, \ldots, m$ or $c = j - 1$ for $j = 1, ..., m + 1$. Due to the recursive structure of the polynomial model, it turns out that there is a systematic relationship between the null limit approximations for different values of c. This relationship is exploited for an efficient delivery of the null limit distributions for different c values.

Under $\mathcal{H}_{0,m}^{(j)} : \gamma_* = c$ we have

$$\mathbb{E}[y_t|x_t, d_t] = \sum_{i=0,\ i\neq c}^{m} \alpha_{i*} x_t^i + (\alpha_{c*} + \beta_*) x_t^c + d_t' \eta_*,$$

and then neither α_{c*} nor β_* is separately identified without imposing some additional condition, although $(\alpha_{c*} + \beta_*)$ is an identified composite coefficient. Thus, imposing every possible additional condition for the model identification we examine how the resulting null limit distributions are associated with each other. As will become apparent, this process derives the desired limit distribution under $\mathcal{H}_{0,m}^{(j)}$.

Our analysis is conducted in three steps. First, we let β be unidentified and fix its value so that α_{c*} is identified. Through this identification scheme (conditional on the fixed value β), we obtain the null limit distribution for that fixed value β. Similarly, we select another value of β and iterate the same steps, examining how the separately obtained null limit distributions are associated with each other. By this process, we can characterize the null limit distribution of the QLR test statistic when β is fixed. Second, we modify the identification scheme by fixing the value α_c so that β_* is identified. By iterating steps analogous to the β-fixed case, we can characterize the null limit distribution. Finally, we examine how the two characterized null limit distributions are associated with each other, as obtained in the first two sequence of steps, which leads us to derive the null limit distribution under $\mathcal{H}_{0,m}$. The schema is described in full in what follows.

2.2.1 When β_* is Not Identified

We first fix β and approximate the constrained quasi-likelihood (CQL) with respect to the other identified parameters $(\alpha_*', \eta_*')'$. Let the following be the CQL function:

$$L_n(\gamma; \beta) := L_n(\widehat{\alpha}_n(\gamma; \beta), \widehat{\eta}_n(\gamma; \beta), \beta, \gamma),$$

where $(\widehat{\alpha}_n(\gamma; \beta), \widehat{\eta}_n(\gamma; \beta)')' := \arg\max_{\alpha, \eta} L_n(\alpha, \beta, \gamma, \eta)$. Upon calculation the CQL is given by the explicit formula

$$L_n(\gamma; \beta) = -\{Y - \beta X(\gamma)\} M \{Y - \beta X(\gamma)\},$$

where $M := I_n - Z(Z'Z)^{-1}Z'$, and $X(\gamma) := (x_1^\gamma \ldots x_n^\gamma)'$. Note that $MY = MU$ under $\mathcal{H}_{0,m}$, where $U := (u_1, u_2, \ldots, u_n)'$. For notational simplicity, define

$$A_c := [x_1^c \log(x_1), \ldots, x_n^c \log(x_n)]', \quad B_c := [x_1^c \log^2(x_1), \ldots, x_n^c \log^2(x_n)]',$$

and apply a second-order Taylor expansion to obtain

$$\sup_\gamma \{L_n(\gamma; \beta) - L_n(c; \beta)\} = \frac{\{\beta A_c' M U\}^2}{\beta^2 A_c' M A_c - \beta B_c' M U} + o_{\mathbb{P}}(1) = \frac{\{A_c' M U\}^2}{A_c' M A_c} + o_{\mathbb{P}}(1), \tag{3}$$

using the fact that $B_c' M U = o_{\mathbb{P}}(n)$ under Assumptions 1 and 2. This result follows mainly from the simple form of the derivatives $(d/d\gamma) L_n(c; \beta) = 2\beta A_c' M U$ and $(d^2/d\gamma^2) L_n(c; \beta) = 2\beta B_c' M U - 2\beta^2 A_c' M A_c$. We thus obtain the following null limit approximation of the QLR test statistic

$$QLR_n^{(\gamma=c;\beta)} := \sup_\beta \sup_\gamma n \left\{1 - \frac{L_n(\gamma; \beta)}{L_n(c; \beta)}\right\} = \frac{1}{\widehat{\sigma}_{n,0}^2} \frac{\{A_c' M U\}^2}{\{A_c' M A_c\}} + o_{\mathbb{P}}(1). \tag{4}$$

This representation implies that the optimization process with respect to β in (4) is asymptotically innocuous in obtaining the null limit distribution. In (4), the notation $QLR_n^{(\gamma=c;\beta)}$ is used to denote the QLR test statistic that tests $\mathcal{H}_{0,m}^{(j)} : \gamma_* = c$ by fixing β first and subsequently maximizing with respect to γ and β.

2.2.2 When α_{c*} is Not Identified

We next fix α_c first and use the notation α_{-c} to signify the vector α with all elements except α_c. If α_c is fixed, the other parameters $(\alpha_{-c*}', \beta_*, \eta_*')' := (\alpha_{0*}, \alpha_{1*}, \ldots, \alpha_{(c-1)*}, \alpha_{(c+1)*}, \ldots, \alpha_{m*}, \beta_*, \eta_*')'$ are identified under the null. Therefore, we first optimize the QL function with respect to $(\alpha_{-c}', \beta, \eta')'$ in the first stage and then maximize the QL function with respect to γ and finally with respect to α_c. For this purpose, we let $(\widehat{\alpha}_{-c,n}(\gamma;\alpha_c)', \widehat{\beta}_n(\gamma;\alpha_c), \widehat{\eta}_n(\gamma;\alpha_c)')' := \arg\max_{\alpha_{-c},\beta,\eta} L_n(\alpha,\beta,\gamma,\eta)$, whose specific form is

$$(\widehat{\alpha}_{-c,n}(\gamma;\alpha_c)', \widehat{\beta}_n(\gamma;\alpha_c), \widehat{\eta}_n(\gamma;\alpha_c)')' = [Q_c(\gamma)'Q_c(\gamma)]^{-1}Q_c(\gamma)'P_c(\alpha_c),$$

where $P_c(\alpha_c) := Y - \alpha_c X(c)$, $Q_c(\gamma) := [X(0), \ldots, X(c-1), X(\gamma), X(c+1), \ldots, X(m), D]$, and $D := [d_1, \ldots, d_n]'$. The CQL is also obtained as

$$L_n(\gamma;\alpha_c) = -P_c(\alpha_c)'\{I_n - Q_c(\gamma)[Q_c(\gamma)'Q_c(\gamma)]^{-1}Q_c(\gamma)'\}P_c(\alpha_c).$$

We approximate this CQL by a second-order Taylor expansion with respect to γ at c. Some algebra delivers the following first-two derivatives:

$$L_n^{(1)}(c;\alpha_c) = 2(\alpha_{c*} - \alpha_c)A_c'MU + 2U'E_c(Z'Z)^{-1}Z'U - U'Z(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}Z'U,$$

where $E_c := (d/d\gamma)Q_c(c) = [0_{n\times c} \vdots A_c \vdots 0_{n\times(m-c+k)}]$, and

$$\begin{aligned} L_n^{(2)}(c;\alpha_c) =& 2(Z\kappa_c + U)'\{E_c(Z'Z)^{-1}E_c' + Z(Z'Z)^{-1}F_c'\}(Z\kappa_c + U) \\ &- 4(Z\kappa_c + U)'Z(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}E_c'(Z\kappa_c + U) \\ &- (Z\kappa_c + U)'Z(Z'Z)^{-1}(2E_c'E_c + Z'F_c + F_c'Z)(Z'Z)^{-1}Z'(Z\kappa_c + U) \\ &+ 2(Z\kappa_c + U)'Z(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}Z'(Z\kappa_c + U), \end{aligned}$$

where $F_c := (d^2/d\gamma^2)Q_c(c) = [0_{n\times c} \vdots B_c \vdots 0_{n\times(m-c+k)}]$, and $\kappa_c := [\alpha_{0*}, \ldots, \alpha_{(c-1)*}, (\alpha_{c*} - \alpha_c), \alpha_{(c+1)*}, \ldots, \alpha_{m*}, \eta_*']'$. These first-two derivatives are derived in the Appendix. Their null limit behavior is given in the following result.

Lemma 1. *Given Assumptions 1, 2, and $\mathcal{H}_{0,m}^{(j)}$, for each $c = 0, 1, \ldots, m$, we have:*

(i) $L_n^{(1)}(c;\alpha_c) = 2(\alpha_{c*} - \alpha_c)A_c'MU + o_{\mathbb{P}}(\sqrt{n})$*; and*

(ii) $L_n^{(2)}(c;\alpha_c) = -2(\alpha_{c*} - \alpha_c)^2 A_c'MA_c + o_{\mathbb{P}}(n)$. □

The proof of Lemma 1 is given in the Appendix. Using Lemma 1 and a second-order Taylor expansion, it follows that

$$QLR_n^{(\gamma=c;\alpha_c)} := \sup_{\alpha_c}\sup_{\gamma} n\left\{1 - \frac{L_n(\gamma;\alpha_c)}{L_n(c;\alpha_c)}\right\} = \frac{1}{\widehat{\sigma}_{n,0}^2}\frac{\{A_c'MU\}^2}{\{A_c'MA_c\}} + o_{\mathbb{P}}(1). \quad (5)$$

Here, $QLR_n^{(\gamma=c;\alpha_c)}$ is used to denote the QLR test statistic that tests $\mathcal{H}_{0,m}^{(c+1)} : \gamma_* = c$ by fixing α_c first and subsequently maximizing with respect to γ and α_c.

Some remarks are warranted. First, although α_c is treated as an unidentified nuisance parameter, it asymptotically cancels out in the ratio limit, just as in the β-fixed case. Thus, the final optimization process in (5) with respect to α_c does not affect the null limit distribution. Second, the null approximation given on the right side of (5) is asymptotically identical to the right side of (4), implying that the limit obtained by fixing β first is identical to that obtained by fixing α_c first, and that the limit approximation of the QLR test statistic under $\mathcal{H}_{0,m}^{(j)}$ is identical irrespective of whether β or α_j is optimized in the final stage. This property leads directly to the following lemma.

Lemma 2. *Given Assumptions 1 and 2,* $QLR_n^{(\gamma=c)} = \{A_c' MU\}^2 / \{\widehat{\sigma}_{n,0}^2 (A_c' M A_c)\} + o_{\mathbb{P}}(1)$ *under* $\mathcal{H}_{0,m}^{(j)} : \gamma_* = c$*, where* $c = 0, 1, \ldots, m$. □

Here, $QLR_n^{(\gamma=c)}$ denotes the QLR test statistic that tests $\mathcal{H}_{0,m}^{(j)} : \gamma_* = c$. BCP obtained the same result for the special case $m = 1$. Third, the null limit approximation in Lemma 2 has a regular pattern across different null hypotheses. Thus, for a different index (say, $c' = j' - 1$) the limit approximation under $\mathcal{H}_{0,m}^{(j')}$ is obtained by simply replacing A_c in Lemma 2 with $A_{c'} := [x_1^{c'} \log(x_1), \ldots, x_n^{c'} \log(x_n)]'$. This simple regular pattern is produced because of the recursive structure of the polynomial model. Fourth, the derivation of Lemma 2 is virtually an immediate consequence of a second-order Taylor expansion, and this is a very convenient feature of the power transform in comparison with other approaches as we now explain.

Cho, Ishida, and White (2011, 2014) and White and Cho (2012) examined testing linear model hypotheses by adding an analytic function to the linear model following the framework of Bierens (1990) and Stinchcombe and White (1998). They showed that higher-order Taylor expansions are necessary in deriving the null limit distribution of the QLR test statistic. If the so-called no-zero condition holds for the analytic function, a fourth-order Taylor expansion is needed; and if the no-zero condition does not hold, sixth-, eighth-, or even higher-order Taylor expansions are needed, depending on the property of the analytic function in use. This consequence is further aggravated if a polynomial model is the null model. Then, a further higher-order Taylor expansion is needed even when the no-zero condition holds, depending on the polynomial degree under the null model condition. On the other hand, the power transform simplifies the model approximation because at most a second-order Taylor expansion is needed. This feature explains the advantage of using the power transform instead of other nonlinear functions for detecting further neglected nonlinearity.

Finally, the augmented Box-Cox transform in BCP can be further generalized to be adapted to the polynomial model. Note that if we modify the Box-Cox transform for use in the present context as

$$ABC_t(\gamma; c) := \begin{cases} (x_t^\gamma - x_t^c)/(\gamma - c), & \text{if } \gamma \neq c; \\ x_t^c \log[x_t], & \text{if } \gamma = c \end{cases}$$

by noting that

$$\lim_{\gamma \to c} \frac{x_t^\gamma - x_t^c}{\gamma - c} = x_t^c \log[x_t], \tag{6}$$

this formulation generalizes the augmented Box-Cox transformation of BCP, in which $c = 1$. Note that the right side of (6) is a typical element of A_c. This implies that if the conditional mean $\mathbb{E}[y_t|x_t, d_t]$ is parameterized as $x_t(m)'\alpha_* + d_t'\eta_* + \xi_* x_t^c \log[x_t]$. Then testing the hypothesis that $\xi_* = 0$ is equivalent to testing $\mathcal{H}_{0,m}^{(j)}$ in our context, where $j = 1, 2, \ldots, m+1$.

2.3 Limit Distribution of the QLR Test Statistic under $\mathcal{H}_{0,m}^{(m+2)} : \beta_* = 0$

We consider the limit behavior of the QLR test statistic under $\mathcal{H}_{0,m}^{(m+2)} : \beta_* = 0$. As γ_* is not identified under $\mathcal{H}_{0,m}^{(m+2)}$, we first approximate the model with respect to the other parameters $(\alpha', \beta, \eta')'$ and maximize the QL function with respect to γ in the final stage. For notational simplicity, we let the CQL function be denoted by $L_n(\beta; \gamma) := L_n(\widehat{\alpha}_n(\beta; \gamma), \widehat{\eta}_n(\beta; \gamma), \beta, \gamma)$, where $(\widehat{\alpha}_n(\beta; \gamma), \widehat{\eta}_n(\beta; \gamma)')' := \arg\max_{\alpha,\eta} L_n(\alpha, \eta, \beta, \gamma)$. The CQL has the following specific form:

$$L_n(\beta; \gamma) = -\{Y - \beta X(\gamma)\}' M \{Y - \beta X(\gamma)\}. \tag{7}$$

Using this, we obtain the following limit approximation of the QLR test:

$$QLR_n^{(\beta=0)} := \sup_{\gamma} \sup_{\beta} n \left\{1 - \frac{L_n(\beta; \gamma)}{L_n(0; \gamma)}\right\} = \sup_{\gamma} \frac{1}{\widehat{\sigma}_{n,0}^2} \frac{\{X(\gamma)' M U\}^2}{X(\gamma)' M X(\gamma)}. \tag{8}$$

Here, $QLR_n^{(\beta=0)}$ is used to denote the QLR test statistic that tests the hypothesis $\mathcal{H}_{0,m}^{(m+2)} : \beta_* = 0$.

Some remarks are in order to highlight this approximation. Note that the approximation in (8) has the same form as that in BCP. Therefore, we can apply the functional central limit theorem (FCLT) and the uniform law of large numbers (ULLN) to $n^{-1/2} X(\cdot)' M U$ and $n^{-1} \widehat{\sigma}_{n,0}^2 X(\cdot)' M X(\cdot)$, respectively as in BCP. Nevertheless, we further note that for $c = 0, 1, \ldots, m$, $\operatorname{plim}_{\gamma \to c} X(\gamma)' M X(\gamma) = 0$ and $\operatorname{plim}_{\gamma \to c} X(\gamma)' M U = 0$ because $\lim_{\gamma \to c} X(\gamma) = [x_1^c, x_2^c, \ldots, x_n^c]'$ and M is the idempotent matrix formed from the regressor $z_t := (x_t(m)', d_t')'$. As these limits are those of the numerator and denominator constituting (8), the probability limit

$$\underset{\gamma\to c}{\text{plim}}\,\frac{1}{\widehat{\sigma}_{n,0}^2}\frac{\{X(\gamma)'MU\}^2}{X(\gamma)'MX(\gamma)}$$

is an indeterminate form. Applying l'Hôpital's rule we obtain $\text{plim}_{\gamma\to c}\, 2\{X(\gamma)'MU\}\,\{(d/d\gamma)X(\gamma)'MU\} = 0$ and $\text{plim}_{\gamma\to c}\, 2\,\{(d/d\gamma)X(\gamma)'MX(\gamma)\} = 0$, which imply that a first-order application of l'Hôpital's rule is insufficient to determine the probability limit. Moving to the next order, the probability limits from the second-order derivatives are $\text{plim}_{\gamma\to c}(d^2/d\gamma^2)X(\gamma)'MU = 2\{A_c'MU\}^2$ and $\text{plim}_{\gamma\to c}(d^2/d\gamma^2)X(\gamma)'M X(\gamma) = 2A_c'MA_c$. It follows that for $c = 0, 1, \ldots, m$, we have the following limit

$$\underset{\gamma\to c}{\text{plim}}\,\frac{1}{\widehat{\sigma}_{n,0}^2}\frac{\{X(\gamma)'MU\}^2}{X(\gamma)'MX(\gamma)} = \frac{1}{\widehat{\sigma}_{n,0}^2}\frac{\{A_c'MU\}^2}{A_c'MA_c}. \tag{9}$$

Some regularity conditions are needed to justify the limit behavior of this ratio as $n\to\infty$. Specifically, we need conditions for applying the central limit theorem (CLT) and FCLT to $n^{-1/2}[A_0'MU, A_1'MU, \ldots, A_m'MU]'$ and $n^{-1/2}X(\cdot)'MU$, respectively. In a similar manner, it is necessary to simultaneously apply a law of large numbers (LLN) and ULLN to $n^{-1}[A_0'MA_0, A_1'MA_1, \ldots, A_m'MA_m]'$ and $n^{-1}X(\cdot)'MX(\cdot)$, respectively. The conditions in Assumptions 1 and 2 are sufficient for this purpose. In particular, the quantities $n^{-1/2}\sum G_t(\cdot)'u_t$ and $n^{-1}\sum G_t(\cdot)\,G_t(\cdot)'$ obey the FCLT and ULLN because the components constituting $G_t(\cdot) := [z_t',\ x_t(m)'\log(x_t), x_t^{(\cdot)}]'$ also constitute both $n^{-1/2}[X(\cdot)'MU, A_0'MU, \ldots, A_m'MU]'$ and $n^{-1}[X(\cdot)'MX(\cdot), A_0'MA_0, \ldots, A_m'MA_m]$. As a result, the null limit behavior of the QLR test statistic is obtained as a functional of these components, as shown in the following lemma.

