ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

A LOW-DIMENSION PORTMANTEAU TEST FOR NON-LINEARITY

Jennifer L. Castle and David F. Hendry

Number 471
January 2010

Manor Road Building, Oxford OX1 3UQ

A Low-Dimension Portmanteau Test for Non-linearity

Jennifer L. Castle and David F. Hendry*
Department of Economics, Oxford University.

April 15, 2009

Abstract

A new test for non-linearity in the conditional mean is proposed using functions of the principal components of regressors. The test extends the non-linearity tests based on Kolmogorov–Gabor polynomials (Thursby and Schmidt, 1977, Tsay, 1986, and Teräsvirta, Lin and Granger, 1993), but circumvents problems of high dimensionality, is equivariant to collinearity, and includes exponential functions, so is a portmanteau test with power against a wide range of possible alternatives. A Monte Carlo analysis compares the performance of the test to the optimal infeasible test and to alternative tests. The relative performance of the test is encouraging: the test has the appropriate size and has high power in many situations.

Keywords: Functional Form; Portmanteau test; Non-linearity; Principal Components; Collinearity.

JEL Classification: C51; C52.

1 Introduction

It is a great pleasure to contribute a paper in honor of Phoebus Dhrymes, whose many major publications have helped establish econometrics in its modern form. Our topic is the validity of the functional form of a model, which is an essential component of its correct specification: both are topics on which Phoebus has published (see e.g., Dhrymes, 1966, and Dhrymes, Howrey, Hymans, Kmenta *et al.*, 1972). Nevertheless, we consider that a test for non-linearity is required that can evaluate the 'goodness' of a postulated model against general non-linear alternatives, particularly in the context where there would be more non-linear terms to include than available observations. Evidence of non-linearity implies that an alternative functional form should be utilized. Consequently, we propose a general portmanteau test for non-linearity, which is designed to accommodate large numbers of potential regressors, so is applicable prior to undertaking model estimation or selection.

Our proposed test is based on a third-order polynomial with additional exponential functions, formed from the principal components of the original variables, which allows us to obtain a highly flexible non-linear approximation yet in a parsimonious formulation. If there are (say) $n = 10$ linear variables, there are 55 quadratic functions and 220 cubics, adding 275 variables to test for a general third-order polynomial, whereas our proposed test would require just 30 functions yet check exponentials as well; for 30 linear regressors, an unmanageable 5425 additional terms would be required–as against 90.

*We thank the Editors and two anonymous referees of the *Journal of Econometrics*, Jurgen A. Doornik, participants of the 2006 Royal Economic Society Conference, Econometric Society Australasian Meeting, Econometric Society European Meeting and the 4th Oxmetrics Users Conference for helpful comments and suggestions on an earlier version. Financial support from the ESRC under grants RES-000-23-0539 and RES-062-23-0061 is gratefully acknowledged.

Many tests of linearity have been proposed in the literature: see Granger and Teräsvirta (1993, ch.6) for an overview. Ramsey (1969) proposed tests for specification errors in regression, including unmodeled non-linearity, based on adding powers of the fitted values: Doornik (1995) provided a careful evaluation of both the numerical and statistical properties of the RESET test. Thursby and Schmidt (1977) extended the RESET test by proposing the use of powers and cross products of explanatory variables, as well as another variant that used principal components of the explanatory variables: simulation studies across a range of specifications suggested the preferred test was one based on square, cubic and quartic powers of the explanatory variables. Lee, White and Granger (1993) utilized these generalized RESET tests, which took principal components of the polynomials, in their simulation study to control for collinearity. Keenan (1985) developed a univariate test for detecting non-linearity based on a Volterra expansion, an analogue of the Tukey (1949) one degree-of-freedom test for non-additivity: the test checks whether the squared fit is correlated with the estimated residuals, so is a variant of the Ramsey (1969) test. Tsay (1986) improved on the power of the Keenan (1985) test by allowing for disaggregated non-linear variables (all $n\,(n+1)\,/2$ cross products for n linear variables), based on a Kolmogorov–Gabor polynomial. Bierens (1990) developed a conditional moment test of functional form.

White (1989) and Lee *et al.* (1993) proposed a neural network test for neglected non-linearity, in which they undertook a Lagrange multiplier test for hidden activation functions, typically logistic functions, after pre-filtering using an AR(p) model selected by BIC. As the hidden unit activations, denoted $\boldsymbol{\Psi}_t$, tended to be collinear with the regressors, they suggested using principal components of the $\boldsymbol{\Psi}_t$, and we compare a variant of this idea to our proposed test. Teräsvirta *et al.* (1993) showed that such a test had low power in many situations, and was not robust to the inclusion of an intercept. They proposed a related test based on a third-order polynomial that approximated the single hidden-layer artificial neural network model of Lee *et al.* (1993), but had a more general application. This test will suffer when there are many regressors, as the number of non-linear functions, $C_n = n\,(n+1)\,(n+5)\,/6$, increases rapidly with n.

White (1982, 1992, ch.9) proposed a range of dynamic information matrix tests based on the covariance of the conditional score functions. The tests have power against mis-specifications that induce autocorrelation in the conditional scores. Brock, Dechert and Scheinkman (1987) developed a non-parametric test for serial independence based on the correlation integral of a scalar series, which emerged from the chaos theory literature, and McLeod and Li (1983) developed a test against ARCH which has power against non-linearity in the conditional mean, and which is asymptotically equivalent to the LM ARCH test developed by Engle (1982). Finally, the White (1980) test for heteroskedasticity can be considered a test of linearity in which the alternative hypothesis is a doubly stochastic model given by:

$$y_t = \boldsymbol{\alpha}_t'\mathbf{x}_t + \epsilon_t, \qquad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{1}$$

with $\mathbf{x}_t = (x_{1,t}, \ldots, x_{n,t})'$, $\boldsymbol{\alpha}_t \sim \mathsf{IID}_n\left[\boldsymbol{\alpha}, \boldsymbol{\Sigma}_\alpha\right]$, where $\boldsymbol{\Sigma}_\alpha$ is positive definite and $\mathsf{E}\left[\boldsymbol{\alpha}_t\epsilon_s\right] = 0, \forall s, t$. The test adds all squares of regressors, or squares and cross-products, to test for heteroskedasticity, implicitly testing for omitted non-linearity as well: see Spanos (1986, p.466). This has been investigated by numerous authors, including a recent appraisal in Hendry and Krolzig (2003) in the context of model selection, and has direct parallels to the approach we propose.

Although the literature on non-linearity testing is substantial, our test is designed for a setting that is not yet handled well, namely, high dimensional, possibly relatively collinear, specifications where the functional form of the non-linearity is not known *a priori*. Standard tests based on second-order polynomials such as White (1980) and Tsay (1986) face several practical drawbacks: first, their increasing dimensionality with n; secondly, the potentially high collinearity between powers of regressors that are slowly changing (see Phillips, 2007, for a time series analogue); and third, the possibility that the second derivative is not the source of the departure from linearity. To rectify these potential drawbacks, our test first forms the n standardized, mutually-orthogonal principal components $\mathbf{z}_t$ of the original n linear regressors $\mathbf{x}_t$, with weights given by the eigenvectors of their estimated variance-covariance matrix. Then, for fixed regressors, adding the squares, $z_{i,t}^2$, cubes $z_{i,t}^3$, and exponential functions of the $z_{i,t}$, yields an F-test with $3n$ degrees of freedom, jointly resolving the problems of high dimensionality, collinearity, and restriction to second-order departures. The corresponding 'unrestricted' function of all these terms would involve adding $R_n = C_n + n$ elements, which is often infeasible, and usually low powered for large n. When the regressors are weakly exogenous (see Engle, Hendry and Richard, 1983), an approximate F-test results in stationary models.

The structure of the paper is as follows. Section 2 considers functional-form testing, and proposes a test for situations where many general tests are infeasible. Section 3 discusses the statistical power of the proposed test, computes the non-centrality, and examines its power approximations for scalar polynomials. Section 4 describes the range of alternative tests we will compare against, then section 5 undertakes a set of Monte Carlo experiments to examine the comparative power of the test for various IID data generation process (DGP) designs. Testing for non-linearity in dynamic models is considered in section 6, and section 7 concludes.

2 Testing functional form

Given the general functional relationship for $t = 1, \ldots, T$:

$$y_t = f\left(x_{1,t} \ldots x_{n,t}\right) + \epsilon_t \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{2}$$

the linear approximation is:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_n x_{n,t} + e_t = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_t + e_t \tag{3}$$

where $\mathbf{x}_t$ needs to be weakly exogenous for $\boldsymbol{\beta}$. To evaluate (3) requires testing the validity of the functional form approximation. We first consider the case where $f(\mathbf{x}_t)$ is quadratic, which highlights the key considerations, then consider a more general approximation that avoids the three drawbacks of dimensionality, collinearity and only quadratic departures, by developing an Index test.

2.1 Testing against a quadratic

First, consider the optimal test for linearity when $f(\mathbf{x}_t)$ is the exact quadratic:

$$f\left(\mathbf{x}_t\right) = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_t + \gamma \mathbf{x}_t' \mathbf{A} \mathbf{x}_t \tag{4}$$

and $\mathbf{A}$ is known up to a constant factor and symmetric (without loss of generality), so $\mathbf{x}_t' \mathbf{A} \mathbf{x}_t = u_t$ is known and $\gamma \neq 0$. Then when (4) holds:

$$y_t = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_t + \gamma u_t + \epsilon_t \tag{5}$$

so a t-test of H_0: $\gamma = 0$ in (5) will be the most powerful test for non-linearity.

2.2 A quadratic approximation test

However, as $\mathbf{A}$ is unknown in general, a 'natural' test is to consider the importance of adding all the quadratic terms to (3). Let:

$$\mathbf{w}_t = \textit{vech}\left(\mathbf{x}_t\mathbf{x}_t'\right) \tag{6}$$

where *vech* vectorizes and selects non-redundant elements from the lower triangle including the diagonal of the outer product. Then, an exact test for fixed regressors would be an $\mathsf{F}^{n(n+1)/2}_{T-(n+2)(n+1)/2}$-test of the null $\boldsymbol{\delta}_1 = \mathbf{0}$ in:

$$y_t = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\delta}_1'\mathbf{w}_t + e_t \tag{7}$$

When $\mathbf{A}$ is unknown, (7) provides one operational counterpart (analogous to the third-order polynomial test proposed by Spanos, 1986, p.460). However, for a non-collinear test, let:

$$\mathbf{x}_t \sim \mathsf{D}_n\left[\boldsymbol{\mu}, \boldsymbol{\Omega}\right] \tag{8}$$

where $\boldsymbol{\Omega}$ is the symmetric, positive-definite variance-covariance matrix. Factorize $\boldsymbol{\Omega} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}'$, where $\mathbf{H}$ is the matrix of eigenvectors of $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$ the corresponding eigenvalues, so $\mathbf{H}'\mathbf{H} = \mathbf{I}_n$. Since $\boldsymbol{\Lambda}^{-1/2}\mathbf{H}'\boldsymbol{\Omega}\mathbf{H}\boldsymbol{\Lambda}^{-1/2} = \mathbf{I}_n$, let:

$$\mathbf{z}_t^* = \boldsymbol{\Lambda}^{-1/2}\mathbf{H}'\left(\mathbf{x}_t - \boldsymbol{\mu}\right) \sim \mathsf{D}_n\left[\mathbf{0}, \mathbf{I}\right] \tag{9}$$

Replacing $\mathbf{x}_t$ by $\mathbf{z}_t^*$ in (6) and (7) does not affect the test of $\boldsymbol{\delta}_1 = \mathbf{0}$, merely transforming the data to its complete set of principal components, so is equivariant to collinearity, a property that will prove useful below.

2.3 A low-dimensional quadratic test

Instead of including all of the quadratic terms, let $u_{*,i,t} = (z_{i,t}^*)^2$, then under the null that $\gamma = 0$ in (4), the test of $\boldsymbol{\kappa}_1 = \mathbf{0}$ in:

$$y_t = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\kappa}_1'\mathbf{u}_{*,t} + v_t \tag{10}$$

is an exact F-test with n degrees of freedom, for fixed regressors. There are only n elements in $\mathbf{u}_{*,t}$, so relative to one in u_t in (5), power will be lower. However, n is many fewer terms than $n(n+1)/2$ in $\mathbf{w}_t$ in (7), yet every element in $\mathbf{u}_{*,t}$ potentially depends on squares and cross-products of every $x_{i,t}$. Thus, the first and second objectives–effecting a major dimensionality reduction and formulating a test in terms of non-collinear variables–have been achieved. The test based on (10) is equivariant to whether the linear terms are $\mathbf{x}_t$ or $\mathbf{z}_t^*$, and if transformed to the latter, corresponds to adding their squares (see e.g., White, 1980).

