TESTING LINEARITY
USING POWER TRANSFORMS OF REGRESSORS

by

Yae In Baek, Jin Seo Cho, and Peter C. B. Phillips

COWLES FOUNDATION PAPER NO. 1483

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

2015

http://cowles.yale.edu/

Journal of Econometrics 187 (2015) 376–384

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing linearity using power transforms of regressors☆

Yae In Baek [a], Jin Seo Cho [b,*], Peter C.B. Phillips [c,d,e,f]

[a] Department of Economics, University of California, San Diego, 9500 Gilman Dr. La Jolla, CA, 92093, USA
[b] School of Economics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 120-749, Republic of Korea
[c] Yale University, USA
[d] University of Auckland, New Zealand
[e] Singapore Management University, Singapore
[f] University of Southampton, United Kingdom

ARTICLE INFO

Article history:
Received 15 April 2014
Received in revised form
18 March 2015
Accepted 18 March 2015
Available online 1 April 2015

JEL classification:
C12
C18
C46
C52

Keywords:
Box–Cox transform
Gaussian stochastic process
Neglected nonlinearity
Power transformation
Quasi-likelihood ratio test
Trend exponent
Trifold identification problem

ABSTRACT

We develop a method of testing linearity using power transforms of regressors, allowing for stationary processes and time trends. The linear model is a simplifying hypothesis that derives from the power transform model in three different ways, each producing its own identification problem. We call this modeling difficulty the *trifold identification problem* and show that it may be overcome using a test based on the quasi-likelihood ratio (QLR) statistic. More specifically, the QLR statistic may be approximated under each identification problem and the separate null approximations may be combined to produce a composite approximation that embodies the linear model hypothesis. The limit theory for the QLR test statistic depends on a Gaussian stochastic process. In the important special case of a linear time trend regressor and martingale difference errors asymptotic critical values of the test are provided. Test power is analyzed and an empirical application to crop-yield distributions is provided. The paper also considers generalizations of the Box–Cox transformation, which are associated with the QLR test statistic.

1. Introduction

Linear models are a natural starting point in empirical work. They also relate in a fundamental way to underlying Gaussian assumptions and the use of wide sense conditional expectations. Testing linearity is therefore a familiar practice in applications whenever there is concern over specification and Gaussianity. Such tests fall within the framework of general model specification tests.

Power transformations are especially popular as alternatives to linearity. Tukey (1957, 1977) provides several rationales for the use of power transformations, and Box and Cox (1964) further developed their use in nonlinear modeling. The Box–Cox transformation, in particular, successfully implements the so-called Tukey 'ladder of power' option. In time series applications, some studies (notably, Wu (1981) and Phillips (2007)) considered power transforms of a time trend, providing limit theories that are useful in estimation and inference concerning the relevant parameters.

Power transformations can be used to form tests that deliver consistent power against arbitrary alternatives to linearity. As Stinchcombe and White (1998) showed, any non-polynomial analytic function can be used to construct generically comprehensively revealing (GCR) tests, in the sense that linear projection errors are not necessarily orthogonal to any power transform when the linear model is misspecified. This property motivates use

☆ The co-editor, Jianqing Fan, Associated editor, and two anonymous referees provided helpful comments for which we are grateful. We have also benefited from discussions with Shun-ichiro Bessho, Seonghoon Cho, Stan Hurn, Isao Ishida, Tae-Hwan Kim, Shandre Thangavelu, Byungsam Yoo, Valentin Zelenyuk, and other participants at the NZESG meeting (Auckland, 2013) and the 8th Joint Economics Symposium of Four Leading East Asian Universities (Shanghai, 2014). Baek and Cho acknowledge research support from the second stage BK21 project and research grant (NRF-2010-332-B00025) of the National Research Foundation of Korea, respectively, and Phillips acknowledges support from a Kelly Fellowship and the National Science Foundation of USA under Grant No. SES 12-58258.

* Corresponding author.
E-mail address: jinseocho@yonsei.ac.kr (J.S. Cho).

http://dx.doi.org/10.1016/j.jeconom.2015.03.041
0304-4076/

of power transforms for constructing tests with omnibus power. In spite of this apparently useful property, testing linearity using power transforms is largely undeveloped in the literature, mainly because of the identification problem that arises under the null of linearity. As detailed below, the linear model hypothesis can be deduced from a power transformation in three different ways, each of which involves its own identification problem, a feature that we call the *trifold identification problem*. To our knowledge, this problem has never before been addressed in the literature.

Our primary goal in the present paper is to resolve this complex trifold problem. Our focus is pragmatic and involves constructing mechanisms needed in using power transformations. We focus on models involving power transforms of a strictly stationary (SS) variable or a time trend. While this excludes some possibilities, such as nonlinear transforms of nonstationary variates (e.g. Park and Phillips (1999), and Shi and Phillips (2012)), the range of potential applications is large and includes both microeconometric and time series data.

This paper restricts attention to a particular statistic, the quasi-likelihood ratio (QLR) statistic. As we demonstrate, the QLR statistic may produce a composite form that embodies the linear model hypothesis. An additional benefit from focusing on the QLR test is its relationship to the Box–Cox transformation. The score of the test turns out to be related to an augmented form of Box–Cox transform. Our approach to developing a null approximation of the QLR test extends the methodology of Cho and Ishida (2012), who studied how to test the effects of omitted power transformations. We advance that work and compare our null approximation with the QLR tests that are popular in the artificial neural network (ANN) literature where there is at most a twofold identification problem. Our approach also exploits the properties of time-trend power transforms and regressions studied recently in Phillips (2007). Time trend regressors and their power transforms have very different properties from those of stationary regressors in view of the asymptotic degeneracy of the signal matrix.

The paper is organized as follows. Section 2 examines power transformations of a stationary process and tests linearity. The null approximation and the power properties of the QLR test are developed. Section 3 extends the discussion and asymptotic results to power transforms of a time-trend regressor. Simulations and empirical applications are contained in Sections 4 and 5, respectively. Concluding remarks are given in Section 6. All proofs are collected in an Appendix to the paper which is available as an online supplement (Baek et al., 2014).

2. Testing for neglected power transforms of a stationary regressor

We seek to model the conditional mean $\mathbb{E}[Y_t|\mathbf{W}_t]$ of a dependent variable Y_t given a collection of explanatory variables $\mathbf{W}_t$. We define the class of (parameter dependent) conditional mean functions as $m_t(\boldsymbol{\omega}) := \alpha + \mathbf{W}_t'\boldsymbol{\delta} + \beta X_t^{\gamma} = \mathbb{E}[Y_t|\mathbf{W}_t]$, where the parameter vector $\boldsymbol{\omega} := (\alpha, \boldsymbol{\delta}', \beta, \gamma)' \in \boldsymbol{\Omega} \subset \mathbb{R}^{k+4}$, with $\boldsymbol{\delta} \in \mathbb{R}^{k+1}$ for some $k \in \mathbb{N}$. In this specification, the variables $(Y_t, \mathbf{W}_t)$ comprise a strictly stationary and absolutely regular mixing process, the variable X_t is positively valued, and $\boldsymbol{\Omega}$ is the parameter space of $\boldsymbol{\omega}$. In addition to appearing nonlinearly as X_t^{γ}, the variable X_t also enters linearly in $m_t(\boldsymbol{\omega})$ so that X_t is the first element of $\mathbf{W}_t$. Then $\mathbf{W}_t = \left(X_t, \mathbf{D}_t'\right)'$ for some $\mathbf{D}_t \in \mathbb{R}^k$. Similarly, we partition the parameter vector $\boldsymbol{\delta} := (\xi, \boldsymbol{\eta}')'$, so that $\mathbf{W}_t\boldsymbol{\delta} = \xi X_t + \mathbf{D}_t'\boldsymbol{\eta}$. In Section 3, X_t is a linear time trend and so the conditional mean function includes both a linear and nonlinear (power function) trend.

Our interest is primarily in testing the effective form of X_t in the conditional mean $\mathbb{E}[Y_t|\mathbf{W}_t]$. We consider the following explicit hypotheses. Given that $\mathbb{E}[Y_t|\mathbf{W}_t]$ is linear with respect to the components $(1, \mathbf{W}_t)$, we focus on the null hypothesis $\mathcal{H}_0$: $\exists(\alpha_*, \boldsymbol{\delta}_*)$, $\mathbb{E}[Y_t|\mathbf{W}_t] = \alpha_* + \mathbf{W}_t'\boldsymbol{\delta}_*$ w.p. 1 and the alternative hypothesis $\mathcal{H}_1$: $\forall(\alpha, \boldsymbol{\delta})$, $\mathbb{E}[Y_t|\mathbf{W}_t] = \alpha + \mathbf{W}_t'\boldsymbol{\delta}$ w.p. < 1, which implies that nonlinear elements of X_t appear in the conditional mean that cannot be embodied in $\mathcal{H}_0$. The affix '$*$' is used to parameterize $\mathbb{E}[Y_t|\mathbf{W}_t]$, so that for some α_o and β_o, $(\alpha_*, \beta_*, \gamma_*) \in \{(\alpha, \beta, \gamma) : \alpha + \beta X_t^{\gamma} = \alpha_o \text{ or } \alpha + \beta X_t^{\gamma} = \beta_o X_t\}$ under $\mathcal{H}_0$.

Testing the linear model hypothesis using a maintained model with a nonlinear component is common practice in the literature. Such tests may be regarded as a variant of the Bierens (1990) test. Similarly, Stinchcombe and White's (1998) GCR tests are constructed to test for a nonlinear component. A power transform representation is particularly popular for the nonlinear component. For example, Tukey (1957, 1977) introduced power transform flexible nonlinear models, and Box and Cox (1964) found that their transformation accords with Tukey's (1957) 'ladder of power' and it has been widely applied in empirical work (e.g. Sakia (1992)). The GCR property is delivered by non-polynomial analytic functions that can approximate arbitrary functions by Taylor expansion, so that for some γ_*, $\mathbb{E}[V_t X_t^{\gamma_*}] \neq 0$ in a misspecified linear model, where V_t denotes the linear projection error. This property motivates the construction of power transforms to test linearity. The literature already has related variations of power transforms such as those used in Ramsey's (1969) test which have prefixed power exponents. The general power transforms used here do not fix power exponents, and this flexibility is used to gain powers in testing, as detailed below.

Notwithstanding considerable interest in power transforms, $\mathcal{H}_0$ has not been formally examined in the literature mainly because testing $\mathcal{H}_0$ cannot be conducted in a standard way. There are three different identification problems that arise under $\mathcal{H}_0$. If $\beta_* = 0$, γ_* is not identified and Davies' (1977,1987) identification problem arises. On the other hand, if $\gamma_* = 0$, $\alpha_* + \beta_*$ is identified, but neither α_* nor β_* is separately identified. Furthermore, if $\gamma_* = 1$ and $\boldsymbol{\delta}_*$ is conformably partitioned as $(\xi_*, \boldsymbol{\eta}_*)'$, $\xi_* + \beta_*$ is identified although neither ξ_* nor β_* is identified. Thus, *three* different identification problems arise under the linear model hypothesis. We denote these three hypotheses as $\mathcal{H}_0'$: $\beta_* = 0$; $\mathcal{H}_0''$: $\gamma_* = 0$; and $\mathcal{H}_0'''$: $\gamma_* = 1$ and call this construct the *trifold identification problem*.

The current literature approaches the trifold identification problem only in a limited way. Hansen (1996), for instance, provided a testing methodology that employs the weighted bootstrap to treat $\mathcal{H}_0'$. Alternatively, the power coefficient might be fixed as in Ramsey (1969), so that the identification problems under $\mathcal{H}_0''$ and $\mathcal{H}_0'''$ are avoided. Accordingly, the main goal of the current study is to provide a tractable test that is able to handle the trifold identification problem within a unified framework without losing power.