Lemma 3. *Given Assumptions 1, 2, and $\mathcal{H}_{0,m}^{(m+2)}$,*

(i) $QLR_n^{(\beta=0)} = \sup_{\gamma\in\Gamma}\{X(\gamma)'MU\}^2/\{\widehat{\sigma}_{n,0}^2X(\gamma)'MX(\gamma)\}$; and

(ii) $QLR_n^{(\beta=0)} \Rightarrow \sup_{\gamma\in\Gamma}\mathcal{Z}(\gamma)^2$ as $n\to\infty$, where $\mathcal{Z}(\cdot)$ is a mean-zero Gaussian process whose covariance kernel for each $\gamma, \gamma' \in \Gamma$ is

$$\mathbb{E}[\mathcal{Z}(\gamma)\mathcal{Z}(\gamma')] = \frac{\mathbb{E}[\mathcal{G}(\gamma)\mathcal{G}(\gamma')]}{\sqrt{\sigma^2(\gamma,\gamma)}\sqrt{\sigma^2(\gamma',\gamma')}},$$

with $\sigma^2(\gamma,\gamma) := \sigma_*^2(\mathbb{E}[x_t^{2\gamma}] - \mathbb{E}[x_t^\gamma z_t']\mathbb{E}[z_tz_t']^{-1}\mathbb{E}[z_tx_t^\gamma])$, and $\mathcal{G}(\cdot)$ is a mean-zero Gaussian process with covariance kernel for each $\gamma, \gamma' \in \Gamma$,

$$\begin{aligned}\mathbb{E}[\mathcal{G}(\gamma)\mathcal{G}(\gamma')] = \mathbb{E}[u_t^2x_t^{\gamma+\gamma'}] - \mathbb{E}[u_t^2x_t^\gamma z_t']\mathbb{E}[z_tz_t']^{-1}\mathbb{E}[z_tx_t^{\gamma'}] - \mathbb{E}[u_t^2x_t^{\gamma'}z_t']\mathbb{E}[z_tz_t']^{-1}\mathbb{E}[z_tx_t^\gamma]\\ + \mathbb{E}[x_t^\gamma z_t']\mathbb{E}[z_tz_t']^{-1}\mathbb{E}[u_t^2z_tz_t']\mathbb{E}[z_tz_t']^{-1}\mathbb{E}[z_tx_t^{\gamma'}].\end{aligned}$$ □

Note that $\mathcal{G}(\cdot)$ is the weak limit of $n^{-1/2}X(\cdot)'MU$. Given Lemma 2, the limit result (*ii*) in Lemma 3 is identical in form to that of theorem 1 in BCP, and the null limit behavior of the QLR test statistic is obtained in the same way as for the linear model case. We can use the FCLT in Doukhan, Massart, and Rio (1995) to verify tightness of the process $\{n^{-1/2}X(\cdot)MU\}$ and weak convergence to $\mathcal{G}(\cdot)$, and we can apply the Andrews (1992) ULLN to $\{n^{-1}X(\cdot)'MX(\cdot)\}$. Since z_t is defined using the polynomial terms $x_t^2, x_t^3, \ldots x_t^m$, the covariance kernel of $\mathcal{G}(\cdot)$ differs for different values of m. The proof of Lemma 3 is almost identical to that of theorem 1 in BCP, and is therefore omitted.

An additional feature of interest is worth highlighting. The associated score function in the QLR test statistic is discontinuous at c, where $c = 0, 1, \ldots, m$ although it is smooth elsewhere in Γ. Define $z_n(\cdot) := \{\widehat{\sigma}_{n,0}^2 X(\cdot)'MX(\cdot)\}^{-1/2}\{X(\cdot)'MU\}$, which is the sample analog of $\mathcal{Z}(\cdot)$. For each $c = 0, 1, \ldots, m$, it is not hard to show that $\mathrm{plim}_{\gamma\uparrow c} z_n(\gamma) = -\mathrm{plim}_{\gamma\downarrow c} z_n(\gamma)$. This discontinuity applies also to the weak limit $\mathcal{Z}(\cdot)$, generalizing the observation in BCP for the case where $m = 1$. However, it follows that $\mathrm{plim}_{\gamma\uparrow c}\mathcal{Z}(\gamma)^2 = \mathrm{plim}_{\gamma\downarrow c}\mathcal{Z}(\gamma)^2$. Therefore, if we let $\mathcal{Z}(c)^2$ be defined as $\mathrm{plim}_{\gamma\to c}\mathcal{Z}(\gamma)^2$, $\mathcal{Z}(\cdot)^2$ is continuous at each c. On the other hand, $z_n(\cdot)$ is twice continuously differentiable elsewhere in Γ, a consequence of the fact that the power transform is infinitely smooth for all $\gamma \geq 0$ and positive $x > 0$, which in turn implies second-order differentiability of the covariance kernel of $\mathcal{Z}(\cdot)$ over the same region of Γ. Thus, $\mathcal{Z}(\cdot)^2$ is continuous on Γ almost surely, and $\sup_{\gamma\in\Gamma}\mathcal{Z}(\gamma)^2$ is well defined from the fact that Γ is a compact set.

2.4 Limit Distribution of the QLR Test Statistic under $\mathcal{H}_{0,m}$

We now examine the relationships among the limit approximations obtained under each hypothesis. This examination is conducted to obtain the null limit approximation of the QLR test statistic under $\mathcal{H}_{0,m}$.

The null limit approximations given in Sections 2.2 and 2.3 are those obtained by imposing all possible conditions to produce the null model from $\mathcal{M}_m$. By the definition of the QLR test statistic, the null limit approximation has to be obtained as the maximum of all null approximations, and the null approximation derived under $\mathcal{H}_{0,m}^{(m+2)}$ dominates the other null approximants. The null approximation in (4) is identical to the right side of (9). Thus, for each $c = 0, 1, \ldots, m$, we have

$$\sup_{\gamma\in\Gamma}\frac{1}{\widehat{\sigma}_{n,0}^2}\frac{\{X(\gamma)'MU\}^2}{\{X(\gamma)'MX(\gamma)\}} \geq \frac{1}{\widehat{\sigma}_{n,0}^2}\frac{\{A_c'MU\}^2}{\{A_c'MA_c\}} + o_{\mathbb{P}}(1), \tag{10}$$

where the left side of (10) is the QLR test statistic obtained under $\mathcal{H}_{0,m}^{(m+2)} : \beta_* = 0$, and the right side is the limit approximation of the QLR test statistic derived under $\mathcal{H}_{0,m}^{(c+1)} : \gamma_* = c$ as given in Lemmas 3 and 2, respectively. This fact implies that for every $c = j-1$, $QLR_n^{(\beta=0)}$ dominates $QLR_n^{(\gamma=c)}$, from which we conclude the following result.

Theorem 1. *Given Assumptions 1 and 2, $QLR_n \Rightarrow \sup_{\gamma \in \Gamma} \mathcal{Z}(\gamma)^2$ under $\mathcal{H}_{0,m}$.* $\square$

Note that the covariance kernel of $\mathcal{Z}(\cdot)$ in Lemma 3 depends on the joint distribution of (u_t, z_t), and so a different kernel is derived for each different model and/or conditional variance condition of u_t, which implies that the QLR test statistic is not a distribution-free test statistic. Accordingly, different models yield different asymptotic critical values although they are specified in terms of the same data. As BCP show by simulation, Hansen's (1996) weighted bootstrap is useful for obtaining the asymptotic critical values in this case.

2.5 Asymptotic Power of the QLR Test Statistic

BCP showed that the QLR test statistic possesses omnibus and local power for models with $m = 1$ and this property holds for $m > 1$. To demonstrate we let $\mathbb{E}[y_t|x_t, d_t] = \ x_t(m)'\alpha_* + d_t'\eta_* + s(x_t)$ under the alternative, where $s(\cdot)$ is a continuous nonlinear function whose nonlinearity cannot be represented in terms of an m-th degree polynomial. With this formulation the QLR test statistic can be shown to have power against an arbitrary nonlinear function $s(\cdot)$ that satisfies this property.

Define

$$h_0 := \min_{\alpha,\eta} \mathbb{E}[(y_t - \ x_t(m)'\alpha - d_t'\eta)^2] = \mathbb{E}[u_t^2] + \mathbb{E}[q_t^2],$$

$$h(\gamma) := \min_{\alpha,\eta} \mathbb{E}[(y_t - \ x_t(m)'\alpha - d_t'\eta - \beta x_t^{\gamma})^2] = \mathbb{E}[u_t^2] + \mathrm{var}[q_t] - \mathrm{cov}[u_t(\gamma), q_t]^2/\mathrm{var}[q_t],$$

where $u_t(\gamma) := x_t^{\gamma} - z_t'\mathbb{E}[z_t z_t']^{-1}\mathbb{E}[z_t x_t^{\gamma}]$ and $q_t := s(x_t) - z_t'\mathbb{E}[z_t z_t']^{-1}\mathbb{E}[z_t s(x_t)]$, and note that the probability limit of the QLR test statistic is obtained as

$$\frac{1}{n} QLR_n = \sup_{\gamma \in \Gamma} \left(1 - \frac{h(\gamma)}{h_0}\right) + o_{\mathbb{P}}(1)$$

under the given regularity condition. In view of this representation, the power of the QLR test statistic derives from the fact that $\inf_{\gamma \in \Gamma} h(\gamma)$ is strictly less than h_0 for any arbitrarily selected $s(\cdot)$. In addition, the QLR test statistic has nontrivial local power when the nonlinear component $s(x_t)$ vanishes to zero at the rate $O\left(n^{-1/2}\right)$. These results are formally stated as follows.

Theorem 2. *Given Assumptions 1 and 2,*

(i) if $\mathbb{E}[y_t|x_t, d_t] = \ x_t(m)'\alpha_* + d_t'\eta_* + s(x_t)$ with $\mathbb{E}[s(x_t)^2] < \infty$ and $\mathbb{E}[\log^{4j_*}(x_t)] < \infty$, for some

$\widetilde{\gamma} \in \Gamma$, $h(\widetilde{\gamma}) \in (0, h_0)$ and

$$\frac{1}{n} QLR_n = \left(1 - \frac{h(\widetilde{\gamma})}{h_0}\right) + o_{\mathbb{P}}(1),$$

where $j_* := \min\{j \in \mathbb{N} : \mathbb{E}[v_t \log^j(x_t)] \neq 0\}$, and v_t is the linear projection error obtained by projecting y_t into the space of $(x_t(m)', d_t')'$;

(ii) if $\mathbb{E}[y_t|x_t, d_t] = x_t(m)'\alpha_* + d_t'\eta_* + n^{-1/2}s(x_t)$ with $|s(x_t)| \leq m_t$,

$$QLR_n \Rightarrow \sup_{\gamma \in \Gamma} \left(\mathcal{Z}(\gamma) + \frac{\zeta(\gamma)}{\sigma(\gamma)}\right)^2,$$

where $\zeta(\gamma) := \mathbb{E}[s(x_t)x_t^{\gamma}] - \mathbb{E}[s(x_t)z_t']E[z_t z_t']^{-1}\mathbb{E}[z_t x_t^{\gamma}]$. □

Since $s(\cdot)$ is an arbitrarily selected nonlinear function, Theorem 2(*i*) implies that the QLR test statistic has omnibus power. Therefore, the QLR test statistic has power even when $\mathbb{E}[y_t|x_t, d_t]$ is a polynomial function with respect to x_t with degree that exceeds m. The existence of j_* in Theorem 2(*i*) follows from theorem 2 of Bierens (1982), and Theorem 2 follows as a corollary of theorem 5 of BCP. The proof is therefore omitted.

The intuition underlying the existence of omnibus power in Theorem 2 is straightforward and can be exposited in terms of the Stichcombe and White (1998) approach to testing. First note that the testing factor can be written as $x_t^{\gamma} = \exp(\gamma \log(x_t))$. Here, $\log(\cdot)$ is a one-to-one monotonic and measure preserving mapping, so that the consistent power property of the QLR test is unaffected by the log transformation. Second, $\exp(\cdot)$ is an analytic function, so that it is generically comprehensively revealing (in the terminology of Stichcombe and White, 1998), thereby producing the omnibus power property.

2.6 Sequentially Testing the Polynomial Model

We next examine a sequential testing procedure in which we allow the polynomial degree m to increase and apply a sequence of tests until the null hypothesis is no longer rejected. This procedure provides a natural mechanism for estimating the degree of a polynomial model at some given level of significance $\boldsymbol{\alpha}$. Modifying earlier notation to accommodate sequential testing, we signal polynomial model degree in the QLR test statistic by indexing the degree, so that $QLR_n^{(m)}$ denotes the QLR test statistic computed using a polynomial null model of m-th degree. This modification avoids confusion when computing multiple QLR test statistics.

The testing procedure requires that a maximum degree polynomial model be specified in advance. Accordingly, we define $P_d(\bar{m}) := \{1, 2, \ldots, \bar{m}\}$ to be a subset of Γ such that each element of $P_d(\bar{m})$ is an interior element of Γ and $\bar{m}$ is the upper limit polynomial degree envisaged for implementation. Sequential testing then proceeds as follows:

- **Step 1:** Compute $QLR_n^{(1)}$ using $\mathcal{M}_1'$ such that Γ contains $P_d(\bar{m})$ as its subset. If $QLR_n^{(1)}$ is less than the critical value given by Theorem 1, let $\widehat{m}_n = 1$; otherwise, move to the next step, where $\widehat{m}_n$ denotes the estimate of the unknown polynomial degree.

- **Step 2:** Iterate the above steps for $j = 2, 3, \ldots, \bar{m}$ using $\mathcal{M}_j$ with the same Γ until $QLR_n^{(j)}$ is greater than the asymptotic critical value in Theorem 1. We let $\widehat{m}_n$ be the smallest polynomial degree such that the QLR test statistic does not reject the null hypothesis.

- **Step 3**: If for $j = 1, 2, \ldots, \bar{m}$, $QLR_n^{(j)}$ exceeds the asymptotic critical values in Theorem 1, we conclude that an $\bar{m}$-th degree polynomial model is unable to capture the nonlinearity of $\mathbb{E}[y_t|x_t, d_t]$ with respect to x_t.

Several remarks are in order concerning this sequential procedure. First, as shown earlier, the QLR test statistic is not distribution free, so that the asymptotic critical values in Theorem 1 need to be obtained case-dependently. Thus, for different $j = 1, 2, \ldots, \bar{m}$, different asymptotic critical values need to be applied to $QLR_n^{(j)}$. Use of Hansen's (1996) weighted bootstrap can yield consistent asymptotic critical values in this case. In Section 5, we apply the weighted bootstrap in an empirical illustration of the QLR test statistic to demonstrate this implementation. Second, to elaborate on the procedure, we can let Γ contain $P_d(\bar{m})$, but choose another parameter space Γ for each j: that is, model $\mathcal{M}_j$ can be specified using different Γ_j such that $P_d(j)$ is a subset of Γ_j and each element of $P_d(j)$ is an interior element of Γ_j. For each Γ_j, different asymptotic critical values have again to be used. Finally, using Theorems 1 and 2(*i*), we are able to obtain the following result which ensures size control in the sequential testing procedure.

Corollary 1. *If Assumptions 1 and 2 hold for each* $m \in P_d(\bar{m})$, *if* $m_* \in P_d(\bar{m})$, *where*

$$m_* := \inf\{m \in \mathbb{N} : \exists\ (\alpha, \eta), \mathbb{E}[y_t|x_t, d_t] = x_t(m)'\alpha + d_t'\eta\},$$

then for any $\epsilon > 0$ *and significance level* $\boldsymbol{\alpha}$*,* $\lim_{n\to\infty} \mathbb{P}\left(|\widehat{m}_n - m_*| > \epsilon\right) = \boldsymbol{\alpha}$*.* □

Thus, when the significance level $\boldsymbol{\alpha}$ is given, the estimated polynomial degree is equal to the unknown polynomial degree with probability $(1 - \boldsymbol{\alpha})\%$ at the limit. Here, the unknown polynomial degree m_* is defined as the minimum degree polynomial model out of the correctly specified polynomial models. Note that if m_* exists for $P_d(\bar{m})$, every polynomial model with a degree higher than m_* is correctly specified.

Therefore, m_* signifies the most parsimonious polynomial model that is correctly specified. Corollary 1 implies that we can avoid the data snooping problem despite the application of a number of test statistics to a single data set. But there is a type I error: the estimated $\widehat{m}_n$ has the limiting (size controlled) probability $\boldsymbol{\alpha}$ that $\widehat{m}_n$ differs from m_*. Third, there is the opportunity for consistent estimation by $\widehat{m}_n$ if we control size to depend on n so that $\boldsymbol{\alpha} = \boldsymbol{\alpha}_n \to 0$ slowly as $n \to \infty$. The following theorem provides conditions for such consistent estimation of m_*.

Theorem 3. *Under the same conditions as Corollary 1, if (i) there is a Gaussian process $\mathcal{B}^S(\cdot)$ such that for all γ, $\gamma' \in \Gamma$, for some δ, $\mathrm{cov}(\mathcal{B}^S(\gamma), \mathcal{B}^S(\gamma')) = 1 - |\gamma - \gamma'|^\delta (1 + o(1))$ and $\mathrm{cov}(\mathcal{B}^S(\gamma), \mathcal{B}^S(\gamma')) \leq \mathrm{cov}(\mathcal{Z}^0(\gamma), \mathcal{Z}^0(\gamma'))$, where for each γ, $\mathcal{Z}^0(\gamma) := \mathcal{Z}(\gamma)/\sigma^0(\gamma)$ and $\sigma^0(\gamma) := \mathrm{var}[\mathcal{Z}(\gamma)]^{1/2}$, (ii) $\lim_{n\to\infty} \boldsymbol{\alpha}_n = 0$, and (iii) $\lim_{n\to\infty} \log(\boldsymbol{\alpha}_n)/n = 0$, then for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\left(|\widehat{m}_n - m_*| > \epsilon\right) = 0$.* □

By Theorem 3, $\widehat{m}_n$ consistently estimates m_*. Theorem 3 extends the sequential testing result in Hosoya (1989) in which likelihood ratio test statistics are sequentially applied that marginally follow chi-squared distributions under the null. Although the null limit distribution here is not chi-squared but depends on a stochastic process, we can still obtain the same result as Hosoya (1989) under the conditions given in Theorem 3. These conditions are used to apply a suitable approximation of the distribution of the Gaussian extremum (*c.f.*, Piterbarg, 1996). Details are provided in the proof. In brief, by comparing the covariance kernel of $\mathcal{Z}^0(\cdot)$ in Theorem 1 with that of a certain stationary Gaussian process, $\mathcal{B}^s(\cdot)$, we obtain that a critical value c_n' for which $\mathbb{P}(\sup_{\gamma\in\Gamma} \mathcal{Z}^0(\gamma)^2 \geq c_n') = \boldsymbol{\alpha}_n$ is bounded from above by the Slepian inequality. This critical value can be compared with another critical value c_n such that $\mathbb{P}(\sup_{\gamma\in\Gamma} \mathcal{Z}(\gamma)^2 \geq c_n) = \boldsymbol{\alpha}_n$ and we show that the upper bound for c_n' is also a upper bound for c_n. Theorem 3 is proved by associating the upper bound of c_n with the conditions for $\boldsymbol{\alpha}_n$ in Theorem 3 in a manner that if $-\log(\boldsymbol{\alpha}_n)/n \to 0$ and $\boldsymbol{\alpha}_n \to 0$, then $c_n/n \to 0$ and $c_n \to \infty$. These results are sufficient for $\lim_{n\to\infty} \mathbb{P}(\widehat{m}_n > m_*) = 0$ and $\lim_{n\to\infty} \mathbb{P}(\widehat{m}_n < m_*) = 0$, respectively, given Theorem 2(*i*). This implies that $\lim_{n\to\infty} \mathbb{P}(\widehat{m}_n = m_*) = 1$.

