

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: March, 2017

Abstract

We provide a methodology for testing a polynomial model hypothesis by generalizing the approach and results of Baek, Cho, and Phillips (2015; BCP) which test for neglected nonlinearity using power transforms of regressors against arbitrary nonlinearity. We use the BCP quasi-likelihood ratio test and deal with the new 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 both 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) and Bierens and Ginther (2001). The results confirm that the standard Mincer log earnings equation is readily shown to be misspecified. The applications consider different data sets and examine the impact of nonlinear effects of experience and schooling on earnings, allowing for flexibility in the respective polynomial representations.

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 co-editor, Jonathan Wright, and two anonymous referees provided helpful comments and suggestions for which we are most grateful. We also have 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 and the Kelly Foundation at the University of Auckland.

1 Introduction

Polynomial models are popularly used in empirical work to address departures from linearity, as they are able 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 and a quasi-likelihood ratio (QLR) test implied by this. 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. The approach has a natural application in sieve nonparametric estimation for determining the dimension of a suitable sieve space that is typically estimated by information criteria. In fact, as we discuss below, our methodology does relate to and complement the use of information criteria. Specifically, the QLR statistic is designed to detect any information gain by marginally increasing the dimension of the model, thereby providing a mechanism of evaluating the success of any particular specification, just as in the use of information criteria.

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. Many applications were reported in the earlier paper BCP. In related work 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, some aspects of which are relevant in the current paper and are revisited here.

The approach pursued here extends the linear null model framework of BCP to a more general poly-

mial 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 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. 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 under the null, 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 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 inevitably raises questions concerning the appropriate choice of nonlinear model or polynomial degree in the specification. There are many other instances in economic and financial research where polynomial models can be exploited to resolve empirical issues

concerning uncertainty about the degree of nonlinearity in relationships suggested by theory or by empirical work. Some of these further examples are reviewed in Section 5. In addition to these empirical examples, there has been substantial interest in the use of sieve approximations in nonparametric econometrics, wherein a primary task is the selection of a suitable degree for the approximating sieve space in empirical application.

The second degree polynomial model of the Mincer equation provides a natural platform to apply the testing methodology developed in the current study, particularly given the fact that its form has been questioned by Murphy and Welch (1990) and Lemieux (2006) among others. Accordingly, we apply our methodology 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) and the current population survey data from Bierens and Ginther (2001). Revisiting this application, we conclude that the standard Mincer equation fails to capture nonlinearity in response to years of experience if the model is extended to include other explanatory variables and/or different data are used. Our tests therefore provide confirmatory analytic findings that support Lemieux's (2006) conclusions concerning suitable model specifications for earnings equations.

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 also examines 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. Concluding remarks are given in Section 6. Proofs are contained in the online supplement to this study (Cho and Phillips, 2017). For notational simplicity we use $(d^j/dx^j)f(0)$ 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 $z_t := (x_t(m)', d_t')' := (1, x_t, \dots, x_t^m, d_t')'$ that involve m -th degree polynomial components of some process x_t . The model is formulated as $\mathcal{M}_m := \{\mathbb{E}[y_t|z_t] = \mu_t(\cdot) : \Omega \mapsto \mathbb{R} \text{ with } \mu_t(\alpha, \eta, \beta, \gamma) := x_t(m)'\alpha + d_t'\eta + \beta x_t^\gamma\}$, 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. It is this mechanism that enables an omnibus test of specification to be developed.

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 $\omega := (\alpha', \eta', \beta, \gamma)' := (\alpha_0, \dots, \alpha_m, \eta', \beta, \gamma)'$ 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 $Z = [z_1, \dots, z_n]'$ 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 formulated to facilitate testing the following hypothesis:

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

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 $\{0, 1, \dots, m\}$, there are $(m+2)$ different ways to obtain the null model from \mathcal{M}_m . First, for each $c = 0, 1, \dots, m$, if $\gamma_* = c$, the coefficient of x_t^c becomes $(\alpha_{c*} + \beta_*)$, thereby leading to the null model. Nevertheless, α_{c*} 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:

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

The union of the parameter spaces under the sub-hypotheses must be the null parameter space under $\mathcal{H}_{0,m}$.

In the following subsections, we examine the limit distribution of the QLR test statistic defined as $QLR_n := n(1 - \hat{\sigma}_{n,A}^2/\hat{\sigma}_{n,0}^2)$ under each null hypothesis as in BCP. Here, $\hat{\sigma}_{n,A}^2$ and $\hat{\sigma}_{n,0}^2$ 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 $\hat{\sigma}_{n,A}^2 := -n^{-1} \max_{\alpha, \eta, \beta, \gamma} L_n(\alpha, \eta, \beta, \gamma)$ and $\hat{\sigma}_{n,0}^2 := -n^{-1} \max_{\alpha, \eta, \beta, \gamma} L_n(\alpha, \eta, 0, \gamma)$, where γ in the latter is simply a placeholder whose value is irrelevant under the null.

The null limit distribution of the QLR test statistic is different under each sub-hypothesis. Therefore, the limit distribution of QLR_n under $\mathcal{H}_{0,m}$ has to be obtained by relating it to the limit distribution of the QLR_n under each hypothesis. For this purpose, we clarify our notation. For each $c = 0, 1, \dots, m$, we let $QLR_n^{(\gamma=c)}$ be the QLR test statistic obtained by imposing $\mathcal{H}_{0,m}^{(c+1)}$. Likewise, we let $QLR_n^{(\beta=0)}$ be the QLR test statistic obtained by impose $\mathcal{H}_{0,m}^{(m+2)}$. Given the definition of the QLR test statistic, it has to follow that

$$QLR_n = \max[QLR_n^{(\gamma=0)}, QLR_n^{(\gamma=1)}, \dots, QLR_n^{(\gamma=m)}, QLR_n^{(\beta=0)}], \quad (1)$$

where $\max[\cdot]$ is used to accommodate maximization across the component sub-hypotheses of the null parameter space $\mathcal{H}_{0,m}$. We examine the limit distribution of each QLR test statistic under each hypothesis in Sections 2.2 and 2.3 and their relationship in Section 2.4 so that the null limit distribution required for significance testing can be obtained.

The following conditions are assumed throughout this section to fix ideas and develop an asymptotic theory of inference, and they also generalize those in BCP.

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 coefficients 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 $0, 1, \dots, m$ being interior elements of Γ , and for $i = 0, 1, \dots, m$, \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 $G_t(\gamma) := [z_t', \log(x_t)x_t(m)', x_t^\gamma]'$, 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 $\{z_{t+1}, u_t, z_t, u_{t-1}, \dots\}$; (iii) There is a strictly stationary and ergodic sequence $\{m_t, s_t\}$ such that for $i = 1, 2, \dots, k$, $|d_{t,i}| \leq m_t$, $\mathbb{E}[m_t^{4\rho}] < \infty$, $\mathbb{E}[s_t^8] < \infty$, 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

Here, $\sum_{t=1}^n u_t G_t(\cdot)$ stands for the score component that determines the limit distribution of the QLR test statistic under $\mathcal{H}_{0,m}^{(m+2)}$ as we discuss in Section 2.3. 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 dependence for 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.

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

We first examine the limit behavior of the QLR test statistic under $\mathcal{H}_{0,m}^{(c+1)} : \gamma_* = c$, where $c = 0, 1, \dots, m$. 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 .

Under $\mathcal{H}_{0,m}^{(c+1)}$ 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.

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 limit distribution under $\mathcal{H}_{0,m}$. The schema is described in full in what follows. Although our analysis is parallel to the development in BCP, it generalizes the BCP approach as they do not consider cases with $m > 1$.