Some related identification problems have appeared in the literature. Cho et al. (2011, 2014) test for neglected nonlinearity using ANN models and find that two different identification problems arise under the null of linearity. They show how this twofold identification problem may be addressed using the QLR test. Cho and Ishida (2012) similarly test for effects of power transforms using the same QLR statistic but their focus of interest differs from ours and their model has only a twofold identification problem. None of this work considers nonlinear trend effects.

The approach taken in the current work is to extend the analysis of Cho et al. (2011, 2014) and Cho and Ishida (2012). The maximum order involved in the null approximation used in Cho et al. (2011) is the fourth order, whereas that used in Cho et al. (2014) is the sixth order. They observe that the maximum order depends on the activation function used in constructing the test. On the other hand, Cho and Ishida (2012) use a second-order approximation, as is common in econometric practice. The present paper examines how these approximations are modified by the trifold identification problem.

We follow ongoing practice and examine the QLR test defined as $QLR_n := n(1 - \hat{\sigma}^2_{n,A}/\hat{\sigma}^2_{n,0})$, where $\hat{\sigma}^2_{n,A} := \inf_{\alpha,\beta,\gamma,\delta} \frac{1}{n}\sum_{t=1}^n (Y_t - \alpha - \mathbf{W}_t'\delta - \beta X_t^\gamma)^2$ and $\hat{\sigma}^2_{n,0} := \inf_{\alpha,\delta} \frac{1}{n}\sum_{t=1}^n (Y_t - \alpha - \mathbf{W}_t'\delta)^2$. The following subsections separately examine the asymptotic approximations of the QLR statistic that apply under $\mathcal{H}_0'$, $\mathcal{H}_0''$, and $\mathcal{H}_0'''$.

Before proceeding it is convenient to define the model and assumptions.

Assumption 1. (i) $(Y_t, \mathbf{W}_t')' \in \mathbb{R}^{2+k}$ is an SS and absolutely regular process with mixing coefficients β_ℓ such that for some $r > 1$, $\sum_{\ell=1}^\infty \ell^{1/(r-1)}\beta_\ell < \infty$; $\mathbb{E}[|Y_t|] < \infty$; X_t is positively valued w.p. 1; and $\mathbf{Z}'\mathbf{Z} = \sum_{t=1}^n \mathbf{Z}_t\mathbf{Z}_t'$ is nonsingular w.p. 1, where $\mathbf{Z}_t := (1, \mathbf{W}_t')'$, and n is the sample size; (ii) $\mathbb{E}[Y_t|\mathbf{W}_t]$ is specified as $\mathcal{M}$, where $\boldsymbol{\Omega} := \boldsymbol{A} \times \boldsymbol{\Delta} \times \boldsymbol{B} \times \boldsymbol{\Gamma}$ is the parameter space of $\boldsymbol{\omega}$ such that $\boldsymbol{A}, \boldsymbol{\Delta}, \boldsymbol{B}$, and $\boldsymbol{\Gamma} := [\underline{\gamma}, \bar{\gamma}]$ are convex and compact in $\mathbb{R}$, $\mathbb{R}^{k+1}$, $\mathbb{R}$, and $\mathbb{R}$, respectively; and 0 and 1 are interior elements of $\boldsymbol{\Gamma}$; (iii) $\{U_t, \mathcal{F}_t\}$ is a martingale difference sequence (MDS), where $U_t := Y_t - \mathbb{E}[Y_t|\mathbf{W}_t]$, and $\mathcal{F}_t$ is the adapted smallest σ-field generated by $\{\mathbf{Z}_{t+1}, U_t, \mathbf{Z}_t, U_{t-1}, \ldots\}$. □

2.1. The QLR statistic under $\mathcal{H}_0' : \beta_* = 0$

We examine the asymptotic null approximation of the QLR test under $\mathcal{H}_0'$. As γ_* is not identified under $\mathcal{H}_0'$, we approximate the model with respect to the other parameters and treat γ as an unidentified nuisance parameter as in Davies (1977, 1987). For notational simplicity, let the quasi-likelihood (QL) and concentrated QL (CQL) be denoted as $L_n(\alpha, \beta, \gamma, \boldsymbol{\delta}) := -\sum_{t=1}^n (Y_t - \alpha - \beta X_t^\gamma - \mathbf{W}_t'\boldsymbol{\delta})^2$ and $L_n(\beta; \gamma) := L_n(\hat{\alpha}_n(\beta;\gamma), \beta, \gamma, \hat{\boldsymbol{\delta}}_n(\beta;\gamma))$, respectively, where $(\hat{\alpha}_n(\beta;\gamma), \hat{\boldsymbol{\delta}}_n(\beta;\gamma)')' := \arg\max_{\alpha,\delta} L_n(\alpha, \beta, \gamma, \boldsymbol{\delta})$. The resulting CQL has the form $L_n(\beta;\gamma) = -\{\mathbf{Y} - \beta\mathbf{X}(\gamma)\}'\mathbf{M}\{\mathbf{Y} - \beta\mathbf{X}(\gamma)\}$, where $\mathbf{Y} := (Y_1, \ldots, Y_n)'$, $\mathbf{M} := \mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{X}(\gamma) := (X_1^\gamma \ldots X_n^\gamma)'$, $\mathbf{Z} := [\mathbf{Z}_1', \ldots, \mathbf{Z}_n']'$ with $\mathbf{Z}_t := [1, \mathbf{W}_t']'$. Under $\mathcal{H}_0$, $\mathbf{MY} = \mathbf{MU}$ and $\mathbf{U} := (U_1, \ldots, U_n)'$. We can sequentially maximize the CQL with respect to β and γ:

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

Recall that $\hat{\sigma}^2_{n,0} := \inf_{\alpha,\delta}\frac{1}{n}\sum_{t=1}^n (Y_t - \alpha - \mathbf{W}_t'\boldsymbol{\delta})^2$. This statistic is asymptotically bounded in probability under mild conditions.

Assumption 2. (i) For each $\epsilon > 0$, $\mathbf{A}^{(\beta=0)}(\gamma)$ and $\mathbf{B}^{(\beta=0)}(\gamma)$ are positive definite (PD) uniformly on $\boldsymbol{\Gamma}(\epsilon) := \boldsymbol{\Gamma} \setminus ((-\epsilon, \epsilon) \cup (1-\epsilon, 1+\epsilon))$, where $\mathbf{A}^{(\beta=0)}(\gamma) := \mathbb{E}[\bar{\mathbf{R}}_t(\gamma)\bar{\mathbf{R}}_t(\gamma)']$, and $\mathbf{B}^{(\beta=0)}(\gamma) := \mathbb{E}[U_t^2\bar{\mathbf{R}}_t(\gamma)\bar{\mathbf{R}}_t(\gamma)']$ with $\bar{\mathbf{R}}_t(\gamma) := [X_t^\gamma, \mathbf{Z}_t']'$; (ii) there is a strictly stationary and ergodic (SSE) sequence $\{M_t\}$ such that $\mathbb{E}[M_t^{4\rho}] < \infty$; (iii) $\sup_{\gamma\in\boldsymbol{\Gamma}} |X_t^\gamma| \le M_t$, $\sup_{\gamma\in\boldsymbol{\Gamma}} |X_t^\gamma \log(X_t)| \le M_t$; (iv) for each j, $|D_{t,j}| \le M_t$, $|U_t| \le M_t$; (v) $\rho = r$. □

We can apply the functional central limit theorem (FCLT) and uniform law of large numbers (ULLN) to (1) using Assumptions 1 and 2. Nonetheless, if $\gamma = 0$ or 1, $\mathbf{X}(\gamma)'\mathbf{MU} \equiv 0$ and $\mathbf{X}(\gamma)'\mathbf{MX}(\gamma) \equiv 0$ by definition of $\mathbf{M}$, the idempotent projector constructed from $[1, X_t, \mathbf{D}_t']'$. So $QLR_n^{(\beta=0)}$ may not be well defined under $\mathcal{H}_0'$. For the moment, therefore, we redefine the QLR test as

$$QLR_n^{(\beta=0)}(\epsilon) := \sup_{\gamma\in\boldsymbol{\Gamma}(\epsilon)} \sup_\beta n\left\{1 - \frac{L_n(\gamma;\beta)}{L_n(0;\beta)}\right\} = \sup_{\gamma\in\boldsymbol{\Gamma}(\epsilon)} \frac{\{\mathbf{X}(\gamma)'\mathbf{MU}\}^2}{\hat{\sigma}^2_{n,0}\mathbf{X}(\gamma)'\mathbf{MX}(\gamma)}, \quad (2)$$

which explains the necessity of $\boldsymbol{\Gamma}(\epsilon)$ in Assumption 2(i). From the definition of $QLR_n^{(\beta=0)}(\cdot)$, it is monotonically decreasing, so the test may be more powerful under $\mathcal{H}_1$ as $\epsilon \to 0$. Later in this section, we consider behavior at the limits of the domain of definition as $\epsilon \to 0$ and show that $QLR_n^{(\beta=0)}$ can still be asymptotically bounded in probability under the null.

The main result of this subsection now follows.

Theorem 1. *Given Assumptions 1 and 2, and $\mathcal{H}_0'$, for each $\epsilon > 0$, $QLR_n^{(\beta=0)}(\epsilon) \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}(\epsilon)} \mathcal{Z}(\gamma)^2$, where for each $\gamma \in \boldsymbol{\Gamma}(\epsilon)$, $\mathcal{Z}(\gamma) \sim N(0, \rho(\gamma,\gamma))$, and for each pair (γ, γ'), $\mathbb{E}[\mathcal{Z}(\gamma)\mathcal{Z}(\gamma')] = \rho(\gamma,\gamma') := \kappa(\gamma,\gamma')/\{\sigma^2(\gamma)\sigma^2(\gamma')\}^{1/2}$; $\kappa(\gamma,\gamma') := \mathbb{E}[U_t^2 X_t^{\gamma+\gamma'}] - \mathbb{E}[U_t^2 X_t^\gamma \mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t X_t^{\gamma'}] - \mathbb{E}[U_t^2 X_t^{\gamma'}\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t X_t^\gamma] + \mathbb{E}[X_t^\gamma\mathbf{Z}_t']\ \mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[U_t^2\mathbf{Z}_t\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t X_t^{\gamma'}]$; and $\sigma^2(\gamma) := \sigma_*^2(\mathbb{E}[X_t^{2\gamma}] - \mathbb{E}[X_t^\gamma\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t X_t^\gamma])$.* □

The kernel $\kappa(\cdot,\cdot)$ is composed of analytic functions that satisfy dominated convergence and assure smooth second-order differentiability. This feature is important when obtaining the asymptotic null distribution. The absolutely regular mixing condition is used to demonstrate tightness of $\{\mathbf{X}(\cdot)'\mathbf{MU}\}$. The relatively simple covariance kernel is obtained because U_t is an MDS. If U_t exhibits conditional homoskedasticity, $\kappa(\gamma,\gamma')$ further simplifies to $\sigma_*^2\{\mathbb{E}[X_t^{\gamma+\gamma'}] - \mathbb{E}[X_t^\gamma\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t X_t^{\gamma'}]\}$.

2.2. The QLR statistic under $\mathcal{H}_0'' : \gamma_* = 0$

We next develop the asymptotic null approximation under $\mathcal{H}_0''$. As mentioned earlier, if $\gamma_* = 0$, α_* and β_* are not separately identified. To resolve this difficulty, our discussion proceeds in two ways. First, we may fix β, identify α_*, and obtain the asymptotic null approximation. Alternatively, we may fix α and identify β_*. We examine each case separately in what follows.

