



DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

**A LOW-DIMENSION COLLINEARITY-ROBUST TEST FOR
NON-LINEARITY**

Jennifer L. Castle and David F. Hendry

Number 326

May 2007

Manor Road Building, Oxford OX1 3UQ

A Low-Dimension Collinearity-Robust Test for Non-linearity

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

May 8, 2007

Abstract

A new test for non-linearity is developed using weighted combinations of regressor powers based on the eigenvectors of the variance-covariance matrix. The test extends the ingenious test for heteroskedasticity proposed by White (1980), but both circumvents problems of high dimensionality and collinearity, and allows inclusion of cubic functions to ensure power against asymmetry or skewness. A Monte Carlo analysis compares the performance of the test to the optimal infeasible test and to a variant of White's test. The relative performance of the test is encouraging; the test has the appropriate size and has high power in many situations. Furthermore, collinearity between regressors can increase the power of the test.

Keywords: Functional Form Test; Non-linearity; Collinearity.

JEL Classification: C51; C52.

1 Introduction

The validity of the functional form of a model is an essential component of its specification. Consequently, a test for non-linearity is required that can evaluate the 'goodness' of a postulated model against a general non-linear alternative. Evidence of non-linearity implies that an alternative functional form should be utilized. Non-linear models are becoming increasingly popular, and include switching regression models (Quandt, 1983), with models in this class including smooth-transition regression models (see Granger and Teräsvirta, 1993, and Teräsvirta, 1994), smooth-transition autoregressive models (see Chan and Tong, 1986, and Luukkonen, Saikkonen and Teräsvirta, 1988), threshold autoregressive models (see Tong, 1990), and exponential autoregressive models (see Priestley, 1981); Markov switching models (see Hamilton, 1989); bilinear models (see Subba Rao and Gabr, 1984, and Subba Rao, 1985); chaotic

*We thank participants of the Royal Economic Society Conference 2006, Econometric Society Australasian Meeting 2006, Econometric Society European Meeting 2006 and the 4th Oxmetrics Users Conference 2006 for helpful comments and suggestions. Financial support from the ESRC under grants RES-051-27-0035 and RES-000-23-0539 is gratefully acknowledged.

processes (see Lorenz, 1989, and Puu, 1989); through to neural network models (see White, 1989, 1992, and the related model selection algorithm of Perez-Amaral, Gallo and White, 2003). Given the range of possible non-linear functional forms, correctly specifying a specific parametric alternative *a priori* seems infeasible, so we propose a general test against non-linearity.

Many direct and estimation-based tests of linearity have been proposed in the literature, including Ramsey (1969), White (1980), McLeod and Li (1983), Keenan (1985), Tsay (1986), Brock, Dechert and Scheinkman (1987) and Lee, White and Granger (1993); see Granger and Teräsvirta (1993, ch.6), for an overview of many of the tests.¹ Ramsey proposed tests for specification errors in regression, including unmodelled non-linearity, based on adding powers of the fitted values: Doornik (1995) provides a careful evaluation of both the numerical and statistical properties of the RESET test. Keenan (1985) developed a univariate test for detecting non-linearity (which is a special case of the RESET test) and Tsay (1986) extended Keenan's test to allow for contemporaneous non-linearity. White's test added all squares of regressors, or squares and cross-products, to test for heteroskedasticity, implicitly testing for omitted non-linearity as well. This has been investigated by numerous authors, including a recent appraisal in Hendry and Krolzig (2003) in the context of model selection. Tests against specific alternatives include tests based on chaotic processes such as Brock *et al.* (1987), tests against ARCH such as that developed by McLeod and Li (1983), and neural network tests proposed by Lee *et al.* (1993).

A test of functional form is developed that extends and modifies the test for heteroskedasticity proposed by White (1980), where the heteroskedastic-consistent covariance matrix estimator will differ from the conventional formula when the squares and cross-products of the regressors would be significant if added to the model. There are three practical drawbacks of a test such as White's: first, its high dimensionality; secondly, the potentially high level of collinearity between products of regressors; and third, the possibility that the second derivative is not the source of the departure from linearity, which may depend on asymmetry or skewness, and be better reflected in the third derivative. To rectify these potential drawbacks, our test forms a composite function of all product terms of the regressors, based on weights given by the eigenvalues of the variance-covariance matrix. Thus, for fixed regressors the test

¹The list is by no means exhaustive.

is an exact F-test with n degrees of freedom (where n is the number of regressors) on the standardized, mutually-orthogonal combinations of squares and cross-products of the original data matrix, solving the problems of high dimensionality and collinearity. This version of the test will not have power when the second derivative is zero and higher orders are non-zero, but in such a case, cubic terms of the composite function are easily included as well, yielding an F-test with $2n$ degrees of freedom. The corresponding ‘unrestricted’ function of all terms up to cubic would involve adding about $2n^2$ elements, which would often prove infeasible.

The structure of the paper is as follows. Section 2 considers the general problem of functional form testing, and proposes a new test for settings where more general tests (e.g., White, 1980) are infeasible. Section 3 discusses the power of the proposed test, computing the non-centrality and examining the power approximations for polynomial functions. Section 4 undertakes a set of Monte Carlo experiments to examine the power of the test for various DGP designs, extended in section 5 to some dynamic models. Finally, section 6 concludes.

2 A quadratic approximation test

Consider the relationship:

$$y_t = f(x_{1,t} \dots x_{n,t}) + \epsilon_t, \quad (1)$$

where $(x_{1,t} \dots x_{n,t}) = \mathbf{x}'_t$ are strongly exogenous and $\epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]$. The linear approximation is:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_n x_{n,t} + e_t = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_t + e_t. \quad (2)$$

To evaluate the goodness of (2) requires testing the validity of the functional form approximation. Using the mean value theorem around a fixed point $\mathbf{x}_p$:

$$\begin{aligned} f(\mathbf{x}_t) &= f(\mathbf{x}_p) + \frac{\partial f(\mathbf{x}_t)}{\partial \mathbf{x}'_t} \Big|_{\mathbf{x}_t=\mathbf{x}_p} (\mathbf{x}_t - \mathbf{x}_p) + (\mathbf{x}_t - \mathbf{x}_p)' \frac{\partial^2 f(\mathbf{x}_t)}{\partial \mathbf{x}_t \partial \mathbf{x}'_t} \Big|_{\mathbf{x}_t=\mathbf{x}^*} (\mathbf{x}_t - \mathbf{x}_p) \\ &= f(\mathbf{x}_p) + \boldsymbol{\beta}'_p (\mathbf{x}_t - \mathbf{x}_p) + (\mathbf{x}_t - \mathbf{x}_p)' \mathbf{A}^* (\mathbf{x}_t - \mathbf{x}_p) \\ &= \beta_0 + \boldsymbol{\beta}'_p \mathbf{x}_t + \mathbf{x}'_t \mathbf{A}^* \mathbf{x}_t \end{aligned} \quad (3)$$

where $\mathbf{x}^* \in [\mathbf{x}_t, \mathbf{x}_p]$, and the value of β_p depends on $\mathbf{x}_p$ when $f(\cdot)$ is not linear. Thus, a ‘natural’ test is to consider the importance of the quadratic term $\mathbf{x}_t' \mathbf{A}^* \mathbf{x}_t$ added to (2). Since:

$$\mathbf{x}_t' \mathbf{A} \delta^* \mathbf{x}_t = \text{tr}(\mathbf{x}_t' \mathbf{A}^* \mathbf{x}_t) = \text{tr}(\mathbf{A}^* \mathbf{x}_t \mathbf{x}_t'), \quad (4)$$

for fixed regressors, an exact test would be an F-test of $\delta_1 = \mathbf{0}$ in:

$$y_t = \beta_0 + \beta_p' \mathbf{x}_t + \delta_1' \mathbf{w}_t + e_t \quad (5)$$

where:

$$\mathbf{w}_t = (\mathbf{x}_t \mathbf{x}_t')^{v_e} \quad (6)$$

and v_e vectorizes and selects non-redundant elements from the lower triangle (including the diagonal) of the outer product. This procedure is close to the test for heteroskedasticity proposed by White (1980), where his heteroskedastic-consistent covariance matrix estimator will differ from the conventional formula when the squares and cross-products of the regressors would be significant if added to the model.