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: $L_n(\gamma; \beta) := L_n(\hat{\alpha}_n(\gamma; \beta), \hat{\eta}_n(\gamma; \beta), \beta, \gamma)$, where $(\hat{\alpha}_n(\gamma; \beta), \hat{\eta}_n(\gamma; \beta))' := \arg \max_{\alpha, \eta} L_n(\alpha, \beta, \gamma, \eta)$. 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 $X(\gamma) := (x_1^\gamma \dots x_n^\gamma)'$. Note that $MY = MU$ under $\mathcal{H}_{0,m}$, where $U := (u_1, u_2, \dots, u_n)'$. For notational simplicity, define $A_c := [x_1^c \log(x_1), \dots, x_n^c \log(x_n)]'$, $B_c := [x_1^c \log^2(x_1), \dots, x_n^c \log^2(x_n)]'$, and apply a

second-order Taylor expansion to obtain

$$\sup_{\gamma} \{L_n(\gamma; \beta) - L_n(c; \beta)\} = \frac{\{\beta A'_c MU\}^2}{\beta^2 A'_c M A_c - \beta B'_c MU} + o_{\mathbb{P}}(1) = \frac{\{A'_c MU\}^2}{A'_c M A_c} + o_{\mathbb{P}}(1),$$

using the fact that $B'_c MU = o_{\mathbb{P}}(n)$ under Assumptions 1 and 2. This result follows mainly from the simple form of the derivatives $L_n^{(1)}(c; \beta) := (d/d\gamma)L_n(c; \beta) = 2\beta A'_c MU$ and $L_n^{(2)}(c; \beta) := (d^2/d\gamma^2)L_n(c; \beta) = 2\beta B'_c MU - 2\beta^2 A'_c M A_c$. We thus obtain

$$QLR_n^{(\gamma=c; \beta)} := \sup_{\beta} \sup_{\gamma} n \left\{ 1 - \frac{L_n(\gamma; \beta)}{L_n(c; \beta)} \right\} = \frac{1}{\hat{\sigma}_{n,0}^2} \frac{\{A'_c MU\}^2}{\{A'_c M A_c\}} + o_{\mathbb{P}}(1). \quad (2)$$

This representation implies that the optimization process with respect to β in (2) is asymptotically innocuous in obtaining the null limit distribution. In (2), the notation $QLR_n^{(\gamma=c; \beta)}$ is used to denote the QLR test statistic that tests $\mathcal{H}_{0,m}^{(c+1)} : \gamma_* = c$ 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 $(\alpha'_{-c^*}, \beta_*, \eta'_*)' := (\alpha_{0^*}, \alpha_{1^*}, \dots, \alpha_{(c-1)^*}, \alpha_{(c+1)^*}, \dots, \alpha_{m^*}, \beta_*, \eta'_*)'$ 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 $L_n(\gamma; \alpha_c) := \max_{\alpha_{-c}, \beta, \eta} L_n(\alpha, \beta, \gamma, \eta)$ and approximate this CQL by a second-order Taylor expansion with respect to γ at c . The null limit behaviors of the first-two derivatives are given in the following result.

Lemma 1. *Given Assumptions 1, 2, and $\mathcal{H}_{0,m}^{(c+1)}$, for each $c = 0, 1, \dots, m$, we have: (i) $L_n^{(1)}(c; \alpha_c) = 2(\alpha_{c^*} - \alpha_c)A'_c MU + o_{\mathbb{P}}(\sqrt{n})$; and (ii) $L_n^{(2)}(c; \alpha_c) = -2(\alpha_{c^*} - \alpha_c)^2 A'_c M A_c + o_{\mathbb{P}}(n)$. \square*

Proofs are given in the online supplement. Lemma 1 and a second-order Taylor expansion lead to the following representation of the statistic

$$QLR_n^{(\gamma=c; \alpha_c)} := \sup_{\alpha_c} \sup_{\gamma} n \left\{ 1 - \frac{L_n(\gamma; \alpha_c)}{L_n(c; \alpha_c)} \right\} = \frac{1}{\hat{\sigma}_{n,0}^2} \frac{\{A'_c MU\}^2}{\{A'_c M A_c\}} + o_{\mathbb{P}}(1). \quad (3)$$

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 (3) with respect to α_c does not affect the null limit distribution. Second, the null approximation given on the right side of (3) is asymptotically identical to the right side of (2), implying that the limit approximation under $\mathcal{H}_{0,m}^{(c+1)}$ is identical irrespective of whether β or α_c 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 $\mathcal{H}_{0,m}^{(c+1)}$: $\gamma_* = c$, where $c = 0, 1, \dots, m$.* \square

Here, $QLR_n^{(\gamma=c)}$ denotes the QLR test statistic that tests $\mathcal{H}_{0,m}^{(c+1)}$: $\gamma_* = c$. BCP obtained this result for the special case $m = 1$. Third, for a different index (say, c') the limit approximation under $\mathcal{H}_{0,m}^{(c'+1)}$ is obtained by simply replacing A_c in Lemma 2 with $A_{c'} := [x_1^{c'} \log(x_1), \dots, x_n^{c'} \log(x_n)]'$. 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. In contrast, the power transform approach simplifies the model approximation because at most a second-order Taylor expansion is needed. This feature explains the advantage of using the power transform in detecting further neglected nonlinearity.

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 $\mathcal{H}_{0,m}^{(m+2)}$, where γ_* is not identified. 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: $L_n(\beta; \gamma) = -\{Y - \beta X(\gamma)\}' M \{Y - \beta X(\gamma)\}$, leading to the following limit approximation:

$$QLR_n^{(\beta=0)} := \sup_{\gamma} \sup_{\beta} n \left\{ 1 - \frac{L_n(\beta; \gamma)}{L_n(0; \gamma)} \right\} = \sup_{\gamma} \frac{1}{\widehat{\sigma}_{n,0}^2} \frac{\{X(\gamma)' MU\}^2}{X(\gamma)' M X(\gamma)}. \quad (4)$$

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

Some remarks are in order to highlight this approximation. Note that the approximation in (4) 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 $n^{-1/2}X(\cdot)'MU$ and $n^{-1}\hat{\sigma}_{n,0}^2X(\cdot)'MX(\cdot)$, respectively, as in BCP. Nevertheless, we further note that for $c = 0, 1, \dots, m$, $\text{plim}_{\gamma \rightarrow c} X(\gamma)'MX(\gamma) = 0$ and $\text{plim}_{\gamma \rightarrow c} X(\gamma)'MU = 0$ because $\lim_{\gamma \rightarrow c} X(\gamma) = [x_1^c, x_2^c, \dots, x_n^c]'$ 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 (4), the probability limit of the ratio in the right side of (4) is indeterminate as γ approaches c . Applying a higher-order l'Hôpital rule, we obtain the following for $c = 0, 1, \dots, m$,

$$\text{plim}_{\gamma \rightarrow c} \frac{1}{\hat{\sigma}_{n,0}^2} \frac{\{X(\gamma)'MU\}^2}{X(\gamma)'MX(\gamma)} = \frac{1}{\hat{\sigma}_{n,0}^2} \frac{\{A'_c MU\}^2}{A'_c M A_c}. \quad (5)$$