3 Sequential QLR Testing for Time-Trend Stationary Data

3.1 DGP and the m-th Degree Polynomial Time-Trend Model

We now extend the analysis to include a polynomial time-trend stationary process. The focus is on testing for further neglected nonlinearity in trend when an m-th degree polynomial time-trend model is specified. We suppose that our alternative model for $\mathbb{E}[y_t|d_t]$ is specified as

$$\mathcal{M}_m' := \{\mu_t(\cdot) : \Omega_n \mapsto \mathbb{R} : \mu_t(\alpha_n, \eta, \beta_n, \gamma) := \ s_t(m)'\alpha_n + d_t'\eta + \beta_n s_{n,t}^\gamma\},$$

where $d_t \in \mathbb{R}^k$ ($k \in \mathbb{N}$) is a strictly stationary and ergodic process, y_t is a polynomial time-trend stationary process, and $s_t(m) := [1, s_{n,t}, s_{n,t}^2, \ldots, s_{n,t}^m]'$, where for $t = 1, 2, \ldots, n$, $s_{n,t} := t/n$ is a (normalized) linear time trend. As before, the hypothesis of interest is

$$\widetilde{\mathcal{H}}_0 : \exists(\alpha_*', \eta_*')', \ \ \mathbb{E}[y_t|d_t] = \ s_t(m)'\alpha_{n,*} + d_t'\eta_* \text{ with probability } 1.$$

The model $\mathcal{M}_m'$ is a reparameterized version of the following polynomial time-trend stationary model:

$$\mathcal{M}_m'' := \{\mu_t(\cdot) : \Omega \mapsto \mathbb{R} : \mu_t(\alpha, \eta, \beta, \gamma) := \ t(m)'\alpha + d_t'\eta + \beta t^\gamma\},$$

where $t(m) := [1, t, t^2, \ldots, t^m]'$. The parameters in $\mathcal{M}_m'$ are related to those in $\mathcal{M}_m''$ through the identities $\alpha_n \equiv \mathrm{diag}[1, n, n^2, \ldots, n^m]\alpha$ and $\beta_n \equiv \beta n^\gamma$. Thus, estimating the parameters in $\mathcal{M}_m''$ by least squares is easily converted to least squares using $\mathcal{M}_m'$, and vice versa. This equivalence implies that the QLR test statistic value obtained from $\mathcal{M}_m'$ is identical to that obtained from $\mathcal{M}_m''$.

The null limit distribution has to be deduced from $\mathcal{M}_m'$, although the two models yield the same level of the QLR test statistic. The null limit distribution cannot be easily obtained from $\mathcal{M}_m''$ due to the singularity problem involved in the limit theory (see Phillips, 2007). Specifically, upon normalization, the associated signal matrix of $\mathcal{M}_m''$, viz., $\sum \bar{G}_t(\gamma)\bar{G}_t(\gamma)'$, involves a singular almost sure limit, where $\bar{G}_t(\gamma) := [t(m)', d_t', t(m)'\log(t), t^\gamma]'$ corresponds to $G_t(\gamma)$ in Section 2. For instance, taking moment matrix formed from the outer product of $[t(m)', t(m)'\log(t)]'$ and normalizing appropriately, we obtain

$$F_n^{-1} \sum_{t=1}^{n} \begin{bmatrix} t(m)t(m)' & t(m)t(m)'\log(t) \\ t(m)t(m)'\log(t) & t(m)t(m)'\log^2(t) \end{bmatrix} F_n^{-1} \overset{\text{a.s.}}{\to} \begin{bmatrix} N & N \\ N & N \end{bmatrix},$$

where $F_n := \mathrm{diag}[n^{1/2}, n^{3/2}, \ldots, n^{m+1/2}, n^{1/2}\log(n), n^{3/2}\log(n), \ldots, n^{m+1/2}\log(n)]$, and N is an $(m+1) \times (m+1)$ matrix whose j-th row and i-column element equals $1/(j+i+1)$. The normalizing matrix F_n is selected to ensure a bounded almost sure nontrivial limit for $\sum \bar{G}_t(\gamma)\bar{G}_t(\gamma)'$. This limit matrix is the analogue of $A(\gamma)$ of Section 2 (defined in Assumption 2) in the polynomial time trend context of $\mathcal{M}_m''$. The noninvertibility of this limit matrix makes it awkward to obtain the null limit distribution using $\mathcal{M}_m''$. Phillips (2007) dealt with singularities of this type involving general slowly varying functions such as $\log(t)$ in regression and nonlinear regression models and showed how the use of alternative weak trend formulations such as $\mathcal{M}_m'$ provides a convenient approach to the limit theory. In particular, in the present case the use of the formulation $\mathcal{M}_m'$ removes the limiting singularity and the null limit distribution of the QLR test statistic can be readily analyzed, as is now discussed.

3.2 Asymptotic Null Distribution of the QLR Test Statistic

We assume the following conditions.

Assumption 3. *(i) The time series $\{d_t\}$ is stationary ϕ-mixing with mixing decay rate $-\ell/2(\ell-1)$ with $\ell \geq 2$ or α-mixing with mixing decay rate $-\ell/(\ell-2)$ with $\ell > 2$, and y_t is a time-trend stationary process;*

(ii) The model for $\mathbb{E}[y_t|d_t]$ is specified as $\mathcal{M}'_m := \{\mu_t(\cdot) : \Omega_n \mapsto \mathbb{R} : \mu_t(\alpha_n, \eta, \beta_n, \gamma) := s_t(m)'\alpha_n + d_t'\eta + \beta_n s_{n,t}^{\gamma}\}$, where Ω_n is the parameter space of $\omega_n := (\alpha_n', \eta', \beta_n, \gamma)'$, and n is the sample size;

(iii) $\Omega_n = (\prod_{i=0}^{m} \bar{A}_{i,n}) \times H \times \bar{B}_n \times \Gamma$ such that H and Γ are convex and compact parameter spaces in $\mathbb{R}^k$ and $\mathbb{R}$, respectively, with $0, 1, \cdots, m$ being interior elements of Γ with $\inf\, \Gamma > -1/2$; for $i = 0, 1, \cdots, m$ and for each n, $\bar{A}_{i,n}$ and $\bar{B}$ are also convex and compact spaces in $\mathbb{R}$; and

(iv) $Z'Z = \sum_{t=1}^{n} z_{n,t} z_{n,t}'$ is nonsingular with probability 1, where $z_{n,t} := (s_t(m)', d_t')'$. $\square$

Further conditions are needed to obtain regular null limit behavior of the QLR test statistic. Before imposing them, we introduce the following symmetric matrices to aid notation. For each γ, let

$$\widetilde{A}(\gamma) := \left[\begin{array}{c:cc} \widetilde{A}^{1,1} & \widetilde{A}^{1,2} & \widetilde{A}^{1,3}(\gamma) \\ \hdashline \widetilde{A}^{2,1} & \widetilde{A}_{3,3} & \widetilde{A}_{3,4}(\gamma) \\ \widetilde{A}^{3,1}(\gamma) & \widetilde{A}_{4,3}(\gamma) & \widetilde{A}_{4,4}(\gamma) \end{array}\right] := \left[\begin{array}{cc:cc} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} & \widetilde{A}_{1,3} & \widetilde{A}_{1,4}(\gamma) \\ \widetilde{A}_{2,1} & \widetilde{A}_{2,2} & \widetilde{A}_{2,3} & \widetilde{A}_{2,4}(\gamma) \\ \hdashline \widetilde{A}_{3,1} & \widetilde{A}_{3,2} & \widetilde{A}_{3,3} & \widetilde{A}_{3,4}(\gamma) \\ \widetilde{A}_{4,1}(\gamma) & \widetilde{A}_{4,2}(\gamma) & \widetilde{A}_{4,3}(\gamma) & \widetilde{A}_{4,4}(\gamma) \end{array}\right],$$

where the submatrices are defined as follows: for $i, j = 1, 2, \ldots, m+1$,

$$\underset{(m+1)\times(m+1)}{\widetilde{A}_{1,1}} := \left[\frac{1}{i+j-1}\right], \quad \underset{(m+1)\times k}{\widetilde{A}_{1,2}} := \left[\frac{\mathbb{E}[d_t']}{j}\right], \quad \underset{(m+1)\times(m+1)}{\widetilde{A}_{1,3}} := \left[\frac{-1}{(i+j-1)^2}\right],$$

$$\underset{(m+1)\times 1}{\widetilde{A}_{1,4}(\gamma)} := \left[\frac{1}{\gamma+j}\right], \quad \underset{k\times k}{\widetilde{A}_{2,2}} := \mathbb{E}[d_t d_t'], \quad \underset{k\times(m+1)}{\widetilde{A}_{2,3}} := \left[-\frac{\mathbb{E}[d_t]}{j^2}\right], \quad \underset{k\times 1}{\widetilde{A}_{2,4}(\gamma)} := \left[\frac{\mathbb{E}[d_t]}{\gamma+1}\right],$$

$$\underset{(m+1)\times(m+1)}{\widetilde{A}_{3,3}} := \left[\frac{2}{(i+j-1)^3}\right], \quad \underset{(m+1)\times 1}{\widetilde{A}_{3,4}(\gamma)} := \left[\frac{-1}{(\gamma+j)^2}\right], \quad \text{and} \quad \underset{1\times 1}{\widetilde{A}_{4,4}(\gamma)} := \frac{1}{2\gamma+1}.$$

Since $\widetilde{A}(\gamma)$ is supposed to be symmetric, we let $\widetilde{A}_{2,1} := \widetilde{A}_{1,2}'$, $\widetilde{A}_{3,1} := \widetilde{A}_{1,3}'$, $\widetilde{A}_{4,1}(\gamma) := \widetilde{A}_{1,4}(\gamma)'$, $\widetilde{A}_{2,3} := \widetilde{A}_{3,2}'$, $\widetilde{A}_{2,4}(\gamma) := \widetilde{A}_{4,2}(\gamma)'$, and $\widetilde{A}_{4,3} := \widetilde{A}_{3,4}'$. Observe that $\widetilde{A}(\gamma)$ corresponds to $A(\gamma)$ in Section 2 and is identical to the almost sure limit of $n^{-1}\sum \widetilde{G}_t(\gamma)\widetilde{G}_t(\gamma)'$, where $\widetilde{G}_t(\gamma) := [\, s_t(m)', d_t',\ s_t(m)'\log(s_{n,t}), s_{n,t}^{\gamma}]'$, which exists under mild moment conditions that are assured by Assumption 4 below. We next define

$\widetilde{B}(\gamma, \gamma')$ as follows:

$$\widetilde{B}(\gamma,\gamma') := \left[\begin{array}{c:cc} \widetilde{B}^{1,1} & \widetilde{B}^{1,2} & \widetilde{B}^{1,3}(\gamma') \\ \hdashline \widetilde{B}^{2,1} & \widetilde{B}_{3,3} & \widetilde{B}_{3,4}(\gamma') \\ \widetilde{B}^{3,1}(\gamma) & \widetilde{B}_{4,3}(\gamma) & \widetilde{B}_{4,4}(\gamma,\gamma') \end{array}\right] := \left[\begin{array}{cc:cc} \widetilde{B}_{1,1} & \widetilde{B}_{1,2} & \widetilde{B}_{1,3} & \widetilde{B}_{1,4}(\gamma') \\ \widetilde{B}_{2,1} & \widetilde{B}_{2,2} & \widetilde{B}_{2,3} & \widetilde{B}_{2,4}(\gamma') \\ \hdashline \widetilde{B}_{3,1} & \widetilde{B}_{3,2} & \widetilde{B}_{3,3} & \widetilde{B}_{3,4}(\gamma') \\ \widetilde{B}_{4,1}(\gamma) & \widetilde{B}_{4,2}(\gamma) & \widetilde{B}_{4,3}(\gamma) & \widetilde{B}_{4,4}(\gamma,\gamma') \end{array}\right],$$

where the submatrices are defined below, for $i, j = 1, 2, \ldots, m+1$,

$$\underset{(m+1)\times(m+1)}{\widetilde{B}_{1,1}} := \left[\frac{\mathbb{E}[u_t^2]}{i+j-1}\right], \quad \underset{(m+1)\times k}{\widetilde{B}_{1,2}} := \left[\frac{\mathbb{E}[u_t^2 d_t']}{j}\right], \quad \underset{(m+1)\times(m+1)}{\widetilde{B}_{1,3}} := \left[\frac{-\mathbb{E}[u_t^2]}{(i+j-1)^2}\right],$$

$$\underset{(m+1)\times 1}{\widetilde{B}_{1,4}(\gamma')} := \left[\frac{\mathbb{E}[u_t^2]}{\gamma'+j}\right], \quad \underset{k\times k}{\widetilde{B}_{2,2}} := \mathbb{E}[u_t^2 d_t d_t'], \quad \underset{k\times(m+1)}{\widetilde{B}_{2,3}} := \left[-\frac{\mathbb{E}[u_t^2 d_t]}{j^2}\right], \quad \underset{k\times 1}{\widetilde{B}_{2,4}(\gamma')} := \left[\frac{\mathbb{E}[u_t^2 d_t]}{\gamma'+1}\right],$$

$$\underset{(m+1)\times(m+1)}{\widetilde{B}_{3,3}} := \left[\frac{2\mathbb{E}[u_t^2]}{(i+j-1)^3}\right], \quad \underset{(m+1)\times 1}{\widetilde{B}_{3,4}(\gamma')} := \left[\frac{-\mathbb{E}[u_t^2]}{(\gamma'+j)^2}\right], \quad \underset{1\times 1}{\widetilde{B}_{4,4}(\gamma,\gamma')} := \frac{\mathbb{E}[u_t^2]}{\gamma+\gamma'+1},$$

where $u_t := y_t - \mathbb{E}[y_t|d_t]$. Let $\widetilde{B}(\gamma,\gamma)$ be symmetric, so that $\widetilde{B}_{2,1} := \widetilde{B}_{1,2}'$, $\widetilde{B}_{3,1} := \widetilde{B}_{1,3}'$, $\widetilde{B}_{4,1}(\gamma) := \widetilde{B}_{1,4}(\gamma)'$, $\widetilde{B}_{2,3} := \widetilde{B}_{3,2}'$, $\widetilde{B}_{2,4}(\gamma) := \widetilde{B}_{4,2}(\gamma)'$, and $\widetilde{B}_{4,3} := \widetilde{B}_{3,4}'$. The matrix $\widetilde{B}(\gamma,\gamma)$ corresponds to $B(\gamma)$ in Section 2, and $\widetilde{B}(\gamma,\gamma')$ is the almost sure limit of $n^{-1}\sum u_t^2 \widetilde{G}_t(\gamma)\widetilde{G}_t(\gamma')'$, which again exists under mild moment and other regularity conditions that are assured by the following assumption.

Assumption 4. *(i) For each $\epsilon > 0$, $\widetilde{A}(\cdot)$ and $\widetilde{B}(\cdot,\cdot)$ are positive definite uniformly on $\Gamma(\epsilon)$;*

(ii) $\{u_t, \mathcal{F}_t\}$ is an MDS, where $\mathcal{F}_t$ is the adapted smallest σ-field generated by $\{d_{t+1}, u_t, d_t, u_{t-1}, \cdots\}$;

(iii) There is a strictly stationary and ergodic sequence $\{m_t\}$ such that for $j = 1, 2, \cdots, k$, $|d_{t,i}| \leq m_t$, $|u_t| \leq m_t$, and for some $r > 1$, $\mathbb{E}[m_t^{4r}] < \infty$, where $d_{t,i}$ is the i-th row element of d_t. $\square$

Several remarks are warranted on these conditions. First, Assumption 4 matches assumption 7 of BCP except that $\widetilde{A}(\cdot)$ and $\widetilde{B}(\cdot,\cdot)$ in Assumption 4(*i*) are constructed for an arbitrary polynomial degree m rather than $m = 1$. Second, although m is unspecified, it is not hard to verify that $\widetilde{A}(\cdot)$ is positive definite uniformly on $\Gamma(\epsilon)$ if and only if the covariance matrix of d_t is positive definite. If $\widetilde{A}(\cdot)$ is reorganized into

$$\left[\begin{array}{c:c} \mathbb{A}^{1,1}(\gamma) & \mathbb{A}^{1,2}(\gamma) \\ \hdashline \mathbb{A}^{2,1}(\gamma) & \widetilde{A}_{2,2} \end{array}\right] := \left[\begin{array}{ccc:c} & & & \widetilde{A}_{1,2} \\ & \mathbb{A}^{1,1}(\gamma) & & \widetilde{A}_{3,2} \\ & & & \widetilde{A}_{4,2}(\gamma) \\ \hdashline \widetilde{A}_{2,1} & \widetilde{A}_{2,3} & \widetilde{A}_{2,4}(\gamma) & \widetilde{A}_{2,2} \end{array}\right],$$

where

$$\mathbb{A}^{1,1}(\gamma) := \left[\begin{array}{c:c} \mathbb{A}_{1,1} & \mathbb{A}_{1,2}(\gamma) \\ \hdashline \mathbb{A}_{2,1}(\gamma) & \widetilde{A}_{4,4}(\gamma) \end{array}\right] := \left[\begin{array}{cc:c} \widetilde{A}_{1,1} & \widetilde{A}_{1,3} & \widetilde{A}_{1,4}(\gamma) \\ \widetilde{A}_{3,1} & \widetilde{A}_{3,3} & \widetilde{A}_{3,4}(\gamma) \\ \hdashline \widetilde{A}_{4,1}(\gamma) & \widetilde{A}_{4,3}(\gamma) & \widetilde{A}_{4,4}(\gamma) \end{array}\right],$$

$\mathbb{A}_{1,1}$ is positive definite by the definitions of $\widetilde{A}_{1,1}$, $\widetilde{A}_{1,3}$, and $\widetilde{A}_{3,3}$. This in turn implies that $\mathbb{A}^{1,1}(\gamma)$ is positive definite uniformly for each $\gamma \in \Gamma(\epsilon)$ if and only if $\widetilde{A}_{4,4}(\gamma) - \mathbb{A}_{2,1}(\gamma)\mathbb{A}_{1,1}^{-1}\mathbb{A}_{1,2}(\gamma)$ is positive definite. Some algebra shows that

$$\widetilde{A}_{4,4}(\gamma) - \mathbb{A}_{2,1}(\gamma)\mathbb{A}_{1,1}^{-1}\mathbb{A}_{1,2}(\gamma) = \frac{\prod_{i=0}^{m}(\gamma - i)^4}{(1+2\gamma)\prod_{i=0}^{m}(\gamma+i+1)^4},$$

which is strictly greater than zero for each $\gamma \in \Gamma(\epsilon)$, implying that $\mathbb{A}^{1,1}(\cdot)$ is positive definite uniformly on $\Gamma(\epsilon)$. This further implies that $\widetilde{A}(\cdot)$ is positive definite uniformly on $\Gamma(\epsilon)$ if and only if for each $\gamma \in \Gamma(\epsilon)$, $\widetilde{A}_{2,2} - \mathbb{A}^{2,1}(\gamma)\mathbb{A}^{1,1}(\gamma)^{-1}\mathbb{A}^{1,2}(\gamma)$ is positive definite. Here, every column of $\mathbb{A}^{1,2}(\gamma)$ is a linear transformation of the first column of $\mathbb{A}^{1,1}(\gamma)$, so that $\widetilde{A}_{2,2} - \mathbb{A}^{2,1}(\gamma)\mathbb{A}^{1,1}(\gamma)^{-1}\mathbb{A}^{1,2}(\gamma) = \mathbb{E}[d_t d_t'] - \mathbb{E}[d_t]\mathbb{E}[d_t]'$ that is the covariance matrix of d_t. Therefore, $\widetilde{A}(\cdot)$ is positive definite uniformly on $\Gamma(\epsilon)$ if and only if the covariance matrix of d_t is positive definite, that is provided the elements of d_t are not linearly dependent almost surely. Third, Assumptions 3 and 4 imply that the regressors are now bounded processes in probability, so that the null limit distribution of the QLR test statistic can be analyzed similarly to that of Section 2. In particular, the singular matrix problem no longer arises.