There are three main drawbacks to such a test: first, its high dimensionality; secondly the potentially high collinearity between the elements of $\mathbf{w}_t$; and thirdly, the possibility that the second derivative is not the source of the important departure from linearity, which may depend on asymmetry or skewness, and be better reflected in the third derivative. To rectify these potential drawbacks, we develop an alternative test.

2.1 Improved testing

First, consider the optimal test of $\delta_1 = \mathbf{0}$ when $f(\mathbf{x}_t)$ is the exact quadratic:

$$f(\mathbf{x}_t) = \beta_0 + \beta' \mathbf{x}_t + \mathbf{x}_t' \mathbf{A} \mathbf{x}_t \quad (7)$$

and $\mathbf{A}$ is known and symmetric. Let $\mathbf{A} = \mathbf{K} \mathbf{\Upsilon} \mathbf{K}'$, where $\mathbf{\Upsilon}$ is the matrix of eigenvalues of $\mathbf{A}$ and $\mathbf{g}_t = \mathbf{K}' \mathbf{x}_t$, so that

$$\mathbf{x}_t' \mathbf{A} \mathbf{x}_t = \mathbf{x}_t' \mathbf{K} \mathbf{\Upsilon} \mathbf{K}' \mathbf{x}_t = \mathbf{g}_t' \mathbf{\Upsilon} \mathbf{g}_t = \sum_{i=1}^n \tau_i g_{i,t}^2 = \boldsymbol{\tau}' \mathbf{r}_t \quad (8)$$

where $\boldsymbol{\tau}$ is the vector of diagonal elements of $\mathbf{\Upsilon}$ and $\mathbf{r}_t$ is the $n \times 1$ vector with elements $g_{i,t}^2$. Then:

$$y_t = \beta_0 + \beta' \mathbf{x}_t + \alpha (\boldsymbol{\tau}' \mathbf{r}_t) + \epsilon_t = \beta_0 + \beta' \mathbf{x}_t + \alpha w_t + \epsilon_t \quad (9)$$

say, so a t-test of $H_0: \alpha = 0$ will be the most powerful test for non-linearity when $w_t = \boldsymbol{\tau}' \mathbf{r}_t$.

To provide an operational counterpart when $\mathbf{A}$ is unknown, let:

$$\mathbf{x}_t \sim \mathbf{D}_n [\boldsymbol{\mu}, \boldsymbol{\Omega}] . \quad (10)$$

For the present, we take $\boldsymbol{\Omega}$ to be known such that $\boldsymbol{\Omega} = \mathbf{H} \boldsymbol{\Lambda} \mathbf{H}'$ and $\mathbf{z}_t^* = \mathbf{H}' \mathbf{x}_t$ where $\mathbf{H}' \mathbf{H} = \mathbf{I}_n$ so that:

$$\mathbf{z}_t^* \sim \mathbf{D}_n [\mathbf{H}' \boldsymbol{\mu}, \boldsymbol{\Lambda}] . \quad (11)$$

Finally, take deviations from their means, $z_{i,t}^d = z_{i,t}^* - \overline{z_i^*}$, then scale using the square roots of their corresponding λ_i :

$$\frac{z_{i,t}^d}{\sqrt{\lambda_i}} = z_{i,t}, \quad (12)$$

so:

$$\mathbf{z}_t \underset{app}{\sim} \mathbf{D}_n [\mathbf{0}, \mathbf{I}] . \quad (13)$$

Thus, the $z_{i,t}$ are standardized and mutually orthogonal combinations of squares and cross-products of the original $x_{i,t}$.

As (4) is a scalar:

$$\text{tr} (\mathbf{x}_t' \mathbf{A}^* \mathbf{x}_t) = \text{tr} (\mathbf{x}_t' (\mathbf{A}^*)' \mathbf{x}_t) = \frac{1}{2} \text{tr} (\mathbf{x}_t' [\mathbf{A}^* + (\mathbf{A}^*)'] \mathbf{x}_t) = \text{tr} (\mathbf{x}_t' \mathbf{B} \mathbf{x}_t) \quad (14)$$

where $\mathbf{B}$ is symmetric. Hence, we can take $\mathbf{A}^*$ to be symmetric without loss of generality. Next:

$$\text{tr} (\mathbf{x}_t' \mathbf{A}^* \mathbf{x}_t) = \text{tr} (\mathbf{x}_t' \mathbf{H} (\mathbf{H}' \mathbf{A}^* \mathbf{H}) \mathbf{H}' \mathbf{x}_t) = \text{tr} ((\mathbf{H}' \mathbf{A}^* \mathbf{H}) \mathbf{z}_t \mathbf{z}_t') = \text{tr} (\mathbf{C} \mathbf{z}_t \mathbf{z}_t') . \quad (15)$$

Under the null of a linear function, $\mathbf{C} = \mathbf{0}$, so for a local alternative, exploiting symmetry, we expand $\mathbf{C}$ around $\mathbf{0}$ as $\mathbf{C} = \boldsymbol{\Delta}_n$ for a diagonal $\boldsymbol{\Delta}_n$ ($= \mathbf{0}$ under the null). Hence, from (15):

$$\text{tr} (\mathbf{C} \mathbf{z}_t \mathbf{z}_t') \simeq \text{tr} (\boldsymbol{\Delta}_n \mathbf{z}_t \mathbf{z}_t') = \mathbf{z}_t' \boldsymbol{\Delta}_n \mathbf{z}_t = \sum_{i=1}^n \delta_{1,i} z_{i,t}^2 . \quad (16)$$

Thus, the test has the same form as that in (9), but with $u_{i,t} = z_{i,t}^2$ in place of $w_{i,t}$. Relative to (5), there are only n elements in $\mathbf{u}_t$, as opposed to $n(n+1)/2$, but every element potentially depends on squares and cross-products of every $x_{i,t}$. Thus, the first and second objectives of effecting a major dimensionality

reduction and formulating a test in terms of non-collinear variables are achieved. Additional terms from the next sub and super diagonals of $\mathbf{C}$ could also be used; or going in the opposite direction, a scalar test corresponding to (9) could be constructed using $\sum_{i=1}^n z_{i,t}^2$ as a single regressor.

Under the null, for fixed regressors and $e_t \sim \text{IN} [0, \sigma_e^2]$, the test of $\delta_1 = \mathbf{0}$ in:

$$y_t = \beta_0 + \beta' \mathbf{x}_t + \delta_1' \mathbf{u}_t + e_t \quad (17)$$

where $u_{i,t} = z_{i,t}^2$, is an exact F-test with n degrees of freedom. Under the alternative, the test should have power against quadratic departures that are not orthogonal to the $w_{i,t}$. Finally, it is easy to accommodate the third drawback, and generalize the test for higher-derivative departures from the null by also including $\{z_{i,t}^3\}$ terms. When the additional terms $\sum_{j=1}^n \delta_{2,j} z_{j,t}^3$ are included, the test is an exact F-test, with $2n$ degrees of freedom, under the null of $\delta_1 = \delta_2 = \mathbf{0}$.

3 Test Power

3.1 Non-centrality

The easiest case to consider is (7), as a t-test of $H_0: \delta = 0$ will be the most powerful test for non-linearity in (9). Letting $\beta_0 = 0$ and $\mu = \mathbf{0}$ for simplicity, so all linear terms have means of zero, where $\mathbf{y}' = (y_1 \dots y_T)$ and $\mathbf{X}' = (\mathbf{x}_1 \dots \mathbf{x}_T)$, then:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{w}\delta + \epsilon, \quad (18)$$

with $\epsilon \sim \text{IN} [\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_T]$. Then:

$$\sqrt{T} (\hat{\delta} - \delta) = (T^{-1} \mathbf{w}' \mathbf{Q} \mathbf{w})^{-1} \frac{\mathbf{w}' \mathbf{Q} \epsilon}{\sqrt{T}} \xrightarrow{D} \text{N} [0, \sigma_\delta^2] \quad (19)$$

for:

$$\mathbf{Q} = \mathbf{I}_T - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$$

where:

$$\sigma_\delta^2 = \sigma_\epsilon^2 \text{plim}_{T \rightarrow \infty} (T^{-1} \mathbf{w}' \mathbf{Q} \mathbf{w})^{-1}. \quad (20)$$

After double de-meaning to ensure the squares and cross-products are not highly collinear with the $x_{i,t}$, $\mathbf{Q}\mathbf{w}$ should be approximately $\mathbf{w}$, leading to:²

$$\sigma_{\hat{\delta}}^2 = \sigma_{\epsilon}^2 \text{plim}_{T \rightarrow \infty} (T^{-1} \mathbf{w}' \mathbf{w})^{-1}. \quad (21)$$