First fix β, approximate the CQL with respect to $(\alpha, \boldsymbol{\delta})$ as before, and then optimize the CQL with respect to β in the final step. For this purpose, define the CQL as $L_n(\gamma;\beta) := L_n(\hat{\alpha}_n(\gamma;\beta), \beta, \gamma, \hat{\boldsymbol{\delta}}_n(\gamma;\beta)) = -\{\mathbf{Y} - \beta\mathbf{X}(\gamma)\}'\mathbf{M}\{\mathbf{Y} - \beta\mathbf{X}(\gamma)\}$, where $(\hat{\alpha}_n(\gamma;\beta), \hat{\boldsymbol{\delta}}_n(\gamma;\beta)')' := \arg\max_{\alpha,\delta} L_n(\alpha, \beta, \gamma, \boldsymbol{\delta})$. Here, the nuisance parameter is β, while the nuisance parameter of $L_n(\beta;\gamma)$ is γ. Applying a second-order Taylor expansion to this function and optimizing with respect to γ, to approximate the QLR test, we have

$$QLR_n^{(\gamma=0;\beta)} := \sup_\beta \sup_\gamma n\left\{1 - \frac{L_n(\gamma;\beta)}{L_n(0;\beta)}\right\} = \sup_\beta \frac{\{n^{-1/2}\mathbf{L}_1'\mathbf{MU}\}^2}{\hat{\sigma}^2_{n,0}\{n^{-1}\mathbf{L}_1'\mathbf{ML}_1\}} + o_p(1), \quad (3)$$

where for each $j = 1, 2, \ldots$, $\mathbf{L}_j := [\log^j(X_1), \ldots, \log^j(X_n)]'$. Here, we also used the fact that $L_n(0;\beta) = -n\hat{\sigma}^2_{n,0}$. In particular, the right side of (3) is free of β, which holds when $\mathbf{L}_2'\mathbf{MU} = o_p(n)$. This readily holds under mild regularity conditions by virtue of the MDS property of $\{U_t, \mathcal{F}_t\}$. Therefore, the maximization process with respect to β is innocuous.

Although the approximation (3) is a consequence of a conventional second-order approximation, it differs from those in the ANN literature. Importantly, $(\partial/\partial\gamma)L_n(0;\beta)$ is not necessarily equal to zero. In the ANN literature, it is common to have zero first-order derivatives, so that higher-order approximations are needed (e.g., Cho et al. (2011, 2014); and White and Cho (2012)). This difference mainly arises because the nonlinear functions in the ANN literature have nuisance parameters that are multiplicative to X_t, whereas here the parameter γ enters nonlinearly through the power coefficient. From this feature of the specification, we expect local power properties to be different from those in the ANN literature.

We next identify the model in another way when $\gamma_* = 0$. That is, we can fix α. For this purpose, let $(\hat{\beta}_n(\gamma;\alpha), \hat{\boldsymbol{\delta}}_n(\gamma;\alpha)')' := \arg\max_{\beta,\boldsymbol{\delta}} L_n(\alpha,\beta,\gamma,\boldsymbol{\delta})$ and obtain the CQL as $L_n(\gamma;\alpha) := L_n(\alpha, \hat{\beta}_n(\gamma;\alpha), \gamma, \hat{\boldsymbol{\delta}}_n(\gamma;\alpha)) = -\mathbf{P}(\alpha)'[\mathbf{I} - \mathbf{Q}(\gamma)[\mathbf{Q}(\gamma)'\mathbf{Q}(\gamma)]^{-1}\mathbf{Q}(\gamma)']\mathbf{P}(\alpha)$, where $\mathbf{P}(\alpha) := \mathbf{Y} - \alpha\boldsymbol{\iota}$, $\mathbf{Q}(\gamma) := [\mathbf{X}(\gamma) \vdots \mathbf{W}]$, and $\boldsymbol{\iota}$ is the $n \times 1$ vector of ones. We approximate this CQL function and obtain the following approximation.

$$QLR_n^{(\gamma=0;\alpha)} := \sup_{\alpha}\sup_{\gamma} n\left\{1 - \frac{L_n(\gamma;\alpha)}{L_n(0;\alpha)}\right\} = \sup_{\alpha} \frac{\left\{n^{-1/2}\mathbf{L}_1'\mathbf{MU}\right\}^2}{\hat{\sigma}_{n,0}^2\{n^{-1}\mathbf{L}_1'\mathbf{ML}_1\}} + o_p(1). \quad (4)$$

Note that this is the same final approximation as obtained on the right side of (3), although different approximations were applied. The unidentified parameter α cancels and optimizing with respect to α is inconsequential.

Applying a central limit theorem (CLT) and the ergodic theorem to (3) or (4), we find that $QLR_n^{(\gamma=0)} := \sup_{\alpha,\gamma} n\{1 - L_n(\gamma;\alpha)/L_n(0;\alpha)\}$ weakly converges to a scaled chi-squared variate. For this purpose, we impose the following conditions.

Assumption 3. (i) $\mathbf{A}^{(\gamma=0)}$ and $\mathbf{B}^{(\gamma=0)}$ are PD, where $\mathbf{A}^{(\gamma=0)} := \mathbb{E}[\dot{\mathbf{R}}_t\dot{\mathbf{R}}_t']$, $\mathbf{B}^{(\gamma=0)} := \mathbb{E}[U_t^2\dot{\mathbf{R}}_t\dot{\mathbf{R}}_t']$, and $\dot{\mathbf{R}}_t := [\log(X_t), \mathbf{Z}_t']'$; (ii) for an SSE sequence $\{M_t\}$ and each j, $|W_{t,j}| \le M_t$, $|U_t| \le M_t$, $|\log(X_t)| \le M_t$, and $\mathbb{E}[M_t^4] < \infty$. □

The following theorem formalizes the result.

Theorem 2. *Given Assumptions 1 and 3, and $\mathscr{H}_0''$, $QLR_n^{(\gamma=0)} = \left\{\mathbf{L}_1'\mathbf{MU}\right\}^2/\{\hat{\sigma}_{n,0}^2(\mathbf{L}_1'\mathbf{ML}_1)\} + o_p(1)$ under $\mathscr{H}_0'' : \gamma_* = 0$, and $QLR_n^{(\gamma=0)} \overset{A}{\sim} \mathcal{Z}_0^2$, where $\mathcal{Z}_0 \sim N(0, \kappa_0^2/\sigma_0^2)$; $\kappa_0^2 := \mathbb{E}[U_t^2\log^2(X_t)] - 2\mathbb{E}[U_t^2\log(X_t)\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t\log(X_t)] + \mathbb{E}[\log(X_t)\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[U_t^2\mathbf{Z}_t\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t\log(X_t)]$; and $\sigma_0^2 := \sigma_*^2(\mathbb{E}[\log^2(X_t)] - \mathbb{E}[\log(X_t)\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t\log(X_t)])$.* □

The asymptotic null approximation of the QLR test is driven by $\mathbf{L}_1$, a feature that, intuitively, is associated with the Box–Cox transformation. Passing the parameter of the Box–Cox transform to zero gives $(d/d\gamma)X_t^{\gamma}\big|_{\gamma=0} = \lim_{\gamma\to 0}(X_t^{\gamma} - 1)/\gamma = \log X_t$. Thus, the Box–Cox transform with $\gamma = 0$ is associated with the first-order derivative which forms the primary component constituting the score of the QLR test. Additionally, the Box–Cox transform approximates $\mathbb{E}[Y_t|\mathbf{W}_t] = (\alpha_* + \beta_*) + \xi_* X_t + \mathbf{D}_t'\boldsymbol{\eta}_* + \beta_*\gamma_*(X_t^{\gamma_*} - 1)/\gamma_*$ by $\alpha_* + \xi_* X_t + \mathbf{D}_t'\boldsymbol{\eta}_* + \beta_*\gamma_*\log(X_t)$ when γ_* is sufficiently close to zero. For such a case, $\mathbf{L}_1'\mathbf{MU}$ is the primary score of standard statistics obtained under the null that $\beta_*\gamma_* = 0$. This implies that the Box–Cox transform can be understood as an alternative to the constant function hypothesis.

2.3. The QLR statistic under $\mathscr{H}_0''' : \gamma_* = 1$

We repeat the procedure to obtain the asymptotic null approximation under $\mathscr{H}_0'''$. If $\gamma_* = 1$, β_* and ξ_* are not separately identified as mentioned earlier. The procedure to obtain the asymptotic approximation is similar to that of Section 2.2. As β_* and ξ_* are not separately identified, we first fix β at some particular value and concentrate the QL with respect to $(\alpha, \boldsymbol{\delta}')'$. The CQL obtained in this way is expanded with respect to γ around $\gamma_* = 1$ by a second-order approximation, leading to

$$QLR_n^{(\gamma=1;\beta)} := \sup_{\beta}\sup_{\gamma}\sup_{\alpha,\boldsymbol{\delta}} n\left\{1 - \frac{L_n(\alpha,\beta,\gamma,\boldsymbol{\delta})}{L_n(1;\beta)}\right\} = \sup_{\beta} \frac{\left\{n^{-1/2}\mathbf{C}_1'\mathbf{MU}\right\}^2}{\hat{\sigma}_{n,0}^2\{n^{-1}\mathbf{C}_1'\mathbf{MC}_1\}} + o_p(1), \quad (5)$$

provided that $\mathbf{C}_2'\mathbf{MU} = o_p(n)$, where for each $j = 1, 2, \ldots$, $\mathbf{C}_j := [X_1\log^j(X_1), \ldots, X_n\log^j(X_n)]'$.

We now reverse the plan of identification. We first fix ξ and identify the other parameters $(\alpha_*, \beta_*, \boldsymbol{\eta}_*')'$. For notational simplicity, let $\boldsymbol{\theta} := (\beta, \boldsymbol{\eta}')'$ and $\mathbf{S}_t(\gamma) := (X_t^{\gamma}, \mathbf{D}_t')'$, so that $\boldsymbol{\theta}_* := (\beta_*, \boldsymbol{\eta}_*')'$. We obtain $L_n(\gamma;\xi) := L_n(\widehat{\alpha}_n(\gamma;\xi), \widehat{\boldsymbol{\theta}}_n(\gamma;\xi), \gamma, \xi) = -\widetilde{\mathbf{P}}(\xi)'[\mathbf{I} - \widetilde{\mathbf{Q}}(\gamma)[\widetilde{\mathbf{Q}}(\gamma)'\widetilde{\mathbf{Q}}(\gamma)]^{-1}\widetilde{\mathbf{Q}}(\gamma)']\widetilde{\mathbf{P}}(\xi)$, where $\widetilde{\mathbf{P}}(\xi) := \mathbf{Y} - \xi\mathbf{X}$, $\widetilde{\mathbf{Q}}(\gamma) := [\boldsymbol{\iota} \vdots \mathbf{S}(\gamma)]$, $\mathbf{X} := (X_1, \ldots, X_n)'$, and $\mathbf{S}(\gamma) := [\mathbf{S}_1(\gamma), \ldots, \mathbf{S}_n(\gamma)]'$. We again approximate the CQL, and the asymptotic approximation of the QLR test is simply

$$QLR_n^{(\gamma=1,\xi)} := \sup_{\xi}\sup_{\gamma} n\left\{1 - \frac{L_n(\gamma;\xi)}{L_n(1;\xi)}\right\} = \sup_{\xi} \frac{\left\{n^{-1/2}\mathbf{C}_1'\mathbf{MU}\right\}^2}{\widehat{\sigma}_{n,0}^2\{n^{-1}\mathbf{C}_1'\mathbf{MC}_1\}} + o_p(1). \quad (6)$$

This expression has the same approximate form as that on the right side of (5). Our next assumption provides regularity assumptions for this result to hold.