The main result of this section is contained in the following theorem.

Theorem 4. *Given Assumptions 3, 4, and $\widetilde{\mathcal{H}}_0$, $QLR_n \Rightarrow \sup_{\gamma\in\Gamma} \widetilde{\mathcal{Z}}(\gamma)^2$, where $\widetilde{\mathcal{Z}}(\cdot)$ is a Gaussian process with covariance kernel for each γ and $\gamma' \in \Gamma$ given by*

$$\mathbb{E}[\widetilde{\mathcal{Z}}(\gamma)\widetilde{\mathcal{Z}}(\gamma')] = c_m(\gamma, \gamma')\frac{(1+2\gamma)^{1/2}(1+2\gamma')^{1/2}}{(1+\gamma+\gamma')} \tag{11}$$

with $c_m(\gamma,\gamma') := \prod_{i=0}^{m}(\gamma-i)(\gamma'-i)/|\prod_{i=0}^{m}(\gamma-i)(\gamma'-i)|$. □

The proof of Theorem 4, which is given in the Appendix, proceeds along the following lines. We first show that the QLR test statistic under $\widetilde{\mathcal{H}}_0$ is identical to that obtained under the hypothesis that $\beta_* = 0$. Next, the null limit distribution under the hypothesis that $\beta_* = 0$ is obtained as $\sup_{\gamma\in\Gamma} \widetilde{\mathcal{Z}}(\gamma)^2$. Finally, the covariance kernel in (11) is derived from the sample analog of $\widetilde{\mathcal{Z}}(\cdot)$ denoted as $\widetilde{z}_n(\cdot) := \{\widehat{\sigma}_{n,0}^2 S(\cdot)'MS(\cdot)\}^{-1/2}\{S(\cdot)'M U\}$, where $S(\gamma) := [s_{n,1}^{\gamma}, s_{n,2}^{\gamma}, \ldots, s_{n,n}^{\gamma}]'$, $M := I_n - Z(Z'Z)^{-1}Z'$, and $U := [u_1, u_2, \ldots, u_n]'$. Derivation of the weak limit process proceeds in the same way as Theorem 1. We therefore focus on deriving the covariance kernel of $\widetilde{\mathcal{Z}}(\cdot)$ in Theorem 4. Let $\widetilde{\mathcal{G}}(\cdot)$ and $\widetilde{\sigma}^2(\cdot,\cdot)$ be the weak limit of $n^{-1/2}S(\cdot)'MU$ and the almost sure limit of $n^{-1}\widehat{\sigma}^2_{n,0}S(\cdot)'MS(\cdot)$, respectively. We show that for each γ and γ', we have

$$\widetilde{\sigma}^2(\gamma,\gamma) = \frac{\sigma_*^2 \prod_{i=0}^m (\gamma - i)^2}{(2\gamma+1)\prod_{i=0}^m (\gamma+i+1)^2}, \quad \text{and}$$

$$\mathbb{E}[\widetilde{\mathcal{G}}(\gamma)\widetilde{\mathcal{G}}(\gamma')] = \frac{\sigma_*^2 \prod_{i=0}^m (\gamma - i)(\gamma' - i)}{(\gamma+\gamma'+1)\prod_{i=0}^m (\gamma+i+1)(\gamma'+i+1)},$$

where $\sigma_*^2 := \mathbb{E}[u_t^2]$ as before. Thus, the covariance kernel in (11) is obtained as

$$\frac{\mathbb{E}[\widetilde{\mathcal{G}}(\gamma)\widetilde{\mathcal{G}}(\gamma')]}{\{\widetilde{\sigma}^2(\gamma,\gamma)\}^{1/2}\{\widetilde{\sigma}^2(\gamma',\gamma')\}^{1/2}}.$$

The Gaussian process $\widetilde{\mathcal{Z}}(\cdot)$ is independent of the joint distribution of $\{d_t, u_t\}$, just as in BCP. In particular, Theorem 4 holds irrespective of whether the error is conditionally hetroskedasticity or homoskedastic, viz., the QLR test is a distribution free test. Its applicability is therefore relatively wide. We call the Gaussian process $\widetilde{\mathcal{Z}}(\cdot)$ the *polynomial power Gaussian process*.

The polynomial power Gaussian process is associated with some other useful Gaussian processes. First, the polynomial power Gaussian process generalizes the power Gaussian process in BCP, which is obtained by simply setting $m = 1$. Second, the distribution of the polynomial power Gaussian process differs according to the value of m. Nonetheless, the squared polynomial power Gaussian process has an identical distribution irrespective of m because for any m, $c_m^2(\cdot,\cdot) \equiv 1$. Therefore, the critical values of the QLR test statistic can also be obtained, just as in BCP, by simulating the truncated exponential Gaussian processes in Cho and White (2010) and Cho, Cheong, and White (2011). Specifically, let the truncated exponential Gaussian process be defined as

$$\bar{\mathcal{Z}}_\ell(\gamma) := \sum_{i=2}^{\ell} \left[\frac{\gamma^4}{(\gamma+1)^2(2\gamma+1)}\right]^{-1/2} \left(\frac{\gamma}{\gamma+1}\right)^i G_i,$$

where $G_i \sim_{iid} N(0,1)$ and ℓ is some given large integer. Then, the functional $\sup_{\gamma\in\Gamma} \bar{\mathcal{Z}}_\ell(\gamma)^2$ can be simulated in order to obtain the asymptotic critical values. When ℓ is sufficiently large, the true asymptotic critical values are close to the critical values obtained by simulating $\sup_{\gamma\in\Gamma} \bar{\mathcal{Z}}_\ell(\gamma)^2$.

We tabulate asymptotic critical values obtained in this way for large ℓ. The critical values of BCP should be used only when $m = 1$. Table 1 reports critical values for the QLR test for models with polynomials of degree $m = 2, 3, 4, 5, 6, 7, 8, 9, 10$. With these tabulated results, users can test for neglected nonlinearity up to a 10th degree polynomial null model. The values reported are obtained with $\ell = 1000$ and one million replications. Since this methodology provides more precise critical values than those in BCP, we include the $m = 1$ case in Table 1.[1]

3.3 Asymptotic Power of the QLR Test Statistic

As in the stationary case, the QLR test statistic has power for detecting misspecified time-trend polynomial models. Model misspecification can arise in many ways for time-trend stationary data due to the vast extent of possible nonstationary time trends, and the QLR test statistic does not have omnibus power against all forms of misspecification, although it would have non-trivial power against an analytic transformation of time trend by an analysis analogous to that in Section 2.5. We therefore restrict attention to the power of the QLR test under a set of time-trend alternatives involving misspecified polynomial degree and omitted smoothly slowly varying (SSV) functions. The former alternative is particularly important in constructing a consistent time-trend degree selection algorithm. The latter are important in case of logarithmic and more general power function alternatives. Suppose, for example, that $\mathbb{E}[y_t|d_t] = t(m)'\alpha_* + d_t'\eta_* + s(t)$ under the alternative, where $s(\cdot)$ is an SSV function. Phillips (2007) provides many SSV functions that include powered logarithm functions and iterated logarithmic function that occur in empirical applications and nonlinear regression problems. Since the set of SSV functions is relatively large and typically involves only minor departures from polynomical time trends, the results given below indicate that the QLR test statistic will have power and non-trivial local power against such alternatives, as well as a large number of other time-trend alternatives.

Theorem 5. *Given Assumptions 3 and 4,*

(i) if for some $m_0 > m$, $\mathbb{E}[y_t|d_t] = t(m_0)'\alpha_* + d_t'\eta_*$,

$$\frac{1}{n}QLR_n = \sup_{\gamma\in\Gamma} \frac{\widetilde{\sigma}^2(\gamma, m_0)}{\{\widetilde{\sigma}^2(\gamma,\gamma)\}^{1/2}\{\widetilde{\sigma}^2(m_0, m_0)\}^{1/2}} + o_{\mathbb{P}}(1);$$

(ii) if $\mathbb{E}[y_t|d_t] = t(m)'\alpha_* + d_t'\eta_* + s(t)$ *with* $s(\cdot)$ *being a SSV function, and* $ns'(n) \to c\ (\neq 0)$,

$$\frac{1}{n}QLR_n = \sup_{\gamma\in\Gamma} \left(\frac{c^2\sigma_*^2}{\sigma_*^2 + c^2 q}\right)\left(\frac{p(\gamma)}{\widetilde{\sigma}(\gamma,\gamma)}\right)^2 + o_{\mathbb{P}}(1),$$

where $p(\gamma) := (\gamma-1)(7\gamma+15)/\{4(\gamma+1)^2(\gamma+2)\}$ *and* $q := 91/64$*;*

[1] Interested readers can download the GAUSS program code that generates the null limit distribution. The URL is http://web.yonsei.ac.kr/jinseocho/research.htm. Users can select different values of the lower and upper bounds of Γ, ℓ, and the number of replications in running the code.

(iii) if $\mathbb{E}[y_t|d_t] = t(m)'\alpha_* + d_t'\eta_* + s(t)$ *with* $s(\cdot)$ *being a SSV function, and* $ns'(n) \to \infty$*,*

$$\frac{1}{n}QLR_n = \sup_{\gamma\in\Gamma}\left(\frac{\sigma_*^2}{q}\right)\left(\frac{p(\gamma)}{\widetilde{\sigma}(\gamma,\gamma)}\right)^2 + o_{\mathbb{P}}(1),$$

(iv) if $\mathbb{E}[y_t|d_t] = t(m)'\alpha_* + d_t'\eta_* + s(t)/\{n^{3/2}s'(n)\}$ *with* $s(\cdot)$ *being a SSV function,*

$$QLR_n \Rightarrow \sup_{\gamma\in\Gamma}\left(\widetilde{\mathcal{Z}}(\gamma) + \frac{p(\gamma)}{\widetilde{\sigma}(\gamma)}\right)^2. \qquad \square$$

Part (i) of Theorem 5 gives the power function of the QLR test statistic when the null model is misspecified by setting the polynomial time-trend degree too low. This result is useful in assuring consistency of the sequential testing algorithm discussed below. Polynomial functions do not belong to the SSV function class and parts (*ii*, *iii*, and *iv*) give the power function properties of the QLR test statistic against various SSV function alternatives and these results hold as corollaries of theorem 6 of BCP.

3.4 Sequentially Testing the Polynomial Time-Trend Model

The test procedure can be used sequentially to estimate polynomial degree using the approach in Section 2.6 applied to the time-trend stationary models $\mathcal{M}_j'$ for $j = 1, 2, \ldots, \bar{m}$. The results given in Corollary 1 continue to hold for sequential testing in this context. That is, if we let

$$m_* := \inf\{m \in \mathbb{N} : \exists(\alpha,\eta),\ \mathbb{E}[y_t|d_t] = s_{n,t}(m)'\alpha + d_t'\eta\},$$

as in Corollary 1, then for any $\epsilon > 0$ and significance level $\boldsymbol{\alpha}$, $\lim_{n\to\infty}\mathbb{P}\left(|\widehat{m}_n - m_*| > \epsilon\right) = \boldsymbol{\alpha}$. As before, consistent estimation of m_* is achieved if the significance level tends to zero slowly as $n \to \infty$.

Corollary 2. *Given that Assumptions 3 and 4 hold for each* $m \in P_d(\bar{m})$*, if (i)* $m_* \in P_d(\bar{m})$*, (ii) there is a Gaussian process* $\mathcal{B}^S(\cdot)$ *such that for all* $\gamma, \gamma' \in \Gamma$*, for some* δ*,* $\mathrm{cov}(\mathcal{B}^S(\gamma), \mathcal{B}^S(\gamma')) = 1 - |\gamma-\gamma'|^\delta(1 + o(1))$ *and* $\mathrm{cov}(\mathcal{B}^S(\gamma), \mathcal{B}^S(\gamma')) \leq \mathrm{cov}(\widetilde{\mathcal{Z}}(\gamma), \widetilde{\mathcal{Z}}(\gamma'))$*, (iii)* $\lim_{n\to\infty}\boldsymbol{\alpha}_n = 0$*, and (iv)* $\lim_{n\to\infty}\log(\boldsymbol{\alpha}_n)/n = 0$*, then for any* $\epsilon > 0$*,* $\lim_{n\to\infty}\mathbb{P}\left(|\widehat{m}_n - m_*| > \epsilon\right) = 0$*.* $\square$

The intuition behind Corollary 2 is identical to that of Theorem 3. As $\boldsymbol{\alpha}_n \to 0$, Theorem 4 implies that $\lim_{n\to\infty}\mathbb{P}(\widehat{m}_n > m_*) = 0$. Next, if the asymptotic critical value $c_n = o(n)$, Theorem 5(*i*) implies that $\lim_{n\to\infty}\mathbb{P}(\widehat{m}_n < m_*) = 0$, so that $\lim_{n\to\infty}\mathbb{P}(\widehat{m}_n = m_*) = 1$. This desired result follows just as in the proof of Theorem 3. The only point of difference from Theorem 3 is that we do not have to standardize $\widetilde{\mathcal{Z}}(\cdot)$ as its variance is already unity, as given in Theorem 4, so that Corollary 2 compares the covariance function of $\mathcal{B}^S(\cdot)$ directly with that of $\widetilde{\mathcal{Z}}(\cdot)$ to yield a consistent estimator for m_*.

4 Simulations

We conducted an extensive simulation to assess the performance characteristics of the QLR test statistic. The following simulation design was used for a time-trend stationary process. First, we generated data sets $\{y_t, d_t\}$ according to the scheme

$$y_t = \alpha_{0*} + \alpha_{1*}t + \alpha_{2*}t^2 + \eta_* d_t + u_t,$$

where $u_t := \cos(d_t)v_t$, $d_t := \rho_* d_{t-1} + w_t$ with $d_0 \sim N(0, 1/(1-\rho_*^2))$ such that $(v_t, w_t)' \sim_{iid} N(0_2, \sigma_*^2 I_2)$ and $(\alpha_{0*}, \alpha_{1*}, \alpha_{2*}, \eta_*, \sigma_*^2, \rho_*) = (1, 1, 1, 1, 1, 0.5)$. This design is a typical second degree polynomial time-trend stationary process with conditionally heteroskedastic residuals. Second, we used the following models for testing specification

$$\mathcal{M}'_m := \{\mu_t(\cdot) : \Omega_n \mapsto \mathbb{R} : \mu_t(\alpha_n, \eta, \beta_n, \gamma) := \ s_t(m)'\alpha_n + d_t\eta + \beta_n s_{n,t}^{\gamma}\}$$

with $\gamma \in \Gamma := [0.0, 3.5]$ and $m = 1, 2, 3$. These models have a parameter space Γ that includes the unknown polynomial degree as an interior element. Third, we implemented the sequential testing algorithm at significance levels of 1%, 5%, and 10%. We used sample sizes of 50, 100, 200, 300, 400, and 500, and for each sample size 5,000 replications were performed, enabling estimation of the probability of the sequential procedure leading to a polynomial degree estimate equal to the unknown true polynomial degree $m_* = 2$.

Simulation results are reported in Table 2 and can be summarized as follows. First, when m is less than the unknown polynomial degree 2, the model rejection rates are 100%. Even when the sample size is as small as 50, the rejection rates are 100% for every level of significance, implying that the sequential testing procedure estimates the degree less than the unknown polynomial degree with an extremely low probability. This also implies that the power of the QLR test statistic is high even for small sample sizes. Second, for the given significance level $\boldsymbol{\alpha}$, the predicted probability for the unknown polynomial degree is almost $(1 - \boldsymbol{\alpha})$ even when the sample size is as small as 50. This implies that the overall type I error is controlled efficiently in estimating the polynomial degree.

Before moving to discuss the next simulation, some caveats should be mentioned. First, the procedure assumes that the model is correctly specified with respect to other covariates. If the polynomial degrees of other explanatory variables are incorrectly specified, the estimated polynomial degree by the procedure can be biased. Second, in practice, a higher degree polynomial model can be rejected although a lower degree polynomial model cannot be rejected. Given that the lower degree polynomial model is nested within the higher degree polynomial model, the decision should be made based upon the test outcome for the higher degree polynomial model.

Next, we studied sequential estimation of the polynomial degree. For this purpose, we used the same design environment and applied Corollary 2 with the significance level $\boldsymbol{\alpha}_n$ determined by the sample size so that $\boldsymbol{\alpha}_n \to 0$ and $\log(\boldsymbol{\alpha}_n)/n \to 0$ as n increases. To assess performance, we estimated the empirical probability of $\widehat{m}_n$ equaling $m_* = 2$ for each $\boldsymbol{\alpha}_n$ as follows

$$\widehat{P}_n(\boldsymbol{\alpha}_n) := \frac{1}{r}\sum_{i=1}^{r}\mathbb{I}(\widehat{m}_{n,i} = m_*),$$

where $\mathbb{I}(\cdot)$ is the indicator function, r is the total number of iterations, viz., 5,000, and $\widehat{m}_{n,i}$ denotes the sequential estimator of m_* for the i-th simulation. For each given $\boldsymbol{\alpha}_n$, $\widehat{P}_n(\boldsymbol{\alpha}_n)$ estimates the probability of $\widehat{m}_n = m_*$, so that if $\widehat{m}_n$ estimates m_* consistently, $\widehat{P}_n(\boldsymbol{\alpha}_n) - (1 - \boldsymbol{\alpha}_n)$ should converge to zero as n tends to ∞ because $\boldsymbol{\alpha}_n \to 0$ as $n \to \infty$. We examine how $\widehat{P}_n(\boldsymbol{\alpha}_n)$ evolves as $n \to \infty$.

The simulation results are reported in Table 3. We consider three sequences for the level of significance: $\boldsymbol{\alpha}_n = n^{-1}$, $\boldsymbol{\alpha}_n = n^{-3/4}$, and $\boldsymbol{\alpha}_n = n^{-1/2}$. Note that $\boldsymbol{\alpha}_n \to 0$ and $\log(\boldsymbol{\alpha}_n)/n \to 0$ in each case as $n \to \infty$. If $\boldsymbol{\alpha}_n = n^{-1}$, the significance level approaches zero quickly, whereas the approach to zero is much slower when $\boldsymbol{\alpha}_n = n^{-1/2}$, and $\boldsymbol{\alpha}_n = n^{-3/4}$ provides an intermediate rate of approach. These rates are selected to cover significance levels between 10% and 0%, when the sample size is greater than 100, so that type I errors are neither too large or too small for moderately sized samples. If the level of significance converges to zero more slowly than $n^{-1/2}$, the level of significance becomes too large to use in most practical applications. On the other hand, if the level of significance converges to zero more quickly than n^{-1}, the level of significance is too small for good estimates $\widehat{P}_n(\boldsymbol{\alpha}_n)$.