From (8):

$$T^{-1} \mathbf{w}' \mathbf{w} = T^{-1} \sum_{t=1}^T \left(\sum_{i=1}^n \sum_{j=1}^n \tau_i \tau_j r_{i,t}^2 r_{j,t}^2 \right) \rightarrow \boldsymbol{\tau}' \mathbf{R} \boldsymbol{\tau} \quad (22)$$

and $\mathbf{R}$ is the limit of the matrix with elements $T^{-1} \sum_{t=1}^T r_{i,t}^2 r_{j,t}^2$. Then:

$$\sqrt{T} (\hat{\delta} - \delta) \xrightarrow{D} \mathcal{N} \left[0, \sigma_{\epsilon}^2 (\boldsymbol{\tau}' \mathbf{R} \boldsymbol{\tau})^{-1} \right]. \quad (23)$$

Under the sequence of local alternatives that $\delta = c/\sqrt{T}$:

$$t_{\delta=0}^2 = \frac{T \hat{\delta}^2 (\boldsymbol{\tau}' \mathbf{R} \boldsymbol{\tau})}{\sigma_{\epsilon}^2} \quad (24)$$

has a local non-centrality parameter:

$$\psi = \frac{c^2 (\boldsymbol{\tau}' \mathbf{R} \boldsymbol{\tau})}{\sigma_{\epsilon}^2}. \quad (25)$$

To relate the operational test in (17) to the optimal, and derive the non-centrality corresponding to (25), again for $\beta_0 = 0$, consider (15) for $\mathbf{A} = \mathbf{A}^*$ and symmetric:

$$\mathbf{x}_t' \mathbf{A} \mathbf{x}_t = \mathbf{x}_t' \mathbf{H} (\mathbf{H}' \mathbf{A} \mathbf{H}) \mathbf{H}' \mathbf{x}_t = \mathbf{z}_t' (\mathbf{H}' \mathbf{K} \boldsymbol{\Upsilon} \mathbf{K}' \mathbf{H}) \mathbf{z}_t = \mathbf{z}_t' \mathbf{C} \mathbf{z}_t, \quad (26)$$

where:

$$\mathbf{C} = \mathbf{H}' \mathbf{A} \mathbf{H} = \boldsymbol{\Delta}_n + \mathbf{D}, \quad (27)$$

where $\mathbf{D}$ has a zero diagonal. Then:

$$y_t = \beta_p' \mathbf{x}_t + \mathbf{z}_t' \boldsymbol{\Delta}_n \mathbf{z}_t + \mathbf{z}_t' (\mathbf{C} - \boldsymbol{\Delta}_n) \mathbf{z}_t + \epsilon_t = \beta_p' \mathbf{x}_t + \delta_1' \mathbf{u}_t + \mathbf{z}_t' \mathbf{D} \mathbf{z}_t + \epsilon_t. \quad (28)$$

Thus, the closer $\mathbf{D}$ is to $\mathbf{0}$, the less the power loss. If, for example, $\mathbf{A}$ was positive definite, then $\mathbf{H}$ could jointly diagonalize $\mathbf{A}$ and $\boldsymbol{\Omega}$, such that $\boldsymbol{\Delta}_n = \mathbf{I}_n$ and so $\mathbf{D} = \mathbf{0}$ (see, e.g., Hendry, 1995, p.631).

²See Castle and Hendry (2005).

3.2 Power derivations

To derive the power function analytically, we approximate F_{T-k}^k by a χ^2 with k degrees of freedom:

$$F_{T-k}^k(\varphi_{r,\alpha}) \sim \chi_k^2(\varphi_{r,\alpha}^2). \quad (29)$$

Next, we relate that non-central χ^2 distribution to a central χ^2 using (see, e.g., Hendry, 1995, p.475):

$$\chi_k^2(\varphi_{r,\alpha}^2) = h\chi_m^2(0), \quad (30)$$

such that:

$$h = \frac{k + 2\varphi_{r,\alpha}^2}{k + \varphi_{r,\alpha}^2} \text{ and } m = \frac{k + \varphi_{r,\alpha}^2}{h}. \quad (31)$$

Finally, we calculate the power function of the $\chi_k^2(\varphi_{r,\alpha}^2)$ test in (29) using:

$$P[\chi_k^2(\varphi_{r,\alpha}^2) > c_\alpha \mid H_1] \simeq P[\chi_m^2(0) > h^{-1}c_\alpha] \quad (32)$$

For example, when $k = 20$ and $\varphi_{r,\alpha}^2 = 5$, then $h = 70/45 \simeq 1.56$ and $m \simeq 29$ with $c_\alpha \simeq 31.4$ for $\chi_{20}^2(0)$ so $P[\chi_{29}^2(0) > 20.1] \simeq 0.89$, delivering high power. Reducing the degrees of freedom of the test would increase power at the same $\varphi_{r,\alpha}$, or more generally there is a trade-off between fewer terms and the value of $\varphi_{r,\alpha}$.

3.3 Powerless case

The test will have no power if the departure from linearity is in the direction of $\mathbf{u}_{t,\perp}$, which would require $\Delta_n = \mathbf{0}_n$ when $\mathbf{D} \neq \mathbf{0}$. This may occur if the $x_{i,t}$ s were perfectly orthogonal, such that $\Omega = \sigma_x^2 \mathbf{I}_n$, and the non-linearity entered in the form of a cross-product. In this case, the matrix of eigenvectors would be the identity matrix, and the resulting linear combinations would exclude cross-product terms. However, the practical relevance of such a case is likely to be limited. Furthermore, the test would have no power if the second derivative of $f(\cdot)$ was zero, but the third was non-zero: this is precisely the reason for seeking to include the additional terms $\sum_{j=1}^n \delta_{2,j} z_{j,t}^3$. If the first non-zero derivative is the fourth, the test will have power, as the second derivative will almost certainly be correlated with the fourth; similarly for the third and fifth. Monte Carlo experiments are undertaken to confirm these conjectures.

3.4 Power approximations for a polynomial function

Consider a polynomial DGP given by:

$$y_t = \beta x_t^i + \epsilon_t, \quad (33)$$

where $\epsilon_t \sim \text{IN}[0, 1]$ and $x_t \sim \text{IN}[0, 1]$, for $t = 1, \dots, T$, and $i = 1, \dots, 4$, such that the four cases are a linear, quadratic, cubic and quartic function. The analytic distribution is calculated as:

$$\frac{T\hat{\beta}^2 \text{E}[(x_t^i)^2]}{\sigma_\epsilon^2} \sim \chi_1^2(\varphi_{\alpha,i}^2) \quad (34)$$

with non-centrality:

$$\varphi_{\alpha,i}^2 = \frac{T\beta^2 \text{E}[(x_t^i)^2]}{\sigma_\epsilon^2}, \quad (35)$$

where $\text{E}[(x_t)^2] = 1$; $\text{E}[(x_t^2)^2] = 3$; $\text{E}[(x_t^3)^2] = 15$; and $\text{E}[(x_t^4)^2] = 105$.

Power functions for the four DGPs are recorded in Figure 1, based on 10,000 replications for $\varphi_{\alpha,i} = 1, \dots, 6$, along with the analytic power function calculated using the χ^2 -approximation outlined in Section 3.2. In finite samples (panels a and b, $T = 100$) the divergence between the analytic and Monte Carlo test powers is substantial for cubic and quartic, particularly for intermediate non-centralities. This is probably due to the high skewness and kurtosis of the distribution of x_t^n , which impacts on the distribution of the t-statistic under the alternative, as the sample mean can be far from $\text{E}[(x_t^n)^2]$.