Assumption 4. (i) $\mathbf{A}^{(\gamma=1)}$ and $\mathbf{B}^{(\gamma=1)}$ are PD, where $\mathbf{A}^{(\gamma=1)} := \mathbb{E}[\ddot{\mathbf{R}}_t\ddot{\mathbf{R}}_t']$ and $\mathbf{B}^{(\gamma=1)} := \mathbb{E}[U_t^2\ddot{\mathbf{R}}_t\ddot{\mathbf{R}}_t']$ with $\ddot{\mathbf{R}}_t := [X_t\log(X_t), \mathbf{Z}_t']'$; (ii) for an SSE sequence $\{M_t, S_t\}$ and each j, $|D_{t,j}| \le M_t$, $\mathbb{E}[M_t^{4\rho}] < \infty$, $\mathbb{E}[S_t^8] < \infty$, and (ii.a) $|U_t| \le M_t$, $|X_t| \le S_t$, and $|\log[X_t]| \le S_t$; (ii.b) $|X_t| \le M_t$, $|U_t| \le S_t$, and $|\log[X_t]| \le S_t$; or (ii.c) $|\log[X_t]| \le M_t$, $|X_t| \le S_t$, and $|U_t| \le S_t$; (iii) $\rho = 1$. □

Note that the moment condition in Assumption 4(ii.a) does not imply Assumption 4(ii.b or ii.c) or vice versa. If at least one of these separate conditions holds, however, the desired results follow as given below.

Theorem 3. *Given Assumptions 1 and 4, and $\mathscr{H}_0'''$, $QLR_n^{(\gamma=1)} = \left\{\mathbf{C}_1'\mathbf{MU}\right\}^2/\{\hat{\sigma}_{n,0}^2(\mathbf{C}_1'\mathbf{MC}_1)\} + o_p(1)$, where $QLR_n^{(\gamma=1)}$ denotes the QLR statistic testing $\mathscr{H}_0'''$, and $QLR_n^{(\gamma=1)} \overset{A}{\sim} \mathcal{Z}_1^2$; $\mathcal{Z}_1 \sim N(0, \kappa_1^2/\sigma_1^2)$; $\kappa_1^2 := \mathbb{E}[U_t^2X_t^2\log^2(X_t)] - 2\mathbb{E}[U_t^2X_t\log(X_t)\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_tX_t\log(X_t)] + \mathbb{E}[X_t\log(X_t)\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[U_t^2\mathbf{Z}_t\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_tX_t\log(X_t)]$; and $\sigma_0^2 := \sigma_*^2(\mathbb{E}[X_t^2\log^2(X_t)] - \mathbb{E}[X_t\log(X_t)\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_tX_t\log(X_t)])$.* □

The asymptotic null distribution is driven by $\mathbf{C}_1$ and, as before, this link can be associated with the Box–Cox transformation. In particular $(d/d\gamma)X_t^{\gamma}\big|_{\gamma=1} = \lim_{\gamma\to 1}(X_t^{\gamma} - X_t)/(\gamma - 1)$. So modifying the Box–Cox transform as

$$ABC_t(\gamma) := \begin{cases} (X_t^{\gamma} - X_t)/(\gamma - 1), & \text{if } \gamma \neq 1; \\ X_t\log[X_t], & \text{if } \gamma = 1, \end{cases}$$

we see that $X_t\log(X_t)$ is the typical element of $\mathbf{C}_1$, implying an interpretation of the test in terms of the Box–Cox transformation. That is, when γ_* is believed to be sufficiently close to one in $\mathbb{E}[Y_t|\mathbf{W}_t] = \alpha_* + (\xi_* + \beta_*)X_t + \mathbf{D}_t'\boldsymbol{\eta}_* + \beta_*(\gamma_* - 1)\{(X_t^{\gamma_*} - X_t)/(\gamma_* - 1)\}$, the augmented Box–Cox transformation approximates the mean function by $\alpha_* + (\xi_* + \beta_*)X_t + \mathbf{D}_t'\boldsymbol{\eta}_* + \beta_*(\gamma_* - 1)X_t\log(X_t)$. For such a case, the primary score of standard statistics is constructed using $\mathbf{C}_1'\mathbf{MU}$ under the null that $\beta_*(\gamma_* - 1) = 0$. This implies that the given transformation can be understood as an alternative to the linearity hypothesis.

2.4. Interrelationships of the QLR statistics under $\mathcal{H}_0$

The separate weak limits obtained in the previous subsections are not independent. The stochastic relationships can be studied by letting γ converge to zero and unity in the test studied in Section 2.1. To wit, define $N_n(\gamma)$ and $D_n(\gamma)$ as $N_n(\gamma) := \{\mathbf{X}(\gamma)'\mathbf{MU}\}^2$ and $D_n(\gamma) := \widehat{\sigma}^2_{n,0}\mathbf{X}(\gamma)'\mathbf{MX}(\gamma)$, representing the numerator and denominator of (1), respectively. First, consider the case where $\gamma \to 0$, which shows that $\text{plim}_{\gamma\to 0}N_n(\gamma) = 0$ and $\text{plim}_{\gamma\to 0}D_n(\gamma) = 0$ because $\text{plim}_{\gamma\to 0}\mathbf{X}(\gamma) = \boldsymbol{\iota}$ and $\mathbf{M}$ is the idempotent projector constructed from $[1, X_t, \mathbf{D}_t']'$. First order use of l'Hôpital's rule also fails due to the further degeneracy: $\text{plim}_{\gamma\to 0}(d/d\gamma)N_n(\gamma) = 0$, $\text{plim}_{\gamma\to 0}(d/d\gamma)D_n(\gamma) = 0$ by the same reasoning. It also follows that $\text{plim}_{\gamma\to 1}N_n(\gamma) = \text{plim}_{\gamma\to 1}(d/d\gamma)N_n(\gamma) = 0$ and $\text{plim}_{\gamma\to 1}D_n(\gamma) = \text{plim}_{\gamma\to 1}(d/d\gamma) D_n(\gamma) = 0$. Hence, it is necessary to apply l'Hôpital's rule a further time to remove the degeneracy.

The required further derivatives are provided in the following lemma.

Lemma 1. *Given Assumption 1,* (i) $\text{plim}_{\gamma\to 0}N_n^{(2)}(\gamma) = 2\{\mathbf{L}_1\mathbf{MU}\}^2$ *and* $\text{plim}_{\gamma\to 0}D_n^{(2)}(\gamma) = 2\widehat{\sigma}^2_{n,0}\mathbf{L}_1\mathbf{ML}_1$; *and* (ii) $\text{plim}_{\gamma\to 1}N_n^{(2)}(\gamma) = 2\{\mathbf{C}_1\mathbf{MU}\}^2$ *and* $\text{plim}_{\gamma\to 1}D_n^{(2)}(\gamma) = 2\widehat{\sigma}^2_{n,0}\mathbf{C}_1\mathbf{MC}_1$, *where for* $j = 1, 2, \ldots, N_n^{(j)}(\gamma) := (\partial^j/\partial\gamma^j)N_n(\gamma)$ *and* $D_n^{(j)}(\gamma) := (\partial^j/\partial\gamma^j) D_n(\gamma)$. □

Lemma 1 implies that $\text{plim}_{\gamma\to 0}N_n(\gamma)/D_n(\gamma) = \{\mathbf{L}_1\mathbf{MU}\}^2/\widehat{\sigma}^2_{n,0}\mathbf{L}_1\mathbf{M}\mathbf{L}_1$ and $\text{plim}_{\gamma\to 1}N_n(\gamma)/D_n(\gamma) = \{\mathbf{C}_1\mathbf{MU}\}^2/\widehat{\sigma}^2_{n,0}\mathbf{C}_1\mathbf{MC}_1$. That is, the asymptotic null approximations provided in Theorems 2 and 3 can be combined with the null approximation in Theorem 1. For this purpose, we combine the regularity conditions of Theorems 2 and 3 as in the following assumption.

Assumption 5. For each $\epsilon > 0$, $\mathbf{A}(\gamma)$ and $\mathbf{B}(\gamma)$ are PD uniformly on $\boldsymbol{\Gamma}(\epsilon)$, where $\mathbf{A}(\gamma) := \mathbb{E}[\mathbf{R}_t(\gamma)\mathbf{R}_t(\gamma)']$, $\mathbf{B}(\gamma) := \mathbb{E}[U_t^2\mathbf{R}_t(\gamma)\mathbf{R}_t(\gamma)']$, and $\mathbf{R}_t(\gamma) := [X_t^{\gamma}, X_t\log(X_t), \log(X_t), \mathbf{Z}_t']'$. □

Assumption 5 is stronger than Assumptions 2–4, each of which separately holds under Assumption 5. Using these conditions we have the following result.

Theorem 4. *Given Assumptions 1 and 2(iii, v), 4(ii), 5, and $\mathcal{H}_0$,* $QLR_n = \sup_{\gamma\in\boldsymbol{\Gamma}}\left\{\mathbf{X}(\gamma)'\mathbf{MU}\right\}^2/\{\hat{\sigma}^2_{n,0}\mathbf{X}(\gamma)'\mathbf{MX}(\gamma)\}$, *and* $QLR_n \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}}\mathcal{Z}(\gamma)^2$. □

This result gives the asymptotic approximation of the QLR test under $\mathcal{H}_0$ and its limiting form as a functional of a Gaussian process $\mathcal{Z}(\cdot)$. Importantly, $\mathcal{Z}(\cdot)$ is discontinuous at $\gamma = 0$ and 1 w.p. 1. Defining $Z_n(\gamma) := \mathbf{X}(\gamma)'\mathbf{MU}/\{\widehat{\sigma}^2_{n,0}\mathbf{X}(\gamma)'\mathbf{MX}(\gamma)\}^{1/2}$, we can regard $\mathcal{Z}(\cdot)$ as the weak limit of $Z_n(\cdot)$. Observe that $\lim_{\gamma\downarrow 0}Z_n(\gamma) = -\lim_{\gamma\uparrow 0}Z_n(\gamma)$ and $\lim_{\gamma\downarrow 1}Z_n(\gamma) = -\lim_{\gamma\uparrow 1}Z_n(\gamma)$ w.p. 1, so that $\lim_{\gamma\downarrow 0}\mathcal{Z}(\gamma) = -\lim_{\gamma\uparrow 0}\mathcal{Z}(\gamma)$ and $\lim_{\gamma\downarrow 1}\mathcal{Z}(\gamma) = -\lim_{\gamma\uparrow 1}\mathcal{Z}(\gamma)$ w.p. 1. In view of these limits, the squared process $\mathcal{Z}(\gamma)^2$ has equal left-hand and right-hand side limits as γ tends to 0 and 1. If $\mathcal{Z}(0)^2$ and $\mathcal{Z}(1)^2$ are defined by these limits, it follows that $\mathcal{Z}(\cdot)^2$ is continuous on $\boldsymbol{\Gamma}$ w.p. 1.

Theorem 4 has the following main implications. First, the asymptotic null approximation addresses the trifold identification problem and, under the regularity conditions for each case, ensures that the limiting null distribution exists for each form of the null hypothesis. Second, the QLR test simultaneously satisfies these separate conditions, thereby accommodating the trifold identification issues. With this property, the QLR test has the capacity to test linearity within a unified framework. Finally, the null approximation is obtained by using only second-order approximations, thereby ensuring that the QLR test has a $\sqrt{n}$ convergence rate under H_0'' and H_0'''. This property differs from the ANN literature and leads the QLR test to have nontrivial power against an $n^{-1/2}$-local alternative, as verified in the next subsection.