Some regularity conditions are needed to justify the limit behavior of this ratio as $n \rightarrow \infty$. We first need conditions for applying the central limit theorem (CLT) and FCLT to $n^{-1/2}[A'_0 MU, A'_1 MU, \dots, A'_m MU]'$ and $n^{-1/2}X(\cdot)'MU$, respectively. In addition, a law of large numbers (LLN) and ULLN need to be applied to $n^{-1}[A'_0 M A_0, A'_1 M A_1, \dots, A'_m M A_m]'$ and $n^{-1}X(\cdot)'MX(\cdot)$, respectively. The conditions in Assumptions 1 and 2 are sufficient for this. It also follows from these that $n^{-1/2} \sum u_t G_t(\cdot)$ and $n^{-1} \sum G_t(\cdot) G_t(\cdot)'$ obey the FCLT and ULLN, respectively because each component of $G_t(\cdot) := [z_t', \log(x_t)x_t(m)', x_t^{(\cdot)}]'$ constitutes $n^{-1/2}[X(\cdot)'MU, A'_0 MU, \dots, A'_m MU]'$ and $n^{-1}[X(\cdot)'MX(\cdot), A'_0 M A_0, \dots, A'_m M A_m]$. Note, for example, that $n^{-1}[X(\cdot)'MX(\cdot), A'_0 M A_0, \dots, A'_m M A_m]$ is the covariance matrix estimator of the prediction error obtained by projecting $[\log(x_t)x_t(m)', x_t^{(\cdot)}]'$ on z_t . As a result, the null limit of the QLR test statistic is 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) $QLR_n^{(\beta=0)} = \sup_{\gamma \in \Gamma} \{X(\gamma)'MU\}^2 / \{\hat{\sigma}_{n,0}^2 X(\gamma)'MX(\gamma)\}$; and (ii) $QLR_n^{(\beta=0)} \Rightarrow \sup_{\gamma \in \Gamma} \mathcal{Z}(\gamma)^2$ as $n \rightarrow \infty$, where $\mathcal{Z}(\cdot)$ is a mean-zero Gaussian process whose covariance kernel for each $\gamma, \gamma' \in \Gamma$ is $\mathbb{E}[\mathcal{Z}(\gamma)\mathcal{Z}(\gamma')] = \mathbb{E}[\mathcal{G}(\gamma)\mathcal{G}(\gamma')] / \{\sigma^2(\gamma, \gamma)\sigma^2(\gamma', \gamma')\}^{1/2}$, with $\sigma^2(\gamma, \gamma') := \sigma_*^2(\mathbb{E}[x_t^{2\gamma}] - \mathbb{E}[x_t^\gamma z_t']\mathbb{E}[z_t z_t']^{-1}\mathbb{E}[z_t x_t^{\gamma'}])$, and $\mathcal{G}(\cdot)$ is a mean-zero Gaussian process with covariance kernel for each $\gamma, \gamma' \in \Gamma$, $\mathbb{E}[\mathcal{G}(\gamma)\mathcal{G}(\gamma')] = \mathbb{E}[u_t^2 x_t^{\gamma+\gamma'}] - \mathbb{E}[u_t^2 x_t^\gamma z_t']\mathbb{E}[z_t z_t']^{-1}\mathbb{E}[z_t x_t^{\gamma'}] - \mathbb{E}[u_t^2 x_t^{\gamma'} z_t']\mathbb{E}[z_t z_t']^{-1}\mathbb{E}[z_t x_t^\gamma] + \mathbb{E}[x_t^\gamma z_t']\mathbb{E}[z_t z_t']^{-1}\mathbb{E}[u_t^2 z_t z_t']\mathbb{E}[z_t z_t']^{-1}\mathbb{E}[z_t x_t^{\gamma'}]$. \square*

Note that $\mathcal{G}(\cdot)$ is the weak limit of $n^{-1/2}X(\cdot)'MU$, and that $\mathbb{E}[\mathcal{G}(\gamma)\mathcal{G}(\gamma')]$ simplifies to $\sigma^2(\gamma, \gamma')$ if u_t is conditionally homoskedastic on z_t . Given Lemma 2, the limit result in Lemma 3(ii) 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. The proof of Lemma 3 is also 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 $c = 0, 1, \dots, m$ although it is smooth elsewhere in Γ . Define $z_n(\cdot) := \{\hat{\sigma}_{n,0}^2 X(\cdot)'MX(\cdot)\}^{-1/2} \{X(\cdot)'MU\}$, which is the sample analog of $\mathcal{Z}(\cdot)$. For each $c = 0, 1, \dots, m$, it is not hard to show that $\text{plim}_{\gamma \uparrow c} z_n(\gamma) = -\text{plim}_{\gamma \downarrow c} z_n(\gamma)$. 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 $\text{plim}_{\gamma \uparrow c} \mathcal{Z}(\gamma)^2 = \text{plim}_{\gamma \downarrow c} \mathcal{Z}(\gamma)^2$. Therefore, if we let $\mathcal{Z}(c)^2$ be defined as $\text{plim}_{\gamma \rightarrow c} \mathcal{Z}(\gamma)^2$, $\mathcal{Z}(\cdot)^2$ 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 to obtain the null limit approximation of the QLR test statistic under $\mathcal{H}_{0,m}$.

By definition of the QLR test statistic and the equality in (1), 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 (2) is identical to the right side of (5), implying that

$$\sup_{\gamma \in \Gamma} \frac{1}{\hat{\sigma}_{n,0}^2} \frac{\{X(\gamma)'MU\}^2}{\{X(\gamma)'MX(\gamma)\}} \geq \max_{c=0,1,\dots,m} \frac{1}{\hat{\sigma}_{n,0}^2} \frac{\{A'_c MU\}^2}{\{A'_c M A_c\}} + o_{\mathbb{P}}(1).$$

This fact implies that for every c , $QLR_n^{(\beta=0)}$ dominates $QLR_n^{(\gamma=c)}$ under $\mathcal{H}_{0,m}$, leading to the following result.

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

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 . Even when u_t exhibits conditional homoskedasticity on z_t , the kernel form is still dependent upon the distribution of z_t . This 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.

One of the referees suggested extending the procedure to a Wald-type test using a heteroskedasticity-consistent covariance matrix. However, direct use of this approach does not produce a test of $\mathcal{H}_{0,m}$ due to the multifold identification problem. In particular, given Theorem 1, one might consider a Wald test using the statistic $\widehat{\beta}_n(\gamma) := \arg \max_{\beta} L_n(\beta; \gamma)$. However, for $c = 0, 1, \dots, m$, $\widehat{\beta}_n(c)$ is not defined due to multicollinearity, leading to a failure in constructing a well defined Wald statistic.

2.5 Sequentially Testing the Polynomial Model

We examine a sequential testing procedure in which we allow the polynomial degree m to increase and apply a sequence of tests until the null is no longer rejected. This procedure provides a natural mechanism for estimating the degree of a polynomial model at some given level of significance α . 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 $P_d(\bar{m}) := \{1, 2, \dots, \bar{m}\}$ 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 and significance level α , 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 $j = 2, 3, \dots, \bar{m}$ using \mathcal{M}_j with the same Γ until $QLR_n^{(j)}$ is greater than the asymptotic critical value in Theorem 1 and the same significance level α . 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 $j = 1, 2, \dots, \bar{m}$, $QLR_n^{(j)}$ exceeds the asymptotic critical values in Theorem 1 and the same significance level α , 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, testing multiple hypotheses using multiple test statistics requires caution to avoid data-snooping problems. Thus, when a lower degree polynomial model is not rejected, the lower degree polynomial model should be selected, which enables estimation of a parsimonious model and avoids errors involved in testing both lower and higher degree

polynomial models. Type-II errors become negligible for large n under test consistency and type-I error control is primarily important, especially when both lower and higher degree polynomial models are rejected by the above testing procedure. We discuss below how to control these errors in practice.

Second, as shown earlier, the QLR test statistic is not distribution free, so that for different $j = 1, 2, \dots, \bar{m}$, different asymptotic critical values apply. Use of Hansen's (1996) weighted bootstrap can provide consistent asymptotic critical values in this case. In Section 5, we apply this procedure in our empirical work.

Third, to elaborate on the details, we can let Γ contain $P_d(\bar{m})$, but choose another parameter space for each j , say Γ_j . In particular, 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 the asymptotic power of the test, we are able to obtain the following result on size control in model selection.