Since $\boldsymbol{\Omega}$ is unknown, a symmetric, positive-definite estimate $\widehat{\boldsymbol{\Omega}}$ is used when calculating principal components, denoted $\mathbf{z}_t$ so:

$$\mathbf{z}_t = \widehat{\boldsymbol{\Lambda}}^{-1/2}\widehat{\mathbf{H}}'\left(\mathbf{x}_t - \widehat{\boldsymbol{\mu}}\right) \underset{\widetilde{app}}{} \mathsf{D}_n\left[\mathbf{0}, \mathbf{I}\right] \tag{11}$$

Then the $z_{i,t}$ in (11) are the standardized, mutually-orthogonal combinations of the original $x_{i,t}$ to be used for the operational quadratic version of the test via $u_{1,i,t} = z_{i,t}^2$.

2.4 Testing against more general alternatives

To accommodate the third drawback, and generalize the test for higher-derivative departures from the null, we also include $u_{2,i,t} = z_{i,t}^3$ to capture skewness. Since polynomial approximations may be slow to converge when the approximated series involves exponentials, and tend not to be a parsimonious approximation for other functional forms, Abadir (1999) suggested confluent hypergeometric functions as a more general class. We examined simple exponential functions such as $\{e^{z_{i,t}}\}$, $\{e^{z_{i,t}} z_{i,t}\}$ and $\{e^{-|z_{i,t}|}\}$ but these yielded almost no additional information over the polynomial approximation for a range of alternatives. However, a product like $u_{3,i,t} = e^{-|z_{i,t}|} z_{i,t}$ does yield additional flexibility as well as help capture overall asymmetry, with the possible disadvantage that it is not differentiable. Since the $z_{i,t}$ are standardized as in (11), both signs occur equally often on average, so adding $u_{1,i,t}$, $u_{2,i,t}$, $u_{3,i,t}$ as in (12) yields the low-dimensional portmanteau $\mathsf{F}^{3n}_{T-(4n+1)}$-test of $\boldsymbol{\kappa}_1 = \boldsymbol{\kappa}_2 = \boldsymbol{\kappa}_3 = \mathbf{0}$ in:

$$y_t = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\kappa}_1'\mathbf{u}_{1,t} + \boldsymbol{\kappa}_2'\mathbf{u}_{2,t} + \boldsymbol{\kappa}_3'\mathbf{u}_{3,t} + e_t \tag{12}$$

Under the null, $e_t = \epsilon_t$, so the test is distributed as F for fixed regressors. Under the alternative:

$$\begin{aligned} y_t &= \beta_0 + \sum_{i=1}^{n} \left(\beta_i z_{i,t} + \kappa_{1,i} z_{i,t}^2 + \kappa_{2,i} z_{i,t}^3 + \kappa_{3,i} z_{i,t} e^{-|z_{i,t}|} \right) + e_t \\ &\approx \beta_0 + \sum_{i=1}^{n} \left(\beta_i z_{i,t} + \gamma_i z_{i,t} \left(1 - |z_{i,t}|\right) + \kappa_{1,i} z_{i,t}^2 + \kappa_{2,i} z_{i,t}^3 + \theta_i z_{i,t}^3 \left(1 - \frac{1}{3} |z_{i,t}|\right) \right) + e_t \end{aligned} \tag{13}$$

Thus, the approximation is quite flexible: the model is constant if $\boldsymbol{\beta} = \boldsymbol{\kappa}_1 = \boldsymbol{\kappa}_2 = \boldsymbol{\kappa}_3 = \mathbf{0}$; linear if $\boldsymbol{\kappa}_1 = \boldsymbol{\kappa}_2 = \boldsymbol{\kappa}_3 = \mathbf{0}$; quadratic if $\boldsymbol{\kappa}_2 = \boldsymbol{\kappa}_3 = \mathbf{0}$; cubic if $\boldsymbol{\kappa}_3 = \mathbf{0}$; 'bi-linear with bi-quadratic' if $\boldsymbol{\beta} = \boldsymbol{\kappa}_2 = \mathbf{0}$; and fairly general if all coefficients are non-zero. The non-linear bases included in (13) are recorded in figure 1: a wide range of possible non-linearities can be detected by combinations of the three proposed here. $e^{-\phi z_i^2} z_i$ was considered as well, but yielded almost no additional information over the exponential product included, as can be seen from replacing $|z_{i,t}|$ by $z_{i,t}^2$ in (13).[1]

We call this joint F-test the Index-test, as only $3n$ additional regressors are needed, against R_n if all terms were entered unrestrictedly. Despite including terms for quadratic, cubic and exponential departures as shown, Index-test remains computable provided $4n < T$, whereas $R_n < T$ would be needed unrestrictedly (e.g., allowing $n \simeq 24$ versus $n \simeq 6$ at $T = 100$). We next investigate its null distribution, then consider its power in several canonical settings.

2.5 Null distributions

Since $\mathbf{z}_t$ is just an orthogonal transformation of $\mathbf{x}_t$, Index-test needs the same assumptions as almost all the other tests considered in section 4 (other than RESET: see Caceres, 2007). To check that Index-test is distributed as an F^{3n}_{T-4n-1} in finite samples even when there would be more variables than observations in the unrestricted test, so that $R_n > T$, we examined its QQ-plots under the null for the static experiments in section 5.1. The results closely corresponded to an F-distribution. Collinearity between regressors did not affect the null distribution. We also checked the null distributions for fixed regressors of those alternative tests, including the RESET test, V23 (Teräsvirta *et al.*, 1993), principal components of V23, denoted PCV23, V2 (Tsay, 1986), and principal components of V2, PCV2, confirming all these tests are

[1] As $e^{z_{i,t}} = 1 + z_{i,t} + \frac{1}{2} z_{i,t}^2 + \frac{1}{6} z_{i,t}^3 + \frac{1}{24} z_{i,t}^4 + O\left(z_{i,t}^5\right)$, and $z_{i,t}^2$ and $z_{i,t}^4$ tend to be highly correlated, $\exp(z_{i,t})$ adds little to a cubic approximation.

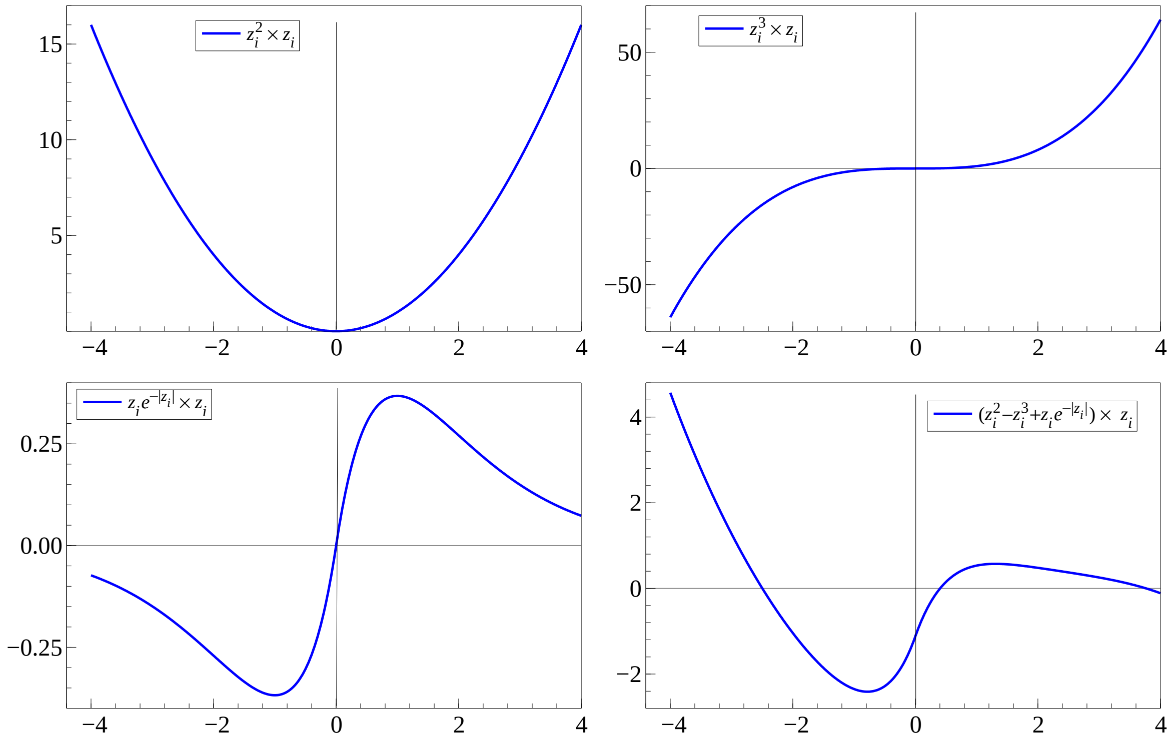

Figure 1: Quadratic, cubic, exponential and composite functions

distributed as F under the null with the appropriate degrees of freedom.[2] In stationary dynamic processes with weakly exogenous regressors, Index-test is only an F^{3n}_{T-4n-1} in large samples.

3 Test power

Thus, the key issue is the comparative powers of the various tests to reject the null when it is false in a range of states of nature. We first consider scalar polynomials, then a vector of quadratic regressors, which also highlights when the quadratic Index-test would have no power, and finally turn to the general portmanteau test.

3.1 Power approximations for polynomials

The simplest polynomial DGPs are given by:

$$y_t = \beta_j x_t^j + \epsilon_t \tag{14}$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$ and $x_t \sim \mathsf{IN}\left[\mu, 1\right]$, for $t = 1, ..., T$, and $j = 1, ..., 4$, such that the four cases are a linear, quadratic, cubic and quartic function. The distribution of the t^2-test of H_0: $\beta_j = 0$ conditional on $\mathbf{x}_t$ is:

$$\frac{\widehat{\beta}_j^2 \sum_{t=1}^T (x_t^j)^2}{\widehat{\sigma}_\epsilon^2} \underset{app}{\sim} \chi_1^2\left(\varphi_j^2\right) \tag{15}$$

[2] The labels follow the notation of Teräsvirta *et al.* (1993), where V refers to a Volterra expansion.

with non-centrality parameter:

$$\varphi_j^2 = \frac{\beta_j^2 \sum_{t=1}^{T} (x_t^j)^2}{\sigma_\epsilon^2} \tag{16}$$

Rather than use the specific value of $\sum_{t=1}^{T}(x_t^j)^2$ in the simulations, we have approximated it by $T\mathsf{E}[(x_t^j)^2]$ when calculating φ_j^2 where, using normality, $\mathsf{E}[(x_t)^2] = 1 + \mu^2$, $\mathsf{E}[(x_t^2)^2] = 3 + 6\mu^2 + \mu^4$, $\mathsf{E}[(x_t^3)^2] = 15 + 45\mu^2 + 15\mu^4 + \mu^6$, and $\mathsf{E}[(x_t^4)^2] = 105 + 420\mu^2 + 210\mu^4 + 28\mu^6 + \mu^8$.

Simulated power functions when $T = 100$ and $\sigma_\epsilon^2 = 1$ for the four DGPs are recorded in Figure 2, based on 10,000 replications, for each value of $\varphi_j = 1, ..., 6$, along with the analytic power functions, calculated by approximating F_{T-k}^{k} by a χ^2 with k degrees of freedom, $\mathsf{F}_{T-k}^{k}\left(\varphi^2\right) \underset{app}{\widetilde{\ }} \chi_k^2\left(\varphi^2\right)$, and relating that non-central χ^2 distribution to a central χ^2 (as in e.g., Hendry, 1995, p.475). When $\mu = 0$ (panels a and b), the divergence between the analytic and Monte Carlo test powers is notable for cubic and quartic, particularly for intermediate non-centralities. This is due to the high skewness and kurtosis of the distribution of x_t^j, which impacts on the distribution of the t-statistic under the alternative, as the sample mean of the $(x_t^j)^2$ can be far from $\mathsf{E}[(x_t^j)^2]$ in (16). Because of this effect, at a given non-centrality, it is more difficult to detect a higher-order term–even if its form is known–when means are zero (or y_t depends on $(x_t - \mu)^j$). Although the mean is arbitrary in most economic data series, (14) is not equivariant to the mean, unlike (3), as even for a quadratic:

$$y_t = \beta_2 x_t^2 + \epsilon_t = \beta_2\mu^2 + 2\beta_2\mu\left(x_t - \mu\right) + \beta_2\left(x_t - \mu\right)^2 + \epsilon_t.$$

Panels c and d show the much better match for $\mu \neq 0$. Also, while (16) still holds, non-centralities diverge faster at higher powers as μ increases. Combinations of polynomial functions like (14) lead to F-tests, and in these scalar cases are identical to using z_t.