To be more specific on the implications, we contrast the result in Theorem 4 with the tests in the prior literature. By following Stinchcombe and White's (1998) Theorem 2.3, we define Bierens (1990) conditional moment (CM) test as $CM_n := \sup_{\gamma\in\boldsymbol{\Gamma}}\{\widehat{W}_n(\gamma)/\widetilde{\sigma}_n(\gamma)\}^2$, where for each γ,

$$\widehat{W}_n(\gamma) := n^{-1/2}\sum_{t=1}^{n}(Y_t - \widetilde{\alpha}_n - \mathbf{W}_t'\widetilde{\boldsymbol{\delta}}_n)\exp(\gamma X_t)$$

under the linear model context, and where $\widetilde{\sigma}_n(\gamma)^2 := n^{-1}\widehat{\sigma}^2_{n,o}\mathbf{E}(\gamma)'\mathbf{ME}(\gamma)$ with $\mathbf{E}(\gamma) := [\ldots, \exp(\gamma X_t), \ldots]'$. Here, $(\widetilde{\alpha}_n, \widetilde{\boldsymbol{\delta}}_n')$ is obtained by regressing Y_t on $(1, \mathbf{W}_t')$. Therefore, $\widehat{W}_n(0) \equiv 0$ and $\widehat{W}_n(\gamma)^2 = n^{-1}\{\mathbf{U}'\mathbf{ME}(\gamma)\}^2$ under the null, so that

$$CM_n = \sup_{\gamma\in\boldsymbol{\Gamma}}\frac{(\mathbf{U}'\mathbf{ME}(\gamma))^2}{\widehat{\sigma}^2_{n,o}\mathbf{E}(\gamma)'\mathbf{ME}(\gamma)}.$$

Note that the only difference between the QLR_n and CM_n statistics is that $\mathbf{X}(\gamma)$ in QLR_n replaces $\mathbf{E}(\gamma)$, where the parameter γ of $\mathbf{X}(\gamma)$ exists as an integral parametric part of model, whereas γ of $\mathbf{E}(\gamma)$ is an auxiliary parameter that is introduced specifically for defining the CM test. Although Bierens (1990) does not explain how the CM test is defined when $\gamma = 0$ (note that $(\widehat{W}_n(0)/\widetilde{\sigma}(0))^2 = (0/0)^2$), the current paper shows that the QLR test has the capability of jointly testing $\beta_* = 0, \gamma_* = 0$, and/or $\gamma_* = 1$ using a second-order Taylor expansion that differs from the expansion orders in Cho et al. (2011, 2014). As another conditional moment test, Bierens and Ploberger (1997) define the integrated conditional moment (ICM) test as

$$ICM_n := \int_{\gamma\in\boldsymbol{\Gamma}}\widehat{W}_n^2(\gamma)d\mu(\gamma),$$

where $\mu(\cdot)$ is a probability measure on $\boldsymbol{\Gamma}$. Instead of the uniform norm, the L_2-norm is used to construct this test, and $\widehat{W}_n(\gamma)$ is no longer standardized as for the CM test. Due to this fact, $\lim_{\gamma\to 0}\widehat{W}_n^2(\gamma) = 0$, which is now different from the CM test. So, we can no longer link the ICM test to the score that tests $\gamma_* = 0$. By virtue of this fact, the ICM test can be said to test only $\beta_* = 0$, and the second-order Taylor expansion is enough for testing $\beta_* = 0$.

2.5. Power examination of the QLR test

The omnibus power of the QLR test derives from the GCR property. Under H_1, for any non-polynomial analytic function, $\psi(\cdot)$ say, $\mathbb{E}[V_t\psi(X_t)] \neq 0$ when V_t is the linear projection error. This also implies that for some non-negative integer j_*, $\mathbb{E}[V_t\log^{j_*}(X_t)] \neq 0$ by Bierens's (1982) Theorem 2, given that $\log(\cdot)$ is a one-to-one mapping. The omnibus power of the QLR test is associated with this property.

For a specific examination of the power of the QLR test, we suppose that $\mathbb{E}[Y_t|\mathbf{W}_t] = \alpha_* + \mathbf{W}_t'\boldsymbol{\delta}_* + m(X_t)$ and that there is possibly no parameter vector (β_*, γ_*) such that $m(X_t) = \beta_* X_t^{\gamma_*}$ w.p. 1., so that the class $\mathcal{M}$ may not be able to deliver a consistent estimate of $\mathbb{E}[Y_t|\mathbf{W}_t]$. By usual least squares projection algebra we find that

$$\min_{\alpha,\boldsymbol{\delta},\beta}\mathbb{E}[(Y_t - \alpha - \mathbf{W}_t'\boldsymbol{\delta} - \beta X_t^{\gamma})^2]$$
$$= h(\gamma) := \mathbb{E}[U_t^2] + \text{var}[Q_t] - \text{cov}[U_t(\gamma), Q_t]^2/\text{var}[Q_t],$$

where $U_t(\gamma) := X_t^{\gamma} - \mathbf{Z}_t'\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t X_t^{\gamma}]$ and $Q_t := m(X_t) - \mathbf{Z}_t'\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t m(X_t)]$. Thus, if it happens that for some (β_*, γ_*),

$m(X_t) = \beta_* X_t^{\gamma_*}$ w.p. 1 and $\gamma_* \in \boldsymbol{\Gamma}$, then $h(\cdot)$ is minimized as $\mathbb{E}[U_t^2]$ by letting $\gamma = \gamma_*$. Note that if $h_0 := \min_{\alpha,\delta} \mathbb{E}[(Y_t - \alpha - \mathbf{W}_t'\boldsymbol{\delta})^2] = \mathbb{E}[U_t^2] + \mathrm{var}[Q_t]$, we have $QLR_n/n = (1 - h(\gamma_*)/h_0) + o_p(1)$, and $h(\gamma_*)/h_0 < 1$. Therefore, the QLR test has consistent power. This property remains true even if there is no such (β_*, γ_*).

Theorem 5. *Given Assumptions 1 and 2(iii, v), 4(ii), and 5, (i) if $\mathbb{E}[Y_t|\mathbf{W}_t] = \alpha_* + \mathbf{W}_t'\boldsymbol{\delta}_* + m(X_t)$ with $\mathbb{E}[m(X_t)^2] < \infty$ and $\mathbb{E}[\log^{4j_*}(X_t)] < \infty$, for some $\widetilde{\gamma} \in \boldsymbol{\Gamma}$, $h(\widetilde{\gamma}) \in (0, h_0)$ and $QLR_n/n = (1 - h(\widetilde{\gamma})/h_0) + o_p(1)$; (ii) if $\mathbb{E}[Y_t|\mathbf{W}_t] = \alpha_* + \mathbf{W}_t'\boldsymbol{\delta}_* + m(X_t)/\sqrt{n}$ with $|m(X_t)| \le M_t$, $QLR_n \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}}\{\mathcal{Z}(\gamma) + \mu(\gamma)/\sigma(\gamma)\}^2$, where $\mu(\gamma) := \mathbb{E}[m(X_t)X_t^{\gamma}] - \mathbb{E}[m(X_t)\mathbf{Z}_t']\mathbb{E}[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbb{E}[\mathbf{Z}_t X_t^{\gamma}]$.* □

Theorem 5(i) follows by showing that $h(\cdot)$ is not a constant function on $\boldsymbol{\Gamma}$ if $\mathbb{E}[V_t \log^{j_*}(X_t)] \neq 0$, and Theorem 5(ii) derives the local power of the test.

3. Testing for power transforms of a trend regressor

We now extend the discussion to the case where Y_t is a trend stationary process with a deterministic time trend. This type of model is particularly important in analyzing nonstationary time series and in trend removal procedures. We suppose that $\mathbb{E}[Y_t|\mathbf{D}_t]$ is a function of both t and $\mathbf{D}_t$, where $\{\mathbf{D}_t\}$ is, as before, strictly stationary. Primary attention now focuses on testing whether $\mathbb{E}[Y_t|\mathbf{D}_t]$ is a linear function of $(1, \mathbf{D}_t', t)'$. For such a test, we consider $\mathcal{M}' := \{m_t(\cdot) : \boldsymbol{\Omega} \mapsto \mathbb{R} : m_t(\alpha, \boldsymbol{\delta}, \beta, \gamma) := \alpha + \mathbf{D}_t'\boldsymbol{\eta} + \xi t + \beta t^{\gamma}\}$. The only difference between $\mathcal{M}$ and $\mathcal{M}'$ arises from the replacement of X_t with t. The regressor $\mathbf{D}_t$ may be used to capture temporal dependence in the data that is not embodied in t^{γ}.

In spite of this correspondence with the earlier model, the QLR test cannot be straightforwardly analyzed because Assumption 5 no longer holds. The PD matrix condition in Assumption 5(i) fails and the (implied) regressors are asymptotically collinear. The following lemma states this property in a precise way.

Lemma 2. *If $\{\mathbf{D}_t\}$ is SSE such that for each j, $\mathbb{E}[D_{t,j}^2] < \infty$, then for each $\gamma \in \boldsymbol{\Gamma}(\epsilon)$, $\mathbf{F}_n^{-1}\sum_{t=1}^n \mathbf{H}_t(\gamma)\mathbf{H}_t(\gamma)'\mathbf{F}_n^{-1}$ almost surely converges to a singular matrix, where $\mathbf{H}_t(\gamma) := [t^{\gamma}, t\log(t), \log(t), 1, t, \mathbf{D}_t']'$, $\mathbf{F}_n := \mathrm{diag}[n^{\frac{1}{2}+\gamma}, n^{\frac{3}{2}}\log(n), n^{\frac{1}{2}}\log(n), n^{\frac{1}{2}}, n^{\frac{3}{2}}, n^{\frac{1}{2}}\boldsymbol{\iota}_k]$, and $\boldsymbol{\iota}_k$ is a $k \times 1$ vector of ones.* □

Note that $\mathbf{F}_n^{-1}\sum_{t=1}^n \mathbf{H}_t(\gamma)\mathbf{H}_t(\gamma)'\mathbf{F}_n^{-1}$ is a (matrix normalized) sample analog of $\mathbf{A}(\gamma)$ in Assumption 5(i). Since time trends are involved, the scaling rates of the components are different from the standard stationary variable case and are parameter dependent on γ. As the limit of the square signal matrix in Lemma 2 is a singular matrix, the QLR test cannot be analyzed as in Section 2. Importantly, this singularity does not imply that the asymptotic null distribution of the QLR test does not necessarily exist and that rotating the regressor space is required for testing the null hypothesis (e.g. Park and Phillips (1988) and Phillips (1989)). It is convenient to use the approach based on smoothly slowly varying (SSV) functions in Phillips (2007).

The asymptotic null distribution of the QLR test can most conveniently be found by reformulation. Instead of $\mathcal{M}'$, we use the following 'weak trend' specification involving the trend fraction $s_{n,t} := \frac{t}{n}$ and power functions of $s_{n,t}$: $\mathcal{M}'' := \{m_t(\cdot) : \boldsymbol{\Omega} \mapsto \mathbb{R} : m_t(\alpha, \boldsymbol{\delta}, \beta, \gamma) := \alpha + \mathbf{D}_t'\boldsymbol{\eta} + \xi_n s_{n,t} + \lambda_n(\beta, \gamma)s_{n,t}^{\gamma}\}$, where $\xi_n := \xi n$ and $\lambda_n(\beta, \gamma) := \beta n^{\gamma}$. This weak trend has asymptotics closely related to those of a stationary regressor. Linearity is obtained from $\mathcal{M}''$ by setting $\lambda_n(\cdot) = 0$ for any n, $\gamma = 0$, or $\gamma = 1$. Furthermore, $\beta = 0$ if and only if $\lambda_n(\cdot) = 0$. Thus, when the null is given as $\widetilde{\mathcal{H}}_0$: $\exists(\alpha_*, \boldsymbol{\eta}_*, \xi_*)$, $\mathbb{E}[Y_t|\mathbf{D}_t] = \alpha_* + \mathbf{D}_t'\boldsymbol{\eta}_* + \xi_* t$ w.p. 1, it can be formulated in terms of $\widetilde{\mathcal{H}}_0'$: $\lambda_n(\beta_*, \gamma_*) = 0$; $\widetilde{\mathcal{H}}_0''$: $\gamma_* = 0$; and $\widetilde{\mathcal{H}}_0'''$: $\gamma_* = 1$.