4 Power Simulations

Monte Carlo experiments are undertaken to examine the power of the test (hereafter denoted the index test) for static DGPs, for varying degrees of collinearity and numbers of regressors. Four tests are undertaken, namely: the optimal F-test on the known non-linear functions in (9), which has maximum power; White's test, which is an F-test on the squares and cross-products of all the regressors; the index test, which computes the F-test for the orthogonalized quadratic functions; and the extended index test including orthogonalized cubic functions. To summarize, the four fitted models on which the respective tests are based have the forms:

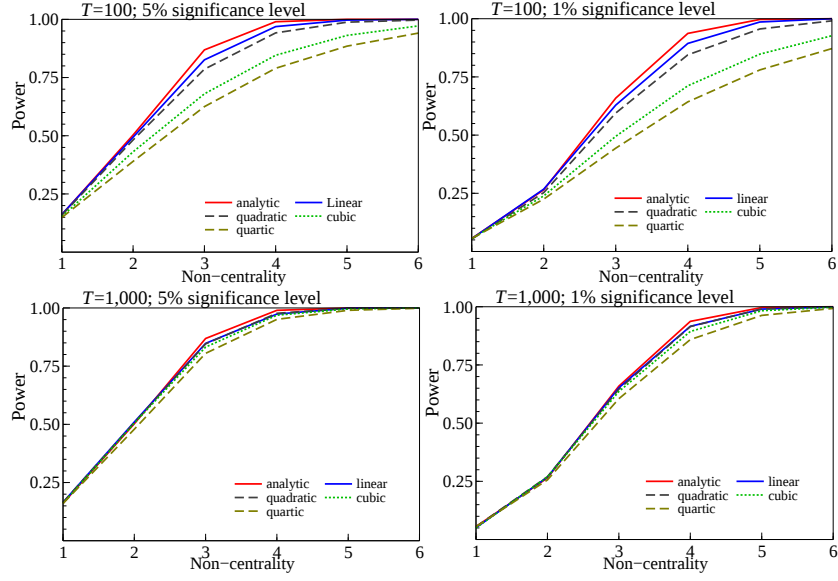


Figure 1: Analytic and Monte Carlo power functions for a polynomial function

1. Optimal test, $H_0 : \phi = 0$ for:

$$y_t = \sum_{i=1}^n \beta_i x_{i,t} + \phi g(x_{1,t} \dots x_{n,t}) + v_t. \quad (36)$$

2. White's test, $H_0 : \delta = \mathbf{0}$ for:

$$y_t = \sum_{i=1}^n \beta_i x_{i,t} + \sum_{j=1}^n \sum_{k=1}^n \delta_{j,k} x_{j,t} x_{k,t} + u_t. \quad (37)$$

3. Index test, $H_0 : \gamma = \mathbf{0}$ for:

$$y_t = \sum_{i=1}^n \beta_i x_{i,t} + \sum_{j=1}^n \gamma_j z_{j,t}^2 + e_t. \quad (38)$$

4. Extended index test, $H_0 : \psi = \theta = \mathbf{0}$ for:

$$y_t = \sum_{i=1}^n \beta_i x_{i,t} + \sum_{j=1}^n \psi_j z_{j,t}^2 + \sum_{k=1}^n \theta_k z_{k,t}^3 + \eta_t. \quad (39)$$

The data are generated by:

$$\mathbf{x}_t \sim \text{IN}_n[\mathbf{0}, \mathbf{\Omega}] \quad (40)$$

where $V[x_{i,t}] = 1, \forall i$, and $\text{cov}[x_{i,t} x_{j,t}] = \rho, \forall i \neq j$. n is the number of linear regressors in each model, increasing from two to twelve in the general unrestricted model, of which only two ($x_{1,t}$ and $x_{2,t}$) are in fact relevant.

4.1 Quadratic DGP

Consider the DGP given by:

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{1,t}^2 + \epsilon_t, \quad (41)$$

where $\epsilon_t \sim \text{IN}[0, 1]$ for $t = 1, \dots, T$.

The optimal test of $H_0 : \beta_3 = 0$, has a non-centrality of $\varphi_{r,\alpha}^2 = 3T\beta_3^2$, which is independent of ρ , the degree of collinearity.

The power of White's test for (41) will depend on:

$$\delta_1 x_{1,t}^2 + \delta_2 x_{2,t}^2 + \delta_3 x_{1,t} x_{2,t}, \quad (42)$$

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

$$\varphi_{r,\alpha}^2 = T [3\delta_1^2 + 2\rho^2 \delta_1 \delta_2 + 12\rho \delta_1 \delta_2 + 3\delta_2^2 + 12\rho \delta_2 \delta_3 + \rho^2 \delta_3^2]. \quad (43)$$

As $\delta_2 = \delta_3 = 0$, the test non-centrality collapses to the optimal test non-centrality of $3T\delta_1^2$. Hence, the difference between the optimal test and White's test will be a function of the number of degrees of freedom alone.

The index test for (41) is based on $\gamma_1 z_{1,t}^2 + \gamma_2 z_{2,t}^2$, where:

$$z_{1,t} = x_{1,t} + \kappa_1 x_{2,t} \quad (44)$$

$$z_{2,t} = x_{2,t} + \kappa_2 x_{1,t} \quad (45)$$

where κ is based on the eigenvalues, Λ , of Ω (assuming a zero mean, the $z_{i,t}$ are given by $(\mathbf{H}'\mathbf{x}_t) \Lambda^{-\frac{1}{2}}$). If $\Omega = \mathbf{I}_2$, this implies $\Lambda = \mathbf{I}$, and $\mathbf{H} = \mathbf{I}_2$, such that $z_{1,t}^2 = x_{1,t}^2$ and $z_{2,t}^2 = x_{2,t}^2$. In practice, sampling errors will imply $\hat{\Omega} \neq \mathbf{I}_2$ such that $\mathbf{z}_t$ comprises a linear combination of the $\mathbf{x}_t$ s:

$$z_{1,t}^2 = x_{1,t}^2 + \kappa_1^2 x_{2,t}^2 + 2\kappa_1 x_{1,t} x_{2,t} \quad (46)$$

$$z_{2,t}^2 = x_{2,t}^2 + \kappa_2^2 x_{1,t}^2 + 2\kappa_2 x_{1,t} x_{2,t} \quad (47)$$

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

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

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

so the index test power will depend on:

$$\gamma_1 (x_{1,t}^2 + \kappa_1^2 x_{2,t}^2 + 2\kappa_1 x_{1,t} x_{2,t}) + \gamma_2 (x_{2,t}^2 + \kappa_2^2 x_{1,t}^2 + 2\kappa_2 x_{1,t} x_{2,t}) \quad (48)$$

which can be solved using:

$$\gamma_1 + \gamma_2 \kappa_2^2 \approx \beta \quad (49)$$

$$\gamma_1 \kappa_1^2 + \gamma_2 \approx 0 \quad (50)$$

$$2\gamma_1 \kappa_1 + 2\gamma_2 \kappa_2 \approx 0. \quad (51)$$

Under orthogonality, the analytic non-centrality collapses to that of the optimal test, $3T\gamma_1^2$, but with fewer degrees of freedom than White's test. Collinearity impacts on the non-centrality via the κ weighting.

Two sample sizes are considered, $T = 100$ and 300 , and $M = 10,000$ replications are undertaken. We set $\beta_1 = \beta_2 = 0.3$ and $\beta_3 = 0.1732$ in (41), which results in non-centralities of 3 for all regressors under orthogonality, for $T = 100$. The coefficients are fixed for $T = 300$ to assess the impact of increasing sample size. The results are reported in Figure 2, where the number of linear regressors is recorded along the horizontal axis and power is recorded on the vertical axis. The top two panels record results for $T = 100$, and the bottom two panels record results for $T = 300$. The divergence between the analytic and optimal test is evident, as discussed above. Both the White and index test have a similar power for $n = 2$, but White's power declines rapidly as n increases because the degrees of freedom increase by $\frac{n(n+1)}{2}$. The decline in the index test power as n increases is not as sharp, because the degrees of freedom increase by n . The extended index test (including cubic terms) has a lower power than the index test using squares, although the magnitude of the loss in power is not substantial and is fairly constant across n and ρ . The index test has a higher power under collinearity due to interactions between sampling error and collinearity; as $\hat{\Omega} \neq \Omega$, under orthogonality, the cross-products would be contaminated, whereas a high ρ would increase the power to detect non-linearity via collinear squares and cross-products. As T increases to 300, the power of all tests increase, with a unit power for the analytic calculation, and near unit for the infeasible optimal test.