Corollary 1. *If for each $m \in P_d(\bar{m})$, Assumptions 1 and 2 hold and $m_* \in P_d(\bar{m})$, where $m_* := \inf\{m \in \mathbb{N} : \exists (\alpha, \eta), \mathbb{E}[y_t|x_t, d_t] = x_t(m)'\alpha + d_t'\eta\}$, for any $\epsilon > 0$ and α , $\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{m}_n - m_*| > \epsilon) = \alpha$. \square*

Thus, for a given significance level α , the estimated polynomial degree is equal to the unknown polynomial degree with probability $(1 - \alpha)\%$ at the limit. Here, the unknown polynomial degree m_* is defined as the most parsimonious polynomial model out of the correctly specified polynomial models. Corollary 1 implies that data-snooping problems are avoided by the above test procedure, but there is a type-I error: the estimated \hat{m}_n has the limiting (size controlled) probability α that \hat{m}_n differs from m_* .

There is the opportunity for consistent estimation by \hat{m}_n if we control size to depend on n so that $\alpha = \alpha_n \rightarrow 0$ slowly as $n \rightarrow \infty$. The following theorem provides conditions for such consistent estimation of m_* .

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

By Theorem 2, \hat{m}_n consistently estimates m_* . Although the null limit distribution here depends on a stochastic process, we can still obtain the same result as Hosoya (1989) under the conditions given by Theorem 2. 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 online supplement. To explain in brief here, 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 establish that the critical value c'_n for which $\mathbb{P}(\sup_{\gamma \in \Gamma} \mathcal{Z}^0(\gamma)^2 \geq c'_n) = \alpha_n$ is bounded from above by the Slepian inequality. This critical value can be compared with another critical value c_n such that $\mathbb{P}(\sup_{\gamma \in \Gamma} \mathcal{Z}(\gamma)^2 \geq c_n) = \alpha_n$ and we show that the upper bound for c'_n is also an upper bound for c_n . Theorem 2 is proved by associating the upper bound of c_n with the conditions for α_n in Theorem 2 in a manner that if $-\log(\alpha_n)/n \rightarrow 0$ and $\alpha_n \rightarrow 0$, then $c_n/n \rightarrow 0$ and $c_n \rightarrow \infty$. These results are sufficient to ensure the consistency result that $\lim_{n \rightarrow \infty} \mathbb{P}(\hat{m}_n > m_*) = 0$ and $\lim_{n \rightarrow \infty} \mathbb{P}(\hat{m}_n < m_*) = 0$.

The polynomial degree estimation given by the sequential testing methodology is related to order estimation by information criteria. Akaike's (1973, 1974) information criterion (AIC) and the Bayesian information criterion (BIC) are popular methods in practical work and there is a relationship between our approach and these methods. Heuristically, the QLR statistic tests for a decreasing form of information criterion as a function of polynomial degree when $m_* \notin P_d(\bar{m})$. For example, if we let $BIC_{n,j}$ be the BIC level from the j -th degree polynomial model, it follows that $BIC_{n,j+1} - BIC_{n,j} = \log(\hat{\sigma}_{n,j+1}^2/\hat{\sigma}_{n,j}^2) + \log(n)/n$, where $\hat{\sigma}_{n,j}^2$ is the mean of the squared residuals from the j -th degree polynomial model. If we further suppose that $m_* \notin P_d(\bar{m})$, it is desired that for every j , $BIC_{n,j+1} - BIC_{n,j} < 0$ almost surely, so that it can be used successfully as a test basis. Note that it is necessary for this negative difference that for every j , $\lim_{n \rightarrow \infty} \mathbb{P}((1 - \hat{\sigma}_{n,j+1}^2/\hat{\sigma}_{n,j}^2) > \log(n)/n) = 1$ by noting that $(1 - \hat{\sigma}_{n,j+1}^2/\hat{\sigma}_{n,j}^2)$ is the dominant component of $\log(\hat{\sigma}_{n,j+1}^2/\hat{\sigma}_{n,j}^2)$. Nevertheless, if $BIC_{n,j}$ is weakly decreasing with respect to j , this property does not necessarily hold even when $m_* \notin P_d(\bar{m})$ because for some j the almost sure limit of $\hat{\sigma}_{n,j+1}^2/\hat{\sigma}_{n,j}^2$ can be unity. Our sequential testing methodology usefully complements this feature of BIC by noting that $QLR_n^{(j:j+1)} := n(1 - \hat{\sigma}_{n,j+1}^2/\hat{\sigma}_{n,j}^2)$ is another form of QLR statistic; by using $QLR_n^{(j)}$ instead of $QLR_n^{(j:j+1)}$ omnibus power can be achieved and, for every j , $\lim_{n \rightarrow \infty} \mathbb{P}(n^{-1}QLR_n^{(j)} > \log(n)/n) = 1$ if $m_* \notin P_d(\bar{m})$, as indeed is desired in terms of strictly decreasing BIC values with respect to j . This omnibus power is a marginal benefit of our approach that is obtained at a marginal cost requiring the use of critical values that are obtained in a nonstandard way as in Theorem 1. Similar interpretations also apply for other information criteria. In Section 4, we conduct simulations and compare the performances of the sequential testing procedure and the information criteria by supposing that $m_* \leq \bar{m}$.

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. 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 $s_t(m) := [1, s_{n,t}, s_{n,t}^2, \dots, s_{n,t}^m]'$, where for $t = 1, 2, \dots, n$, $s_{n,t} := t/n$ is a (normalized) linear time trend. As before, the hypothesis of interest is

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

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 $t(m) := [1, t, t^2, \dots, t^m]'$. The parameters in \mathcal{M}'_m are related to those in \mathcal{M}''_m through the identities $\alpha_n \equiv \text{diag}[1, n, n^2, \dots, n^m] \alpha$ 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 and BCP).

3.2 Asymptotic Null Distribution of the QLR Test Statistic

We assume the following conditions which generalize those used in BCP.

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 $0, 1, \dots, m$ being interior elements of Γ with $\inf \Gamma > -1/2$; for $i = 0, 1, \dots, m$ 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 $\tilde{A}(\gamma)$ be the almost sure limit of $n^{-1} \sum \tilde{G}_t(\gamma) \tilde{G}_t(\gamma)'$, where $\tilde{G}_t(\gamma) := [s_t(m)', d_t, \log(s_{n,t}) s_t(m)', s_{n,t}^\gamma]'$, which exists under mild moment conditions that are assured by Assumption 4 below. We next define $\tilde{B}(\gamma, \gamma')$ to be

the almost sure limit of $n^{-1} \sum u_t^2 \tilde{G}_t(\gamma) \tilde{G}_t(\gamma)'$, where $u_t := y_t - \mathbb{E}[y_t | d_t]$, which again exists under mild moment and other regularity conditions that are assured by the following assumption.

Assumption 4. (i) For each $\epsilon > 0$, $\tilde{A}(\cdot)$ and $\tilde{B}(\cdot, \cdot)$ 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 $\{d_{t+1}, u_t, d_t, u_{t-1}, \dots\}$; (iii) There is a strictly stationary and ergodic sequence $\{m_t\}$ such that for $i = 1, 2, \dots, k$, $|d_{t,i}| \leq m_t$, $|u_t| \leq m_t$, 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 $\tilde{A}(\cdot)$ and $\tilde{B}(\cdot, \cdot)$ in Assumption 4(i) are constructed for an arbitrary polynomial degree m rather than $m = 1$. Second, some of the components in $\tilde{A}(\cdot)$ and $\tilde{B}(\cdot, \cdot)$ can be explicitly derived because the covariates other than d_t are nonrandom. This aspect conversely implies that $\tilde{A}(\cdot)$ is positive definite uniformly on $\Gamma(\epsilon)$ if and only if the covariance matrix of d_t is positive definite, although m is unspecified. We separately show this property in the online supplement along with the forms of $\tilde{A}(\cdot)$ and $\tilde{B}(\cdot, \cdot)$. 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.