The main results of Table 3 can be summarized as follows. First, the distance between $\widehat{P}_n(\boldsymbol{\alpha}_n)$ and $(1 - \boldsymbol{\alpha}_n)$ is close to zero for every selection of $\boldsymbol{\alpha}_n$. This outcome suggests that m_* can be successfully estimated by the sequential estimation procedure. Second, as the sample size increases, the distance between $\widehat{P}_n(\boldsymbol{\alpha}_n)$ and $(1 - \boldsymbol{\alpha}_n)$ shows evidence of convergence to zero for every selection of $\boldsymbol{\alpha}_n$, indicating as expected that degree estimation by $\widehat{m}_n$ becomes more precise in large samples. Third, the distance between $\widehat{P}_n(\boldsymbol{\alpha}_n)$ and $(1 - \boldsymbol{\alpha}_n)$ is relatively small when $\boldsymbol{\alpha}_n = n^{-1}$ and this choice of $\boldsymbol{\alpha}_n$ appears to deliver more desirable sequential estimation results than the other choices.

We compare these estimation results with standard information criterion-based estimators using the same

DGP. Three information criteria are examined, viz., Akaike's (1973, 1974) information criterion (AIC), the Bayesian information criterion (BIC), and small sample-size corrected AIC. These methods are applied to the following models

$$\mathcal{M}'_{0,m} := \{\mu_t(\cdot) : \Omega_n \mapsto \mathbb{R} : \mu_t(\alpha_0, \ldots, \alpha_m, \eta) := \alpha_0 + \alpha_1 t + \ldots + \alpha_m t^m + d_t \eta\},$$

with $m = 1, 2, 3$. Note that $\mathcal{M}'_{0,m}$ differs from $\mathcal{M}'_m$ in the fact that the power transform of the time trend is omitted from the right side of the model. The motivation for using $\mathcal{M}'_{0,m}$ lies in the fact that these information criteria are typically defined to apply to identified models, whereas if $\mathcal{M}'_m$ were attempted for use with $m = m_*$, the model would be unidentified. Instead, to apply the information criteria as degree selectors, we first follow the usual procedure of working with identified models. We let $\widetilde{m}_n$ be the polynomial degree estimated by the smallest information criterion value out of $m = 1, 2, 3$.

The penultimate lower panel of Table 3 shows simulation results based on the information criteria. The performances of the information criteria are measured by

$$\widetilde{P}_n := \frac{1}{r} \sum_{i=1}^{r} \mathbb{I}(\widetilde{m}_{n,i} = m_*),$$

where $\widetilde{m}_{n,i}$ is the estimator of m_* for the i-th simulation using the information criteria. The results are as follows. First, the performance measure $\widetilde{P}_n \times 100$ converges to 100% for BIC as the sample size increases, whereas those for AIC and AICc do not converge to 100% as fast as BIC. Second, BIC performs overall better than AIC and AICc. If the sample size is as high as 1,000, most of the estimates obtained give $m_* = 2$. In fact, 99.06% of 5,000 iterations are correctly estimated. Third, the overall performance of the BIC-based estimator is, nevertheless, inferior to those of the sequential test procedure. In particular, if $\boldsymbol{\alpha}_n = n^{-3/4}$ or n^{-1}, the sequential estimation of the polynomial degree is more often precise than the BIC-based estimator, whereas if $\boldsymbol{\alpha}_n = n^{-1/2}$, the BIC-based estimator shows better performance than the sequential estimation procedure. These results show that the sequential estimation procedure generally estimates polynomial degree better than information criteria, especially when faster approach rates to zero are selected for $\boldsymbol{\alpha}_n$.

We also apply the information criteria to $\mathcal{M}'_m$ despite the presence of the identification problem and report the simulation results in the lower panel of Table 3. To distinguish the earlier information criteria, we added the superscript '$'$' to the information criteria labels. The overall simulation results differ from the results using $\mathcal{M}'_{0,m}$. First, the performance measures steadily converge to 100% for all of the criteria AIC$'$, BIC$'$, and AICc$'$, as the sample size increases. Second, it is not recommended to use these information criteria in small samples. If the sample size is less than 500, performance of all of the information criteria is poor. On the other hand, if the sample size is as large as 600, performance of these criteria are more or less similar to those performed by AIC, BIC, and AICc. Third, the best performing information criterion is BIC$'$, although it is inferior to BIC, implying that the sequential estimation procedure outperforms BIC$'$ when $\boldsymbol{\alpha}_n = n^{-1}$ or $n^{-3/4}$, and the dominance of the sequential procedure now applies even when $n^{-1/2}$.

5 Empirical Applications

Since Mincer (1958, 1974) first introduced the earnings equation using schooling years and potential work experience, the following equation has been the most influential empirical model for human capital earnings:

$$\log(w_t) = \alpha_{0*} + \eta_* s_t + \alpha_{1*} x_t + \alpha_{2*} x_t^2 + u_t, \tag{12}$$

where w_t is earnings, s_t is schooling years, and x_t is potential work experience of individual t. Most empirical models on earnings data since Mincer (1958, 1974) are specified by adding more explanatory variables to the right side of (12) or by modifying the model in (12) into a structural equation.[2] Unless structural interpretations are involved, the unknown parameters are estimated by least squares method for most available earnings data across countries. The main reasons for the popularity of this model are its power to fit earnings data well despite its simple structure and its useful theory underpinnings. According to Card (1999), about 20–35% of earnings variation are explained by this simple equation .

Against this background persistent questions have been raised over the possibility that the earnings equation in (12) is misspecified. Murphy and Welch (1990) empirically examined the usefulness of the functional form in (12) using the current population survey (CPS) data from 1964 to 1987 and concluded that the quadratic functional form in (12) is unacceptable and argued instead for a quartic functional form in the experiences variable. Heckman, Lochner, and Todd (2006) and Lemieux (2006), motivated by the same question, both conclude that recent earnings data do not fit the Mincer equation as well as 1960's and 1970's earnings data. In particular, Lemieux (2006) shows that the quadratic function is not flexible enough to capture empirically the relationship between earnings and experiences. The quartic model is also preferred by Lemieux (2006), who points out that the Mincer equation in recent years needs to accommodate different cohort effects and potential misspecification in terms of schooling years that may be corrected by including squared schooling years. Heckman, Lochner, and Todd (2006) estimated the earnings equation nonparametrically, so that a polynomial degree was not estimated.

[2] See Card (1999) for a survey of the empirical literature on Mincer's equation.

Following a similar motivation to Murphy and Welch (1990), we revisit the Mincer equation using the QLR statistic to test specification. The data set used for this study is the same as in Card (1995) and examines the causal relationship between earnings and schooling years. The national longitudinal survey (NLS) data constructed by Card (1995) were drawn from 24–36 aged men in 1976, so that different cohort effects do not affect estimation of the Mincer equation. The sample size is 3,010, of which 2,707 individuals are white males. For more information on the data, readers are refered to Card (1995).

We focus on estimating the following models including the conventional Mincer equation in (12):

$$\log(w_t) = \alpha_{0*} + \eta_{1*} s_t + \alpha_{1*} x_t + \alpha_{2*} x_t^2 + \eta_{2*} b_t + \eta_{3*} m76_t + u_t, \tag{13}$$

$$\log(w_t) = \alpha_{0*} + \eta_{1*} s_t + \alpha_{1*} x_t + \alpha_{2*} x_t^2 + \eta_{2*} b_t + \eta_{3*} m76_t + \eta_{4*} m66_t + \sum_{j=5}^{11} \eta_{j*} r_{j,t} + u_t, \tag{14}$$

where b_t is a dummy variable for black/white, $m76_t$ is a dummy variable for residence in the South and in a metropolitan area in the year of 1976, $r_{j,t}$ is an indicator for region of residence in 1966; and $m66_t$ is a dummy variable for residence in the South and in a metropolitan area in the year of 1966. The year 1966 is treated as an important base year because the NLS data survey started in the same year. These models are the first two Mincer equation models estimated by Card (1995) modified by features of the NLS data. Note that all variables besides experience and schooling years are dummy variables, so that the functional form in the conditional mean equation is otherwise linear. In addition to these models, Card (1995) estimated various other models by including additional explanatory variables, but we focus here on the models in (12), (13), and (14) as the other model estimation results are very similar.

We apply the QLR test in the following manner. First, we test for further neglected nonlinearity with respect to experience x_t. We let the parameter space of the power coefficient be $[-0.25, 5.00]$, so that we can test up to fifth degree polynomial models as the null model. Hansen's (1996) weighted bootstrap is applied to our QLR test to obtain the p-values of the QLR tests. The bootstrap iteration number is 500. While computing the test statistics, we extend the null models in (12), (13), and (14) to including polynomial terms in schooling years. This modification accommodates the possibility that the QLR test may reject the null model because of nonlinearity with respect to schooling years, which was one of Lemieux's (2006) concerns. The models in (12), (13), and (14) are therefore extended as follows

$$\log(w_t) = \alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + u_t, \tag{15}$$

$$\log(w_t) = \alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + \eta_{1*} b_t + \eta_{2*} m76_t + u_t, \tag{16}$$

$$\log(w_t) = \alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + \eta_{1*} b_t + \eta_{2*} m76_t + \eta_{3*} m66_t + \sum_{j=4}^{10} \eta_{j*} r_{j,t} + u_t, \tag{17}$$

with m_1, $m_2 = 1, 2, \ldots, 5$. These models are treated as the null specification in our tests. Second, we reverse the roles of schooling years and experience and conduct the same testing procedures in the first step. That is, we test for further neglected nonlinearity with respect to schooling years s_t using the same parameter space for power coefficient.

The test results are contained in Tables 4 and 5. The left- and right-side panels report the p-values from testing for further neglected nonlinearity with respect to experience and schooling years, respectively. Inferences depend on the data, models, and levels of significance. Despite these differences, we can draw some consistent features of the data from these specification tests. We summarize the findings as follows.

First, the major implication of these tests on the specification of the Mincer equation is that all models that are linear in experience are rejected when testing for the neglected nonlinearity in experience at the 1% level of significance, confirming the presence of nonlinearity in this variable and the need for squared or higher degree polynomial terms in experience in the earnings equation. The nonlinearity in experience specification is further affirmed by testing the null models with respect to schooling years. All p-values in the right-side panels of Tables 4 and 5 imply that neglected nonlinear terms with respect to schooling years are hard to detect if squared or higher degree terms in experience are included in the regression, although its reversed relationship is not found. That is, even if schooling years are squared or further higher terms are included, the models are still nonlinear with respect to experience as observed for all models and hold also for white men. This finding differs from what Lemieux (2006) discovered from more recent CPS data.

Second, the results in Tables 4 and 5 imply that the original Mincer's hypothesis is statistically supported by the sequential estimation procedure. For the original Mincer equation, we focus on (15) and sequentially estimate m_1 and m_2 in the following manner using the first-left panels of Tables 4 and 5: for given m_2 say 1, we sequentially test $m_1 = 1, 2, \ldots, 5$ at the 1% level of significance. If we cannot accept the null for all $m_1 = 1, 2, \ldots, 5$, we increase m_2 to the next higher level, say 2 for this case, and continue testing with respect to $m_1 = 1, 2, \ldots, 5$, until the hypothesis cannot be rejected. We let m_2 increase from 1 to 5. The first-left panels of Table 4 and 5 show that m_1 and m_2 estimated by this sequential estimation are 1 and 2, respectively, and these are the same degrees as asserted by the Mincer equation. Furthermore, we also note that Mincer equation holds for white men data even when models are extended to Models (16) and (17). This finding is consistent with Heckman, Lochner, and Todd (2006) and Lemieux (2006)'s conclusion that the Mincer equation fits well 1960's and 1970's earnings data.

Third, the evidence suggests that different polynomial models for different set of explanatory variables are required to address nonlinearity in specification. The original Mincer equation does not include explanatory variables other than schooling years and experience. Models (16) and (17) are specified by including additional explanatory variables. Our empirical findings using black and white men data and the same sequential testing procedure evidently show that m_1 and m_2 in Model (16) need to be at least 2 and 3 in order to eliminate need for further nonlinearity in schooling years and experience, respectively. On the other hand, Model (17) estimates 1 and 3 for m_1 and m_2, respectively. These estimations show that the respective degrees of polynomial nonlinearity with respect to schooling years and experience in the original Mincer equation are not invariant to the inclusion of other explanatory variables in the model, thereby indicating the need for some flexibility in treating potential nonlinearity in these key variables, as is possible with flexible polynomial specifications and, more generally, with sieve approximants.

6 Conclusion

Testing for misspecification is now a standard feature of empirical econometric work. The methodology developed here provides a convenient mechanism for testing for an arbitrary presence of neglected nonlinearity in models that already involve polynomial functions of covariates or time trends. Given the extensive use of such polynomial specifications in empirical applications, it is especially useful to have simple tools to test directly for omitted nonlinearities. Our approach relies on QLR statistics that are constructed explicitly to evaluate the impact of including additional power transforms of the regressors in the regression. This approach provides for convenient implementation to assess specification in practice and further enables direct estimation of polynomial degree along with its consistent power against arbitrary alternatives. While the methods have been developed here for parametric models, they may be used in the context of nonparametric sieve approximations in assessing choice of a polynomial approximant degree.

Of particular interest is the fact that the null limit distribution of the QLR statistic resolves the multifold identification problem inherent in polynomial and power transform regressions. Moreover, when the prediction errors in the equation form an MDS the QLR test statistic is asymptotically distribution free for testing further neglected nonlinearity with respect to time trends, so is well suited for convenient application in models where the nature of the time trend is uncertain. Simulations confirm that these tests have good finite sample performance and relate well to the limit theory. The sequential testing procedure for consistently estimating unknown polynomial degree also works well in simulations, comparing favorably with and frequently dominating the performance of information criteria. Simulations show that this procedure controls overall type I error efficiently. Empirical application of these methods to earnings data studied by Card (1995) show that the methods are informative about specification weaknesses in conventional Mincer equation modeling, indicating that more flexible specifications are needed to capture the impact of schooling and experience on earnings.

7 Appendix

Before proving the main results, we provide the following supplementary lemma to assist in the derivations.

Lemma A1. *Given Assumptions 1 and 2,*

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

(ii) $A_c'Z = O_{\mathbb{P}}(n)$, $Z'Z = O_{\mathbb{P}}(n)$, $E_c'Z = O_{\mathbb{P}}(n)$*;*

(iii) $A_c'A_c = O_{\mathbb{P}}(n)$, $A_c'E_c = O_{\mathbb{P}}(n)$, $B_c'U = O_{\mathbb{P}}(n)$, $B_c'Z = O_{\mathbb{P}}(n)$, $E_c'Z = O_{\mathbb{P}}(n)$, $E_c'E_c = O_{\mathbb{P}}(n)$, $F_c'U = O_{\mathbb{P}}(n)$*, and* $F_c'Z = O_{\mathbb{P}}(n)$*; and*

(iv) $B_c'U = o_{\mathbb{P}}(n)$ *and* $F_c'U = o_{\mathbb{P}}(n)$. □

Proof of Lemma A1: (*i*) By the definition of $E_c := [0_{n\times c} \vdots A_c \vdots 0_{n\times(m-c+k)}]$, we note that if $A_c'U = O_{\mathbb{P}}(\sqrt{n})$, then $E_c'U = O_{\mathbb{P}}(\sqrt{n})$. Therefore, we focus on proving that $A_c'U = O_{\mathbb{P}}(\sqrt{n})$.

By the definition of $A_c'U$, $n^{-1/2}A_c'U = [n^{-1/2}\sum x_t^c \log(x_t)u_t]$, so that if $\mathbb{E}[x_t^{2c}\log^2(x_t)\ u_t^2] < \infty$, we can apply the CLT. When we apply the Cauchy-Schwarz inequality, we obtain: (*a*) $\mathbb{E}[x_t^{2c}\log^2(x_t)u_t^2] \leq \mathbb{E}[x_t^{4c}\log^4(x_t)]^{1/2}\mathbb{E}[u_t^4]^{1/2} \leq \mathbb{E}[x_t^{8c}]^{1/4}\mathbb{E}[\log^8(x_t)]^{1/4}\mathbb{E}[u_t^4]^{1/2}$; (*b*) $\mathbb{E}[x_t^{2c}\log^2(x_t)u_t^2] \leq \mathbb{E}[u_t^4\log^4(x_t)]^{1/2}\mathbb{E}[x_t^{4c}]^{1/2} \leq \mathbb{E}[u_t^8]^{1/4}\mathbb{E}[\log^8(x_t)]^{1/4}\mathbb{E}[x_t^{4c}]^{1/2}$; and (*c*) $\mathbb{E}[x_t^{2c}\log^2(x_t)u_t^2] \leq \mathbb{E}[x_t^{4c}u_t^4]^{1/2}\mathbb{E}[\log^4(x_t)]^{1/2} \leq \mathbb{E}[x_t^{8c}]^{1/4}\ \mathbb{E}[u_t^8]^{1/4}\mathbb{E}[\log^4(x_t)]^{1/2}$. We now note that the elements in the right side of (*a*), (*b*), and (*c*) are finite by Assumption 2(*iii*), respectively.

As for $Z'U$, $n^{-1/2}Z'U = n^{-1/2}\sum z_{t,i}u_t$ obeys a CLT if $\mathbb{E}[z_{t,i}^2u_t^2] < \infty$. We note that $\mathbb{E}[z_{t,i}^2u_t^2] \leq \mathbb{E}[z_{t,i}^4]^{1/2}\mathbb{E}[u_t^4]^{1/2}$ by the Cauchy-Schwarz inequality. If $\mathbb{E}[z_{t,i}^4] < \infty$ and $\mathbb{E}[u_t^4] < \infty$, the desired results follow. These conditions are already required in Assumption 2.

(*ii*) As in (*i*), if $A_c'Z = O_{\mathbb{P}}(n)$, $E_c'Z = O_{\mathbb{P}}(n)$ by the definition of E_c. For $A_c'Z = [\sum x_t^c \log(x_t)z_{t,i}]$, this obeys the LLN if $\mathbb{E}[|x_t^c\log(x_t)z_{t,i}|] < \infty$. We consider two cases separately: for some ℓ, when $z_{t,i} = d_{t,\ell}$ and when $z_{t,i} = x_t^\ell$.

Take the case: $z_{t,i} = d_{t,\ell}$. Note that $\mathbb{E}[x_t^c\log(x_t)z_{t,i}] = \mathbb{E}[x_t^c\log(x_t)d_{t,\ell}]$. Therefore, (*a*) $\mathbb{E}[x_t^c\log(x_t)d_{t,\ell}] \leq \mathbb{E}[x_t^{2c}\log^2(x_t)]]^{1/2}\mathbb{E}[d_{t,\ell}^2]^{1/2} \leq \mathbb{E}[x_t^{4c}]^{1/4}\mathbb{E}[\log^4(x_t)]^{1/4}\mathbb{E}[d_{t,\ell}^2]^{1/2}$; (*b*) $\mathbb{E}[x_t^c\log(x_t)d_{t,\ell}] \leq \mathbb{E}[d_{t,\ell}^2\log^2(x_t)]^{1/2}\mathbb{E}[x_t^{2c}]^{1/2} \leq \mathbb{E}[d_{t,\ell}^4]^{1/4}\mathbb{E}[\log^4(x_t)]^{1/4}\mathbb{E}[x_t^{2c}]^{1/2}$; (*c*) $\mathbb{E}[x_t^c\log(x_t)d_{t,\ell}] \leq \mathbb{E}[x_t^{2c}d_{t,\ell}^2]^{1/2}\mathbb{E}[\log^2(x_t)]^{1/2}$

$\leq \mathbb{E}[x_t^{4c}]^{1/4}\mathbb{E}[d_{t,\ell}^4]^{1/4}\mathbb{E}[\log^2(x_t)]^{1/2}$ by the Cauchy-Schwarz inequality. All these bounds are finite by Assumption 2(*iii*).