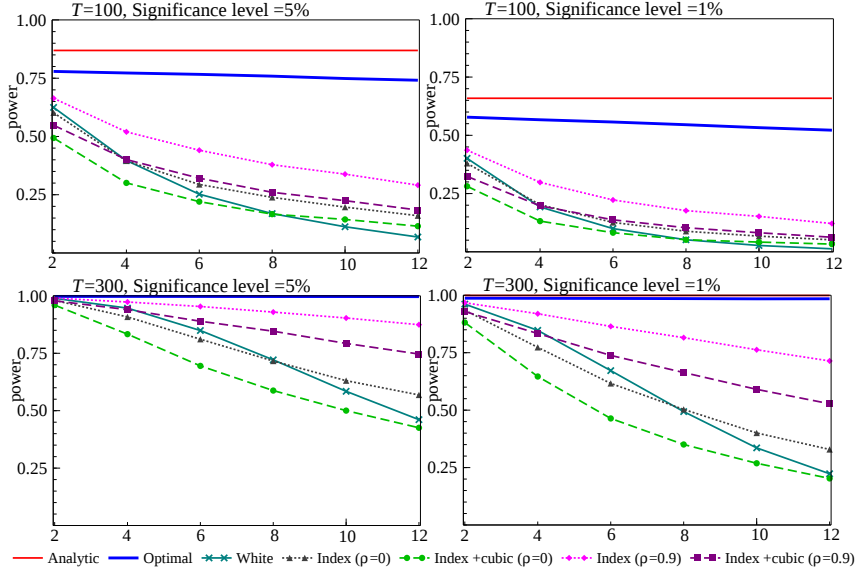


Figure 2: Power of non-linearity tests for a quadratic function

4.2 Quadratic DGP including a cross-product

We next consider augmenting the DGP with a cross-product term, given by:

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{1,t}^2 + \beta_4 x_{1,t} x_{2,t} + \epsilon_t, \quad (52)$$

where $\epsilon_t \sim \text{IN}[0, 1]$ and $\mathbf{x}_t$ is generated by (40). We set $\beta_1 = \beta_2 = \beta_4 = 0.2$ and $\beta_3 = 0.1155$, to result in non-centralities of 2 for the individual regressors 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 (52) 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:

$$\varphi_{\alpha,i}^2 = T [3\beta_3^2 + \beta_4^2 \rho^2 + 12\beta_3 \beta_4 \rho]. \quad (53)$$

Hence, the non-centrality of the joint F-test is $\varphi_{r,\alpha} = 5.7$ for $\rho = 0.9$, delivering a much higher power for all tests under collinearity.

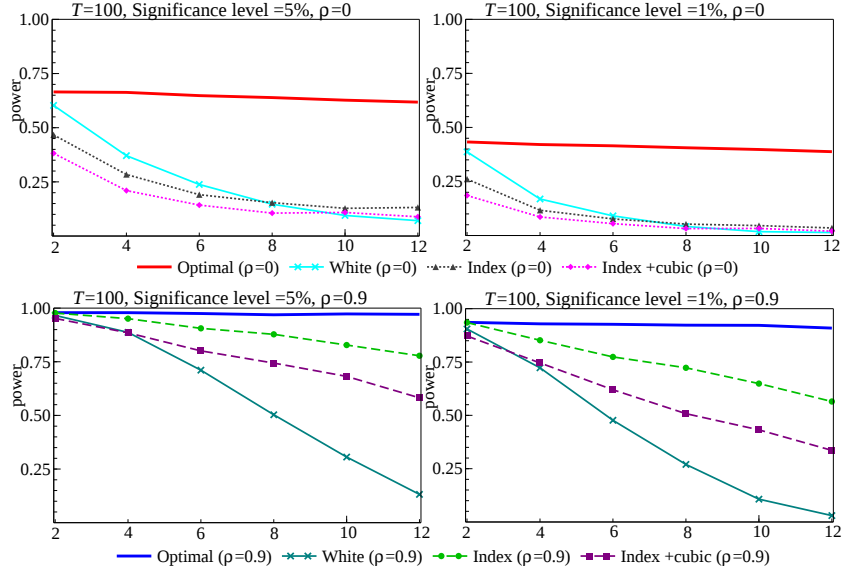


Figure 3: Power of non-linearity tests for a quadratic and cross-product function

Under orthogonality, the test must have low power against the cross-product term in (52). However, as the index is based on $\hat{\Omega}$, sampling errors will result in a non-zero off-diagonal as in (46) and (47). For the test to have power against (52), we require a low weight on the $x_{2,t}^2$ term in the $z_{1,t}^2$ equation. Consider a weighted average of the regressors:

$$z_{1,t}^2 + \delta z_{2,t}^2 = (1 + \delta \kappa_2^2) x_{1,t}^2 + (\delta + \kappa_1^2) x_{2,t}^2 + 2(\kappa_1 + \delta \kappa_2) x_{1,t} x_{2,t}. \quad (54)$$

Hence, a negative δ would down weight the $x_{2,t}^2$, yielding the highest power for the index test: regression should select such a weighting. Nevertheless, there is no close approximation to $\beta_3 x_{1,t}^2 + \beta_4 x_{1,t} x_{2,t}$ since $\delta = -\kappa_1^2$ yields (normalizing on $z_{1,t}^2$):

$$z_{1,t}^2 - \kappa_1^2 z_{2,t}^2 = (1 - \kappa_1^2 \kappa_2^2) x_{1,t}^2 + 2\kappa_1 (1 - \kappa_1 \kappa_2) x_{1,t} x_{2,t} \simeq x_{1,t}^2 + 2\kappa_1 x_{1,t} x_{2,t}, \quad (55)$$

which will be close only when $\beta_4 = 2\kappa_1 \beta_3$ for small κ_1 .

Under orthogonality, White's test has a higher power than the index test for small n , with the degrees of freedom advantages of the index test only resulting in a higher power when n is larger than 8. Under collinearity, the index test has a much higher power, outperforming White's test for all n .

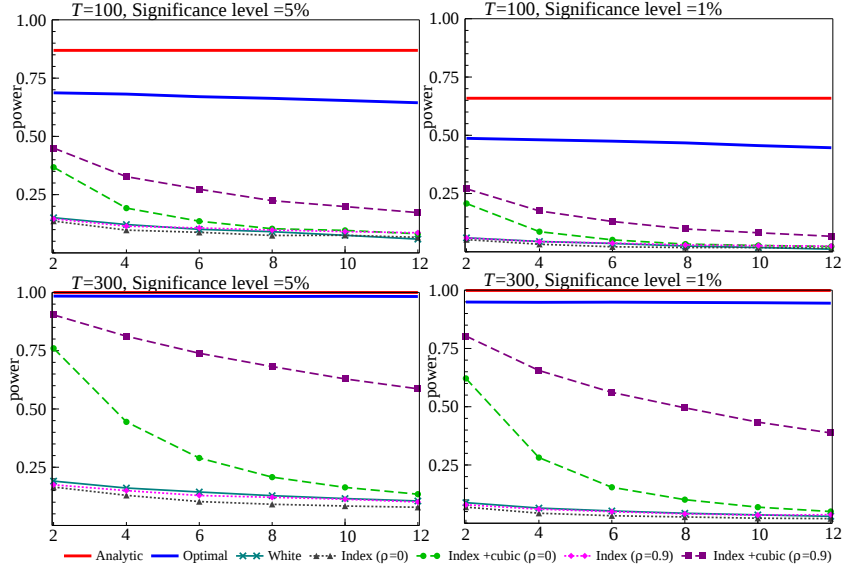


Figure 4: Power of non-linearity tests for a cubic function

4.3 Cubic DGP

We next consider the case in which a cubic term is included in the DGP:

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{1,t}^3 + \epsilon_t, \quad \epsilon_t \sim \text{IN}[0, 1]. \quad (56)$$

Under orthogonality, the non-centralities are:⁴

$$\varphi_{r,\alpha}^2(\beta_1) = \frac{\beta_1^2 \mathbb{E}[\mathbf{x}_1' \tilde{\mathbf{Q}} \mathbf{x}_1]}{\sigma_\epsilon^2} = 0.4T\beta_1^2, \quad (57)$$

$$\varphi_{r,\alpha}^2(\beta_2) = T\beta_2^2 \quad (58)$$

$$\varphi_{r,\alpha}^2(\beta_3) = \frac{\beta_3^2 \mathbb{E}[\mathbf{x}_1^{3'} \mathbf{Q} \mathbf{x}_1^3]}{\sigma_\epsilon^2} = 6T\beta_3^2. \quad (59)$$

such that we can analyze $x_{2,t}$ independently.

We set $\beta_1 = 0.4743$, $\beta_2 = 0.3$ and $\beta_3 = 0.1225$, which results in non-centralities of 3 for all regressors for $T = 100$ under orthogonality, and again hold the coefficients constant to assess the impact of increasing the sample size to $T = 300$.