3.2 Vector of quadratic regressors

For a vector of regressors, let $\beta_0 = 0$ and take $\boldsymbol{\mu} = \mathbf{0}$ in (8), so all linear terms have means of zero (or are deviations from sample means). Under the alternative that $\gamma \neq 0$ in (4), the test of $\boldsymbol{\kappa}_1 = \mathbf{0}$ in (10) will have power against quadratic departures that are not orthogonal to u_t in (5) as follows. Since $\boldsymbol{\Omega} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}'$, from (9):

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \gamma\mathbf{x}_t'\mathbf{A}\mathbf{x}_t + \epsilon_t = \boldsymbol{\beta}'\mathbf{x}_t + \gamma(\mathbf{z}_t^*)'\left(\boldsymbol{\Lambda}^{1/2}\mathbf{H}'\mathbf{A}\mathbf{H}\boldsymbol{\Lambda}^{1/2}\right)\mathbf{z}_t^* + \epsilon_t \tag{17}$$

Let $\boldsymbol{\Lambda}^{1/2}\mathbf{H}'\mathbf{A}\mathbf{H}\boldsymbol{\Lambda}^{1/2} = \boldsymbol{\Upsilon}_* + \mathbf{D}_*$ where $\boldsymbol{\Upsilon}_*$ is diagonal and $\mathbf{D}_*$ non-diagonal with a zero diagonal, so:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \gamma(\mathbf{z}_t^*)'\left(\boldsymbol{\Upsilon}_* + \mathbf{D}_*\right)\mathbf{z}_t^* + \epsilon_t \tag{18}$$

then:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \gamma(\mathbf{z}_t^*)'\boldsymbol{\Upsilon}_*\mathbf{z}_t^* + \gamma(\mathbf{z}_t^*)'\mathbf{D}_*\mathbf{z}_t^* + \epsilon_t = \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\kappa}_1'\mathbf{u}_{*,t} + v_t \tag{19}$$

(say), yielding the test of $\boldsymbol{\kappa}_1 = \mathbf{0}$ in (10). Such a test is not optimal if $\mathbf{D}_* \neq \mathbf{0}$, as then cross-products of the $z_{i,t}^*$ matter, although by construction only those components that are orthogonal to $\mathbf{u}_{*,t}$ will be omitted from (19). The 'closer' $\mathbf{D}_*$ is to $\mathbf{0}$, the less the power loss. Additional terms from the next sub and super diagonals could be added as a check when n is small; or going in the opposite direction, a scalar test could be constructed using $\sum_{i=1}^{n}(z_{i,t}^*)^2$ as a single regressor, as in Tukey (1949). The operational equivalent replaces $\mathbf{z}_t^*$ by $\mathbf{z}_t$ so tests $\boldsymbol{\kappa}_1 = \mathbf{0}$ in:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \gamma\mathbf{z}_t'\boldsymbol{\Upsilon}\mathbf{z}_t + e_t = \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\kappa}_1'\mathbf{u}_{1,t} + e_t. \tag{20}$$

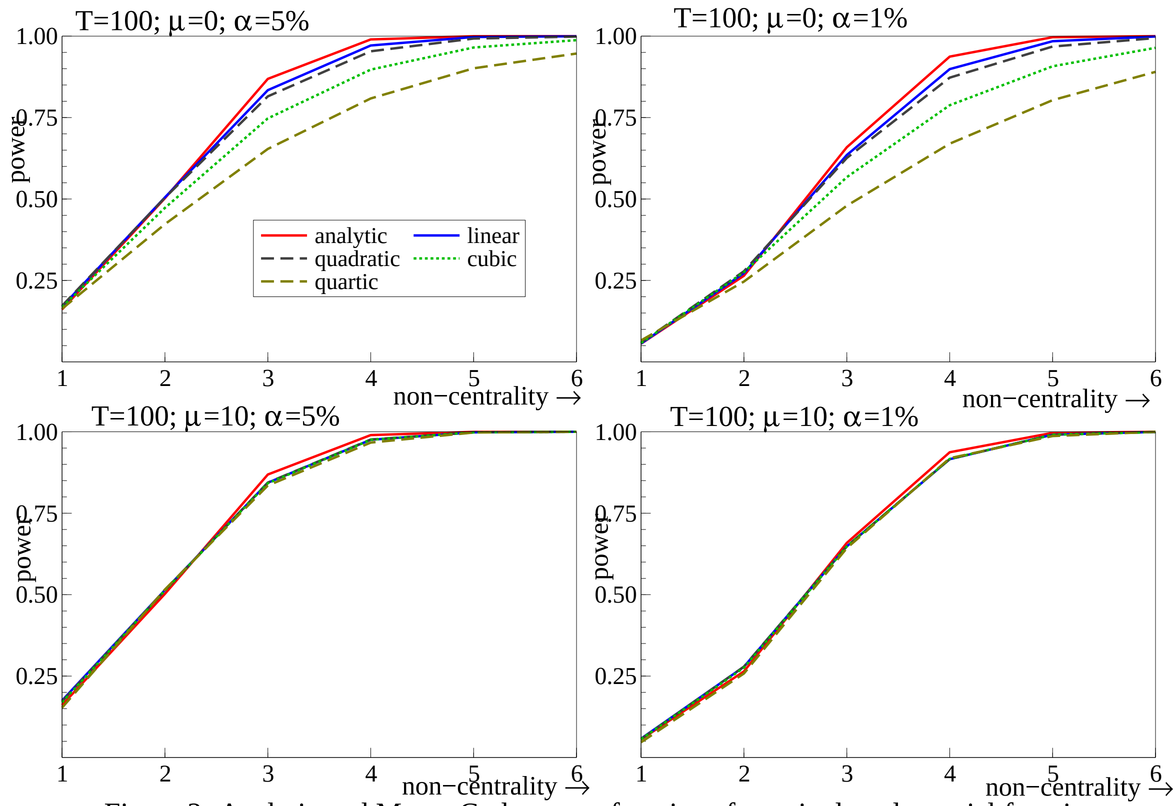

Figure 2: Analytic and Monte Carlo power functions for a single polynomial function

3.3 Powerless cases

The quadratic Index-test will have no power when the departure from linearity is in the direction of $\mathbf{u}_{1,t,\perp}$, which requires $\boldsymbol{\Upsilon} = \mathbf{0}_n$ when $\mathbf{D} \neq \mathbf{0}$ in (20). This may occur if the $x_{i,t}$s are perfectly orthogonal, such that $\boldsymbol{\Omega} = \boldsymbol{\Lambda}$, and the non-linearity only enters in the form of a cross product, whereas the principal components would not include cross-product terms. However, the practical relevance of such a case seems limited. Note that the form of any non-centrality due to non-linear functions of $\mathbf{x}_t$ being omitted from (3) changes with the mapping to $\mathbf{z}_t$, maintaining the test's equivariance to collinearity.

The quadratic Index-test would also have no power when the second derivative of $f(\cdot)$ was zero, but the third was non-zero, a case to which we now turn.

3.4 General Index-test power

The possibility that the first non-zero derivative is the third is precisely the reason for including the additional terms $\mathbf{u}_{2,t}$ and $\mathbf{u}_{3,t}$ in (12). Even if the first non-zero derivative is the fourth, the general Index-test in (12) will have power, as the second derivative will almost certainly be correlated with the fourth; similarly for the third with the fifth.

When the non-linearity takes the form of a 'squashing function' like an ogive, the exponential component is likely to prove useful. There is a trade-off between the degrees of freedom and the value of the test's non-centrality parameter. When the non-centralities of individual terms are small and n is large, a test with R_n degrees of freedom would have very low power. For the general Index-test, the overall power depends on combinations of these non-centralities for each variable and each polynomial power, as well as the exponential term, an issue we now explore by Monte Carlo experiments.

4 Alternative tests

There are a wide range of alternative linearity tests, see Granger and Teräsvirta (1993) for a summary. We consider a subset that aims to test against a general non-linear alternative. Consider a univariate process y_t which is a function of a set of n potential linear regressors $x_{i,t}$, $i = 1, \ldots, n$, and a set of non-linear functions, denoted $g(\cdot)$. The non-linearity tests computed include:

1. Optimal test, $\mathsf{H}_0 : \phi = 0$ for:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \phi g(\cdot) + \nu_t. \tag{21}$$

2. A second-order Kolmogorov–Gabor polynomial test, which applying the Frisch and Waugh (1933) theorem, is identical to the Tsay (1986) test and is analogous to the White (1980) test, $\mathsf{H}_0 : \boldsymbol{\delta} = \mathbf{0}$ for:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \sum_{j=1}^{n} \sum_{k=j}^{n} \delta_{jk} x_{j,t} x_{k,t} + \eta_t. \tag{22}$$

   This test has $n(n+1)/2$ degrees of freedom, and is denoted V2.

3. A third-order Kolmogorov–Gabor polynomial test, proposed by Spanos (1986, p.460) and Teräsvirta *et al.* (1993), $\mathsf{H}_0 : \boldsymbol{\delta} = \boldsymbol{\vartheta} = \mathbf{0}$ for:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \sum_{j=1}^{n} \sum_{k=j}^{n} \delta_{jk} x_{j,t} x_{k,t} + \sum_{j=1}^{n} \sum_{k=j}^{n} \sum_{l=k}^{n} \vartheta_{jkl} x_{j,t} x_{k,t} x_{l,t} + \zeta_t. \tag{23}$$

   This test has $n(n+1)(n+5)/6$ degrees of freedom, and is denoted V23.

4. Principal components of the second-order polynomial test, using the first k principal components of (6), denoted $\widetilde{\mathbf{w}}_t$, and test $\mathsf{H}_0 : \boldsymbol{\pi} = \mathbf{0}$ for:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \sum_{j=1}^{k} \pi_j \widetilde{w}_{j,t} + \xi_t. \tag{24}$$

   We set $k = 3n$ to correspond to Index-test and the test is denoted PCV2.

5. Principal components of the third-order polynomial test based on the first k principal components of $\mathbf{s}_t = vech\left[(\mathbf{x}_t \mathbf{x}_t') \otimes \mathbf{x}_t'\right]$, denoted $\widetilde{\mathbf{s}}_t$: $\mathsf{H}_0 : \boldsymbol{\tau} = \mathbf{0}$ for:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \sum_{j=1}^{k} \tau_j \widetilde{s}_{j,t} + \varsigma_t. \tag{25}$$

   We set $k = 3n$ to compare to Index-test, and the test is denoted PCV23.

6. Index-test, $\mathsf{H}_0 : \boldsymbol{\gamma} = \boldsymbol{\theta} = \boldsymbol{\xi} = \mathbf{0}$ for:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \boldsymbol{\gamma}' \mathbf{u}_{1,t} + \boldsymbol{\theta}' \mathbf{u}_{2,t} + \boldsymbol{\xi}' \mathbf{u}_{3,t} + e_t, \tag{26}$$

   with $3n$ degrees of freedom. Variants of the test with just the quadratic, cubic, and exponential functions are also considered, each with n degrees of freedom. In practice, these would require knowledge of the functional form of the DGP. The benefit of Index-test is that it is a portmanteau test that has power over a range of possible non-linear DGPs.

7. RESET test: $\mathsf{H}_0 : \kappa_1 = \kappa_2 = \kappa_3 = 0$ for:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \kappa_1 \widehat{y}_t^2 + \kappa_2 \widehat{y}_t^3 + \kappa_3 \widehat{y}_t^4 + \upsilon_t, \tag{27}$$

which has 3 degrees of freedom when $\widehat{y}_t$ is the fitted value from the linear model.

4.1 Principal components of non-linear functions

The number of parameters for V2 and V23 quickly becomes large as more linear regressors are included, such that the tests are infeasible for large n. This implies that the principal components of V2 and V23 must also be restricted to small n. For a sample size of $T = 100$, a maximum of 5 linear regressors can be included for the V23 test and a maximum of 13 for the V2 test. Index-test is feasible over a wider range of n, as only $3n$ degrees of freedom are needed for the most general test.