Next consider the case when $z_{t,i} = x_t^\ell$. Then, $\mathbb{E}[x_t^c \log(x_t) z_{t,i}] = \mathbb{E}[x_t^{c+\ell}\log(x_t)]$, which is bounded by $\mathbb{E}[x_t^{2(c+\ell)}]^{1/2}\mathbb{E}[\log^2(x_t)]^{1/2}$. We note that Assumption 2(*iii*) then ensures the required finite bound.

As for $Z'Z$, $n^{-1}Z'Z = n^{-1}\sum z_{t,i}z_{t,\ell}$ obeys an LLN if $\mathbb{E}[|z_{t,i}z_{t,\ell}|] < \infty$. We note that $\mathbb{E}[|z_{t,i}z_{t,\ell}|] \leq \mathbb{E}[z_{t,i}^2]^{1/2}\mathbb{E}[z_{t,\ell}^2]^{1/2}$ by the Cauchy-Schwarz inequality. If $\mathbb{E}[z_{t,i}^2] < \infty$, the desired results follows as it is assumed in Assumption 2(*iii*).

(*iii*) By the definitions of E_c and $F_c := [0_{n\times c} \vdots B_c \vdots 0_{n\times(1+m-j+k)}]$, if $A_c'A_c = O_{\mathbb{P}}(n)$, $B_c'U = O_{\mathbb{P}}(n)$, $B_c'Z = O_{\mathbb{P}}(n)$, and $A_c'Z = O_{\mathbb{P}}(n)$ then $A_c'E_c = O_{\mathbb{P}}(n)$, $F_c'U = O_{\mathbb{P}}(n)$, $F_c'Z = O_{\mathbb{P}}(n)$, $E_c'E_c = O_{\mathbb{P}}(n)$, and $E_c'Z = O_{\mathbb{P}}(n)$. We have already shown that $A_c'Z = O_{\mathbb{P}}(n)$ in (*ii*). We, therefore, focus on proving $A_c'A_c = O_{\mathbb{P}}(n)$, $B_c'U = O_{\mathbb{P}}(n)$, and $B_c'Z = O_{\mathbb{P}}(n)$.

We examine each case in turn. (*a*) We note that $n^{-1}A_c'A_c = n^{-1}\sum x_t^{2c}\log^2(x_t)$, so that if $\mathbb{E}[x_t^{2c}\log^2(x_t)] < \infty$, the LLN holds. We note that $\mathbb{E}[x_t^{2c}\log^2(x_t)] \leq \mathbb{E}[x_t^{4c}]^{1/2}\mathbb{E}[\log^4(x_t)]^{1/2}$, and the right side is finite by Assumption 2(*iii*).

(*b*) Note that $n^{-1}B_c'U = n^{-1}\sum x_t^c \log^2(x_t)u_t$ and, if $\mathbb{E}[|x_t^c\log^2(x_t)u_t|] < \infty$, the LLN holds. We also note that (*b.i*) $\mathbb{E}[x_t^c\log^2(x_t)u_t] \leq \mathbb{E}[x_t^{2c}\log^4(x_t)]^{1/2}\ \mathbb{E}[u_t^2]^{1/2} \leq \mathbb{E}[x_t^{4c}]^{1/4}\mathbb{E}[\log^8(x_t)]^{1/4}\ \mathbb{E}[u_t^2]^{1/2}$; (*b.ii*) $\mathbb{E}[x_t^c\log^2(x_t)u_t] \leq \mathbb{E}[u_t^2\log^4(x_t)]^{1/2}\mathbb{E}[x_t^{2c}]^{1/2} \leq \mathbb{E}[u_t^4]^{1/4}\mathbb{E}[\log^8(x_t)]^{1/4}\mathbb{E}[x_t^{2c}\ ]^{1/2}$; and (*b.iii*) $\mathbb{E}[x_t^c\log^2(x_t)u_t] \leq \mathbb{E}[u_t^2 x_t^{2c}]^{1/2}\mathbb{E}[\log^2(x_t)]^{1/2} \leq \mathbb{E}[u_t^4]^{1/4}\mathbb{E}[x_t^{4c}]^{1/4}\mathbb{E}[\log^2(x_t)]^{1/2}$. Thus, each of the elements forming the right side is finite by Assumption 2(*ii.a*), 2(*ii.b*), and 2(*ii.c*), respectively.

(*c*) Finally, we examine $n^{-1}B_c'Z = [n^{-1}\sum x_t^c\log^2(x_t)z_{t,i}]$. As before, there are two separate cases: for some ℓ, $z_{t,i} = d_{t,\ell}$ or $z_{t,i} = x_t^\ell$. We first consider $z_{t,i} = d_{t,\ell}$. Note that $\mathbb{E}[|x_t^c\log^2(x_t)z_{t,i}|] = \mathbb{E}[|x_t^c\log^2(x_t)d_{t,\ell}|]$. Therefore, (*c.i*) $\mathbb{E}[|x_t^c\log^2(x_t)d_{t,\ell}|] \leq \mathbb{E}[x_t^{2c}\log^4(x_t)]^{1/2}\mathbb{E}[d_{t,\ell}^2]^{1/2} \leq \mathbb{E}[x_t^{4c}]^{1/4}\mathbb{E}[\log^8(x_t)]^{1/4}\ \mathbb{E}[d_{t,\ell}^2]^{1/2}$; (*c.ii*) $\mathbb{E}[x_t^c\log^2(x_t)x_{t,i}] \leq \mathbb{E}[d_{t,\ell}^2\log^4(x_t)]^{1/2}\mathbb{E}[x_t^{2c}]^{1/2} \leq \mathbb{E}[d_{t,\ell}^4]^{1/4}\mathbb{E}[\log^8(x_t)]^{1/4}\mathbb{E}[x_t^{2c}]^{1/2}$; and (*c.iii*) $\mathbb{E}[|x_t^c\log^2(x_t)d_{t,\ell}|] \leq \mathbb{E}[d_{t,\ell}^2\ x_t^{2c}]^{1/2}\mathbb{E}[\log^4(x_t)]^{1/2} \leq \mathbb{E}[d_{t,\ell}^4]^{1/4}\ \mathbb{E}[x_t^{4c}]^{1/4}\mathbb{E}[\log^4(x_t)]^{1/2}$. Then, the right sides are finite by Assumption 2(*iii.a*), 2(*iii.b*), and 2(*iii.c*), respectively.

Next consider $z_{t,i} = x_t^\ell$. Then, $\mathbb{E}[|x_t^c\log^2(x_t)z_{t,i}|] = \mathbb{E}[|x_t^{c+\ell}\log^2(x_t)|] \leq \mathbb{E}[|x_t^{2j-2+2\ell}|]^{1/2}\ \mathbb{E}[|\log^4(x_t)|]^{1/2}$. This bound is also finite by Assumption 2(*iii*).

(*iv*) By the definition of F_c, if $B_c'U = o_{\mathbb{P}}(n)$, it follows that $F_c'U = o_{\mathbb{P}}(n)$. We already proved that $B_c'U = O_{\mathbb{P}}(n)$ in (*iii*), and applying the LLN and the MDS condition in Assumption 2(*ii*) implies that $B_c'U = o_{\mathbb{P}}(n)$. This completes the proof. ■

Proof of Lemma 1: (*i*) To show the stated claim, we first derive the first-order derivative of $L_n(\gamma;\alpha_c)$ with respect to γ. Note that

$$L_n^{(1)}(\gamma;\alpha_c) = \; 2P_c(\alpha_c)'Q_c(\gamma)(Q_c(\gamma)'Q_c(\gamma))^{-1}\left[(d/d\gamma)Q_c(\gamma)'P_c(\alpha_c)\right] \\ + P_c(\alpha_c)'Q_c(\gamma)\{(d/d\gamma)(Q_c(\gamma)'Q_c(\gamma))^{-1}\}Q_c(\gamma)'P_c(\alpha_c),$$

$Q_c(c) = Z$ from $Q_c(\gamma) := [X(0),\ldots,X(j-2),X(\gamma),X(j),\ldots,X(m),D]$ and $(d/d\gamma)Q_c(\gamma) = E_c$. Next, $P_c(\alpha_c) = Y - \alpha_c X(c) = Z[\alpha_{0*},\ldots,\alpha_{j-2},(\alpha_{c*}-\alpha_c),\alpha_j,\ldots,\alpha_{m*},\eta_*']' + Z'U = Z\kappa_c + U$, so that $P_c(\alpha_c) = Z\kappa_c + U$. Finally, we obtain that

$$(d/d\gamma)(Q_c(\gamma)'Q_c(\gamma))^{-1}_{\gamma=c} = -(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1} \tag{18}$$

and collect all these separate derivations in $(d/d\gamma)L_n(\gamma;\alpha_c)$. This yields that

$$L_n^{(1)}(c;\alpha_c) = 2(Z\kappa_c+U)'Z(Z'Z)^{-1}E_c'(Z\kappa_c+U) \\ -(Z\kappa_c+U)'Z(Z'Z)^{-1}(Z'E_c+E_c'Z)(Z'Z)^{-1}Z'(Z\kappa_c+U).$$

We further rearrange the terms on the right side. The first component is the sum of four other components : (a) $2\kappa_c'Z'Z(Z'Z)^{-1}E_c'Z\kappa_c = 2\kappa_c'E_c'Z\kappa_c$; (b) $2\kappa_c'E_c'U$; (c) $2U'Z(Z'Z)^{-1}E_c'Z\kappa_{cy} = 2\kappa_c'Z'E_c(Z'Z)^{-1}Z'$ U; and (d) $2U'Z(Z'Z)^{-1}E_c'U$. Next, the second component is the sum of four components: (a) $-\kappa_c'Z'E_c\kappa_c$ $-\kappa_c'E_c'Z\kappa_c = -2\kappa_c'E_c'Z\,\kappa_c$; (b) $-U'Z(Z'Z)^{-1}Z'E_c\,\kappa_c - \kappa_c'E_c'Z(Z'Z)^{-1}Z'U = -2\kappa_c'E_c'Z(Z'Z)^{-1}Z'\,U$; (c) $-U'Z(Z'Z)^{-1}E_c'Z\kappa_c - \kappa_c'Z'\,E_c(Z'Z)^{-1}Z'U = -2\kappa_c'Z'E_c(Z'Z)^{-1}Z'U$; and (d) $-U'Z(Z'Z)^{-1}(Z'E_c$ $+E_j'Z)(Z'Z)^{-1}Z'U$. If we collect these eight different components according to their order of convergence, they can be classified into the following three different terms:

- (a) $2\kappa_c'E_c'Z\kappa_c - 2\kappa_c'E_c'Z\kappa_c = 0$;
- (b, c) $2\kappa_c'\{E_c' + Z'E_c(Z'Z)^{-1}Z' - E_c'Z(Z'Z)^{-1}Z' - Z'E_c(Z'Z)^{-1}Z'\}U = 2(\alpha_{c*}-\alpha_c)A_c'MU$ because $Z'E_c = A_c'$;
- (d) $2U'Z(Z'Z)^{-1}E_c'U - U'Z(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}Z'U$,

so that the first-order derivative is now obtained as

$$L_n^{(1)}(c;\alpha_c) = 2(\alpha_{c*}-\alpha_c)A_c'MU + 2U'E_c(Z'Z)^{-1}Z'U - U'Z(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}Z'U,$$

and this is the desired first-order derivative. Given this derivative, Lemma A1(*i* and *ii*) implies that the second and third terms in the right side are $o_{\mathbb{P}}(n)$, so that the desired result follows from this.

(*iii*) We next examine the second-order derivative. In the same way, we obtain that

$$\begin{aligned} L_n^{(2)}(c;\alpha_c) =& 2(P_c(\alpha_c)'E_c)(Z'Z)^{-1}(E_c'P_c(\alpha_c)) + 4(P_c(\alpha_c)'Z)\{(d/d\gamma)[Q_c c'Q_c c]^{-1}\}E_c'P_c(\alpha_c) \\ &+ 2(P_c(\alpha_c)'Z)(Z'Z)^{-1}F_c'P_c(\alpha_c) + (P_c(\alpha_c)'Z)\{(d^2/d\gamma^2)[Q_c c'Q_c c]^{-1}\}Z'P_c(\alpha_c). \end{aligned}$$

We note that (18) already provides the form of $(d/d\gamma)[Q(\gamma)'Q(\gamma)]^{-1}_{\gamma=c}$, and

$$\begin{aligned} (d^2/d\gamma^2)[Q(\gamma)'Q(\gamma)]^{-1}_{\gamma=c} =& 2Z(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}Z' \\ &- (Z'Z)^{-1}(2E_c'E_c + Z'F_c + F_c'Z)(Z'Z)^{-1}. \end{aligned}$$

Using these and the previous definitions, the second-order derivative is obtained as

$$\begin{aligned} L_n^{(2)}(c;\alpha_c) = & 2(Z\kappa_c + U)'\{E_c(Z'Z)^{-1}E_c' + Z(Z'Z)^{-1}F_c'\}(Z\kappa_c + U) \\ &- 4(Z\kappa_c + U)'Z(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}E_c'(Z\kappa_c + U) \\ &- (Z\kappa_c + U)'Z(Z'Z)^{-1}(2E_c'E_c + Z'F_c + F_c'Z)(Z'Z)^{-1}Z'(Z\kappa_c + U) \\ &+ 2(Z\kappa_c + U)'Z(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}Z'(Z\kappa_c + U). \end{aligned}$$

Finally, we rearrange the right side according to their order of convergence and obtain that

- $2\kappa_c'\{Z'E_c'(Z'Z)^{-1}E_c' + F_c'\}Z\kappa_c - 4\kappa_c'(Z'E_c + E_c'Z)(Z'Z)^{-1}E_c'Z\kappa_c + 2\kappa_c'(Z'E_c + E_c'Z)(Z'Z)^{-1}Z'$ $E_c + E_c'Z)\kappa_c - \kappa_c'(2E_c'E_c + Z'F_c + F_c'Z)\ \kappa_c = 2\kappa_c'E_c'Z(Z'Z)^{-1}Z'E_c\kappa_c - 2\kappa_c'E_c'E_c\kappa_c = -2(\alpha_{c*} - \alpha_c)^2 A_c'M\ A_c$;
- $4\kappa_c'Z'E_c(Z'Z)^{-1}E_c'U - 4\kappa_c'(Z'E_c + E_c'Z)(Z'Z)^{-1}E_c'U - 4\kappa_c'Z'E_c(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}Z'$ $U + 2\kappa_c'F_c'U + 2\kappa_c'Z'F_c(Z'Z)^{-1}Z'U + 4\kappa_c'(Z'E_c + E_c'Z)(Z'\ Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}Z'U -$ $2\kappa_c'(2E_c'E_c + Z'F_c + F_c'Z)(Z'Z)^{-1}Z'\ U = 2(\alpha_{c*} - \alpha_c)[B_c'MU - 2A_c'ME_c(Z'Z)^{-1}Z'U - 2A_c'Z(Z'$ $Z)^{-1}E_c'MU]$; and
- $2[U'E_c(Z'Z)^{-1}E_c'U + U'F_c(Z'Z)^{-1}Z'U - 2U'E_c(Z'Z)^{-1}(Z'E_c + E_c'Z)(Z'Z)^{-1}Z'\ U] + 2U'Z(Z'$ $Z)^{-1}[(Z'E_c + E_c'Z)(Z'Z)^{-1}(Z'E_c + E_c'Z) - E_c'E_c - Z'F_c](Z'Z)^{-1}Z'\ U$.

We now apply Lemma A1 to each of these terms. First, Lemma A1(*ii* and *iii*) imply that $A_c'MA_c = A_c'A_c - AcZ(Z'Z)^{-1}Z'A_c = O_{\mathbb{P}}(n)$. Second, $B_c'MU = B_c'U - B_c'Z(Z'Z)^{-1}Z'U$, and Lemma A1 (*ii* and *iii*) implies that $B_c'MU = O_{\mathbb{P}}(n)$. Furthermore, Lemma A1(*iv*) implies that $B_c'MU = o_{\mathbb{P}}(n)$.

Third, $A_c' M E_c = A_c' E_c - A_c' Z(Z'Z)^{-1} Z' E_c$. Assumption 2 and Lemma A1(*ii*, *iii*, and *iv*) imply that $A_c' M E_c (Z'Z)^{-1} Z'U = o_{\mathbb{P}}(n)$. Fourth, $E_c' MU = E_c U - E_c Z(Z'Z)^{-1} Z'U = o_{\mathbb{P}}(n)$ by Lemma A1(*i* and *iv*), so that $A_c' Z(Z'Z)^{-1} E_c' MU = o_{\mathbb{P}}(n)$. Therefore, $B_c' MU - 2A_c' M E_c (Z'Z)^{-1} Z'U - 2A_c' Z(Z'Z)^{-1} E_c' MU = o_{\mathbb{P}}(n)$. Finally, we combine all terms in Lemma A1 and obtain that

$$\begin{aligned}
&2[U' E_c (Z'Z)^{-1} E_c' U + U' F_c (Z'Z)^{-1} Z'U - 2U' E_c (Z'Z)^{-1} (Z' E_c + E_c' Z)(Z'Z)^{-1} Z'U] \\
&\quad + 2U' Z(Z'Z)^{-1} [(Z' E_c + E_c' Z)(Z'Z)^{-1} (Z' E_c + E_c' Z) - E_c' E_c - Z' F_c](Z'Z)^{-1} Z'U = o_{\mathbb{P}}(n).
\end{aligned}$$

All of these facts imply that $L_n^{(2)}(c; \alpha_c) = -2(\alpha_{c*} - \alpha_c)^2 A_c' M A_c + o_{\mathbb{P}}(n)$. ■

Proof of Lemma 2: It is proved in the text. ■

Proof of Lemma 3: Given Lemma 2, the proof is almost identical to the proof of theorem 1 of BCP. ■

Proof of Theorem 1: In fact, (10) implies that $QLR_n = QLR_n^{(\beta=0)}$ under $\mathcal{H}_{0,m}$, and Lemma 3(*ii*) implies that $QLR_n^{(\beta=0)} \Rightarrow \sup_{\gamma \in \Gamma} \mathcal{Z}(\gamma)^2$. The desired result follows. ■

Proof of Theorem 2: (*i* and *ii*) Assumptions 1 and 2 satisfy the regularity assumptions 1, 2(*iii*, *v*), 4(*ii*), and 5 of BCP. Furthermore, we can let $[x_t, x_t^2, \ldots, x_t^m]$ be a part of d_t of BCP. From these two facts, the assumptions in theorem 5 of BCP are satisfied. Therefore, the BCP results apply to Theorem 2 with $m(x_t)$ of BCP being $s(x_t)$ in the current paper. ■

Proof of Theorem 3: Before proving the claim, we let $\underline{\gamma}$ and $\bar{\gamma}$ be the lower and upper limit of Γ such that $\Gamma_j := [\gamma_j, \gamma_{j+1}]$ such that $\gamma_0 := \underline{\gamma}$, $\gamma_{\bar{m}+1} := \bar{\gamma}$, and for $j = 1, 2, \ldots, \bar{m}$, $\gamma_j := j$.