Figure 4 records the results. The cost of including the cubic term in the index, in terms of degrees of freedom, is small compared to the loss in power if the cubic terms are excluded. Both the index test

⁴Using partitioning (see, e.g., Hendry, 1995, p.700), where we define $\mathbf{Q} = \mathbf{I} - \mathbf{x}_1(\mathbf{x}_1' \mathbf{x}_1)^{-1} \mathbf{x}_1'$, and $\tilde{\mathbf{Q}} = \mathbf{I} - \mathbf{x}_1^3(\mathbf{x}_1^{3'} \mathbf{x}_1^3)^{-1} \mathbf{x}_1^{3'}$.

including squares and White's test have similar powers, and there is no benefit to using the index test as n increases. The extended index test naturally declines in power as n increases, so there may be benefits to specifying a parsimonious model, although care is required if the non-linear function does not also enter linearly. A high degree of collinearity is again beneficial; the linear combination of the $\mathbf{x}_t$ s for $\mathbf{z}_t$ is:

$$z_{1,t}^3 = x_{1,t}^3 + 3\kappa_1 x_{1,t}^2 x_{2,t} + 3\kappa_1^2 x_{1,t} x_{2,t}^2 + \kappa_1^3 x_{2,t}^3 \quad (60)$$

$$z_{2,t}^3 = x_{2,t}^3 + 3\kappa_2 x_{2,t}^2 x_{1,t} + 3\kappa_2^2 x_{2,t} x_{1,t}^2 + \kappa_2^3 x_{1,t}^3 \quad (61)$$

Therefore, if $\rho = 0.9$, the index test will gain power to detect $x_{1,t}^3$ via the linear combinations, $x_{1,t}^2 x_{2,t}$ and $x_{1,t} x_{2,t}^2$. The gap between the analytic and optimal test is larger than for the quadratic DGP, in keeping with our previous analysis. Furthermore, for small n , the extended index test has a power close to that of the optimal test, especially at $T = 300$, despite the test including all the irrelevant quadratic terms as well.

4.4 Quartic DGP

While a quartic function is somewhat extreme, and the small sample distribution of the t-statistic is poor, we investigate whether the index test based on quadratic functions has power against quartic functions due to the collinearity between the two. Hence, we assess a quartic DGP given by:

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{1,t}^4 + \eta_t, \quad (62)$$

where $\beta_1 = \beta_2 = 0.3$ and $\beta_3 = 0.0293$, which delivers non-centralities of 3 for all regressors under orthogonality, for $T = 100$.

The results are recorded in Figure 5, where the substantial gap between the analytic and optimal test is evident due to regressor kurtosis. While the power is lower than that for the quadratic function, both White's test and the index test do have power against a quartic function. The patterns exhibited by the power functions correspond to those for the quadratic function, although the power does not decline to the same extent as n increases. Including the cubic term in the index test does not improve the power as there is no correlation between the third and fourth order polynomials.

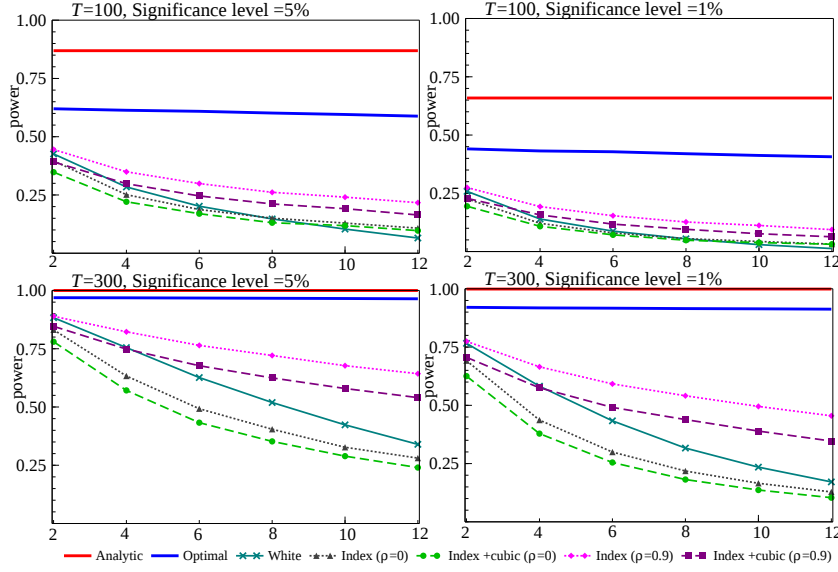


Figure 5: Power of non-linearity tests for a quartic function

4.5 LSTR model

There are many feasible non-linear functional forms for the experimental design. We focus on the LSTR model 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. The Monte Carlo is necessarily equation specific, but it is indicative of the performance of the index test both against White’s test and absolutely to detect this type of departure. The DGP is given by:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + (\delta_0 + \delta_1 x_{1,t} + \delta_2 x_{2,t}) [1 + \exp(-\gamma(x_{1,t} - c))]^{-1} + \epsilon_t, \quad (63)$$

where $\epsilon_t \sim \text{IN}[0, 1]$ and the $\mathbf{x}_t$ process is generated by (40). Two parametrizations are considered, given by $\beta_0 = 0.2, \delta_0 = 0.8, \beta_1 = \delta_1 = 0.3, \beta_2 = \delta_2 = 0.4, \gamma = 2.5$, and $c = 0.5$; and alternatively, $\beta_0 = 0, \delta_0 = 0.2, \beta_1 = \beta_2 = 0.2, \delta_1 = \delta_2 = 0.4, \gamma = 2$, and $c = 0.1$. The first parameterization delivers ‘strong non-linearity’, in the sense that the two regimes are clearly distinct. The second parameterization delivers ‘weak non-linearity’, as the two regimes are closer in mean, and the transition function is less rapid. $T = 100$, and is increased to $T = 200$ and $T = 300$ for the ‘strong’ and ‘weak’ cases respectively, for $M = 10,000$ replications.

While the optimal infeasible test based on the LSTR specification is not computed, we do compute

the power of the feasible test based on a third-order Taylor approximation. Replacing the transition function by:

$$[1 + \exp \{-\gamma (x_{1,t} - c)\}]^{-1} \simeq \frac{1}{2} + \frac{\gamma (x_{1,t} - c)}{4} - \frac{(\gamma (x_{1,t} - c))^3}{48}, \quad (64)$$

results in the approximation of (63):

$$y_t \simeq \theta_0 + \theta_1 x_{1,t} + \theta_2 x_{2,t} + \theta_3 x_{1,t}^2 + \theta_4 x_{1,t}^3 + \theta_5 x_{1,t}^4 + \theta_6 x_{1,t} x_{2,t} + \theta_7 x_{1,t}^2 x_{2,t} + \theta_8 x_{1,t}^3 x_{2,t} + \varepsilon_t. \quad (65)$$

Hence, the Taylor approximation test is highly parameterized, and there may be degrees of freedom gains from the index test.

The results are recorded in Figures 6 and 7 for the two alternative DGPs. All tests have high power under ‘strong non-linearity’, indicating that polynomial approximations do capture non-linearities that are generated by a smooth transition model. The power under collinearity is substantially and consistently higher than the power under orthogonality for all tests. The index test including cubic terms does not deliver a higher power than just including the quadratic terms, indicating that the third order terms in the approximation do not contain substantial extra information on the non-linearity despite (65). Under weaker non-linearity, the power function for White’s test exhibits much higher power for $T = 300$ when $\rho = 0.9$ than would be anticipated given the results for $T = 100$. Alternative LSTR specifications are needed to draw more precise conclusions, but to the extent that (65) is a reasonable approximation, our proposed test will have power against (63). Conversely, rejecting an initial specification does not entail that the alternative must be a polynomial function.