Whilst taking principal components of the linear functions and then computing non-linear functions delivers similar power to taking principal components of the non-linear functions directly for small n, there are drawbacks to the latter approach. No dimension reduction is achieved as the number of principal components will equal the number of non-linear functions. Hence, an arbitrary k is selected, either using formal tests such as the Scree test (Cattell, 1966) or the Kaiser (1960) criterion, or based on a preference for the degrees of freedom. If a batch of tests is computed in which k varies, the critical values would need to be corrected for the joint procedure–which would require a tighter significance level for the principal components tests than Index-test. Most variance is collected by the first few principal components, so a small k is often beneficial, but discarding PCs with smaller eigenvalues can be problematic if the relevant non-linear regressors are highly correlated with y_t but not with other regressors. By undertaking the dimension reduction first and then computing the non-linear transformations, no arbitrary reduction is needed. Further, principal components of the non-linear functions can be computationally demanding for large n compared to the proposed Index-test.

5 Power simulations for static DGPs

5.1 Experimental design

A range of simulations was undertaken in Ox (see Doornik, 2007) using $M = 1000$ replications to examine the powers of Index-test for varying degrees of collinearity and numbers of regressors. The unmodeled variables' DGP is:

$$\mathbf{x}_t \sim \mathsf{IN}_n [\boldsymbol{\mu}, \boldsymbol{\Omega}] \tag{28}$$

where $\mu_i = 0$ or 10, $\mathsf{V}[x_{i,t}] = 1$, $\forall i$, and $\mathsf{cov}[x_{i,t} x_{j,t}] = \rho$, $\forall i \neq j$. Here, n is the number of linear regressors in each model, increasing from 2 to 20 in the general unrestricted model, of which only two ($x_{1,t}$ and $x_{2,t}$) are relevant. We consider up to 18 irrelevant variables, as happens in model selection experiments (see e.g. Hoover and Perez, 1999), so some tests are infeasible. The $\{y_t\}$ DGPs considered are listed in table 1, where (see section 5.3.7 for the 'Quadratic z' case):

$$y_t = f(\cdot) + \epsilon_t, \quad \epsilon_t \sim \mathsf{IN}[0, 1]. \tag{29}$$

We consider two magnitudes of correlation, $\rho = 0, 0.9$, and two sample sizes, $T = 100$ and 300, with the parameters of the DGP held constant to assess the impact on power of increasing the sample size. Parameters are set for the non-linear functions such that the non-centrality of an individual t-test for $T = 100$, $\rho = 0$ and $\boldsymbol{\mu} = \mathbf{0}$ is given by ψ in Table 1. Results are mainly reported for $T = 100$, 5% significance and $\boldsymbol{\mu} = \mathbf{0}$, due to space constraints, but are available on request.

DGP	specification: $f(\cdot)$	coefficients and non-centralities
Quadratic	$\beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{1,t}^2$	$\beta_1 = \beta_2 = 0.3, \beta_3 = 0.1732$; $\boldsymbol{\psi} = 3$
Cubic	$\beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{1,t}^3$	$\beta_1 = 0.4743, \beta_2 = 0.3, \beta_3 = 0.1225$; $\boldsymbol{\psi} = 3$
Quartic	$\beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{1,t}^4$	$\beta_1 = \beta_2 = 0.3, \beta_3 = 0.0293$; $\boldsymbol{\psi} = 3$
Composite	$\beta_1 x_{1,t}^2 + \beta_2 x_{2,t}^3 + \beta_3 x_{2,t} e^{-\lvert x_{2,t}\rvert}$	$\beta_1 = 3.6, \beta_2 = 0.285, \beta_3 = 0.195$; $\boldsymbol{\psi} = 3$
Cross-product	$\beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{1,t}^2 + \beta_4 x_{1,t} x_{2,t}$	$\beta_1 = \beta_2 = 0.2, \beta_3 = 0.1155, \beta_4 = 0.2$; $\boldsymbol{\psi} = 2$
Exponential	(a) $\beta_1 \left\{x_{1,t} e^{-\lvert x_{1,t}\rvert}\right\}$ or (b) $\beta_1 \left\{x_{1,t} e^{-\lvert x_{2,t}\rvert}\right\}$	(a) $\beta_1 = 2.04$; $\boldsymbol{\psi} = 6$ or (b) $\beta_1 = 2.04$; $\boldsymbol{\psi} = 6$
LSTR	$\beta_1 x_{1,t} + \beta_2 x_{2,t} + (\delta_0 + \delta_1 x_{1,t} + \delta_2 x_{2,t}) \times$	$\beta_1 = \beta_2 = 0.3, \delta_0 = 0.9$
	$[1 + \exp(-\gamma(x_{1,t} - c))]^{-1}$	$\delta_1 = \delta_2 = 0.6, \gamma = 2.5, c = 0.3$
Quadratic z	(a) $\beta_1 z_{1,t}^2$ or (b) $\beta_2 z_{2,t}^2$	(a) $\beta_1 = 0.32$; $\boldsymbol{\psi} = 3$ or (b) $\beta_2 = 3.802$; $\boldsymbol{\psi} = 3$

Table 1: Simulation experiments for static DGPs. ψ refers to the non-centrality of an optimal individual t-test for $T = 100$, $\rho = 0$ and $\boldsymbol{\mu} = \mathbf{0}$, for both the linear and non-linear parameters.

5.2 Results for $n = 2$

Although a scenario where the model coincides with a low-dimensional DGP is unlikely, we first consider that case to compare the set of alternative tests.

5.2.1 Quadratic DGP

For the quadratic DGP, the optimal test of $\mathsf{H}_0 : \beta_3 = 0$ when $n = 2$ has a non-centrality of $\varphi_{1,\alpha}^2 = 3T\beta_3^2$, which is independent of ρ, the degree of collinearity.

The power of the V2 test in (22) for (29) will depend on:

$$\delta_1 x_{1,t}^2 + \delta_2 x_{2,t}^2 + \delta_3 x_{1,t} x_{2,t} \tag{30}$$

and the non-centrality of the test is given by:[3]

$$\varphi_{V_2}^2 = T\left(3\delta_1^2 + 3\delta_2^2 + \left(1 + 2\rho^2\right)\left[\delta_3^2 + 2\delta_1\delta_2\right] + 6\rho\delta_3\left[\delta_1 + \delta_2\right]\right)$$

When $\delta_2 = \delta_3 = 0$, the test non-centrality collapses to the optimal test non-centrality of $3T\delta_1^2$. Hence, the difference between the optimal and V2 test will be a function of the number of degrees of freedom alone. This applies as well to V23, and to the principal components versions PCV2 and PCV23 if k spans the DGP non-linearity.

Index-test is based on $\gamma_1 z_{1,t}^2 + \gamma_2 z_{2,t}^2$, where:

$$\begin{aligned} z_{1,t} &= x_{1,t} + \varrho_1 x_{2,t} \\ z_{2,t} &= x_{2,t} + \varrho_2 x_{1,t} \end{aligned} \tag{31}$$

where $\boldsymbol{\varrho}$ depends on the eigenvalues, $\widehat{\boldsymbol{\Lambda}}$, of $\widehat{\boldsymbol{\Omega}}$. Under perfect orthogonality, $\boldsymbol{\Omega} = \mathbf{I}_2$, which implies $\boldsymbol{\Lambda} = \mathbf{I}_2$, and $\mathbf{H} = \mathbf{I}_2$, such that $z_{1,t}^2 = x_{1,t}^2$ and $z_{2,t}^2 = x_{2,t}^2$. When $\boldsymbol{\Omega} \neq \mathbf{I}_2$, $\mathbf{z}_t$ comprises a linear combination of the $\mathbf{x}_t$s:

$$z_{1,t}^2 = x_{1,t}^2 + \varrho_1^2 x_{2,t}^2 + 2\varrho_1 x_{1,t} x_{2,t} \tag{32}$$

$$z_{2,t}^2 = x_{2,t}^2 + \varrho_2^2 x_{1,t}^2 + 2\varrho_2 x_{1,t} x_{2,t} \tag{33}$$

[3]Using the fact that the fourth cumulant of a normal is zero, Hannan (1970, p.23) shows that:

$$\mathsf{E}[w_{1,t}w_{2,t}w_{3,t}w_{4,t}] = \mathsf{E}[w_{1,t}w_{2,t}]\,\mathsf{E}[w_{3,t}w_{4,t}] + \mathsf{E}[w_{1,t}w_{3,t}]\,\mathsf{E}[w_{2,t}w_{4,t}] + \mathsf{E}[w_{1,t}w_{4,t}]\,\mathsf{E}[w_{2,t}w_{3,t}]$$

where $x_{1,t}^3 x_{2,t}$ comprise the four $w_{i,t}$.

so Index-test power will depend on:

$$\gamma_1 \left(x_{1,t}^2 + \varrho_1^2 x_{2,t}^2 + 2\varrho_1 x_{1,t} x_{2,t} \right) + \gamma_2 \left(x_{2,t}^2 + \varrho_2^2 x_{1,t}^2 + 2\varrho_2 x_{1,t} x_{2,t} \right). \tag{34}$$

As the optimal test will directly test for the significance of $x_{1,t}^2$, Index-test will have highest power when:

$$\begin{aligned} \gamma_1 + \gamma_2 \varrho_2^2 &\approx \beta_3 \\ \gamma_1 \varrho_1^2 + \gamma_2 &\approx 0 \\ 2\gamma_1 \varrho_1 + 2\gamma_2 \varrho_2 &\approx 0. \end{aligned} \tag{35}$$

This will give low weight to the $x_{2,t}^2$ and $x_{1,t}x_{2,t}$ terms in the linear combination. Under orthogonality, the analytic non-centrality collapses to that of the optimal test, $3T\delta_1^2$, but with fewer degrees of freedom than the V2 test. Collinearity impacts on the non-centrality via the $\boldsymbol{\varrho}$ weighting.

The non-centrality of the RESET test will also be $3T\delta_1^2$, with 3 degrees of freedom.

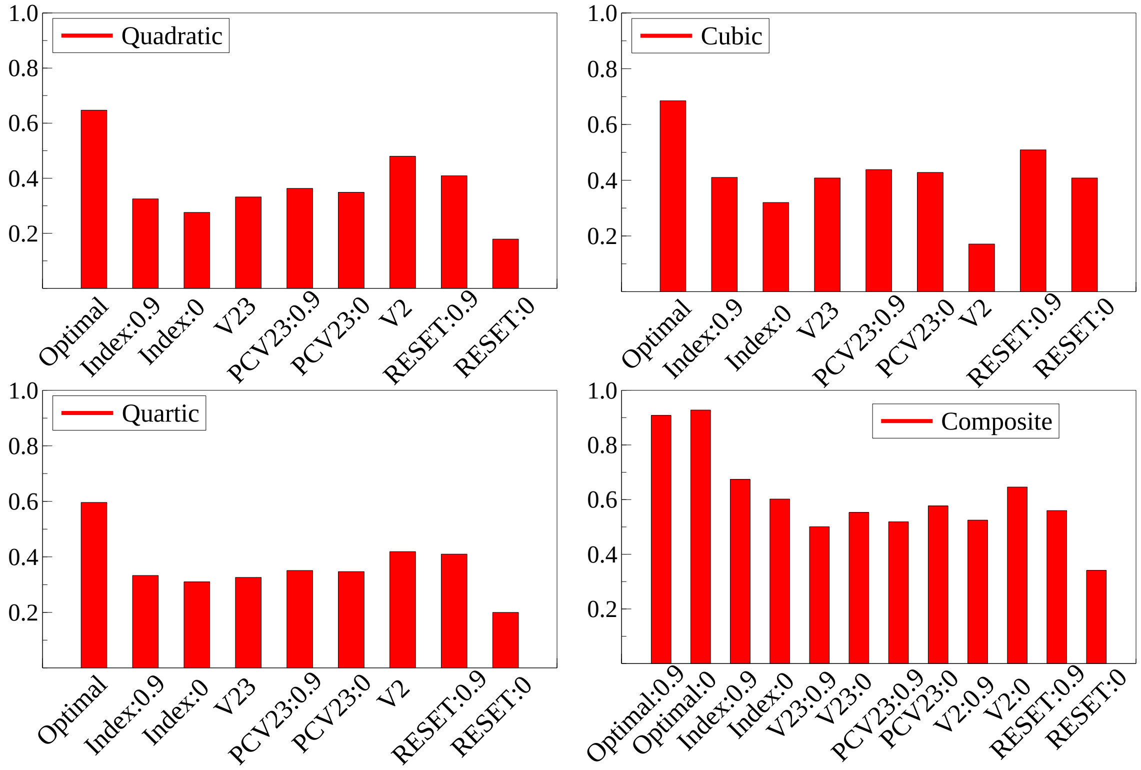

Figure 3: Powers of non-linearity tests for $n = 2$

Figure 3, panels a-d record the powers of the tests for the first four DGPs for $n = 2$, so there are no irrelevant regressors. Figures are labelled *a*, *b*, *d*, *c* in rows from top left. The degree of correlation, ρ, is reported by either :0 or :0.9 after the test name unless the test is invariant to ρ. We consider panel a here, and panels b-d in the following sub-sections. The optimal test power is around 0.60, which compares to the analytic power of 0.87 (see section 3.1). The V2 test (and PCV2) has the highest power for the quadratic as the test is designed to detect that form of non-linearity.[4] As Index-test includes additional

[4] V2 and PCV2 are identical for $n = 2, 4$. PCV2 will include $\min\{k = 3n; n(n+1)/2\}$ principal components. For $n = 2$, there are 3 non-linear functions for V2 and PCV2 will include all 3 principal components of V2. For $n = 4$ there are 10 non-linear functions so all 10 principal components are included and the tests are identical. PCV2=V2 in figure 3.

cubic and exponential terms, this reduces its power by approximately 15% when compared to Index-test with just quadratic $z_{i,t}$s with n degrees of freedom, but it insures against a broader range of non-linear DGPs. Also, Index-test has a higher power to detect non-linearity when there is collinearity, than under orthogonality, as the collinear variables will proxy the relevant non-linear functions.