Using this modification of the model, the asymptotic null behavior of the QLR test can be obtained under appropriate conditions. We start with the following assumptions:

Assumption 6. (i) $(Y_t, \mathbf{D}_t')' \in \mathbb{R}^{1+k}$ $(k \in \mathbb{N})$ is given, and $\{\mathbf{D}_t\}$ is a ϕ-mixing process with mixing decay rate $-m/2(m-1)$ with $m \ge 2$ or an α-mixing process with mixing decay rate $-m/(m-2)$ with $m > 2$, and Y_t is a time-trend stationary process; $\mathbf{Z}'\mathbf{Z} = \sum_{t=1}^n \mathbf{Z}_{n,t}\mathbf{Z}_{n,t}'$ is nonsingular w.p. 1, where $\mathbf{Z}_{n,t} := (1, s_{n,t}, \mathbf{D}_t')'$, and n is the sample size; (ii) $\mathbb{E}[Y_t|\mathbf{D}_t]$ is specified as $\mathcal{M}''$, where $\boldsymbol{\Omega}_n := \boldsymbol{A} \times \mathbf{H} \times \boldsymbol{\Xi}_n \times \boldsymbol{\Lambda}_n \times \boldsymbol{\Gamma}$ is the parameter space of $\boldsymbol{\omega}_n := (\alpha, \boldsymbol{\eta}', \xi_n, \lambda_n, \gamma)'$ such that $\boldsymbol{A}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\Gamma}$ are convex and compact in $\mathbb{R}$, $\mathbb{R}^k$, and $\mathbb{R}$, respectively, such that 0 and 1 are interior elements of $\boldsymbol{\Gamma} := [\underline{\gamma}, \bar{\gamma}]$ with $\gamma_o := \inf \boldsymbol{\Gamma} > -1/2$, and for each n, $\boldsymbol{\Xi}_n$ and $\boldsymbol{\Lambda}_n$ are convex and compact in $\mathbb{R}$. □

Assumption 7. (i) For each $\epsilon > 0$, $\widetilde{\mathbf{A}}(\gamma)$ and $\widetilde{\mathbf{B}}(\gamma)$ are PD uniformly on $\boldsymbol{\Gamma}(\epsilon)$, where $\widetilde{\mathbf{A}}(\gamma) := \mathbb{E}[\widetilde{\mathbf{H}}_t(\gamma)]$, $\widetilde{\mathbf{B}}(\gamma) := \mathbb{E}[U_t^2\widetilde{\mathbf{H}}_t(\gamma)]$, $U_t := Y_t - \mathbb{E}[Y_t|\mathbf{D}_t]$, and

$$\widetilde{\mathbf{H}}_t(\gamma) := \begin{bmatrix} \frac{1}{2\gamma+1} & -\frac{1}{(\gamma+2)^2} & -\frac{1}{(\gamma+1)^2} & \frac{1}{\gamma+1} & \frac{1}{\gamma+2} & \frac{1}{1+\gamma}\mathbf{D}_t' \\ -\frac{1}{(\gamma+2)^2} & \frac{2}{27} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{9} & -\frac{1}{4}\mathbf{D}_t' \\ -\frac{1}{(\gamma+1)^2} & \frac{1}{4} & 2 & -1 & -\frac{1}{4} & -\mathbf{D}_t' \\ \frac{1}{\gamma+1} & -\frac{1}{4} & -1 & 1 & \frac{1}{2} & \mathbf{D}_t' \\ \frac{1}{\gamma+2} & -\frac{1}{9} & -\frac{1}{4} & \frac{1}{2} & \frac{1}{3} & \frac{1}{2}\mathbf{D}_t' \\ \frac{1}{\gamma+1}\mathbf{D}_t & -\frac{1}{4}\mathbf{D}_t & -\mathbf{D}_t & \mathbf{D}_t & \frac{1}{2}\mathbf{D}_t & \mathbf{D}_t\mathbf{D}_t' \end{bmatrix};$$

(ii) $\{U_t, \mathcal{F}_t\}$ is an MDS; and there is an SSE sequence $\{M_t\}$ such that for each j, $|D_{t,j}| \le M_t$, $|U_t| \le M_t$, and for some $r > 1$, $\mathbb{E}[M_t^{4r}] < \infty$. □

Some discussion of Assumptions 6 and 7 is warranted. First, the mixing condition in Assumption 1 is relaxed in Assumption 6. Since the time trend is nonstochastic, tightness of the statistic trivially holds even under the current mixing condition. Second, $\widetilde{\mathbf{A}}(\gamma)$ and $\widetilde{\mathbf{B}}(\gamma)$ are the probability limits of $n^{-1}\sum \mathbf{G}_{n,t}(\gamma)\mathbf{G}_{n,t}(\gamma)'$ and $n^{-1}\sum U_t^2\mathbf{G}_{n,t}(\gamma)\mathbf{G}_{n,t}(\gamma)'$, where $\mathbf{G}_{n,t}(\gamma) := [s_{n,t}^{\gamma}, s_{n,t}\log(s_{n,t}), \log(s_{n,t}), 1, s_{n,t}, \mathbf{D}_t']'$. Third, the nonsingularity of $\widetilde{\mathbf{A}}(\gamma)$ is identical to the condition that $\mathbf{D}_t$ has a nonsingular covariance matrix. The first five principal minors of $\widetilde{\mathbf{A}}(\gamma)$ have strictly positive determinants. Thus, $\widetilde{\mathbf{A}}(\gamma)$ is PD if and only if $\mathbb{E}[\mathbf{D}_t\mathbf{D}_t'] - \widetilde{\mathbf{A}}^{(2,1)}(\gamma)\{\widetilde{\mathbf{A}}^{(1,1)}(\gamma)\}^{-1}\widetilde{\mathbf{A}}^{(1,2)}(\gamma)$ is PD, where we partition $\widetilde{\mathbf{A}}(\gamma)$ as

$$\widetilde{\mathbf{A}}(\gamma) \equiv \begin{bmatrix} \widetilde{\mathbf{A}}^{(1,1)}(\gamma) & \widetilde{\mathbf{A}}^{(1,2)}(\gamma) \\ \widetilde{\mathbf{A}}^{(2,1)}(\gamma) & \mathbb{E}[\mathbf{D}_t\mathbf{D}_t'] \end{bmatrix}.$$

The final entry is the covariance matrix of $\mathbf{D}_t$ by the definition of $\widetilde{\mathbf{A}}(\gamma)$. Finally, the QLR tests obtained by using $\mathcal{M}'$ and $\mathcal{M}''$ are identical by the invariance principle of maximum likelihood: reparameterization does not modify the level of the maximized quasi-likelihood.

Our main result now follows.

Theorem 6. *Given Assumptions 6 and 7, (i) $QLR_n \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}} \widetilde{\mathcal{Z}}(\gamma)^2$ under $\widetilde{\mathcal{H}}_0$, where $\widetilde{\mathcal{Z}}(\cdot)$ is a Gaussian process with zero mean and covariance kernel $\widetilde{\kappa}(\gamma, \gamma')$ such that for each $\gamma, \gamma' \in \boldsymbol{\Gamma} \setminus \{0, 1\}$, $\widetilde{\kappa}(\gamma, \gamma') := c(\gamma, \gamma')(1 + 2\gamma)^{1/2}(1 + 2\gamma')^{1/2}/(1 + \gamma + \gamma')$, where for each $\gamma, \gamma' \in \boldsymbol{\Gamma}$, $c(\gamma, \gamma') := \gamma\gamma'(\gamma - 1)(\gamma' - 1)/|\gamma\gamma'(\gamma - 1)(\gamma' - 1)|$; (ii) when $\mathbb{E}[Y_t|\mathbf{D}_t] = \alpha_* + \mathbf{D}_t'\eta_* + \xi_* t$*

$+\ \beta_* t^{\gamma_*}$ *with* $\bar{\gamma} < \gamma_*$, *and* (ii.a) *if* $0 < \gamma_*$, $QLR_n/n = g^2(\gamma_*, \bar{\gamma})/\{g(\bar{\gamma}, \bar{\gamma})g(\gamma_*, \gamma_*)\} + o_p(1)$, *where* $g(\gamma_*, \gamma) := 1/(\gamma + \gamma_* + 1) + K/\{4(\gamma_* + 1)(\gamma + 1)\}$ *with* $K := \mathbb{E}[\widetilde{\mathbf{D}}_t]'\widetilde{\mathbf{Q}}^{-1}\mathbb{E}[\widetilde{\mathbf{D}}_t]$, $\widetilde{\mathbf{D}}_t := [1, \mathbf{D}_t']'$, *and* $\widetilde{\mathbf{Q}} := \mathbb{E}[\widetilde{\mathbf{D}}_t\widetilde{\mathbf{D}}_t'] - \frac{3}{4}\mathbb{E}[\widetilde{\mathbf{D}}_t]\mathbb{E}[\widetilde{\mathbf{D}}_t']$; (ii.b) *if* $-\frac{1}{2} < \gamma_* < 0$, $QLR_n/n^{1+2\gamma_*} = \beta_*^2 g^2(\gamma_*, \bar{\gamma})/\{\sigma_*^2 g(\bar{\gamma}, \bar{\gamma})\} + o_p(1)$; (iii) *when* $\mathbb{E}[Y_t|\mathbf{D}_t] = \alpha_* + \mathbf{D}_t'\eta_* + \xi_* t + m(t)$ *with* $m(\cdot)$ *being SSV, and* (iii.a) *if* $nm'(n) \to c(\neq 0)$, $QLR_n/n = \sup_{\gamma\in\boldsymbol{\Gamma}}\{cp(\gamma)\}^2/\{g(\gamma, \gamma)(\sigma_*^2 + c^2 q)\} + o_p(1)$, *where* $p(\gamma) := (\gamma - 1)\{(3\gamma + 5)/(\gamma + 1) - \frac{5}{4}K\}/\{4(\gamma + 1)(\gamma + 2)\}$ *and* $q := \frac{29}{16} - \frac{25}{64}K$; (iii.b) *if* $nm'(n) \to \infty$, $QLR_n/n = \sup_{\gamma\in\boldsymbol{\Gamma}} p(\gamma)^2/\{qg(\gamma, \gamma)\}$; (iv) *if* $\mathbb{E}[Y_t|\mathbf{D}_t] = \alpha_* + \mathbf{D}_t'\eta_* + \xi_* t + (\beta_*/n^{\gamma_*+1/2})t^{\gamma_*}$, $QLR_n \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}}(\widetilde{\mathcal{Z}}(\gamma) + \beta_* g(\gamma_*, \gamma)/\{\sigma(\gamma)\})^2$; (v) *if* $\mathbb{E}[Y_t|\mathbf{D}_t] = \alpha_* + \mathbf{D}_t'\eta_* + \xi_* t + m(t)/\{n^{3/2}m'(n)\}$, $QLR_n \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}}(\widetilde{\mathcal{Z}}(\gamma) + p(\gamma)/\{\sigma(\gamma)\})^2$. □

The proof of Theorem 6(i) is similar to those of Theorems 4 and 5. Note that $\{s_{n,t}\}$ is a sequence of non-random positive numbers uniformly bounded by unity and the MDS U_t satisfies the mixing condition of Assumption 6. From this, $\{n^{-1/2}\sum s_{n,t}^{(\cdot)}U_t\}$ is tight.

We focus here on two alternative nonstationary time trends — power functions and SSV trends. Theorem 6(ii–v) shows that the QLR test has consistent global and local powers under the maintained assumptions. For example, if $nm'(n) \to 0$ then the QLR test is not consistent under the assumptions of Theorem 4(iii).