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

Theorem 3. Given Assumptions 3, 4, and $\tilde{\mathcal{H}}_0$, $QLR_n \Rightarrow \sup_{\gamma \in \Gamma} \tilde{\mathcal{Z}}(\gamma)^2$, where $\tilde{\mathcal{Z}}(\cdot)$ is a Gaussian process with covariance kernel for each γ and $\gamma' \in \Gamma$ given by $\mathbb{E}[\tilde{\mathcal{Z}}(\gamma) \tilde{\mathcal{Z}}(\gamma)'] = c_m(\gamma, \gamma')(1 + 2\gamma)^{1/2}(1 + 2\gamma')^{1/2}(1 + \gamma + \gamma')^{-1}$ with $c_m(\gamma, \gamma') := \prod_{i=0}^m (\gamma - i)(\gamma' - i) / \prod_{i=0}^m (\gamma - i)(\gamma' - i)$. \square

The proof of Theorem 3 can proceed along the following lines: the QLR test statistic under $\tilde{\mathcal{H}}_0$ is identical to that obtained under the hypothesis that $\beta_* = 0$ whose null limit distribution is obtained as $\sup_{\gamma \in \Gamma} \tilde{\mathcal{Z}}(\gamma)^2$ as in Section 2. Furthermore, the covariance kernel of $\tilde{\mathcal{Z}}(\cdot)$ is also derived from the sample analog of $\tilde{\mathcal{Z}}(\cdot)$ denoted as $\tilde{z}_n(\cdot) := \{\hat{\sigma}_{n,0}^2 S(\cdot)' M S(\cdot)\}^{-1/2} \{S(\cdot)' M U\}$, where $S(\gamma) := [s_{n,1}^\gamma, s_{n,2}^\gamma, \dots, s_{n,n}^\gamma]'$, $M := I_n - Z(Z'Z)^{-1}Z'$, and $U := [u_1, u_2, \dots, u_n]'$. As the weak limit process is derived in the same way as Theorem 1, we focus on deriving the covariance kernel of $\tilde{\mathcal{Z}}(\cdot)$ in the proof: if we let $\tilde{\mathcal{G}}(\cdot)$ and $\tilde{\sigma}^2(\cdot, \cdot)$ be the weak limit of $n^{-1/2} S(\cdot)' M U$ and the almost sure limit of $n^{-1} \hat{\sigma}_{n,0}^2 S(\cdot)' M S(\cdot)$, respectively, it follows that for each γ and γ' ,

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

where $\sigma_*^2 := \mathbb{E}[u_t^2]$ as before. Using these results, we derive the covariance kernel of $\tilde{\mathcal{Z}}(\cdot)$ in the proof of the Theorem.

The Gaussian process $\tilde{\mathcal{Z}}(\cdot)$ is independent of the joint distribution of $\{d_t, u_t\}$, just as in BCP. In particular, Theorem 3 holds irrespective of whether the error is conditionally heteroskedasticity or homoskedastic, viz., the QLR test is a distribution free test. Its applicability is therefore relatively wide. We call the Gaussian process $\tilde{\mathcal{Z}}(\cdot)$ 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 depends on 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 process in Cho and White (2010) and Cho, Cheong, and White (2011) that is defined by the truncated representation

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

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. We tabulate asymptotic critical values obtained in this way for large ℓ and various assumptions on Γ in the online supplement.

3.3 Sequentially Testing the Polynomial Time-Trend Model

The test procedure can be used sequentially to estimate polynomial degree using the approach in Section 2.5 applied to \mathcal{M}'_j for $j = 1, 2, \dots, \bar{m}$. The results given in Corollary 1 continue to hold: if we let $m_* := \inf\{m \in \mathbb{N} : \exists(\alpha, \eta), \mathbb{E}[y_t|d_t] = s_{n,t}(m)'\alpha + d_t'\eta\}$, for any $\epsilon > 0$ and significance level α , $\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{m}_n - m_*| > \epsilon) = \alpha$. Furthermore, consistent estimation of m_* is achieved if $\alpha = \alpha_n$ tends to zero slowly as $n \rightarrow \infty$.

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 , $\text{cov}(\mathcal{B}^S(\gamma), \mathcal{B}^S(\gamma')) = 1 - |\gamma - \gamma'|^\delta (1 + o(1))$ and $\text{cov}(\mathcal{B}^S(\gamma), \mathcal{B}^S(\gamma')) \leq \text{cov}(\tilde{\mathcal{Z}}(\gamma), \tilde{\mathcal{Z}}(\gamma'))$, (iii) $\lim_{n \rightarrow \infty} \alpha_n = 0$, and (iv) $\lim_{n \rightarrow \infty} \log(\alpha_n)/n = 0$, then for any $\epsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{m}_n - m_*| > \epsilon) = 0$. \square*

The intuition behind Corollary 2 is identical to that of Theorem 2. The only point of difference from Theorem 2 is that we do not have to standardize $\tilde{\mathcal{Z}}(\cdot)$ as its variance is already unity, as given in Theorem 3.

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 errors. Here, the conditional variance of u_t on d_t is arbitrarily selected. As pointed out in Section 3.2, the QLR test statistic is distribution free for the polynomial time-trend model, so that the critical values generated by the truncated exponential Gaussian process can be applied to DGPs with arbitrary forms of conditional heteroskedasticity. The given DGP is used to affirm this property. 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 1 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 underestimates the degree of the unknown polynomial 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 α , the predicted probability of ‘accepting’ the correct polynomial degree is almost $(1 - \alpha)$ 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. Although 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

lower degree polynomial model. This is mainly because the type-II error associated with the lower degree polynomial model becomes negligible as the sample size tends to infinity. On the other hand, the type-I error probability associated with the higher degree polynomial model does not vanish to zero, so that it needs to be controlled as detailed in what follows.

Next, we studied how well the type-I error probability is controlled by the sequential estimation. For this purpose, we used the same design environment and applied Corollary 2 with the significance level α_n determined by the sample size so that $\alpha_n \rightarrow 0$ and $\log(\alpha_n)/n \rightarrow 0$ as n increases. To assess performance, we estimated the empirical probability of \hat{m}_n equaling $m_* = 2$ for each α_n as $\hat{P}_n(\alpha_n) := r^{-1} \sum_{i=1}^r \mathbb{I}(\hat{m}_{n,i} = m_*)$, where $\mathbb{I}(\cdot)$ is the indicator function, r is the total number of iterations, viz., 5,000, and $\hat{m}_{n,i}$ denotes the sequential estimator of m_* for the i -th simulation. For each given α_n , $\hat{P}_n(\alpha_n)$ estimates the probability of $\hat{m}_n = m_*$, so that if \hat{m}_n estimates m_* consistently, $\hat{P}_n(\alpha_n) - (1 - \alpha_n)$ should converge to zero as n tends to ∞ . We examine how $\hat{P}_n(\alpha_n)$ evolves as $n \rightarrow \infty$.

The simulation results are reported in Table 2. We consider three sequences for the level of significance: $\alpha_n = n^{-1}$, $\alpha_n = n^{-3/4}$, and $\alpha_n = n^{-1/2}$. If $\alpha_n = n^{-1}$, the significance level approaches zero quickly, whereas the approach to zero is much slower when $\alpha_n = n^{-1/2}$, and $\alpha_n = n^{-3/4}$ 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 $\hat{P}_n(\alpha_n)$.