5.2.2 Cubic DGP

We next consider the cubic DGP in (29). Under orthogonality, the non-centralities are:

$$\begin{aligned}
\varphi^2_{r,\alpha}(\beta_1) &= \frac{\beta_1^2 \mathsf{E}\left[\mathbf{x}_1' \widetilde{\mathbf{Q}} \mathbf{x}_1\right]}{\sigma^2_\epsilon} = 0.4T\beta_1^2 \\
\varphi^2_{r,\alpha}(\beta_2) &= T\beta_2^2 \\
\varphi^2_{r,\alpha}(\beta_3) &= \frac{\beta_3^2 \mathsf{E}\left[\mathbf{x}_1^{(3)\prime} \mathbf{Q} \mathbf{x}_1^{(3)}\right]}{\sigma^2_\epsilon} = 6T\beta_3^2
\end{aligned}$$

where $\mathbf{Q} = 1 - \mathbf{x}_1 \left(\mathbf{x}_1'\mathbf{x}_1\right)^{-1} \mathbf{x}_1'$, and $\widetilde{\mathbf{Q}} = 1 - \mathbf{x}_1^{(3)}(\mathbf{x}_1^{(3)\prime}\mathbf{x}_1^{(3)})^{-1}\mathbf{x}_1^{(3)\prime}$ when $\mathbf{x}_1^{(3)}$ denotes the vector of $\{x^3_{1,t}\}$. Parameter values are chosen such that $\psi = 3$ for $T = 100$ under orthogonality.

Figure 3b records the results for $n = 2$. V2 and PCV2 have very low power against a cubic DGP: the benefits of a portmanteau test such as Index-test are seen by comparing across experiments. V23 has high power for parsimonious specifications. The RESET test performs well for this DGP specification as the simple non-linear function is easily picked up by the parsimonious fitted value. The RESET test has a higher power to detect cubic non-linearity than quadratic. Index-test suffers as usual under orthogonality, and a high degree of collinearity is again beneficial. The linear combination of the $\mathbf{x}_t$s for $\mathbf{z}_t$ is:

$$\begin{aligned}
z^3_{1,t} &= x^3_{1,t} + 3\kappa_1 x^2_{1,t}x_{2,t} + 3\kappa_1^2 x_{1,t}x^2_{2,t} + \kappa_1^3 x^3_{2,t} \\
z^3_{2,t} &= x^3_{2,t} + 3\kappa_2 x^2_{2,t}x_{1,t} + 3\kappa_2^2 x_{2,t}x^2_{1,t} + \kappa_2^3 x^3_{1,t} \qquad (36)
\end{aligned}$$

Therefore, if $\rho = 0.9$, Index-test will gain power to detect $x^3_{1,t}$ via the linear combinations, $x^2_{1,t}x_{2,t}$ and $x_{1,t}x^2_{2,t}$. The gap between the analytic and optimal test is larger than for the quadratic DGP, in keeping with our previous analysis.

5.2.3 Quartic DGP

While a quartic function is somewhat extreme, and the small-sample distribution of even the 'optimal' t-statistic is poor under zero means, we investigate whether Index-test based on quadratic functions has power against quartic functions due to the collinearity between them. Again $\psi = 3$ for $T = 100$.

The results are recorded in Figure 3c for $n = 2$. The substantial gap between the analytic power (0.87) and optimal test power (0.60) is evident, due to regressor kurtosis. The power is only marginally lower than that for the quadratic function, and both tests based on Volterra expansions and Index-test do have power against a quartic function. The patterns exhibited by the power functions correspond to those for the quadratic function. The RESET test requires a high degree of collinearity and all linear functions of non-linear regressors to be included in the DGP for reasonable power.

5.2.4 Composite DGP

As Index-test is designed as a portmanteau test, it has power against a wide range of alternatives. We next consider its performance for a composite DGP in which the non-linearity enters in several ways, such as quadratic, cubic and exponential jointly. The results are recorded in Figure 3, panel d, and confirm that this composite DGP favors Index-test, especially when there is collinearity.

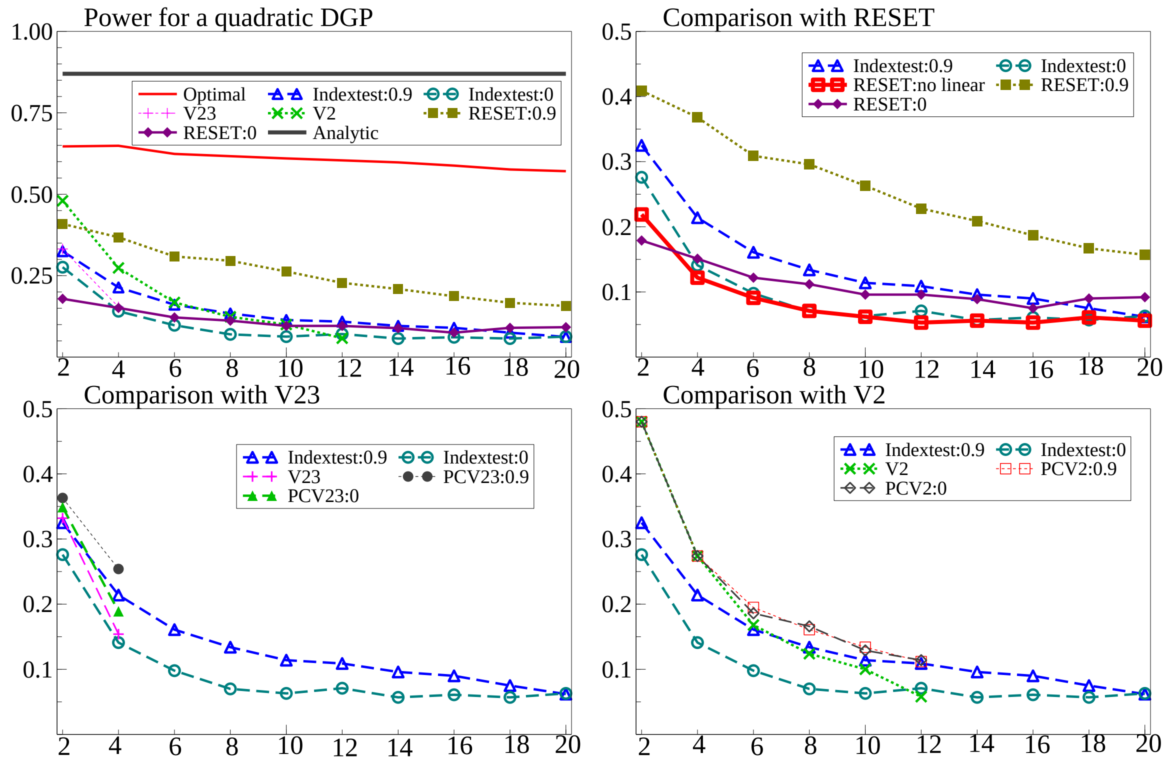

Figure 4: Powers of non-linearity tests for a quadratic function as n increases

5.3 DGPs for $n > 2$

Generally, models will not coincide with DGPs, so we now increase n by adding irrelevant regressors.

5.3.1 Quadratic DGP

Figure 4 reports the results for the quadratic DGP as n increases from 2 to 20. The number of linear regressors is recorded along the horizontal axis with the empirical rejection frequency on the vertical axis. Index-test does not have the highest power, but is robust to a range of non-linearities. Panel a compares the power to a range of alternative tests. As the DGP is a simple form of non-linearity, detecting the non-linearity is difficult and all tests have low power for large n compared to the optimal test. The RESET test has the highest power when $\rho = 0.9$, but panel b compares this to the case where the linear functions do not enter the DGP (i.e. $\beta_1 = \beta_2 = 0$), when the RESET test has a lower power than Index-test for all n, as powers of $\widehat{y}_t$ will not contribute much to detecting non-linearity. The V23 (panel c) and V2 (panel d) tests, and their PCs, have high power at $n = 2$, but the power declines sharply as n increases due to degrees of freedom. These tests are infeasible for many values of n here. Again, Index-test has higher power under collinearity, as a high ρ increases power to detect non-linearity via collinear squares and cross-products. As T increases to 300, the powers of all tests increase, with a unit power for the analytic calculation, and near unit power for the optimal infeasible test. As before, V2 and PCV2 deliver the highest power for the case they target.

5.3.2 Cross-product DGP

For the DGP with a cross-product term, all individual regressors have non-centralities of 2 when $T = 100$ under orthogonality. A conventional t-test of each null would, therefore, have power of approximately 0.5 at 5% when the specification in (29) was known. As the power of the optimal test depends on:

$$\beta_3 x_{1,t}^2 + \beta_4 x_{1,t} x_{2,t},$$

the non-centrality of the optimal F-test is:

$$\varphi_F^2 = T\left[3\beta_3^2 + 6\beta_3\beta_4\rho + \beta_4^2\left(1 + 2\rho^2\right)\right].$$

Hence, the non-centrality of the joint F-test is $\varphi_F = 5.2$ for $\rho = 0.9$, delivering a high power for all tests under collinearity.

Under orthogonality, Index-test must again have low power against a single cross-product term in (29). For the test to have power against (29), we require a low weight on the $x_{2,t}^2$ term in the $z_{1,t}^2$ equation, but then there is no close approximation to $\beta_3 x_{1,t}^2 + \beta_4 x_{1,t} x_{2,t}$.

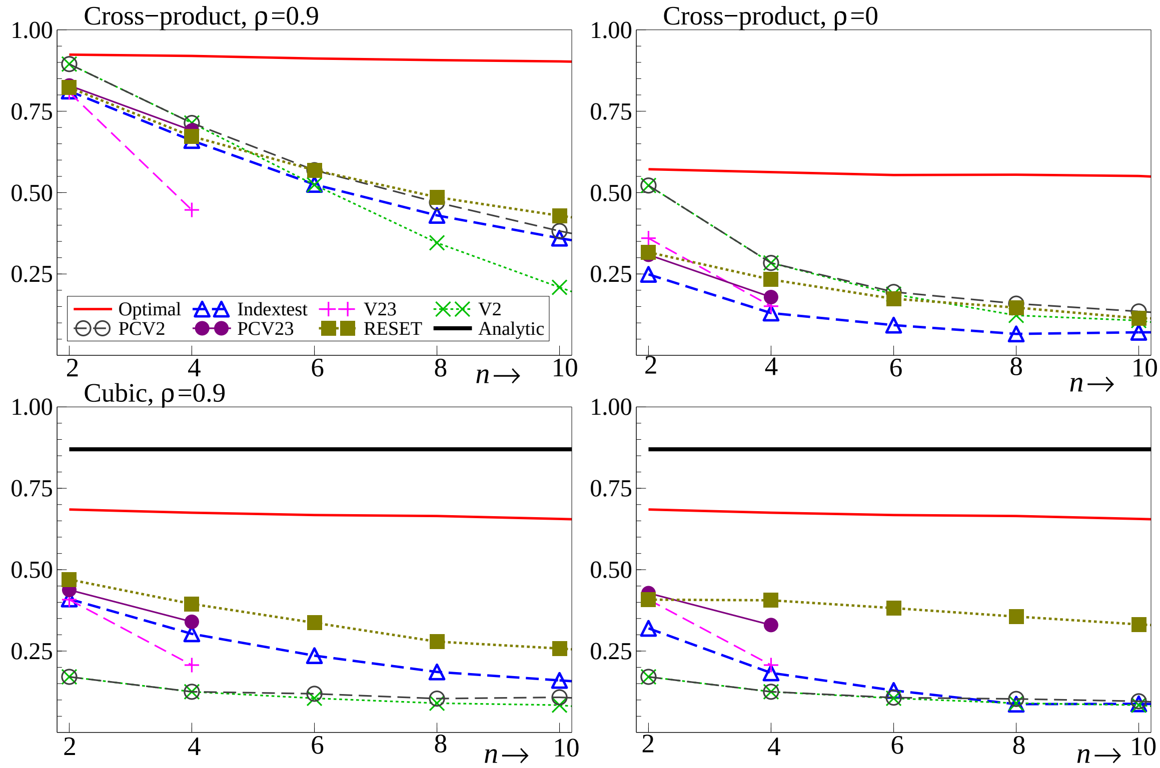

Figure 5: Powers of non-linearity tests for a cross-product function and a cubic function

Results for the cross-product experiment are reported in figure 5, panel a for $\rho = 0.9$ and panel b for $\rho = 0$ to a maximum of $n = 10$, as the power functions are nearly horizontal for $n > 10$. PCV2 has the highest power, but as n increases it declines sharply, whereas Index-test power declines more slowly, as does RESET. The degrees of freedom are costly for V23 in terms of sharply declining power. Under orthogonality, Index-test again has low power relative to the alternative tests. If the non-linear terms offset each other, all tests have low power. This form of non-linearity is particularly difficult to detect: increasing collinearity yields lower power as the high correlations cancel the relevant non-linear terms. Thus, higher correlations are not always advantageous.