We now prove the stated claim. First, $\lim_{n\to\infty} \mathbb{P}(\widehat{m}_n > m_*) = \lim_{n\to\infty} \boldsymbol{\alpha}_n = 0$ by virtue of the size decay condition (ii). Second, Theorem 2(*i*) implies that if $c_n = o(n)$, for any $j < m_*$, $\lim_{n\to\infty} \mathbb{P}(QLR_n^{(j)} > c_n) = 1$. This implies that if $\boldsymbol{\alpha}_n$ is selected to yield $c_n = o(n)$, the desired result follows. We note the following six properties (*i* to *vi*): (*i*) $\sup_{\gamma\in\Gamma_j} \mathcal{Z}(\gamma)^2 = \sup_{\gamma\in\Gamma_j} \{\max[0, \mathcal{Z}(\gamma)]^2 + \max[0, \mathcal{Z}(\gamma)]^2\} \leq \sup_{\gamma\in\Gamma_j} \max[0, \mathcal{Z}(\gamma)]^2 + \sup_{\gamma\in\Gamma_j} \max[0, \mathcal{Z}(\gamma)]^2\}$, so that for any $u > 0$,

$$\begin{aligned}
\mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \mathcal{Z}(\gamma)^2 \geq u^2\right) &\leq \mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \max[0, \mathcal{Z}(\gamma)]^2 \geq \frac{u^2}{2}\right) + \mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \min[0, \mathcal{Z}(\gamma)]^2 \geq \frac{u^2}{2}\right) \\
&= \mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \mathcal{Z}(\gamma) \geq \frac{u}{\sqrt{2}}\right) + \mathbb{P}\left(\inf_{\gamma\in\Gamma_j} \mathcal{Z}(\gamma) \leq -\frac{u}{\sqrt{2}}\right) = 2\mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \mathcal{Z}(\gamma) \geq \frac{u}{\sqrt{2}}\right)
\end{aligned}$$

by the fact that $\mathbb{P}(\inf_{\gamma\in\Gamma_j} \mathcal{Z}(\gamma) \leq -u/\sqrt{2}) = \mathbb{P}(\sup_{\gamma\in\Gamma_j} \mathcal{Z}(\gamma) \geq u/\sqrt{2})$, where the last equality holds from the symmetry of Gaussian process distribution. Therefore, for any $u > 0$,

$$\mathbb{P}\left(\sup_{\gamma\in\Gamma} \mathcal{Z}(\gamma)^2 \geq u^2\right) \leq 2\sum_{j=1}^{\bar{m}+1} \mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \mathcal{Z}(\gamma) \geq \frac{u}{\sqrt{2}}\right). \tag{19}$$

(*ii*) Given the conditions, if we let $\sigma_* := \sup_{\gamma\in\Gamma} \mathrm{var}[\mathcal{Z}(\gamma)]^{1/2}$, for any γ, $|\mathcal{Z}(\gamma)/\sigma_*| \leq |\mathcal{Z}(\gamma)/\sigma^0(\gamma)| = |\mathcal{Z}^0(\gamma)|$, so that for any $u > 0$,

$$\mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \frac{\mathcal{Z}(\gamma)}{\sigma_*} \geq u\right) \leq \mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \mathcal{Z}^0(\gamma) \geq u\right). \tag{20}$$

(*iii*) Lemma 7.1 of Piterbarg (1996) implies that as $u \to \infty$,

$$\mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \mathcal{B}^S(\gamma) \geq u\right) = H_\delta \mu(\Gamma_j) u^{2/\delta}(1 - \Phi(u))(1 + o(1)), \tag{21}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal random variable, $\mu(\cdot)$ is the Lebesgue measure of the given argument, $H_\delta := \lim_{\bar{\gamma}\to\infty} H(\bar{\gamma})/\bar{\gamma}$, and $H(\bar{\gamma}) := E[\exp(\max_{\gamma\in[0,\bar{\gamma}]} \mathcal{B}^F(\gamma))]$. Here, $\mathcal{B}^F(\cdot)$ is a fractional Brownian motion with mean $-|\gamma|^\delta$ and $\mathrm{cov}(\mathcal{B}^F(\gamma), \mathcal{B}^F(\gamma')) = |\gamma|^\delta + |\gamma'|^\delta - |\gamma - \gamma'|^\delta$ on Γ.

(*iv*) The Slepian inequality implies that for any v, $\mathbb{P}(\sup_\gamma \mathcal{Z}^0(\gamma) \geq v) \leq \mathbb{P}(\sup_\gamma \mathcal{B}^S(\gamma) \geq v)$ (*e.g.*, Piterbarg, 1996, p.6). Therefore, the Slepian inequality, (20), and (21) imply that as $u \to \infty$,

$$\mathbb{P}\left(\sup_{\gamma\in\Gamma_j} \mathcal{Z}(\gamma) \geq \frac{u}{\sqrt{2}}\right) \leq H_\delta \mu(\Gamma_j) \left(\frac{u}{\sqrt{2}\sigma_*}\right)^{2/\delta} \left(1 - \Phi\left(\frac{u}{\sqrt{2}\sigma_*}\right)\right)(1 + o(1)), \tag{22}$$

so that it follows that

$$\mathbb{P}\left(\sup_{\gamma\in\Gamma} \mathcal{Z}(\gamma)^2 \geq u^2\right) \leq 2H_\delta \mu_* \left(\frac{u}{\sqrt{2}\sigma_*}\right)^{2/\delta} \left(1 - \Phi\left(\frac{u}{\sqrt{2}\sigma_*}\right)\right)(1 + o(1))$$

by (19), where $\mu_* := \mu(\Gamma)$.

(*v*) We further note that $1 - \Phi(\cdot) = \frac{1}{2}\mathrm{erfc}((\cdot)/\sqrt{2}) \leq \frac{1}{2}\exp(-(\cdot)^2/2)$. Hence, if $u \to \infty$, it follows from

$$\mathbb{P}\left(\sup_{\gamma\in\Gamma} \mathcal{Z}(\gamma)^2 \geq u^2\right) \leq H_\delta \mu_* \left(\frac{u^2}{2\sigma_*^2}\right)^{1/\delta} \exp\left(-\frac{u^2}{4\sigma_*^2}\right)(1 + o(1)). \tag{23}$$

(*vi*) Finally, if we let the left side of (23) and u^2 be the significance level $\boldsymbol{\alpha}_n$ and its associated critical value c_n, respectively, then

$$-\frac{\log(\boldsymbol{\alpha}_n)}{n} \geq -\frac{1}{\delta}\frac{\log(c_n)}{n} + \frac{1}{4\sigma_*^2}\frac{c_n}{n} + o(1)$$

by noting that $\left\{\log\left(H_\delta \mu_*\right) - \frac{1}{\delta}\log(2\sigma_*^2)\right\} = O(1)$. We now note that

$$-\frac{1}{\delta}\frac{\log(c_n)}{n} + \frac{1}{4\sigma_*^2}\frac{c_n}{n} = \frac{1}{4\sigma_*^2}\frac{c_n}{n}\left(1 - \frac{4\sigma_*^2}{\delta}\frac{\log(c_n)}{c_n}\right) = \frac{1}{4\sigma_*^2}\frac{c_n}{n}\left(1 + o(1)\right)$$

as $c_n \to \infty$. Therefore, if $\log(\boldsymbol{\alpha}_n) = o(n)$, as is assumed in condition (iii), it follows that $c_n = o(n)$. This completes the proof. ■

Proof of Theorem 4: Weak convergence of the QLR test statistic is proved in the same way as that of Theorem 1, so we only derive the covariance kernel of $\widetilde{\mathcal{Z}}(\cdot)$.

First, note that applying Theorem 1 implies that $QLR_n = \sup_{\gamma\in\Gamma}\{S(\gamma)'MU\}^2/\{\widehat{\sigma}_{n,0}^2 S(\gamma)'MS(\gamma)\}$ under $\widetilde{\mathcal{H}}_0$. Next, applying the ULLN to $n^{-1}S(\cdot)'MS(\cdot)$ shows that $\sup_{\gamma\in\Gamma}|n^{-1}\widehat{\sigma}_{n,0}^2 S(\gamma)'MS(\gamma) - \widetilde{\sigma}^2(\gamma,\gamma)| \stackrel{\text{a.s.}}{\to} 0$, where for each γ,

$$\widetilde{\sigma}^2(\gamma,\gamma) := \sigma_*^2\{\widetilde{A}_{4,4}(\gamma) - \widetilde{A}^{3,1}(\gamma)(\widetilde{A}^{1,1})^{-1}\widetilde{A}^{1,3}(\gamma)\} = \frac{\sigma_*^2\prod_{i=0}^m(\gamma-i)^2}{(2\gamma+1)\prod_{i=0}^m(\gamma+i+1)^2}.$$

Also note that for each γ,

$$\frac{1}{\sqrt{n}}(S(\gamma)'MU) = \frac{1}{\sqrt{n}}\sum u_t s_{n,t}^{\gamma} - \widetilde{A}^{3,1}(\gamma)(\widetilde{A}^{1,1})^{-1}\frac{1}{\sqrt{n}}\sum u_t z_{n,t} + o_{\mathbb{P}}(1),$$

so that, if we let $\widetilde{\mathcal{G}}(\cdot)$ be the weak limit of $n^{-1/2}S(\gamma)'MU$, we have

$$\begin{aligned}\mathbb{E}[\widetilde{\mathcal{G}}(\gamma)\widetilde{\mathcal{G}}(\gamma')] &= \widetilde{B}_{4,4}(\gamma,\gamma') - \widetilde{A}^{3,1}(\gamma)(\widetilde{A}^{1,1})^{-1}\widetilde{B}^{1,3}(\gamma') \\ &\qquad - \widetilde{A}^{3,1}(\gamma')(\widetilde{A}^{1,1})^{-1}\widetilde{B}^{1,3}(\gamma) + \widetilde{A}^{3,1}(\gamma)(\widetilde{A}^{1,1})^{-1}\widetilde{B}^{1,1}(\widetilde{A}^{1,1})^{-1}\widetilde{A}^{1,3}(\gamma') \\ &= \frac{\sigma_*^2\prod_{i=0}^m(\gamma-i)(\gamma'-i)}{(\gamma+\gamma'+1)\prod_{i=0}^m(\gamma+i+1)(\gamma'+i+1)}.\end{aligned}$$

This implies that

$$\mathbb{E}[\widetilde{\mathcal{Z}}(\gamma)\widetilde{\mathcal{Z}}(\gamma')] = \frac{\{\prod_{i=0}^m(\gamma-i)(\gamma'-1)\}(2\gamma+1)^{1/2}(2\gamma'+1)^{1/2}}{\{\prod_{i=0}^m|\gamma-i|\cdot|\gamma'-i|\}(\gamma+\gamma'+1)} = c_m(\gamma,\gamma')\frac{(2\gamma+1)^{1/2}(2\gamma'+1)^{1/2}}{(\gamma+\gamma'+1)}$$

by the definition of $c_m(\gamma,\gamma') := \prod_{i=0}^m(\gamma-i)(\gamma'-i)/|\prod_{i=0}^m(\gamma-i)(\gamma'-i)|$, as desired. ■

Proof of Theorem 5: **Part (i):** Given that $m_0 > m$, if we define $G(m_0) := \sum_{j=m+1}^{m_0} \alpha_{j*}[1^j, 2^j, \ldots, t^j, \ldots, (n-1)^j, n^j]'$, then

$$\widehat{\sigma}_{n,0}^2 - \widehat{\sigma}_{n,A}^2 = \sup_{\gamma \in \Gamma} \frac{\{n^{-1}(U + G(m_0))' MS(\gamma)\}^2}{(n^{-1}S(\gamma)' MS(\gamma))}.$$

Here, we note that $\sup_{\gamma} |n^{-1}U'MS(\gamma)| = o_{\mathbb{P}}(1)$. Furthermore, $G(m_0) = O(n^{m_0})$ and $n^{-m_0}G(m_0) = \alpha_{m_0*}S(m_0) + o(1)$, so that $n^{-1}G(m_0)'MS(\gamma) = \alpha_{m_0*}n^{m_0-1}S(m_0)'MS(\gamma) + O_{\mathbb{P}}(n^{m_0-2})$. This implies that $\sup_{\gamma\in\Gamma} |n^{-1-m_0}G(m_0)'MS(\gamma) - \alpha_{m_0*}n^{-1}S(m_0)'MS(\gamma)| = o_{\mathbb{P}}(1)$, so it follows that

$$\widehat{\sigma}_{n,0}^2 - \widehat{\sigma}_{n,A}^2 = \sup_{\gamma \in \Gamma} \alpha_{m_0*}^2 n^{2m_0} \frac{\{n^{-1}S(m_0)'MS(\gamma)\}^2}{(n^{-1}S(\gamma)'MS(\gamma))} + o_{\mathbb{P}}(n^{2m_0}). \tag{24}$$

We next note that $\widehat{\sigma}_{n,0}^2 = n^{-1}(U + G(m_0))'M(U + G(m_0))$. Hence,

$$\widehat{\sigma}_{n,0}^2 = \sigma_*^2 + \alpha_{m_0*}^2 n^{2m_0} n^{-1} S(m_0)'MS(m_0) + o_{\mathbb{P}}(n^{2m_0}). \tag{25}$$

With these results in hand, (24) and (25) imply that

$$\begin{aligned}
\frac{1}{n} QLR_n = \frac{\widehat{\sigma}_{n,0}^2 - \widehat{\sigma}_{n,A}^2}{\widehat{\sigma}_{n,0}^2} &= \sup_{\gamma\in\Gamma} \frac{\{n^{-1}S(\gamma)'MS(m_0)\}^2}{(n^{-1}S(\gamma)'MS(\gamma))(n^{-1}S(m_0)'MS(m_0))} + o_{\mathbb{P}}(1) \\
&= \sup_{\gamma\in\Gamma} \frac{\widetilde{\sigma}^2(\gamma, m_0)}{\{\widetilde{\sigma}^2(\gamma,\gamma)\}^{1/2}\{\widetilde{\sigma}^2(m_0,m_0)\}^{1/2}} + o_{\mathbb{P}}(1),
\end{aligned}$$

by noting that $\widetilde{\sigma}^2(\cdot,\cdot)$ is the almost sure limit of $n^{-1}\widehat{\sigma}_{n,0}^2 S(\cdot)'MS(\cdot)$.

Parts (ii, iii, and iv): In our context, we can let $\sigma_*^2 g(\gamma, \widetilde{\gamma})$ and K of theorem 6 in BCP be $\widetilde{\sigma}(\gamma, \widetilde{\gamma})$ and 1, respectively. The desired results then follow from theorem 6(*ii.a*, *ii.b*, *v*). ■

References

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

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

AKAIKE, H. (1974): "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.

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

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

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

BILLINGSLEY, P. (1995): *Probability and Measure*. New York: John Wiley & Sons.

BOX, G. AND COX, D. (1964): "An Analysis of Transformations," *Journal of the Royal Statistical Society, Series B*, 26, 211–252.

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

CARD, D. (1999): "The Causal Effects of Education on Easrnings," Chapter 30 in *Handbook of Labor Economics*. Eds. O. Ashenfelter and D. Card. Vol 3, Part 3, 1801–1863. Amsterdam: Elsevier B.V.

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

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

CHO, J.,S. ISHIDA, I., AND WHITE, H. (2011): "Revisiting Tests for Neglected Nonlinearity Using Artificial Neural Network," *Neural Computation*, 23, 1133–1186.

CHO, J.S., ISHIDA, I., AND WHITE, H. (2014): "Testing for Neglected Nonlinearity Using Twofold Unidentified Models under the Null and Hexic Expansions," in *Essays on Nonlinear Time Series Econometrics: A Festschrift in Honor of Timo Teräsvirta*, 3–27. Eds. Niels Haldrup, Mika Meitz, and Pentti Saikkonen. Oxford: Oxford University Press.

CHO, J.S. AND WHITE, H. (2010): "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 157 458–480.

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

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

DING, Z., GRANGER, C.W.J., AND ENGLE, R.F. (1993): "Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

DOUKHAN, P., MASSART, P., AND RIO, E. (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincaré, Probabilites et Statistiques*, 31 393–427.

DUAN, J. C. (1997). "Augmented GARCH(p,q) Process and Its Diffusion Limit," *Journal of Econometrics*, 79, 97–127

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

HECKMAN, J., LOCHNER, L., AND TODD, P. (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," Chapter 7 in *Handbook of the Economics of Education*. Eds. Eric A. Hanushek and Finis Welch. Vol 1, 309–458. Amsterdam: Elsevier B.V.

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

LEMIEUX, T. (2006): "The"Mincer Equation" Thirty Years After *Schooling, Experience, and Earnings*," in *Jacob Mincer: A Pioneer of Modern Labor Economics*. Ed. Shoshana Grossbard. 127–145. New York: Springer.

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

MINCER, J. (1974): *Schooling, Experience and Earnings*. New York, NY: Columbia University Press.

MURPHY, K. AND WELCH, F. (1990): "Empirical Age-Earnings Profiles," *Journal of Labor Economics*, 8, 202–229.

PHILLIPS, P.C.B. (2007): "Regression with Slowly Varying Regressors and Nonlinear Trends," *Econometric Theory*, 23, 557–614.

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

RAMSEY, J.B. (1969): "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of Royal Statistical Society*, *Series B*, 31, 350–371.

STINCHCOMBE, M. AND WHITE, H. (1998): "Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative," *Econometric Theory*, 14, 295–324.

TUKEY, J. (1957): "On the Comparative Anatomy of Transformations," *Annals of Mathematical Statistics*, 28, 602–632.

TUKEY, J. (1977): *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

WHITE, H. AND CHO, J.S. (2012): "Higher-Order Approximations for Testing Neglected Nonlinearity," *Neural Computation*, 24, 273–287.

WU, C.-F. (1981): "Asymptotic Theory of Nonlinear Least Squares Estimation," *Annals of Statistics*, 9, 501–513.