The main results of Table 2 can be summarized as follows. First, the distance between $\hat{P}_n(\alpha_n)$ and $(1 - \alpha_n)$ is close to zero for every selection of α_n . This outcome suggests that m_* can be successfully estimated by the sequential estimation procedure. Second, as the sample size increases, the distance between $\hat{P}_n(\alpha_n)$ and $(1 - \alpha_n)$ shows evidence of convergence to zero for every selection of α_n , indicating, as expected, that degree estimation by \hat{m}_n becomes more precise in large samples. Third, the distance between $\hat{P}_n(\alpha_n)$ and $(1 - \alpha_n)$ is relatively small when $\alpha_n = n^{-1}$ and this choice of α_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., AIC, BIC, and small sample-size corrected AIC. These methods are applied to the following models: $\mathcal{M}'_{0,m} := \{\mu_t(\cdot) : \Omega_n \mapsto \mathbb{R} : \mu_t(\alpha_0, \dots, \alpha_m, \eta) := \alpha_0 + \alpha_1 t + \dots + \alpha_m t^m + d_t \eta\}$ 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 \tilde{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 2 shows simulation results based on the information criteria. The performances of the information criteria are measured by $\tilde{P}_n := r^{-1} \sum_{i=1}^r \mathbb{I}(\tilde{m}_{n,i} = m_*)$, where $\tilde{m}_{n,i}$ is the estimator of m_* for the i -th simulation using the information criteria. The results are as follows. First, the performance measure $\tilde{P}_n \times 100$ 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 $\alpha_n = n^{-3/4}$ or n^{-1} , the sequential estimation of the polynomial degree is more often precise than the BIC-based estimator, whereas if $\alpha_n = n^{-1/2}$, 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 α_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 2. 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. But 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 $\alpha_n = n^{-1}$ or $n^{-3/4}$, and the dominance of the sequential procedure now applies even when $n^{-1/2}$.

5 Empirical Applications

Economic hypotheses that involve nonlinear responses are commonly evaluated by using polynomial model specifications in empirical work. For example, Mincer (1958, 1974) provides a theoretical background that specified a quadratic function between log of earnings and potential work experience. Murphy and

Welch (1990) and Lemieux (2006) empirically examine this hypothesis by estimating polynomial models. Another example is the environmental Kuznets curve hypothesis relating the level of pollution to GDP per capita, whose shape is typically inferred by estimating a quadratic function. The inverse-U curve Kuznets hypothesis has been questioned by the more recent re-linking hypothesis that posits instead an N-curve relationship in which pollution levels deteriorate again as GDP per capita increases further (*e.g.*, De Bruyn and Opschoor, 1997). Empirical work on this hypothesis typically estimates cubic models to capture shape. Many other empirical studies estimate polynomial models to assess the implications of various economic theories about shape characteristics. The U-bent hypothesis of life-long welfare and the J-curve hypothesis between trade balance and the real exchange rate are other common examples in the economic literature. In all of these models concerning shape features in the data, important empirical issues arise concerning the selection of polynomial model degree and whether there is neglected nonlinearity in chosen polynomial model specification. Our methodology serves to address these issues.

This section examines one such empirical examples using our methods. 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

$$\log(w_t) = \alpha_{0*} + \eta_* s_t + \alpha_{1*} x_t + \alpha_{2*} x_t^2 + u_t, \quad (6)$$

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 (6) or by modifying the model in (6) into a structural equation.¹ Unless structural interpretations are involved, the unknown parameters are estimated by least squares 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 successful background of empirical work, persistent questions have been raised over the possibility that the earnings equation in (6) is misspecified. Murphy and Welch (1990) empirically examined the usefulness of the functional form in (6) using the current population survey (CPS) data from 1964 to 1987 and concluded that the quadratic functional form in (6) is unacceptable and argued instead for a quartic functional form in the experience variable. Heckman, Lochner, and Todd (2006) and Lemieux (2006), motivated by the same question, both found that recent earnings data do not fit the Mincer equation as well as

¹See Card (1999) for a survey of the empirical literature on Mincer’s equation.

1960's and 1970's earnings data. Lemieux (2006), in particular, recommends a quartic model by graphically showing that most of the quartic model predictions belong to the confidence band on unrestricted profiles that are obtained by estimating an unrestricted earnings equation using dummies of experience and schooling years. He also 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. Finally, Bierens and Ginther (2001) examined the robustness of the model assumptions using least absolute deviation (LAD) estimation and found that even the quartic model is misspecified using 1988 CPS data. With the exception of Heckman, Lochner, and Todd (2006), all these empirical contributions estimate polynomial models.

The need for econometric methods to assist in such empirical modeling exercises with polynomial functions motivates us to revisit recent findings for the Mincer equation. We focus on the following research questions: *(i)* is there any remaining nonlinearity with respect to experience that cannot be detected by the quartic function advocated by Murphy and Welch (1990) and Lemieux (2006)? *(ii)* do different sampling periods of data follow different DGPs? and *(iii)* with recent data, do satisfactory empirical models of earnings require convex function specifications involving schooling years? These questions are addressed to assess the empirical findings by Lemieux (2006) using the methodology of our sequential testing procedure.

In considering these questions, we use the same data sets used in earlier studies and seek to discover whether the empirical results are corroborated using the new methods. Specifically, we use the same data as Card (1995) and Bierens and Ginther (2001). The national longitudinal survey (NLS) data constructed by Card (1995) involve men aged 24–34 years in 1976, so that different cohort effects do not affect estimation of the Mincer equation. The sample size is 3,010. Bierens and Ginther's (2001) CPS data set was drawn from males aged between 18 and 70 with annual income greater than US\$50 in 1992. For our analysis, we select only the samples that satisfy the data conditions in Card (1995), so that different cohort effects do not affect the estimation. The resulting sample size is 8,775. Readers can refer to Card (1995) and Bierens and Ginther (2001) for more information on the data. These two different data sets are intentionally selected to examine whether there is any DGP change between 1976 and 1988.

Our empirical models follow the specifications in Card (1995). We focus on estimating the following model in addition to the conventional Mincer equation in (6)

$$\log(w_t) = \alpha_{0*} + \eta_{1*}s_t + \alpha_{1*}x_t + \alpha_{2*}x_t^2 + \eta_{2*}b_t + sm_t'\eta_{3*} + u_t, \quad (7)$$

where b_t is a dummy variable for black/white, sm_t is a set of dummy variables for residence in the South and in a metropolitan area in the year of 1976. This model is the first Mincer equation model estimated by Card (1995) modified by racial and residential features. 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 regressors, but we focus here on the models in (6) and (7) as the other model estimation results are very similar.²

We apply the QLR test in the following manner. First, we test for further neglected nonlinearity with respect to experience x_t . The parameter space of the power coefficient was set to $[-0.25, 6.50]$, which enables tests up to 6-th degree polynomial models as the null model. Hansen's (1996) weighted bootstrap is applied to our QLR tests to obtain p -values. The bootstrap iteration number is 500. In cases where the p -values obtained this way were close to the level of significance, we increased the bootstrap iterations to 1,000 to obtain more precise p -value estimates. In another extension, we expanded the null models in (6) and (7) to include higher polynomial degrees 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 (6) and (7) are therefore extended as follows:

$$\log(w_t) = \alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + u_t, \quad (8)$$

$$\log(w_t) = \alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + \eta_{1*} b_t + sm_t' \eta_{2*} + u_t, \quad (9)$$

with $m_1, m_2 = 1, 2, 4$, and 6. Here, integer settings of $m_1, m_2 = 3$ and 5 are excluded because theory underlying the log earnings equation does not support for odd-degree polynomials, as pointed out by a referee. The models (8) and (9) are treated as the null specification in our tests. Second, we reverse the roles of schooling years and experience and conduct the same procedures of the first step, testing for further neglected nonlinearity now with respect to schooling years s_t .