5.3.3 Cubic DGP

Figure 5, panels c and d, records powers for the cubic DGP for up to 10 regressors, and shows similar patterns, although RESET is less affected by n increasing.

5.3.4 Exponential DGPs

Many non-linear models include exponential functions such as neural networks with logistic squashing functions, and logistic or exponential smooth transition models. Tests based on polynomials will have power against exponential DGPs due to the exponential approximation in footnote 2, but tests with exponentials included, such as Index-test, should be able to capture this form of non-linearity in a more parsimonious manner.

Figure 6, panels a and b, record the power for an exponential DGP in one linear variable, i.e. $\{x_{1,t}e^{-|x_{1,t}|}\}$, and panels c and d record the power for a non-linear function of a combination of linear regressors $\{x_{1,t}e^{-|x_{2,t}|}\}$. The divergence between the optimal test and the non-linear tests is marked. When the regressors are collinear, Index-test has a higher power to detect this form of non-linearity than the RESET test and the Volterra-expansion based tests. V2 and PCV2 have no power to detect exponentials of this form, but V23 and PCV23 do have power due to the exponential approximation:

$$y_t = \beta_1 \left\{x_{1,t}e^{-|x_{2,t}|}\right\} + \epsilon_t \approx \beta_1 \left(x_{1,t} - x_{1,t}\left|x_{2,t}\right| + \frac{1}{2}x_{1,t}x_{2,t}^2 - \frac{1}{6}x_{1,t}\left|x_{2,t}^3\right|\right) + v_t \tag{37}$$

The powers for V23 and PCV23 increase under orthogonality, which is the opposite of Index-test. The RESET test does not have power for cross-exponentials, but does have a higher power than Index-test under orthogonality for $\{x_{1,t}e^{-|x_{1,t}|}\}$. Including the linear terms in the DGP would favor the RESET test. Index-test with just quadratics (i.e., n degrees of freedom) has a higher power than the general Index-test, suggesting quadratic functions can also 'pick up' exponentials, so the increase in degrees of freedom from n to $3n$ is costly. However, more complex exponential functions favor Index-test.

5.3.5 LSTR DGPs

Having considered an exponential DGP, we generalize this to an LSTR DGP as this nests aspects of threshold models, regime-switching models, Markov-switching models and neural networks, and is therefore representative of a general class of non-linear models. Teräsvirta (1996) proposes a range of non-linear DGPs based on Lee *et al.* (1993) for the time-series domain, but here we focus on cross-section DGPs: section 6 considers dynamic models. The Monte Carlo is necessarily equation specific, but it is indicative of the performance of non-linearity tests to detect this type of departure from linearity.

While the optimal infeasible test based on the LSTR specification is not computed, we do compute the power of one feasible test based on a third-order Taylor approximation. Replacing the transition function by:

$$\left[1 + \exp\left\{-\gamma\left(x_{1,t} - c\right)\right\}\right]^{-1} \simeq \frac{1}{2} + \frac{\gamma\left(x_{1,t} - c\right)}{4} - \frac{\left(\gamma\left(x_{1,t} - c\right)\right)^3}{48}, \tag{38}$$

results in the approximation:

$$y_t \simeq \theta_0 + \theta_1 x_{1,t} + \theta_2 x_{2,t} + \theta_3 x_{1,t}^2 + \theta_4 x_{1,t}^3 + \theta_5 x_{1,t}^4 + \theta_6 x_{1,t}x_{2,t} + \theta_7 x_{1,t}^2 x_{2,t} + \theta_8 x_{1,t}^3 x_{2,t} + \varepsilon_t. \tag{39}$$

Hence, the Taylor approximation test is highly parameterized, and there may be degrees-of-freedom gains for Index-test.

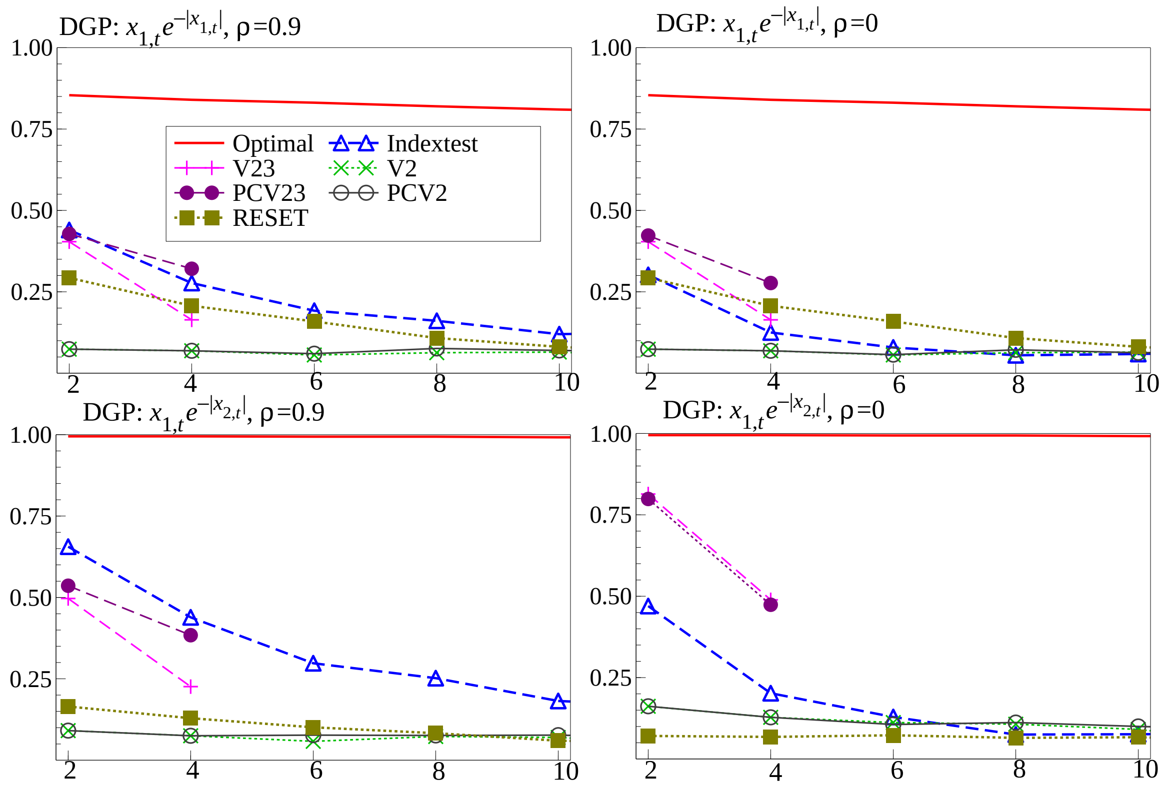

Figure 6: Power of non-linearity tests for an exponential function

The results are recorded in figure 7, panels a and b. Index-test and Volterra-expansion tests have power against an LSTR model under collinearity, indicating that polynomial approximations do capture non-linearities generated by a smooth-transition model. For $n = 2$, most tests outperform the RESET test, but then degrees-of-freedom favor the RESET test. An index of Index-test with comparable degrees of freedom (using $\{\sum_{i=1}^{n} z_{i,t}^2\}$, $\{\sum_{i=1}^{n} z_{i,t}^3\}$ and $\{\sum_{i=1}^{n} e^{-|z_{i,t}|} z_{i,t}\}$) delivers higher power than the RESET test, suggesting that over-parametrized models suffer in an LSTR DGP. Furthermore, including exponential functions in Index-test yields no improvements over Taylor-approximation tests, suggesting that polynomials capture this form of non-linearity well. Orthogonality implies that combinations of the non-linear regressors will have low weight, reducing the power of Index-test relative to PCV2. Alternative LSTR specifications are needed to draw more precise conclusions, but to the extent that (39) is a reasonable approximation, Index-test will have power against (29). Conversely, rejecting an initial specification does not entail that the alternative must be a polynomial function.

5.3.6 Composite DGP

Figure 7, panels c and d confirm that the composite DGP favors Index-test as its power is high for small n when there is collinearity. As n increases, power declines and the degrees-of-freedom benefits of RESET yield a higher power for $n > 7$, where only 3 degrees of freedom are required compared to 21 for Index-test. Increasing the sample size favors Index-test, and its power is higher than RESET for all n at $T = 300$. Index-test outperforms the V2 and V23 tests, but under orthogonality, has lower power than the Volterra-expansion tests and their principal components.

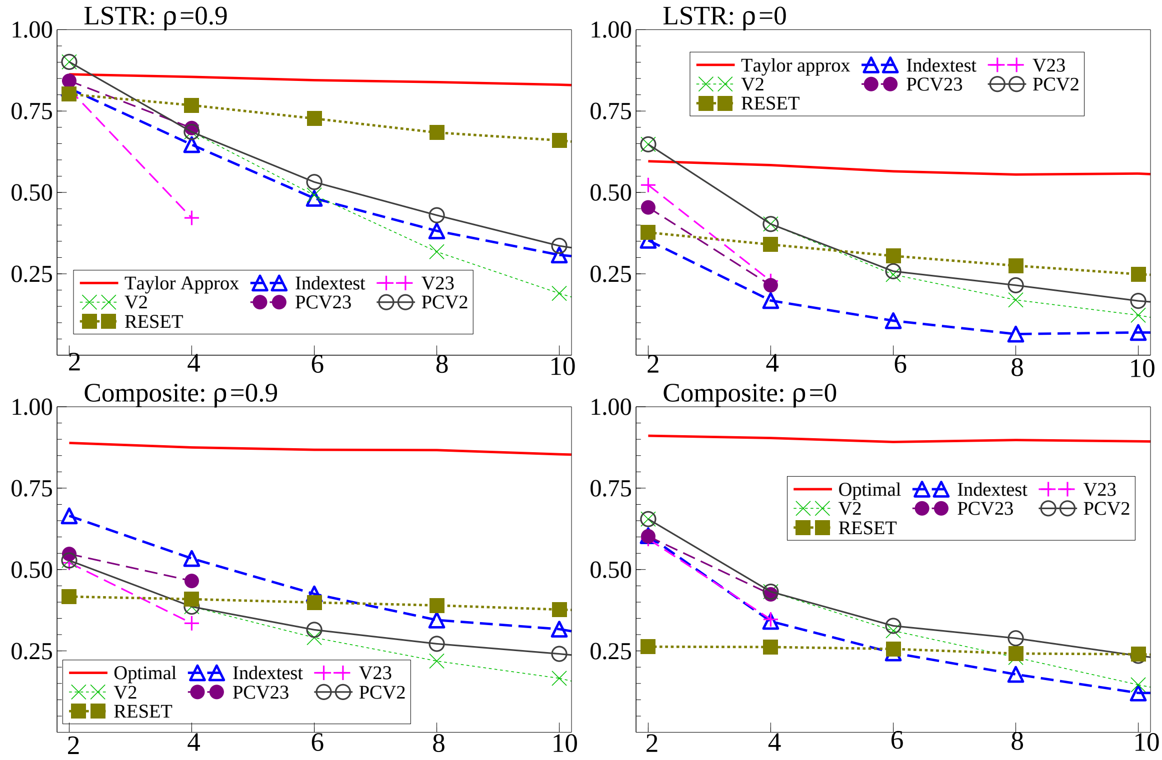

Figure 7: Power of non-linearity tests for an LSTR function and a combination of non-linear functions

5.3.7 Quadratic z DGP

We finally consider a DGP that is specified in terms of the orthogonal linear combinations. From (9) the DGP is:

$$y_t = \beta_1 z_{1,t}^2 + \epsilon_t = \beta_1 \left(4.998x_{1,t}^2 + 4.998x_{2,t}^2 - 9.997x_{1,t}x_{2,t}\right) + \epsilon_t \quad (40)$$

where $\rho = 0.9$ and $\boldsymbol{\mu} = 0$. We set $\beta_1 = 0.32$ so individual non-centralities are $\approx |3|$ for the three non-linear functions. Under orthogonality, however, those weights will be incorrect.