Levels \ Γ	$[-0.20, 1.50]$	$[-0.10, 1.50]$	$[0.00, 1.50]$	$[0.10, 1.50]$
10%	3.7336	3.5869	3.4772	3.4003
5%	5.0114	4.8423	4.7283	4.6434
1%	8.0323	7.8151	7.7430	7.6375
Levels \ Γ	$[-0.20, 2.50]$	$[-0.10, 2.50]$	$[0.00, 2.50]$	$[0.10, 2.50]$
10%	3.8966	3.7750	3.6651	3.5822
5%	5.1831	5.0589	4.9339	4.8459
1%	8.2617	8.1332	7.9663	7.8625
Levels \ Γ	$[-0.20, 3.50]$	$[-0.10, 3.50]$	$[0.00, 3.50]$	$[0.10, 3.50]$
10%	4.0125	3.8996	3.8050	3.7358
5%	5.3049	5.1925	5.0956	5.0150
1%	8.3942	8.2808	8.1330	8.0578
Levels \ Γ	$[-0.20, 4.50]$	$[-0.10, 4.50]$	$[0.00, 4.50]$	$[0.10, 4.50]$
10%	4.0975	3.9859	3.8874	3.8128
5%	5.4021	5.2884	5.1750	5.0841
1%	8.5032	8.3619	8.2586	8.1464
Levels \ Γ	$[-0.20, 5.50]$	$[-0.10, 5.50]$	$[0.00, 5.50]$	$[0.10, 5.50]$
10%	4.1702	4.0576	3.9581	3.8978
5%	5.4927	5.3664	5.2487	5.1970
1%	8.5837	8.4411	8.3105	8.2641
Levels \ Γ	$[-0.20, 6.50]$	$[-0.10, 6.50]$	$[0.00, 6.50]$	$[0.10, 6.50]$
10%	4.2150	4.1058	4.0209	3.9663
5%	5.5267	5.4220	5.3256	5.2666
1%	8.6134	8.5069	8.4181	8.3524
Levels \ Γ	$[-0.20, 7.50]$	$[-0.10, 7.50]$	$[0.00, 7.50]$	$[0.10, 7.50]$
10%	4.2587	4.1599	4.0652	4.0051
5%	5.5725	5.4723	5.3720	5.2999
1%	8.6938	8.5761	8.4599	8.3650
Levels \ Γ	$[-0.20, 8.50]$	$[-0.10, 8.50]$	$[0.00, 8.50]$	$[0.10, 8.50]$
10%	4.3033	4.1951	4.1135	4.0538
5%	5.6144	5.5156	5.4253	5.3551
1%	8.7141	8.6312	8.4897	8.4218
Levels \ Γ	$[-0.20, 9.50]$	$[-0.10, 9.50]$	$[0.00, 9.50]$	$[0.10, 9.50]$
10%	4.3351	4.2366	4.1557	4.0880
5%	5.6507	5.5505	5.4726	5.3905
1%	8.7754	8.6351	8.5425	8.4747
Levels \ Γ	$[-0.20, 10.50]$	$[-0.10, 10.50]$	$[0.00, 10.50]$	$[0.10, 10.50]$
10%	4.3652	4.2769	4.1752	4.1244
5%	5.6828	5.5892	5.4841	5.4492
1%	8.8038	8.7053	8.5877	8.5292

Table 1: ASYMPTOTIC CRITICAL VALUES OF THE QLR TEST STATISTIC. This table contains the asymptotic critical values obtained by generating the truncated exponential Gaussian process 1,000,000 times.

$\boldsymbol{\alpha}$	$m \setminus n$	50	100	200	300	400	500	600	700	800	900	1,000
10%	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2^*	89.42	90.90	90.76	91.66	90.46	90.18	91.22	91.06	91.96	91.40	92.02
	3	8.08	7.00	7.02	6.36	7.48	7.54	6.86	7.36	6.24	6.72	6.28
	≥ 4	2.50	2.10	2.22	1.98	2.06	2.28	1.92	1.58	1.80	1.88	1.70
5%	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2^*	94.68	95.20	95.42	95.98	95.32	95.18	95.48	95.80	96.00	95.46	95.64
	3	4.28	3.90	3.82	3.24	3.84	3.98	3.82	3.76	3.46	3.90	3.74
	≥ 4	1.04	0.90	0.76	0.78	0.84	0.84	0.70	0.44	0.54	0.64	0.62
1%	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2^*	98.98	99.08	99.16	99.06	99.22	99.04	99.08	99.08	99.20	99.26	98.92
	3	0.84	0.80	0.70	0.86	0.68	0.80	0.86	0.88	0.70	0.68	0.96
	≥ 4	0.18	0.12	0.14	0.08	0.10	0.16	0.06	0.04	0.10	0.06	0.12

Table 2: ESTIMATED POLYNOMIAL DEGREES BY THE QLR TEST STATISTIC (IN PERCENT). NUMBER OF ITERATIONS: 5,000. This table shows the portion of the estimated polynomial degrees by sequentially applying the QLR test statistic. DGP: $y_t = \alpha_{0*} + \alpha_{1*}t + \alpha_{2*}t^2 + \eta_* d_t + \cos(d_t)v_t$, $d_t := \rho_* d_{t-1} + w_t$, and $(v_t, w_t)' \sim_{iid} N(0, \sigma_*^2 I_2)$ such that $(\alpha_{0*}, \alpha_{1*}, \alpha_{2*}, \eta_*, \sigma_*^2, \rho_*) = (1, 1, 1, 1, 1, 0.5)$. MODEL: $\mathcal{M}'_m := \{\mu_t(\cdot) : \Omega_n \mapsto \mathbb{R} : \mu_t(\alpha_n, \eta, \beta_n, \gamma) := \ s_t(m)'\alpha_n + d_t\eta + \beta_n s_{n,t}^{\gamma}\}$, where $m = 1, 2, 3$, and $\gamma \in \Gamma := [0.0, 3.5]$.

Methods $\setminus$ n	50	100	200	300	400	500	600	700	800	900	1,000
Seqnt. Estmtn.	85.56	90.90	93.44	95.22	95.32	95.74	96.12	96.78	97.26	97.06	97.00
with $\boldsymbol{\alpha}_n = n^{-1/2}$	(85.85)	(90.00)	(92.92)	(94.22)	(95.00)	(95.52)	(95.91)	(96.22)	(96.46)	(96.66)	(96.83)
Seqnt. Estmtn.	94.34	97.28	98.22	98.78	99.02	99.10	99.20	99.28	99.44	99.58	99.32
with $\boldsymbol{\alpha}_n = n^{-3/4}$	(94.68)	(96.83)	(98.11)	(98.61)	(98.88)	(99.05)	(99.17)	(99.26)	(99.33)	(99.39)	(99.43)
Seqnt. Estmtn.	97.78	99.08	99.58	99.64	99.84	99.82	99.82	99.86	99.88	99.92	99.90
with $\boldsymbol{\alpha}_n = n^{-1}$	(98.00)	(99.00)	(99.50)	(99.66)	(99.75)	(99.80)	(99.83)	(99.85)	(99.87)	(99.88)	(99.90)
AIC	81.70	83.12	83.42	84.10	83.32	83.88	83.80	84.24	84.28	83.38	84.46
BIC	93.94	96.18	97.58	98.16	98.34	98.76	98.58	99.06	99.20	99.16	99.06
AICc	85.94	84.86	84.80	84.78	83.74	84.38	84.18	84.52	84.52	83.64	84.70
AIC′	1.94	2.08	2.42	2.76	5.80	81.00	85.60	82.56	82.20	85.50	86.76
BIC′	0.44	0.16	0.18	0.10	0.12	0.94	89.66	92.18	90.76	94.86	95.82
AICc′	1.50	1.60	2.22	2.54	5.44	81.20	85.76	82.72	82.40	85.76	86.98

Table 3: PORTION OF SEQUENTIALLY ESTIMATED POLYNOMIAL DEGREES BY THE QLR TEST STATISTIC (IN PERCENT). NUMBER OF ITERATIONS: 5,000. This table shows the percentages of the correctly estimated polynomial degree by the sequential application of the QLR test statistic and information criteria: $\widehat{P}_n(\boldsymbol{\alpha}_n) \times 100$ and $\widetilde{P}_n \times 100$. Figures in parentheses denote $(1 - \boldsymbol{\alpha}_n) \times 100$. The level of significance $\boldsymbol{\alpha}_n$ is a function of the sample size n that satisfies the conditions in Corollary 2, and $\widehat{P}_n(\boldsymbol{\alpha}_n) := \frac{1}{r}\sum_{i=1}^{r} \mathbb{I}(\widehat{m}_{n,i} = m_*)$, where r is the number of iterations, and $\widehat{m}_{n,i}$ is the sequential estimator of m_* for the i-th simulation. Similarly, $\widetilde{P}_n := \frac{1}{r}\sum_{i=1}^{r} \mathbb{I}(\widetilde{m}_{n,i} = m_*)$, where $\widetilde{m}_{n,i}$ is the information criterion-based estimator of m_* for the i-th simulation. DGP: $y_t = \alpha_{0*} + \alpha_{1*}t + \alpha_{2*}t^2 + \eta_* d_t + \cos(d_t)v_t$, $d_t := \rho_* d_{t-1} + w_t$, and $(v_t, w_t)' \sim_{iid} N(0, \sigma_*^2 I_2)$ such that $(\alpha_{0*}, \alpha_{1*}, \alpha_{2*}, \eta_*, \sigma_*^2, \rho_*) = (1, 1, 1, 1, 1, 0.5)$. MODEL: $\mathcal{M}'_m := \{\mu_t(\cdot) : \Omega_n \mapsto \mathbb{R} : \mu_t(\alpha_n, \eta, \beta_n, \gamma) := \ s_t(m)'\alpha_n + d_t\eta + \beta_n s_{n,t}^{\gamma}\}$, where $m = 1, 2, 3$, and $\gamma \in \Gamma := [0.0, 3.5]$. AIC, BIC, and AICc are applied to $\mathcal{M}'_{0,m} := \{\mu_t(\cdot) : \Omega_n \mapsto \mathbb{R} : \mu_t(\alpha_0, \ldots, \alpha_m, \eta) := s_t(m)'\alpha_n + d_t\eta\}$, and AIC′, BIC′, and AICc′ are applied to $\mathcal{M}'_m$, where $m = 1, 2, 3$.

Null Model 1: $\alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j$											
$m_2 \setminus m_1$	1	2	3	4	5	$m_1 \setminus m_2$	1	2	3	4	5
1	**54.95**	**38.11**	**38.36**	**31.66**	**31.12**	1	**19.31**	1.41	2.28	2.60	2.86
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**		**(0.00)**	(64.00)	(37.80)	(31.40)	(27.80)
2	3.71	4.88	3.46	3.68	3.25	2	3.14	5.91	3.94	1.61	2.04
	(7.00)	(9.40)	(6.20)	(2.20)	(5.20)		(36.00)	(23.00)	(26.00)	(56.60)	(51.20)
3	2.21	3.09	3.71	2.27	2.01	3	12.41	3.55	5.32	5.65	2.23
	(19.20)	(9.80)	(5.40)	(13.80)	(18.60)		(2.60)	(28.00)	(12.80)	(12.40)	(41.40)
4	2.38	3.44	4.64	3.43	2.51	4	0.64	0.13	0.07	0.02	0.01
	(20.40)	(5.40)	(2.80)	(7.60)	(16.40)		(65.40)	(95.00)	(97.60)	(99.60)	(100.0)
5	1.53	1.30	2.08	7.92	1.13	5	0.31	0.22	0.05	0.21	0.02
	(33.40)	(26.40)	(19.80)	(46.40)	(36.80)		(86.00)	(91.80)	(99.00)	(89.60)	(99.40)

Null Model 2: $\alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + \eta_{1*} b_t + \eta_{2*} m76_t$											
$m_2 \setminus m_1$	1	2	3	4	5	$m_1 \setminus m_2$	1	2	3	4	5
1	**51.27**	**43.96**	**43.80**	**38.70**	**38.13**	1	7.72	0.84	0.08	0.02	0.06
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**		(1.00)	(70.20)	(98.80)	(99.80)	(99.00)
2	**5.90**	**5.41**	**5.23**	**5.40**	**5.09**	2	0.37	0.63	0.10	0.13	0.36
	(0.80)	**(0.60)**	**(0.80)**	**(0.40)**	**(0.40)**		(89.20)	(79.80)	(98.40)	(96.40)	(90.60)
3	**4.86**	5.20	5.18	4.03	3.84	3	7.06	1.18	2.54	2.13	1.11
	(0.60)	(1.20)	(1.40)	(2.80)	(2.20)		(3.80)	(58.00)	(23.80)	(34.80)	(55.80)
4	5.43	4.53	5.53	4.06	4.48	4	0.25	0.06	0.01	0.01	0.00
	(1.00)	(4.00)	(2.00)	(4.60)	(3.00)		(85.60)	(99.00)	(99.80)	(99.80)	(100.0)
5	2.35	1.77	1.89	2.28	1.63	5	0.12	0.07	0.02	0.17	0.05
	(12.40)	(20.00)	(17.60)	(12.20)	(20.40)		(95.00)	(98.80)	(99.60)	(90.40)	(98.00)

Null Model 3: $\alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + \eta_{1*} b_t + \eta_{2*} m76_t + \eta_{3*} m66_t + \sum_{j=4}^{10} \eta_{j*} r_{j,t}$											
$m_2 \setminus m_1$	1	2	3	4	5	$m_1 \setminus m_2$	1	2	3	4	5
1	**54.16**	**46.16**	**45.97**	**41.09**	**40.42**	1	**8.48**	0.84	0.06	0.01	0.06
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**		**(0.80)**	(73.20)	(99.20)	(100.0)	(99.40)
2	**6.52**	**6.14**	**5.80**	**5.68**	**5.61**	2	0.81	0.87	0.11	0.11	0.26
	(0.60)	**(0.60)**	**(0.80)**	**(0.80)**	**(0.40)**		(73.40)	(70.00)	(96.40)	(97.20)	(91.60)
3	5.02	4.93	5.13	3.75	2.66	3	7.31	1.11	2.27	1.68	0.66
	(1.20)	(2.20)	(1.40)	(2.20)	(7.80)		(4.80)	(56.40)	(31.20)	(39.00)	(74.80)
4	4.46	**5.51**	5.83	3.27	3.83	4	0.43	0.08	0.00	0.01	0.00
	(2.20)	**(0.80)**	(1.60)	(6.00)	(5.20)		(71.40)	(96.80)	(100.0)	(99.60)	(100.0)
5	2.18	1.96	1.92	1.74	1.38	5	0.11	0.21	0.02	0.40	0.04
	(13.00)	(18.80)	(18.40)	(17.40)	(24.40)		(96.40)	(91.40)	(99.60)	(81.20)	(98.80)

Table 4: INFERENCES OF THE MINCER EQUATION USING ALL OBSERVATIONS. This table shows the QLR test statistic and its p-values that are obtained by the data set in Card (1995). The sample size is 3,010. Figures are the QLR test statistics, and figures in parentheses are the p-values of the QLR tests measured in percent that are computed by the weighted bootstrap with 500 number of bootstrap iterations. The left- and right-side panels test for neglected polynomial degrees with respect to experiences and schooling years, respectively. Boldface p-values indicate significance levels less than 0.01.

Null Model 1: $\alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j$											
$m_2 \setminus m_1$	1	2	3	4	5	$m_1 \setminus m_2$	1	2	3	4	5
1	**40.00**	**33.34**	**33.31**	**29.12**	**28.75**	1	6.72	0.48	0.08	0.16	0.15
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**		(1.40)	(86.00)	(98.40)	(95.00)	(96.00)
2	1.61	1.13	1.05	0.70	0.59	2	0.42	1.54	0.41	0.31	0.17
	(27.60)	(41.00)	(38.20)	(37.00)	(44.40)		(83.80)	(56.00)	(87.40)	(88.40)	(94.20)
3	3.19	3.34	3.39	1.84	1.94	3	9.61	3.93	5.36	2.30	1.41
	(5.80)	(3.60)	(2.80)	(14.80)	(10.60)		(3.80)	(15.40)	(9.20)	(30.60)	(48.60)
4	2.42	1.53	2.42	1.44	2.35	4	0.07	0.00	0.00	0.00	0.00
	(10.00)	(22.20)	(8.20)	(20.00)	(5.40)		(99.40)	(99.80)	(100.0)	(100.0)	(99.60)
5	2.33	1.45	2.30	1.43	2.26	5	0.01	0.01	0.01	0.02	0.00
	(15.00)	(32.20)	(14.60)	(18.60)	(10.60)		(100.0)	(99.20)	(99.60)	(99.60)	(99.80)

Null Model 2: $\alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + \eta_{1*} m76_t$											
$m_2 \setminus m_1$	1	2	3	4	5	$m_1 \setminus m_2$	1	2	3	4	5
1	**37.02**	**33.05**	**32.69**	**29.06**	**28.88**	1	4.40	0.82	0.29	0.06	0.07
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**		(6.40)	(7.08)	(89.40)	(98.60)	(99.40)
2	1.98	1.33	1.33	1.64	1.54	2	0.73	0.46	0.63	0.77	0.35
	(15.00)	(13.60)	(26.60)	(20.60)	(19.80)		(70.20)	(85.00)	(75.40)	(62.40)	(86.20)
3	**5.47**	1.00	**5.38**	3.89	3.90	3	5.78	3.65	3.91	1.61	1.38
	(0.80)	(27.80)	**(0.80)**	(3.00)	(2.40)		(8.00)	(17.80)	(15.80)	(45.00)	(48.60)
4	3.17	4.44	3.94	3.82	3.89	4	0.02	0.00	0.00	0.01	0.01
	(8.00)	(2.40)	(2.60)	(2.60)	(3.60)		(99.40)	(100.0)	(100.0)	(100.0)	(100.0)
5	2.10	1.75	4.11	1.79	2.55	5	0.00	0.02	0.00	0.04	0.00
	(18.20)	(23.80)	(5.20)	(13.80)	(6.80)		(100.0)	(99.80)	(100.0)	(98.00)	(99.80)

Null Model 3: $\alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + \eta_{1*} m76_t + \eta_{2*} m66_t + \sum_{j=3}^{9} \eta_{j*} r_{j,t}$											
$m_2 \setminus m_1$	1	2	3	4	5	$m_1 \setminus m_2$	1	2	3	4	5
1	**39.82**	**34.88**	**34.66**	**31.06**	**30.89**	1	5.06	0.75	0.12	0.04	0.0
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**		(2.80)	(74.60)	(97.60)	(99.60)	(99.80)
2	2.72	1.83	1.84	2.07	1.42	3	0.39	0.58	0.22	0.58	0.14
	(10.00)	(16.20)	(19.00)	(12.00)	(21.60)		(88.20)	(79.20)	(93.60)	(75.00)	(96.40)
3	**5.57**	0.42	5.61	4.12	4.00	4	3.97	2.03	3.45	0.99	0.98
	(0.40)	(27.20)	(1.20)	(3.60)	(3.40)		(19.00)	(37.40)	(20.40)	(56.80)	(58.80)
4	2.87	2.50	4.45	3.50	4.31	4	0.03	0.00	0.00	0.02	0.01
	(10.60)	(12.40)	(1.00)	(6.00)	(2.60)		(99.20)	(100.0)	(99.80)	(99.60)	(99.40)
5	2.56	2.98	1.55	1.32	1.49	5	0.13	0.00	0.01	0.08	0.00
	(13.20)	(12.00)	(29.20)	(26.00)	(20.60)		(93.00)	(100.0)	(99.60)	(96.40)	(100.0)

Table 5: INFERENCES OF THE MINCER EQUATION USING WHITE YOUNG MEN DATA. This table shows the QLR test statistic and its p-values that are obtained by the data set in Card (1995). The sample size is 2,707. Figures are the QLR test statistics, and figures in parentheses are the p-values of the QLR tests measured in percent that are computed by the weighted bootstrap with 500 number of bootstrap iterations. The left- and right-side panels test for neglected polynomial degrees with respect to experiences and schooling years, respectively. Boldface p-values indicate significance levels less than 0.05.