The test results using 1976 NLS data are contained in Table 3. The left- and right-side panels report the p -values from testing for further neglected nonlinearity with respect to experience and schooling years, respectively. The p -values with affix “†” indicate that the number of bootstrap iterations is 1,000. The inferential findings 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 results as

²For example, we also conducted the same tests using residential dummies in 1966 as additional regressors, just as in Card (1995), and obtained results identical to those using Model (7).

follows.

The first implication of these specification tests is that all models that are linear in experience are rejected when testing for neglected nonlinearity in experience at the 1% level of significance. Nonlinearity in experience is further corroborated by testing the null models with respect to schooling years. All p -values in the right-side panel of Table 3 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. Moreover, even if schooling years are squared or further higher degree terms in schooling are included, all models are still found to be nonlinear with respect to experience. This feature implies that the nonlinear specifications in schooling years in Lemieux (2006) may arise from omitted nonlinearity in experience, given that different cohort effects could be present in his data and that linearity is assumed with respect to either experience or schooling years but not both.

Second, the results in Table 3 imply that the original Mincer hypothesis is statistically supported by sequential estimation in some cases. For the original Mincer equation, we focus on (8) and sequentially estimate both m_1 and m_2 in the following manner. In the first-left panel reported in Table 3, for given $m_2 = 1$, say, we sequentially tested $m_1 = 1, 2, 4, 6$ at the 1% level of significance. If the null is rejected for all $m_1 = 1, 2, 4, 6$, the next level $m_2 = 2$ is considered, and testing continues with respect to $m_1 = 1, 2, 4, 6$ as before until the hypothesis cannot be rejected. In this process, m_2 is raised through the integers 1, 2, 4, and 6. The first-left panel of Table 3 shows that the resulting estimates are $m_1 = 1$ and $m_2 = 2$, which are the same degrees specified by the Mincer equation. The model found in this way is indicated by the italic font in Table 3. This finding is also consistent with Heckman, Lochner, and Todd's (2006) conclusion that the Mincer equation fits well the 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 beyond schooling years and experience. Model (9) is specified with additional explanatory variables. Our findings using the same sequential testing procedure support use of $m_1 = 1$ and $m_2 = 4$ in Model (9) to eliminate the need for further nonlinearity in schooling years and experience. This finding is consistent with the empirical results on the use of a quartic specification by Murphy and Welch (1990) and Lemieux (2006).

Fourth, the quartic specification has economic implications different from those of the quadratic model. In particular, extreme log earning profiles are better fitted by a quartic function than a quadratic because the higher degree polynomial magnifies the effect of experience on earnings. A scatter plot of the data and regression curves endorses this interpretation. Figure 1(a) shows the scatter diagram between experience and

the concentrated log wage³ together with the quadratic and quartic regression curves . The major differences in the regression curves arise in the left- and right-tails. When experience is low, the quadratic fit tends to overestimate the concentrated log wage with more observations below the quadratic fit. When experience is high, the quadratic fit tends to underestimate the concentrated log wage, with more observations above the quadratic fit. The quartic regression appears to correct this unbalanced feature of the quadratic fit near the data boundaries, implying that most experienced (or unexperienced) workers are more highly (or less well) paid than the earning profiles captured by the quadratic model. This finding from sequential testing affirms the earlier empirical conclusion of Lemieux (2006).

We next examine test results obtained from the 1988 CPS data, which are reported in Table 4. The empirical findings are summarized as follows. First, our specification analysis shows that the most parsimonious polynomial orders selected for every model and data set are $m_1 = 1$ and $m_2 = 4$. If the polynomial degree in experience is less than 4, the null models are rejected, showing that the Mincer equation does not hold for the 1988 CPS data, but that the quartic model advocated by Murphy and Welch (1990) and Lemieux (2006) is supported by our formal testing procedure. In addition, this analysis reveals that the convex functional relationship between log earnings and schooling years found in Lemieux (2006) could arise from nonlinearity in experience, as discussed above.

Second, the respective degrees of polynomial nonlinearity with respect to schooling years and experience in the original Mincer equation are not invariant to the data. Note that the standard Mincer equation was preferred in analyzing the 1976 NLS data in terms of Model (8), whereas this is not the case for the 1988 CPS data. This finding supports the empirical conclusion in Heckman, Lochner, and Todd (2006) that the Mincer equation does not fit recent earnings data well.

Finally, the same economic implications about the impact of experience on earnings are obtained as with the 1976 NLS data. In particular, Figure 1(b) shows the analogous sample plot and fitted regression curves as for Figure 1(a). The same underestimation and overestimation at the boundaries by the quadratic specification is found for the 1988 CPS data, revealing that the quartic curve provides a more balanced fit to these data.

Although we do not report the results here, we conducted the same tests and estimated polynomial degrees for models using all observations in the Bierens and Ginther (2001) data, thereby ignoring different cohort effects and introducing a higher level of heterogeneity. The sample size was 25,631 when only fulltime workers were included. In these regressions, we observed higher levels of nonlinearity to be present

³The concentrated log wage is defined by the prediction errors obtained from regressing log earnings against schooling years, the race dummy (b_t), and dummies for residence (sm_t) (the explanatory variables in Model (9) other than experience) with the setting $m_1 = 1$, which was determined according to the sequential estimation procedure, as discussed.

both for experience and schooling in explaining earnings. The estimated polynomial degrees were $m_2 = 6$ and $m_1 = 4$ with respect to experience and schooling years, results that are consistent with the findings of Bierens and Ginther (2001). The higher degree polynomial function is needed to accommodate the irregular profiles involved in extreme log earnings at the tails, in much the same way as the quartic model compensated for deficiencies in the quadratic fits of tail behavior of log earnings. This additional analysis is reported in the online supplement. These results together with those of the main text together imply that the general conditional mean equation can be different, depending on the data and/or the model employed, indicating the need for some flexibility in treating potential nonlinearity in key explanatory variables. This flexibility is attainable, as we have seen, with polynomial specifications and is possible, more generally, with other sieve space 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 it 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) and Bierens and Ginther (2001) 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.

References

- AKAIKE, H. (1973): "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International 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. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.
- BIERENS, H. AND GINTHER, D. (2001): "Integrated Conditional Moment Testing of Quantile Regression Models," *Empirical Economics*, 26, 307–324.
- CARD, D. (1995): "Using Geographic Variation 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 Earnings," 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 PHILLIPS, P.C.B. (2017): “Online Supplement to “Sequentially Testing Polynomial Model Hypotheses using Power Transforms of Regressors”,” Discussion Paper, School of Economics, Yonsei University.
- 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.
- DE BRUYN, S.M. AND OPSCHOOR, J. (1997): “Developments in the Throughput-Income Relationship: Theoretical and Empirical Observations,” *Ecological Economics*, 20, 255–268.
- 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.

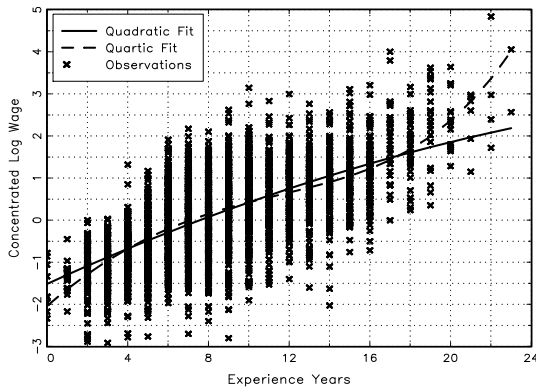
PITERBARG, V. (1996): *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. Translations of Mathematical Monographs, 148. Providence: American Mathematical Society.