The covariance structure of the associated Gaussian process is independent of the joint distribution of $(U_t, \mathbf{D}_t')$. Further, the same covariance structure applies irrespective of whether there is conditional heteroskedasticity in the residuals. We call the Gaussian process $\widetilde{\mathcal{Z}}(\cdot)$ with covariance kernel $\widetilde{\kappa}(\cdot,\cdot)$ the *power Gaussian process*, noting that $\widetilde{\mathcal{Z}}(\cdot)$ is obtained using the power transform of a trend. We note further that $\widetilde{\mathcal{Z}}(\cdot)$ is not continuous at $\gamma = 0$ and 1 as is evident from the functional form of $c(\cdot,\cdot)$.

The null distribution of the QLR test can be represented in terms of another Gaussian process. For this purpose, let $\bar{\mathcal{Z}}(\gamma) := \sum_{j=2}^{\infty}\left[\gamma^4/\{(\gamma+1)^2(2\gamma+1)\}\right]^{-1/2}\{\gamma/(\gamma+1)\}^j G_j$, where $G_j \sim$ IID $N(0, 1)$. When $\gamma > -0.5$, $[\gamma/(1+\gamma)]^j \to 0$ geometrically as $j \to \infty$, so that the covariance structure of this Gaussian process is well defined. This process coincides with the Gaussian process that appeared in Cho and White (2010) and Cho, Cheong, and White (2011) for testing unobserved heterogeneity in duration data. Notice that $\mathbb{E}[\bar{\mathcal{Z}}(\gamma)\bar{\mathcal{Z}}(\gamma')] = (1+2\gamma)^{1/2}(1+2\gamma')^{1/2}/(1+\gamma+\gamma')$. We call the Gaussian process with this covariance kernel the *exponential Gaussian process*. Although the power Gaussian process is different from the exponential Gaussian process, $\bar{\mathcal{Z}}(\cdot)^2$ is distributionally equivalent to $\widetilde{\mathcal{Z}}(\cdot)^2$. The next result immediately follows.

Theorem 7. *Given Assumptions 6 and 7, and* $\widetilde{\mathcal{H}}_0$, $QLR_n \Rightarrow \sup_{\gamma\in\boldsymbol{\Gamma}}\bar{\mathcal{Z}}(\gamma)^2$.

The exponential Gaussian process $\bar{\mathcal{Z}}(\cdot)$ can be easily simulated using a sequence of IID standard normal random variables $\{G_j\}$ and truncating the summation as in $\bar{\mathcal{Z}}_q(\gamma) := \sum_{j=2}^{q}\left[\gamma^4/\{(\gamma+1)^2(2\gamma+1)\}\right]^{-1/2}\{\gamma/(\gamma+1)\}^j G_j$ for some large q. Table 1 reports the asymptotic critical values by implementing this simulation for three levels of significance (1% 5% and 10%) and four different parameter spaces ([−0.20, 1.50], [−0.10, 1.50], [0.00, 1.50], [0.10, 1.50]). Specifically, we let q be 500 and simulate $\sup_{\gamma\in\boldsymbol{\Gamma}}\bar{\mathcal{Z}}_q(\gamma)^2$ 100,000 times to obtain the critical values. Greater powers of the test can be attained when the test is formulated using a space $\boldsymbol{\Gamma}$ that can better capture the alternative.

4. Simulations

We report simulation results to explore the finite sample properties of the QLR test using $\mathcal{M}$.

Table 1
Asymptotic Critical Values of the QLR Test Statistic. This table contains the asymptotic critical values obtained by generating the exponential Gaussian process 100,000 times. A grid search method is used to obtain the maximum of the squared process. The grid distance is 0.01, and q is 500.

Levels \ $\boldsymbol{\Gamma}$	[−0.20, 1.50]	[−0.10, 1.50]	[0.00, 1.50]	[0.10, 1.50]
10%	3.7186	3.6326	3.4669	3.4098
5%	4.9641	4.9065	4.7112	4.6196
1%	7.9861	7.9549	7.7336	7.6404

First, let the data (Y_t, X_t) be generated by $Y_t = 1 + X_t + U_t$, where $X_t := \exp(-H_t)$, $U_t \sim$ IID $N(0, 1)$, and $H_t = 0.5H_{t-1} + G_t$, $G_t = 0$ and E_t w.p. 0.5 and 0.5, respectively; and $E_t \sim$ IID Exp(1) such that U_t is independent of H_t. Given this DGP, we specify $\mathcal{M} = \{m_t(\cdot) : m_t(\alpha, \xi, \beta, \gamma) = \alpha + \xi X_t + \beta X_t^{\gamma}, \gamma \in \boldsymbol{\Gamma}\}$ as the model for $\mathbb{E}[Y_t|X_t]$. We consider the same parameter spaces for $\boldsymbol{\Gamma}$ as used in Table 1. In particular, the parameter space [0.10, 1.50] does not contain zero, reducing the scope of the trifold identification problem, because the number of unidentified model cases is reduced. The associated Gaussian process with the QLR test has the same covariance structure as that of the power Gaussian process mainly from that U_t is IID with conditionally homoskedastic variance and H_t marginally follows an exponential distribution with population mean unity. We also compare the QLR test with Bierens and Ploberger (1997) ICM test defined as $\int_{\boldsymbol{\Gamma}}\widehat{Z}_n(\gamma)^2 d\gamma$, where $\widehat{Z}_n(\gamma) := \frac{1}{\sqrt{n}}\sum\widehat{U}_t\exp(\gamma\Phi(X_t))$, $\widehat{U}_t$ is the residual obtained under the linear model assumption, and $\Phi(\cdot)$ is the standard normal cdf.

Table 2 contains the empirical rejection rates of the null hypothesis in round parentheses obtained from 5000 replications. The significance level is 5% and the findings are as follows. First, for each parameter constellation, the empirical rejection rates approach the nominal levels as n increases. Second, convergence to the nominal levels tends to be slower when the lower bound of $\boldsymbol{\Gamma}$ is closer to −0.50 and the upper bound is the same. Level distortion in the test can therefore be reduced by raising the lower bound of $\boldsymbol{\Gamma}$ from the minimum. Third, convergence to the nominal level improves as the upper bound of $\boldsymbol{\Gamma}$ increases, with the same lower bound. Thus level distortion may be attenuated by using a higher upper bound of $\boldsymbol{\Gamma}$. The Table also provides (in parentheses) the estimated p-values obtained by applying Hansen's (1996) weighted bootstrap and compares these with those of the ICM test. This procedure shows better performance than the asymptotic critical values and strengthens implementation of the test. The overall level distortion of the QLR test is smaller than that of the ICM test.

Next, we compare the global and local powers of the tests. For this purpose, we let $Y_t = 1 + X_t + \log(X_t) + U_t$ and $Y_t = 1 + X_t + \log(X_t)/\sqrt{n} + U_t$ for global and local power examination, respectively. The conditions for X_t and U_t are identical to those of Table 2. Table 3 contains the empirical rejection rates, and the figures in parentheses are local powers. The number of replications used is 2000 (5000) for calculation of global (respectively, local) power. Overall in these experiments, the global and local powers of the QLR test are higher than those of the ICM test. Although these results do not imply that power of the QLR test always dominates that of the ICM test, they do indicate that the QLR test is superior in many cases and is therefore highly competitive test. For brevity, we do not report other simulations that were conducted, including those for time trends and errors exhibiting conditional heteroskedasticity. Interested readers can refer to the earlier version for more details.

5. Empirical applications

A popular empirical topic that comes within the aegis of the present work is the identification detrended crop-yield distributions in agricultural economics. Characterizations of production

Table 2

Levels of the QLR and ICM Test Statistics (Significance Level: 5%). Number of Repetitions: 5000. Model: $Y_t = \alpha + \xi X_t + \beta X_t^{\gamma} + U_t$. DGP: $Y_t = 1 + X_t + U_t$, $X_t := \exp(-H_t)$, $U_t \sim$ IID $N(0, 1)$, $H_t = 0.5H_{t-1} + G_t$, $G_t = 0$ w.p. 0.5 ; $G_t = E_t$ w.p. 0.5; and $E_t \sim$ IID Exp(1) such that U_t is independent of G_t. The figures are the empirical rejection rates obtained by the weighted bootstrap, and the figures in round parentheses are the empirical rejection rates obtained by applying the critical values in Table 1; the number of bootstrap iterations is 500; and the figures are measured in per cent.

Γ	Tests $\backslash n$	50	100	200	300	400	500
[−0.20, 1.50]	QLR	4.90 (4.00)	4.96 (4.42)	4.44 (3.98)	4.84 (4.22)	4.34 (3.86)	4.88 (4.58)
	ICM	6.54	5.84	5.22	5.26	5.42	5.08
[−0.10, 1.50]	QLR	4.52 (3.84)	5.06 (4.48)	4.66 (4.20)	4.78 (4.50)	4.54 (4.26)	4.48 (4.16)
	ICM	6.78	4.96	5.80	5.24	5.30	4.80
[0.00, 1.50]	QLR	4.90 (4.20)	5.08 (4.94)	4.74 (4.82)	5.10 (5.12)	5.22 (5.10)	5.44 (5.44)
	ICM	6.42	5.98	5.42	5.16	5.70	5.20
[0.10, 1.50]	QLR	5.36 (4.78)	5.42 (5.24)	5.24 (5.02)	4.98 (5.00)	4.82 (4.80)	5.60 (5.40)
	ICM	6.44	5.94	4.56	5.54	5.42	4.98

Table 3

Global and Local Powers of the QLR and ICM Test Statistics (Significance Level: 5%). Number of Repetitions: 2000 (resp. 5000). Model: $Y_t = \alpha + \xi X_t + \beta X_t^{\gamma} + U_t$. Global DGP: $Y_t = 1 + X_t + \log(X_t) + U_t$, and the other conditions are identical to those of Table 2. Local DGP: $Y_t = 1 + X_t + \log(X_t)/\sqrt{n} + U_t$, and the others are identical to the global DGP. The figures (resp. in round parentheses) are the empirical rejection rates obtained by the weighted bootstrap under the global (resp. local) DGP; the number of bootstrap iterations is 500; and the figures are measured in per cent.

Γ	Tests $\backslash n$	50 (100)	100 (200)	150 (300)	200 (400)	250 (500)
[−0.20, 1.50]	QLR	71.85 (7.60)	94.15 (7.02)	98.60 (7.34)	99.90 (7.82)	99.95 (7.64)
	ICM	41.75 (6.14)	67.25 (5.96)	84.65 (5.58)	93.55 (6.06)	97.35 (5.92)
[−0.10, 1.50]	QLR	71.85 (7.66)	94.40 (7.60)	99.40 (7.16)	99.70 (7.20)	99.95 (7.54)
	ICM	41.65 (6.20)	66.80 (5.52)	85.65 (5.48)	93.75 (5.56)	97.60 (5.82)
[0.00, 1.50]	QLR	69.00 (7.22)	93.40 (7.22)	99.05 (8.38)	99.80 (7.30)	100.0 (6.80)
	ICM	41.35 (6.26)	67.15 (6.26)	83.70 (6.12)	93.85 (5.88)	97.75 (5.78)
[0.10, 1.50]	QLR	72.15 (7.86)	94.55 (7.58)	98.75 (7.74)	99.95 (7.72)	99.95 (7.50)
	ICM	42.25 (6.22)	66.85 (5.40)	84.55 (5.72)	93.75 (5.98)	97.10 (5.66)

have significant implications for crop insurance and farming business. For this reason, a key focus in the literature has been on identifying whether detrended crop yields follow a normal distribution in order to facilitate convenient use of the mean–variance principle for expected utility maximization. Detrending is an important feature of this process in order to remove technology bias in estimating the underlying distribution.