The DGP has offsetting effects. When collinearity is large, $x_{1,t}x_{2,t}$ will be a proxy for $x_{1,t}^2$ and $x_{2,t}^2$ so the negative coefficient on the cross-product will adversely affect the power. Thus, we also consider the alternative DGP:

$$y_t = \beta_2 z_{2,t}^2 + \epsilon_t = \beta_2 \left(0.263x_{1,t}^2 + 0.263x_{2,t}^2 + 0.526x_{1,t}x_{2,t}\right) + \epsilon_t \quad (41)$$

where $\beta_2 = 3.802$, so both the quadratic and cross-product terms are 'in the same direction' and power will be higher under collinearity.

Figure 8 records the results for (40) in panels a and b, and for (41) in panels c and d. Index-test and V2/PCV2 have the same power for $n = 2$ as the tests are equivalent given the DGP design. V23 has a higher power than its principal component analogue for $n = 4$ under collinearity, which suggests that PCs of the non-linear functions can omit relevant non-linear combinations given the arbitrary selection of the number of PCs. Index-test, V2 and PCV2 have comparative powers for $\rho = 0.9$. Under orthogonality, all tests have higher power due to the off-setting effects of the cross-product term. All tests based on polynomials outperform the RESET test. When the second eigenvector is used to form the DGP weights, all tests have unit power under orthogonality apart from the RESET test, although the power of Index-test does decline as n increases since the weightings are incorrect for $n > 2$.

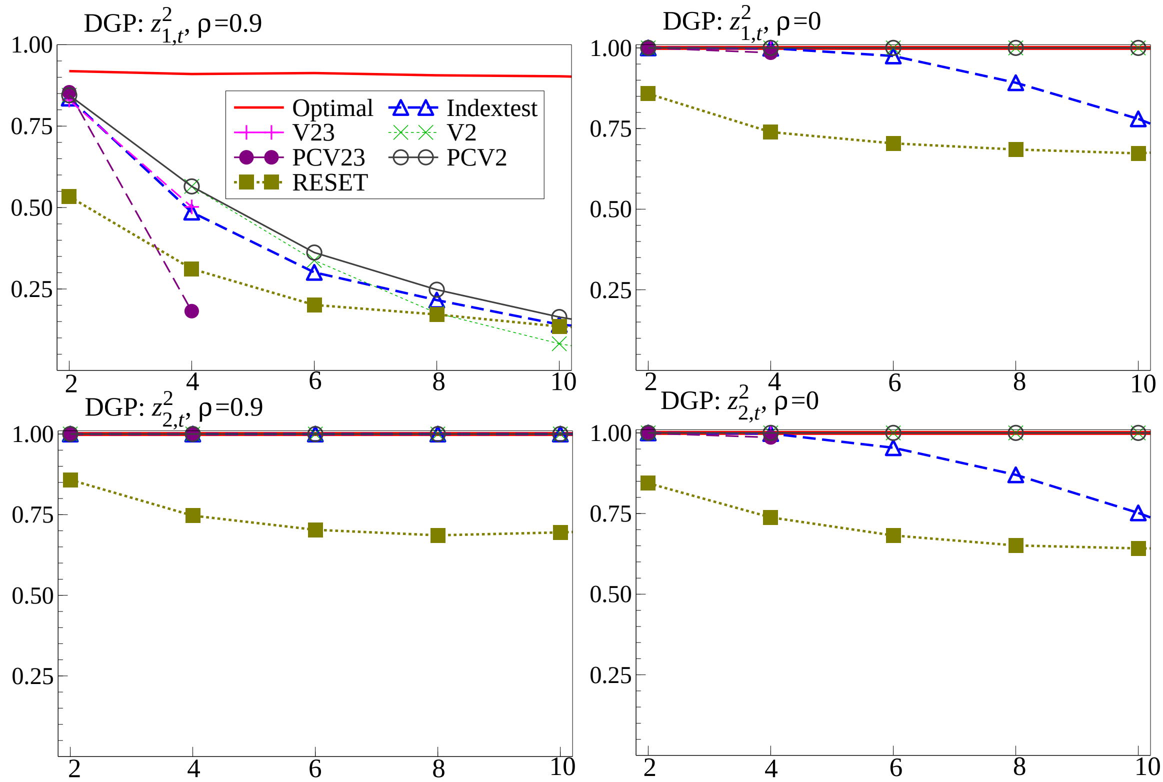

Figure 8: Power of the non-linearity test for a quadratic z_t function

A further case included both orthogonalized regressors, and delivered similar power shapes over n as the single regressor case. A cubic z was also considered, and again results were comparable.

6 Power simulations for dynamic models

6.1 Experiment design

So far we have considered static equations with strongly exogenous regressors, essentially a cross-section context. There are numerous well-known problems in generalizing to a non-linear dynamic context, but some can be addressed. To assess the properties of the test in a dynamic context, we undertake simulation experiments for a first-order autoregressive-distributed lag, ADL(1,1), DGP:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 y_{t-1} + \beta_3 x_{1,t-1} + g(\cdot) + \epsilon_t \tag{42}$$

where $|\beta_2| < 1$ and $\mathbf{x}_t$ is now generated by:

$$\mathbf{x}_t = \mathbf{\Pi}\mathbf{x}_{t-1} + \alpha y_{t-1} + \boldsymbol{\varepsilon}_t \tag{43}$$

where $\mathbf{\Pi}$ is the $(n \times n)$ matrix $\{\pi_{ij}\}$ for $i, j = 1, \ldots, n$, $|\pi_{ii}| < 1, \forall i$, $|\alpha| < 1$, and:

$$\begin{pmatrix} \epsilon_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix} \sim \mathsf{IN}_{1+n}\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \omega_{11} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\omega}_{22} \end{pmatrix}\right]. \tag{44}$$

$\mathsf{E}\left[\epsilon_t^2\right] = \omega_{11}$ is a scalar and $\mathsf{E}\left[\varepsilon_{i,t}, \varepsilon_{j,t}\right] = \omega_{22,.ij}\ \forall i, j$. We generate $n = 1, \ldots, 9$ regressors, $\mathbf{x}_t$, based on (43), of which one (x_1) enters the DGP (42). Commencing with 3 linear regressors $(y_{t-1}, x_{1,t}, x_{1,t-1})$,

Test	$g(\cdot)$	Coefficients
Baseline		$\omega_{11}=1; \omega_{12.i}=0, \forall i; \omega_{22.ii}=1, \forall i; \omega_{22.ij}=0.5, \forall i \neq j; \pi_{ij}=0, \forall i \neq j$
		Strong exogeneity
Size	-	$\beta_0=5; \beta_1=\beta_2=\beta_3=\pi_{ii}=0.5; \alpha=0$
Power	$\beta_4 x_{1,t-1}^2$	$\beta_0=5; \beta_1=\beta_2=\beta_3=\pi_{ii}=0.5; \alpha=0; \beta_4=0.4$
Power	$\beta_4 x_{1,t}^2+\beta_5 x_{1,t-1}^3$	$\beta_0=5; \beta_1=\beta_2=\beta_3=\pi_{ii}=0.5; \alpha=0; \beta_4=0.15; \beta_5=0.1$
Power	$\beta_4 \exp^{x_{1,t-1}^2}$	$\beta_0=5; \beta_1=\beta_2=\beta_3=\pi_{ii}=0.5; \alpha=0; \beta_4=0.4$
		Increasing persistence
Size	-	$\beta_0=5; \beta_1=\beta_3=0.5; \beta_2=\pi_{ii}=0.8; \alpha=0$
Power	$\beta_4 x_{1,t-1}^2$	$\beta_0=5; \beta_1=\beta_3=0.5; \beta_2=\pi_{ii}=0.8; \alpha=0; \beta_4=0.4$
		Relaxing strong exogeneity
Size	-	$\beta_0=5; \beta_1=\beta_2=\beta_3=\pi_{ii}=0.5; \alpha=-0.5$
Power	$\beta_4 x_{1,t-1}^2$	$\beta_0=5; \beta_1=\beta_2=\beta_3=\pi_{ii}=0.5; \alpha=-0.5; \beta_4=0.15$

Table 2: Simulation experiments for dynamic DGPs

we include an additional regressor and its lag sequentially until all 19 regressors $(y_{t-1}, \mathbf{x}_t, \mathbf{x}_{t-1})$ are included in the general model. We discard the first 20 observations when generating the data. Table 2 records the experiments undertaken, and the tests listed in section 4 are computed. We compute the principal components for Index-tests based on the second moment matrix of $(y_{t-1} : \mathbf{x}_t : \mathbf{x}_{t-1})'$. Unless otherwise stated we set $\omega_{11}=1; \omega_{22,ii}=1, \forall i; \omega_{22,.ij}=0.5, \forall i \neq j$; and $\pi_{ij}=0, \forall i \neq j$.

6.2 Results

6.2.1 Strong exogeneity

Strong exogeneity requires the absence of feedback, such that $\alpha = 0$. As the DGP is unknown to the econometrician, a test of non-linearity will require inclusion of all non-linear functions of the information set, $(y_{t-1} : \mathbf{x}_t : \mathbf{x}_{t-1})'$, although we assume the lag length is known at unity here. Figure 9, panel a, records the test size (for nominal sizes of 1% and 5%); and panels b, c and d record the powers (at 5%) for the non-linear functions listed in table 2.[5]

The results show that Index-test has a large-sample actual size close to its nominal, at least at 5% and 1%, although all tests become slightly over-sized as the degree of persistence increases. The powers of the tests based on Volterra expansions are high for small n, but decline rapidly as n increases, and Index-test power declines steadily. All tests outperform the RESET test for small n, although its 3 degrees of freedom help as n increases. A higher correlation between regressors, or a more complex form of non-linearity, should favor Index-test.

6.2.2 Increasing persistence

We next increased the degree of persistence in both the marginals and conditional from $\beta_2 = \pi_{ii} = 0.5, \forall i$ to $\beta_2 = \pi_{ii} = 0.8, \forall i$. The size is reported in figure 10, panel a, and the power for a quadratic DGP in panel b. Increasing persistence yields a higher power for Index-test due to the collinearity arguments noted for the static case. Hence, strongly exogenous dynamic DGPs have the appropriate size and a power close to that of their static counterparts.

[5]We only report results for $T = 100$. As the sample size increases to $T = 300$, the size of the test is similar and the power increases, in keeping with asymptotic theory. We only report powers for a 5% nominal significance level, as the patterns of the power curves are similar at 1%, but sizes are reported for both 1% and 5%. Full results are available on request.