5 Dynamic models

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 based on an ADL(1,1) given by:

$$y_t = \beta_1 x_{1,t} + \beta_2 y_{t-1} + \beta_3 x_{1,t-1} + \beta_4 x_{1,t-1}^2 + \epsilon_t \quad (66)$$

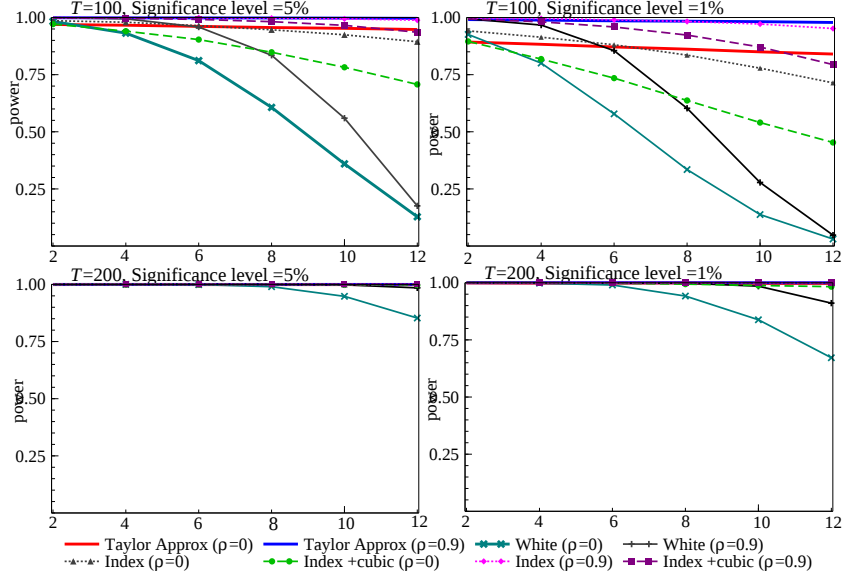


Figure 6: Power of non-linearity tests for an LSTR function: strong non-linearity

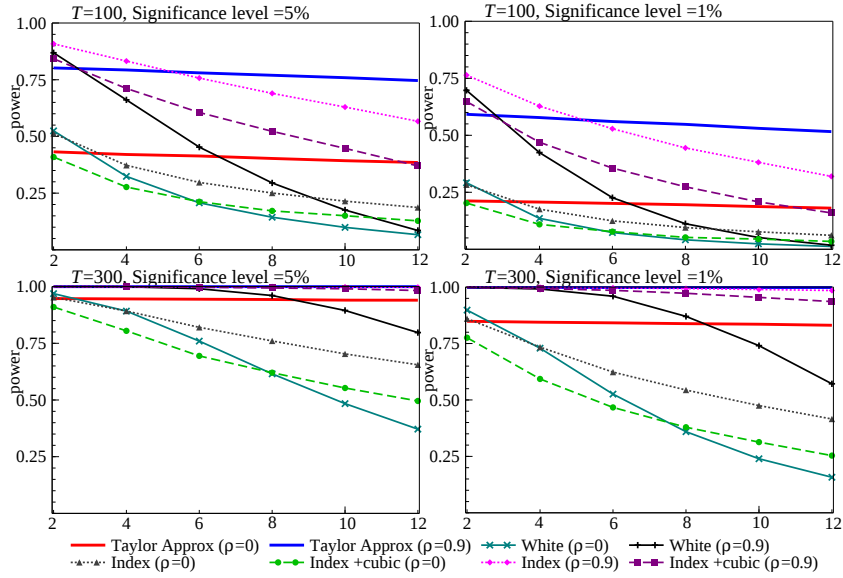


Figure 7: Power of non-linearity tests for an LSTR function: weak non-linearity

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

$$\mathbf{x}_t = \lambda \mathbf{x}_{t-1} + \alpha y_{t-1} + \boldsymbol{\epsilon}_{2,t} \quad (67)$$

where $|\lambda| < 1$ and $|\alpha| < 1$, and

$$\boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \text{IN}_{1+n} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right]. \quad (68)$$

We generate $n = 1, \dots, 5$ regressors, $\mathbf{x}_t$, based on (67), of which one (x_1) enters the DGP (67). Under the null of linearity, $\beta_4 = 0$. Strong exogeneity requires no feedback, such that $\alpha = 0$. The requirement for weak exogeneity is that $\gamma = \sigma_{12.1} = 0$, where $\sigma_{12.1}$ corresponds to $\text{cov}[y, x_1]$.⁵ 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.⁶ White's test will check for the significance of $y_{t-1}^2, x_{i,t}^2, x_{i,t-1}^2, y_{t-1}x_{i,t}, y_{t-1}x_{i,t-1}, x_{i,t}x_{j,t-1}, \forall i, j \leq 5$, resulting in $2n + 1$ regressors and $\frac{(2n+1)(2n+2)}{2}$ degrees of freedom. The index test will have $2n + 1$ degrees of freedom and the extended index test will have $2(2n + 1)$ degrees of freedom.

First consider the case of strongly exogenous dynamic regressors. The experiments are conducted for parameter values of $\alpha = 0, \sigma_{12} = 0, \beta_1 = 0.3, \beta_2 = 0.5, \beta_3 = 0.3, \lambda = 0.5, \sigma_{11} = \sigma_{22.i} = 1$ for $i = 1, \dots, 5$, and $E[\epsilon_{2.i,t}\epsilon_{2.j,t}] = \sigma_{22.ij} = 0.5 \forall i \neq j$. For more persistent dynamics, we set $\beta_2 = \lambda = 0.8$ where all other parameters are held the same. We also consider the impact of increasing the sample size to 300, holding the parameters fixed. Figure 8 records the sizes and powers of the tests.⁷

The results support the contention that the proposed F-test has a large-sample actual size close to its nominal, at least at 5% and 1%. As the degree of persistence in the lagged dependent variable and the exogenous regressors increases, the size becomes slightly distorted as more variables are included. This is particularly evident for the extended index test, which has a size of approximately 6% at a nominal size of 5%. The size is close to the nominal size for less persistent processes. The large-sample power

⁵Under the null, the joint density of the DGP is given by $\mathbf{w}_t | \mathbf{w}_{t-1} \sim \text{N}_2[\boldsymbol{\mu}_t, \boldsymbol{\Omega}]$, where $\mathbf{w}_t = (y_t, x_{1,t})'$ and

$$\boldsymbol{\mu}_t = \begin{pmatrix} \beta_1\alpha + \beta_2 & \beta_1\lambda + \beta_3 \\ \alpha & \lambda \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{1,t-1} \end{pmatrix}, \boldsymbol{\Omega}_{1.1} = \begin{pmatrix} \sigma_{11} + \beta_1^2\sigma_{22.1} + 2\beta_1\sigma_{12.1} & \sigma_{12.1} + \beta_1\sigma_{22.1} \\ \sigma_{12.1} + \beta_1\sigma_{22.1} & \sigma_{22.1} \end{pmatrix}. \quad (69)$$

⁶We assume the lag length is determined in advanced and is fixed for the test.

⁷For brevity, we only report results for $T = 100$. As the sample size increases, the size of the test is similar and the power increases, in keeping with asymptotic theory. We only record powers for a 5% significance level as the powers are similar for the 1% significance level. Full results are available on request.

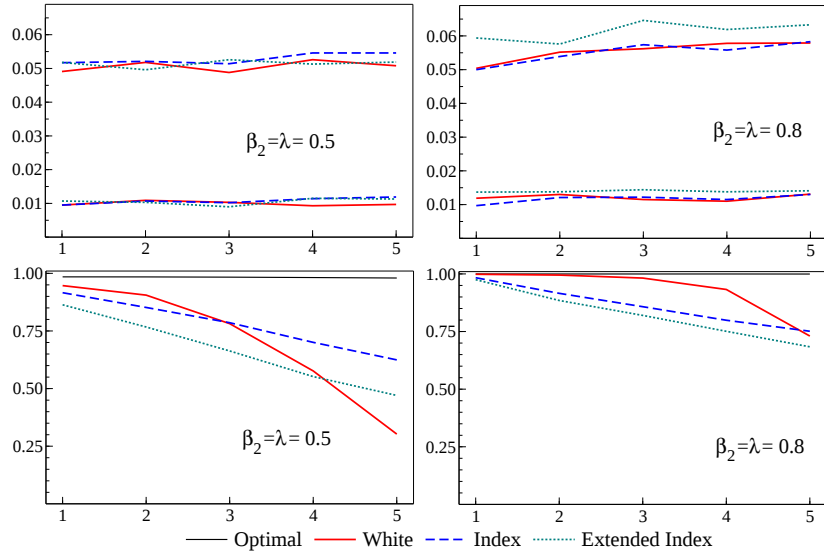


Figure 8: Size and power for ADL(1,1) model with strong exogeneity

is reasonable and delivers similar results to the static cases considered above: White's test has a higher power than the index test initially, given the DGP design, but as additional regressors are included the power declines more rapidly than the index and extended index test. A higher correlation between regressors, or a more complex form of non-linearity, will favour the index test.