Many controversies are present concerning this identification process. These can be classified into two groups. Swinton and King (1991), Raminez et al. (2003) among others report that detrended crop yields are skewed and non-Gaussian. On the other hand, Just and Weninger (1999) point to methodological problems in previous identification work, reporting that Gaussianity cannot be easily rejected when proper corrections are made for trends in the data. They claim the biggest methodological problem is potential misspecification of the time trend model, which is precisely the central concern of Section 3.

We apply our methodology to the data used by Just and Weninger (1999) and explore support for their empirical findings. When Y_t is the crop yield, we specify the following polynomial model for trend: $Y_t = \alpha_* + \sum_{j=1}^{p} \xi_{j*} t^j + U_t$. Just and Weninger (1999) select the integer trend degree p by Akaike's information criterion (AIC) and show that a number of crop yields without trend or with a linear trend have Gaussian errors. We apply the QLR test and, if their trend assumption is refuted, we test normality using residuals obtained by our alternative methods that allow for non-integer trends. Otherwise, we test normality using residuals obtained from the no-trend or linear trend model.

For this purpose, we collect data from the US department of agriculture, following Just and Weninger (1999). They consider time series data on alfalfa, corn, sorghum, soybeans and wheat obtained from Finney, Ford, Gray, and Hodgeman counties in Kansas. Of these 20 series, they report that 9 series do not have trend or have a linear trend. Results for these 9 series are given in Table 4. The first panel provides results using the same sample of data as those in Just and Weninger (1999). The polynomial orders in Table 4 are selected by AIC with a correction for finite sample sizes (AICc), and these are same as those given in Just and Weninger (1999). Next we sequentially apply the QLR test, testing for any neglected trend, so that the null is simply $\mathbb{E}[Y_t] = \alpha_*$. Although the null is different from that in Theorem 7, it is a simple exercise to derive the same asymptotic null distribution for the QLR test in this case. We let the parameter space for γ be [0.00, 2.50], noting that trend is positive mainly due to technological developments. This QLR test and its p-value are denoted as $\text{QLR}^{\dagger 1}$ and $p\text{-value}^{\dagger 1}$, respectively. We test the null using the critical values in Table 1. The null of those series with AICc-based order equal to one is rejected by $\text{QLR}^{\dagger 1}$. For the others, the null is hard to reject. We next test the linear trend null $\mathbb{E}[Y_t] = \alpha_* + \xi_* t$. The QLR test and its p-value are denoted as $\text{QLR}^{\dagger 2}$ and $p\text{-value}^{\dagger 2}$, respectively. We also apply the weighted bootstrap, and its p-value is denoted as $p\text{-value}^{\S}$. We let the parameter space for γ be [−0.20, 2.50]. Note that none of the 9 cases is rejected by the QLR test, implying that the 9 data series have at most a linear trend. The same conclusion is also obtained even when the weighted bootstrap is applied, affirming the trend specification of Just and Weninger (1999). Finally, the Jarque and Bera (1980) statistic is applied to test normality. Their test and its p-value are denoted as $\text{JB}^{\dagger 3}$ and $p\text{-value}^{\dagger 3}$, respectively. Seven cases out of 9 turn out to accept the Gaussian null. This feature generally matches the results obtained by Just and Weninger (1999), although their empirical analysis shows that the alfalfa data of Gray county are also normally distributed.

We next extend the series to currently available samples and repeat the above procedure to see if the previous findings are corroborated in the longer series. The second panel of Table 4 provides the empirical results. Three data sets (Sorghum (Gray), Alfalfa (Finney), and Soybeans (Ford)) turn out to have nonlinear trends according to AICc. The QLR tests also reject the linear trend assumption for Sorghum (Gray) and Alfalfa (Finney) data sets, although the QLR test cannot reject the linear trend assumption for Soybean (Ford) data. When applying the weighted bootstrap, its p-value is 18.3%. This is different from the AICc-based result. We also note that normality is rejected by only the Soybean (Ford) data, so that only one series appears to be non-Gaussian out of these 9 cases. These results reinforce the empirical findings of Just and Weninger (1999) that the majority of the detrended crops examined follow Gaussian

Table 4

Summary of Empirical Findings. Order denotes the polynomial order selected by AICc; QLR$^{\dagger 1}$ and p-val.$^{\dagger 1}$ denote the QLR statistic and its p-value that test the no-trend assumption; QLR$^{\dagger 2}$ and p-val.$^{\dagger 2}$ denote the QLR statistic and its p-value that test the linear trend assumption; p-val.§ denotes the p-value of the QLR$^{\dagger 2}$ obtained by the weighted bootstrap; JB$^{\dagger 3}$ and p-val$^{\dagger 3}$ denote the Jarque and Bera (1980) statistic and its p-value that tests normality; data are obtained from http://quickstats.nass.usda.gov; alfalfa yields are measured in tons per acre, and the other is measured in bushels per acre; p-values are figures in per cent; and the number of bootstrap iterations is 50,000.

Crops	Counties	Sample	Order	QLR$^{\dagger 1}$	p-val.$^{\dagger 1}$	QLR$^{\dagger 2}$	p-val.$^{\dagger 2}$	p-val.§	JB$^{\dagger 3}$	p-val.$^{\dagger 3}$
Sorghum	Finney	80–94	1	4.32	7.01*	2.10	27.3	21.3	1.25	53.3
	Ford	80–94	0	0.26	82.3	0.11	94.8	87.9	0.70	70.2
	Gray	80–94	0	2.30	21.9	1.28	44.2	35.8	0.73	69.3
	Hodgeman	80–94	0	1.19	43.2	2.45	22.4	17.7	2.26	32.2
Alfalfa	Finney	60–93	1	15.19	0.00***	1.41	40.9	33.5	2.85	24.0
	Ford	60–93	1	8.05	1.01**	0.11	94.8	90.9	0.29	86.1
	Gray	60–93	1	7.10	1.70**	1.37	41.9	33.7	6.25	4.38**
Soybeans	Ford	69–94	1	19.41	0.00***	0.32	82.3	70.0	0.19	90.7
	Gray	65–94	1	23.17	0.00***	0.45	75.2	63.3	9.51	0.85***
Sorghum	Finney	80–07	0	0.37	75.7	1.02	52.0	39.0	2.27	32.0
	Ford	80–07	0	1.43	37.1	0.30	83.4	70.7	1.82	40.0
	Gray	80–07	4	9.76	0.40***	4.83	6.04*	4.85**	0.94	62.4
	Hodgeman	80–07	0	0.53	67.2	0.12	94.1	88.8	0.84	65.5
Alfalfa	Finney	60–07	3	12.63	0.10***	4.58	6.90*	4.56**	3.98	13.6
	Ford	60–07	1	12.51	0.11***	1.45	40.0	34.1	1.13	56.5
	Gray	60–07	1	8.71	0.70***	1.21	46.2	37.7	4.41	10.9
Soybeans	Ford	69–02	3	20.00	0.00***	2.41	22.8	18.3	0.50	77.5
	Gray	65–02	1	28.08	0.00***	0.17	91.1	84.3	10.28	0.58***

* Denotes "significant at 10%".
** Denotes "significant at 5%".
*** Denotes "significant at 1%".

distributions. Thus, while the focus of the present methodology is on testing linearity, the current example shows that the methods may be used as an alternative to order selection methods by directly estimating (potentially non-integer) trend degree and conducting tests using the QLR procedure.

6. Conclusion

Linear models continue to be the mainstay of much empirical research, making specification tests of linearity an important feature of model robustness checks. Power transforms offer a natural alternative to linearity and provide a more general framework than simple polynomial specifications. However, tests of linearity in models using power transforms raise critical identification issues, which amount to a trifold identification problem. The approach adopted here resolves these issues by using a QLR test to provide a unified mechanism for capturing the trifold forms of the null hypothesis.

Under some weak conditions, the asymptotic null distribution of the QLR test is shown to be a functional of a Gaussian stochastic process. The limit theory for the stationary regressor case is extended to a model with a time trend and stationary regressors. For such cases, the QLR test has an asymptotic null distribution that takes the form of a functional of a power Gaussian process. Asymptotic critical values of the QLR test are obtained, and simulations confirm the asymptotic theory. An empirical application of our methodology to agricultural crop yields affirms earlier findings by Just and Weninger (1999).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2015.03.041.

References

Baek, Y.I, Cho, J.S., Phillips, P.C.B., 2014, Mathematical Proofs for Testing Linearity using Power Transforms of Regressors. http://web.yonsei.ac.kr/jinseocho/.
Bierens, H., 1982. Consistent model specification tests. J. Econometrics 20, 105–134.
Bierens, H., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443–1458.
Bierens, H., Ploberger, W., 1997. A consistent conditional moment test of functional form. Econometrica 65, 1129–1151.
Box, G., Cox, D., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B 26, 211–252.
Cho, J.S., Cheong, T., White, H., 2011. Experience with the weighted bootstrap in testing for unobserved heterogeneity in exponential and Weibull duration models. J. Econom. Theory and Econom. 22 (2), 60–91.
Cho, J.S., Ishida, I., 2012. Testing for the effects of omitted power transformations. Econom. Lett. 117, 287–290.
Cho, J.S., Ishida, I., White, H., 2011. Revisiting tests for neglected nonlinearity using artificial neural network. Neural Comput. 23, 1133–1186.
Cho, J.S., Ishida, I., White, H., 2014. Testing for neglected nonlinearity using twofold unidentified models under the null and hexic expansions. In: Haldrup, Niels, Meitz, Mika, Saikkonen, Pentti (Eds.), Essays on Nonlinear Time Series Econometrics: A Festschrift in Honor of Timo Teräsvirta. Oxford University Press, Oxford, pp. 3–27.
Cho, J.S., White, H., 2010. Testing for unobserved heterogeneity in exponential and Weibull duration models. J. 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.
Hansen, B., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.
Jarque, C., Bera, A., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econom. Lett. 6, 255–259.
Just, R., Weninger, Q., 1999. Are crop yields normally distributed? Am. J. Agric. Econ. 81, 287–304.
Park, J.Y., Phillips, P.C.B., 1988. Statistical inference in regressions with integrated processes: Part 1. Econometric Theory 4, 468–497.
Park, J.Y., Phillips, P.C.B., 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory 15, 269–298.
Phillips, P.C.B., 1989. Partially identified econometric models. Econometric Theory 5, 181–240.
Phillips, P.C.B., 2007. Regression with slowly varying regressors and nonlinear trends. Econometric Theory 23, 557–614.
Ramsey, J.B., 1969. Tests for specification errors in classical linear least squares regression analysis. J. R. Stat. Soc. Ser. B 31, 350–371.
Raminez, O., Misra, S., Field, J., 2003. Crop-yield distributions revisited. Am. J. Agric. Econ. 85, 108–120.
Sakia, R., 1992. The Box–Cox transformation technique: A review. The Statistician 41, 169–178.
Shi, X., Phillips, P.C.B., 2012. Nonlinear cointegrating regression under weak identification. Econometric Theory 28, 509–547.
Stinchcombe, M., White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory 14, 295–324.
Swinton, S., King, R., 1991. Evaluating robust regression techniques for detrending crop yield data with nonnormal errors. Am. J. Agric. Econ. 73, 446–451.
Tukey, J., 1957. On the comparative anatomy of transformations. Ann. Math. Statist. 28, 602–632.
Tukey, J., 1977. Exploratory data analysis. In: Reading. Addison-Wesley, MA.
White, H., Cho, J.S., 2012. Higher-order approximations for testing neglected nonlinearity. Neural Comput. 24, 273–287.
Wu, C.-F., 1981. Asymptotic theory of nonlinear least squares estimation. Ann. Statist. 9, 501–513.