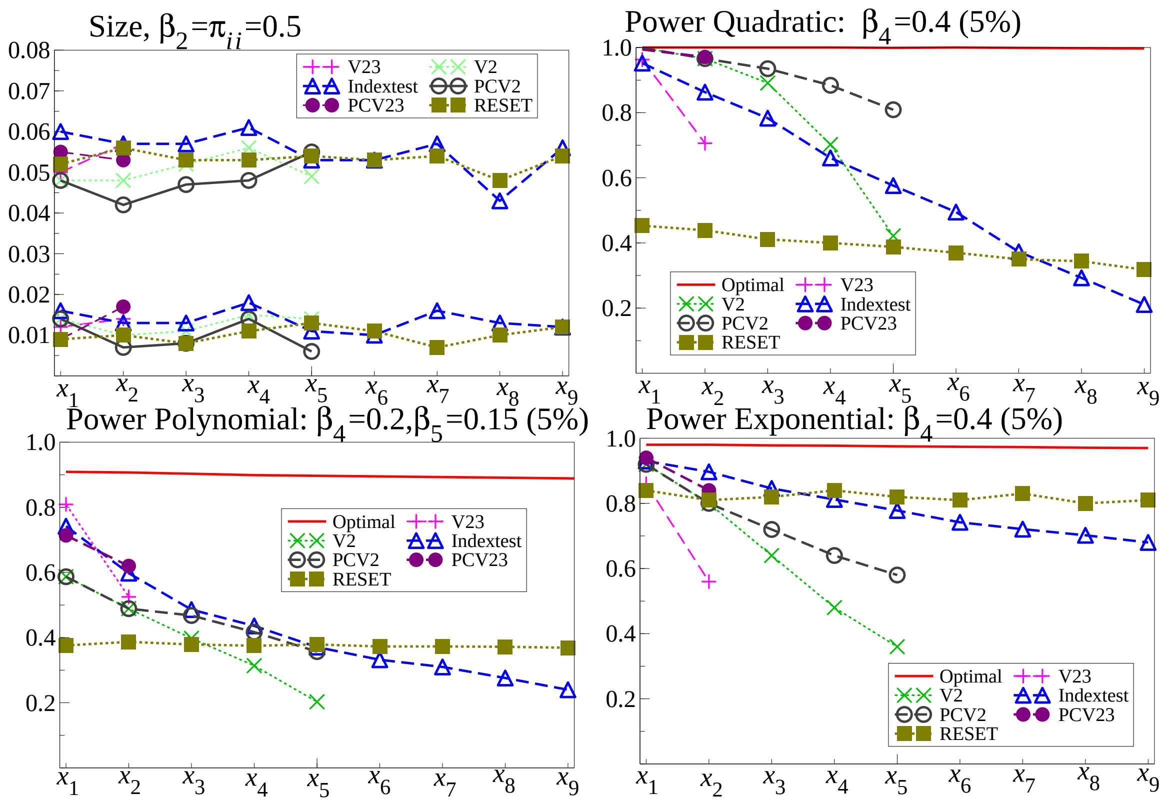

Figure 9: Size and power for ADL(1,1) DGP with strong exogeneity

6.2.3 Weak exogeneity

If the regressors are only weakly exogenous, then non-linearity could induce unstable or chaotic behavior for some parameter configurations. Providing stationarity is maintained, in large samples the null distribution is close to an F, checked here when $\alpha = -0.5$. For power, we set $\beta_4 = 0.15$ to ensure convergence, giving a non-centrality of approximately 3.5. The sizes and powers of the tests are recorded in figure 10, panels c and d, showing similar power behavior to earlier.

6.2.4 Unit roots

If the levels data are integrated, the null distributions of many of the non-linearity tests are non-standard: see Caceres (2007). When the conditional model contains a unit root, resulting in a differenced process estimated in levels, and the strongly exogenous regressors are stationary, Index-test has a size close to the nominal. Non-linear functions of the lagged dependent variable will be non-stationary, but regressions of I(1) variables on I(0) processes do not lead to spurious regression: see Hendry (1995, p.129). If the original formulation is in non-stationary variables, so the non-linear functions have complicated behavior but are nevertheless strongly exogenous, then again the F-distribution is the relevant one based on conditioning, although the power may not be well described by a non-central F, as seen in section 3.1.

7 Conclusion

For scalar DGPs with relatively few regressors, the current tests in the literature perform well. However, if the model contains many potential regressors that are possibly highly correlated, and the form of

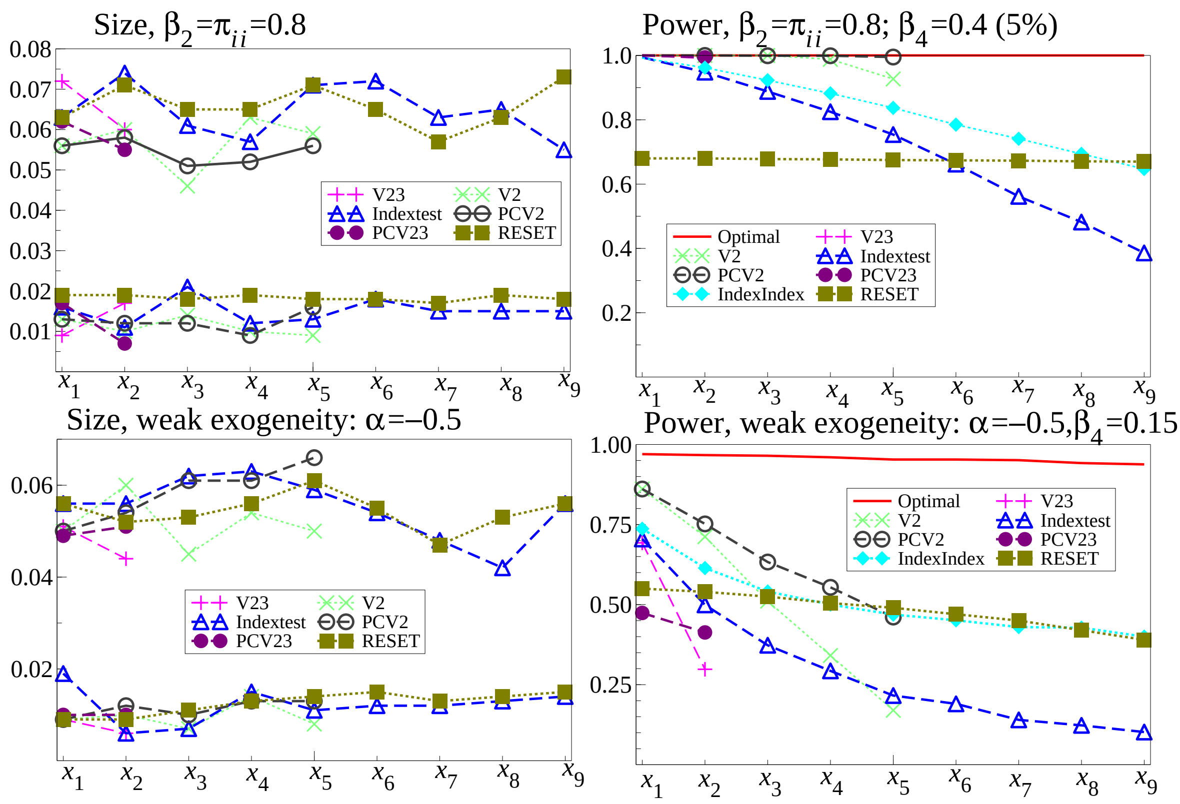

Figure 10: Size and power for ADL(1,1) DGP with increasing persistence or weak exogeneity

non-linearity under the alternative is unknown, many general tests become infeasible: our proposed portmanteau Index-test performs well in such situations. Although it is only distributed exactly as a (non-central) F for fixed regressors, that distribution remains as a large-sample approximation in stationary processes with weakly exogenous regressors.

Tests based on second-order Volterra expansions are ideal if the departure from non-linearity is in the direction of a quadratic and sample sizes are large, but they have little power to detect cubic or exponential non-linearities. Tests based on third-order Volterra expansions have power to detect departures from linearity in more directions, but are only feasible for small numbers of regressors. Principal components of these two tests perform well, but do not uniformly have higher power than their original counterparts (as the quadratic $z_{1,t}$ DGP above demonstrates for V23). Further, they need not achieve a dimension reduction, unlike Index-test, which is computationally simpler. Tests based on Volterra expansions and our proposed Index-test perform comparatively for small n, but the power of Volterra expansion tests declines more sharply as n increases due to the rapidly increasing degrees of freedom.

The RESET test has high power in situations where all the linear components of the non-linear functions enter the DGP, but has low power both when the non-linear functions enter independently of the linear functions, or when many of the linear functions do not enter non-linearly, as then powers of the fitted dependent variable from a linear regression are a poor approximation to the non-linear DGP. Also, non-zero means adversely affect the power of the RESET test–see Teräsvirta (1996) for the impact of the intercept on the RESET test–and outliers also reduce its power. The RESET test has a higher power to detect non-linearity in the form of a cubic than a quadratic, and the degrees of freedom benefits of the RESET test are more apparent for large n.

Index-test has the appropriate size, and is equivariant to collinearity. However, if the non-linear DGP is such that two highly-collinear non-linear functions 'cancel each other', all tests tend to have low power.

More complex DGPs favor Index-test, so when the functional form is unknown and there is a large set of candidate relevant variables, but the specification nests the DGP, Index-test has power to reject a false null in a wide range of circumstances: pure quadratic, pure cubic, pure quartic, exponential, and these in combinations. Recent simulations show that its relative power is also higher when the regressors are non-normal.

All tests (apart from RESET in some cases) decline in power as n increases. While parsimony delivers a higher power, such that selection of relevant variables prior to implementing the test might appear to be beneficial, this may be a hazardous strategy if the linear term is irrelevant, yet enters the DGP in a non-linear function. Thus, there is a trade-off between a higher power after selection and a risk of eliminating variables that are relevant only via a non-linear transformation, resulting in a lower power to detect non-linearity when such a variable is excluded.

By using a portmanteau test, power cannot be uniformly higher than all the alternative tests considered, although Index-test offers some power against a range of possible non-linear functional forms and is feasible for quite large n. Index-test outperforms Volterra-expansion tests and RESET in many such situations, and can even be close to the optimal test. For larger departures from linearity, where several non-linear terms occur for a number of variables, its power will dominate that illustrated here, where the experiments were deliberately chosen with many irrelevant variables to highlight the potential. Moreover, the simulation experiments for dynamic models suggest Index-test has the correct large sample size, and has reasonable power properties for stationary dynamic processes even with weakly exogenous regressors. Thus, it promises to be a useful mis-specification test for examining the functional forms of general models.

References

Abadir, K. M. (1999). An introduction to hypergeometric functions for economists. *Econometric Reviews*, **18**, 287–330.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, **58**, 1443–1458.

Brock, W. A., Dechert, W. D., and Scheinkman, J. A. (1987). A test for independence based on the correlation dimension. SSRI Working Paper No. 8702, Department of Economics, University of Wisconsin.

Caceres, C. (2007). Asymptotic properties of tests for mis-specification. Unpublished doctoral thesis, Economics Department, Oxford University.

Cattell, R. B. (1966). The Scree test for the number of factors. *Multivariate Behavioural Research*, **1**, 245–276.

Dhrymes, P. J. (1966). On the treatment of certain recurrent nonlinearities in regression analysis. *Southern Economic Journal*, **33**, 187–196.

Dhrymes, P. J., Howrey, E. P., Hymans, S. H., Kmenta, J., *et al.* (1972). Criteria for evaluation of econometric models. *Annals of Economic and Social Measurement*, **3**, 291–324.

Doornik, J. A. (1995). Econometric Computing. Ph.D Thesis, University of Oxford, Oxford.

Doornik, J. A. (2007). *Object-Oriented Matrix Programming using Ox* 6th edn. London: Timberlake Consultants Press.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, **50**, 987–1008.

Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304.

Frisch, R., and Waugh, F. V. (1933). Partial time regression as compared with individual trends. *Econometrica*, **1**, 221–223.

Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Hannan, E. J. (ed.)(1970). *Multiple Time Series*. New York: John Wiley & Sons.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F., and Krolzig, H.-M. (2003). New developments in automatic general-to-specific modelling. In Stigum, B. P. (ed.), *Econometrics and the Philosophy of Economics*, pp. 379–419. Princeton: Princeton University Press.

Hoover, K. D., and Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, **2**, 167–191.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, **20**, 141–151.

Keenan, D. M. (1985). A Tukey non-additivity-type test for time series nonlinearity. *Biometrika*, **72**, 39–44.

Lee, T.-H., White, H., and Granger, C. W. J. (1993). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, **56**, 269–290.

McLeod, A. I., and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared residual autocorrelations. *Journal of Time Series Analysis*, **4**, 269–273.

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. *Journal of the Royal Statistical Society B*, **31**, 350–371.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Teräsvirta, T. (1996). Power properties of linearity tests for time series. *Studies in Nonlinear Dynamics & Econometrics, Berkeley Electronic Press*, **1**, 3–10.

Teräsvirta, T., Lin, C.-F., and Granger, C. W. J. (1993). Power of the neural network linearity test. *Journal of Time Series Analysis*, **14**, 309–322.

Thursby, J. G., and Schmidt, P. (1977). Some properties of tests for specification error in a linear regression model. *Journal of the American Statistical Association*, **72**, 635–641.

Tsay, R. S. (1986). Non-linearity tests for time series. *Biometrika*, **73**, 461–466.

Tukey, J. W. (1949). One degree of freedom for non-additivity. *Biometrics*, **5**, 232–42.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–26.

White, H. (1989). An additional hidden unit test for neglected nonlinerarity in multilayer feedforward networks. *Proceedings of The International Joint Conference on Neural Networks, Washington, DC*, **II**, 451–455.

White, H. (1992). *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.