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

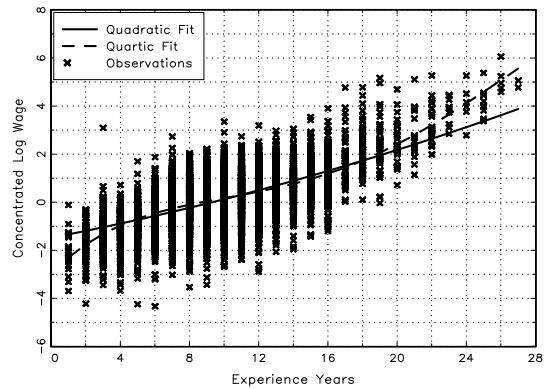
TUKEY, J. (1957): “On the Comparative Anatomy of Transformations,” *Annals of Mathematical Statistics*, 28, 602–632.

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.



(a) 1976 NLS Data



(b) 1988 CPS Data

Figure 1: QUADRATIC, QUARTIC FITS AGAINST EXPERIENCES AND CONCENTRATED LOG WAGES. This figure shows the quadratic and quartic fits of concentrated log wages with respect to experience. The concentrated log wages are the prediction errors obtained by regressing log wages against schooling years and the other dummy variables in (7): b_t and sm_t .

α	$m \setminus n$	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 1: ESTIMATED POLYNOMIAL DEGREES BY THE QLR TEST STATISTIC (IN PERCENT). NUMBER OF REPLICATIONS: 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. with $\alpha_n = n^{-1/2}$	85.56 (85.85)	90.90 (90.00)	93.44 (92.92)	95.22 (94.22)	95.32 (95.00)	95.74 (95.52)	96.12 (95.91)	96.78 (96.22)	97.26 (96.46)	97.06 (96.66)	97.00 (96.83)
Seqnt. Estmtn. with $\alpha_n = n^{-3/4}$	94.34 (94.68)	97.28 (96.83)	98.22 (98.11)	98.78 (98.61)	99.02 (98.88)	99.10 (99.05)	99.20 (99.17)	99.28 (99.26)	99.44 (99.33)	99.58 (99.39)	99.32 (99.43)
Seqnt. Estmtn. with $\alpha_n = n^{-1}$	97.78 (98.00)	99.08 (99.00)	99.58 (99.50)	99.64 (99.66)	99.84 (99.75)	99.82 (99.80)	99.82 (99.83)	99.86 (99.85)	99.88 (99.87)	99.92 (99.88)	99.90 (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	99.18	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 2: PORTION OF SEQUENTIALLY ESTIMATED POLYNOMIAL DEGREES BY THE QLR TEST STATISTIC (IN PERCENT). NUMBER OF REPLICATIONS: 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: $\widehat{P}_n(\alpha_n) \times 100$ and $\widetilde{P}_n \times 100$. Figures in parentheses denote $(1 - \alpha_n) \times 100$. The level of significance α_n is a function of the sample size n that satisfies the conditions in Corollary 2, and $\widehat{P}_n(\alpha_n) := \frac{1}{r} \sum_{i=1}^r \mathbb{I}(\widehat{m}_{n,i} = m_*)$, 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 $\mathcal{M}'_{0,m} := \{\mu_t(\cdot) : \Omega_n \mapsto \mathbb{R} : \mu_t(\alpha_0, \dots, \alpha_m, \eta) := s_t(m)' \alpha_n + d_t \eta\}$, 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	4	6	$m_2 \setminus m_1$	1	2	4	6
1	54.95 (0.00)	38.11 (0.00)	31.86 (0.00)	31.63 (0.00)	1	19.22 (0.00)	3.24 (39.80)	0.52 (67.00)	0.20 (83.60)
2	<i>3.71</i> (<i>9.00</i>)	4.67 (5.00)	3.67 (1.60)	2.45 (5.60)	2	<i>1.41</i> (<i>56.00</i>)	5.67 (23.00)	0.49 (71.40)	0.09 (93.40)
4	3.32 (8.80)	2.95 (15.00)	3.51 (7.00)	2.32 (13.00)	4	2.62 (30.60)	1.59 (57.20)	0.02 (100.0)	0.02 (99.00)
6	0.64 (60.60)	1.16 (49.20)	0.94 (41.00)	1.53 (35.80)	6	2.56 (30.40)	1.97 (50.00)	0.01 (99.80)	0.05 (98.60)

Null Model 2: $\alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + \eta_{1*} b_t + sm_t' \eta_{2*}$									
$m_2 \setminus m_1$	1	2	4	6	$m_2 \setminus m_1$	1	2	4	6
1	51.27 (0.00)	43.96 (0.00)	38.52 (0.00)	38.48 (0.00)	1	7.72 (1.05) [†]	0.61 (77.80)	0.38 (76.00)	0.03 (97.80)
2	5.86 (0.80)	5.41 (0.40)	5.37 (0.60)	4.19 (0.70) [†]	2	0.83 (73.20)	0.73 (77.00)	0.13 (92.00)	0.12 (92.00)
4	<i>3.40</i> (<i>5.60</i>)	5.30 (1.80)	3.19 (4.20)	4.10 (2.00)	4	<i>0.06</i> (<i>99.20</i>)	0.24 (89.80)	0.06 (99.00)	0.02 (99.00)
6	2.81 (9.40)	2.23 (13.60)	2.95 (11.40)	2.81 (5.80)	6	0.07 (97.80)	0.38 (83.60)	0.01 (99.80)	0.02 (98.80)

Table 3: INFERENCES FOR 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 experience and schooling years, respectively. Boldface p -values indicate significance levels less than 1%, and “[†]” indicates the p -values were obtained by 1,000 number of bootstrap iterations.

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	4	6	$m_2 \setminus m_1$	1	2	4	6
1	35.01 (0.00)	38.64 (0.00)	35.26 (0.00)	36.21 (0.00)	1	5.70 (4.60) [†]	11.98 (0.00)	13.60 (0.00)	0.58 (58.20)
2	31.15 (0.00)	26.82 (0.00)	17.01 (0.00)	14.11 (0.00)	2	6.96 (2.20) [†]	13.27 (0.00)	5.93 (3.40) [†]	0.24 (76.40)
4	<i>0.42</i> (<i>91.80</i>)	1.59 (54.00)	0.27 (95.20)	0.25 (91.20)	4	<i>0.65</i> (<i>60.80</i>)	3.62 (12.40)	2.61 (18.40)	0.23 (81.40)
6	0.00 (99.80)	0.00 (99.80)	0.00 (100.0)	0.00 (100.0)	4	0.27 (79.40)	2.72 (16.80)	2.79 (17.20)	0.12 (89.60)

Null Model 2: $\alpha_{0*} + \sum_{j=1}^{m_1} \beta_{j*} s_t^j + \sum_{j=1}^{m_2} \alpha_{j*} x_t^j + \eta_{1*} b_t + sm_t' \eta_{2*}$									
$m_2 \setminus m_1$	1	2	4	6	$m_2 \setminus m_1$	1	2	4	6
1	44.26 (0.00)	35.74 (0.00)	32.23 (0.00)	32.23 (0.00)	1	13.72 (0.00)	9.75 (0.30) [†]	14.20 (0.00)	0.20 (83.20)
2	22.86 (0.00)	24.45 (0.00)	16.19 (0.00)	12.96 (0.20) [†]	2	1.87 (25.40)	10.37 (0.10) [†]	5.01 (7.20)	0.43 (68.00)
4	<i>2.29</i> (<i>40.40</i>)	0.37 (89.60)	0.23 (96.80)	0.26 (90.80)	4	<i>3.15</i> (<i>10.60</i>)	2.43 (19.80)	6.63 (2.70) [†]	0.18 (86.20)
6	0.00 (100.0)	0.00 (100.0)	0.00 (100.0)	0.01 (100.0)	4	2.02 (24.80)	1.94 (25.60)	7.13 (1.60) [†]	0.26 (78.80)

Table 4: INFERENCES FOR 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 Bierens and Ginther (2001). The sample size is 8,775. Refer to Table 3 for more information.