Secondly, if the levels data are non-stationary, the size of the test is distorted. Figure 9 (panel a), records the size for $I(1)$ exogenous regressors ($\lambda = 1$): it is evident that the size increases rapidly as more variables are included due to correlations between the relevant and irrelevant variables driven by the unit roots. However, if the dependent variable contains a unit root but the exogenous regressors are stationary, there are no size distortions. Panel b records the size when $\beta_2 = 1$ and $\lambda = 0.5$. A unit coefficient on the lagged dependent variable results in a differenced process estimated in levels, given the precision of the estimated lagged dependent variable, resulting in a stationary process. Non-linear functions of the lagged dependent variable will be non-stationary, but regression of $I(1)$ variables on a $I(0)$ process does not lead to spurious regression, see Hendry (1995, p.129). Hence, unit roots in the conditional model result in a size equal to the nominal size.

As size distortions can occur due to unit roots in the marginal model, the data can be reduced to $I(0)$ by appropriate differencing and cointegration transformations. This will recreate the strongly exogenous

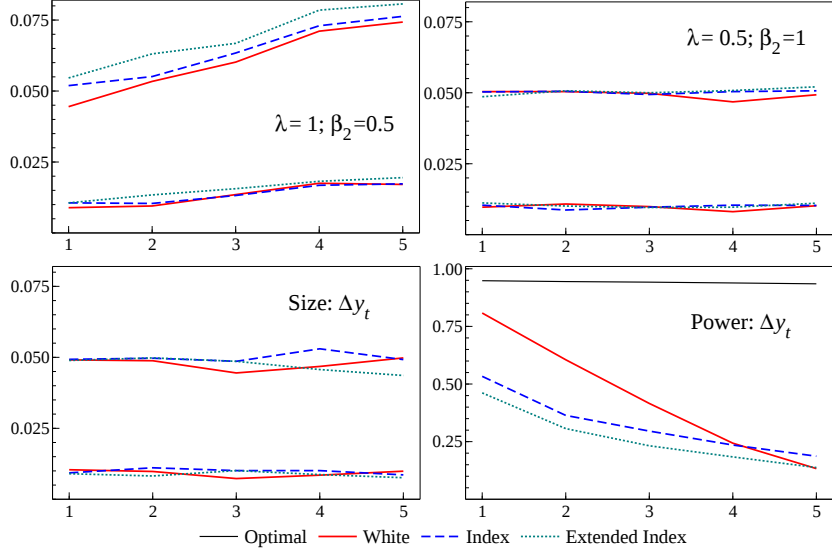


Figure 9: Size and power for strongly exogenous ADL(1,1) model with unit root or differenced model

dynamic case in the transformed variables, assessed above. If $\lambda = \beta_2 = 1$, we can define the DGP under the null as:

$$\Delta y_t = \beta_1 \Delta x_{1,t} + \beta_3 \Delta x_{1,t-1} + \epsilon_t \quad (70)$$

where

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \epsilon_{2,t} \quad (71)$$

The size of the test under appropriate differencing is close to the nominal size, recorded in figure 9c, and the power, recorded in figure 9d at the 5% significance level, reflects that of the stationary dynamic case. If the original formulation is in non-stationary variables, so the non-linear functions have complicated behaviour, 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 with the problems seen in section 3.4.

Thirdly, if the regressors are only weakly exogeneous, then again, the null distribution is close to the F-distribution derived earlier at least in large samples, with similar power behaviour. We confirm this by relaxing strong exogeneity in the above example and set $\alpha = -0.5$. To calculate the power, we set $\beta_4 = 0.15$ to ensure convergence, giving a non-centrality of approximately 3.5. In these experiments,

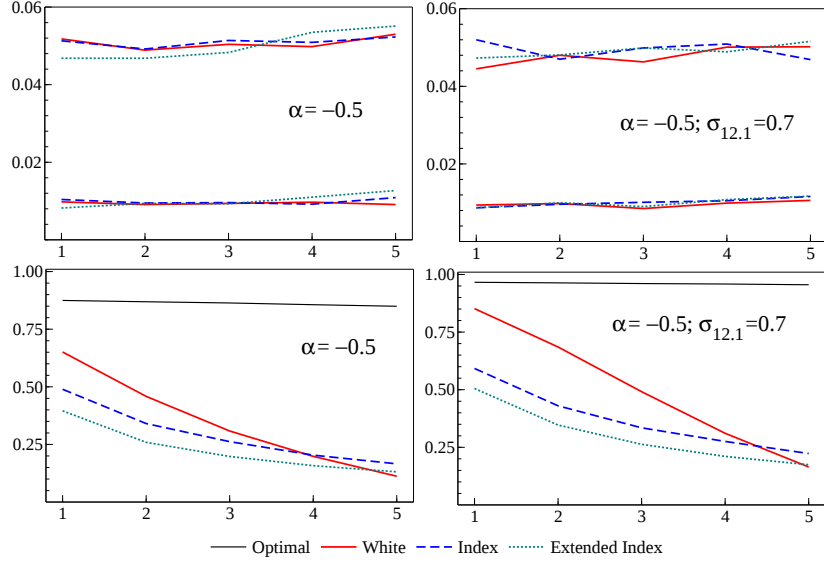


Figure 10: Size and power for a failure of strong and weak exogeneity

we set $\sigma_{22,ij} = 0, \forall i \neq j$. The sizes and powers of the test are recorded in figure 10, panels a and c. Relaxing the assumption of weak exogeneity as well as strong exogeneity, by setting $\sigma_{12,1} = 0.7$ and $\alpha = -0.5$, does not distort the size of the test and furthermore, leads to an increase in power for both White's test and the index test: see panels b and d.

Finally, we consider the choice of the matrix to be factorized in order to orthogonalize the regressors, namely Ω or $\mathbf{Q} = \lambda \mathbf{Q} \lambda' + \Omega$, the standardized second-moment matrix (about means) of the $\mathbf{x}_t$ given by (67). The natural choice would seem to be $\mathbf{Q}$, especially as $\hat{\mathbf{Q}}$ is readily available, and does not depend on postulating a model like (67) for the $\mathbf{x}_t$. In the dynamic case, the $\hat{\Omega}$ matrix may not be feasible as the orthogonalization is required for y_{t-1} and $\mathbf{x}_{t-1}$ as well as $\mathbf{x}_t$. One option would be to specify an autoregressive model for each of the $2n + 1$ regressors, including sufficient lags to ensure the residuals are white noise, then the residuals can be used to formulate the $\hat{\Omega}$ matrix.

We undertake a simulation experiment based on the strongly exogenous dynamic model, in which we orthogonalize using either $\hat{\mathbf{Q}}$ or $\hat{\Omega}$, where $\hat{\Omega}$ is computed as a diagonal matrix with the variance of the residuals from an AR(1) process for each regressor on the diagonals.⁸ In the first experiment, λ

⁸In practice, the lag length of the AR processes should be determined by testing for white-noise residuals, for example by using a general-to-specific lag selection strategy.

is diagonal, so each marginal model is an autoregressive process. Figure 11 records the sizes for both methods of factorization in panels a and b: the sizes are close to the nominal sizes in both cases. As there is little to distinguish between the two options, the factorization based on available data using $\hat{\mathbf{Q}}$ would be recommended as the easiest.

In the second experiment, we examine the case where λ contains non-zero off-diagonal elements, such that the marginal model for the relevant regressor, $x_{1,t}$, is a function of other variables. The DGP is given by (66), but (67) is replaced with:

$$x_{1,t} = \lambda_{1,2}x_{2,t-1} + \lambda_{1,3}x_{3,t-1} + \epsilon_{2.1,t} \quad (72)$$

where

$$x_{j,t} = \gamma x_{j,t-1} + \epsilon_{2.j,t} \quad \text{for } j = 2, \dots, 5 \quad (73)$$

We set $\lambda_{1,2} = \lambda_{1,3} = \gamma = 0.5$. The size for each factorization is recorded in figure 11, c and d. Again, the factorization has no impact on the size of the test, with size close to the nominal size in both cases. Constructing the $\hat{\Omega}$ matrix to be diagonal will give zero weight on the cross product terms in the factorization. Thus, if the DGP does contain linear combinations of the regressors, the $\hat{\mathbf{Q}}$ orthogonalization will clearly be preferable. In the above case, this would be evident if $x_{1,t}$ were excluded.

6 Conclusion

To conclude, we find that White's test and the index test perform comparatively for small n , but the power of White's test declines more sharply as n increases due to the rapidly increasing degrees of freedom. A preferable test may be to use White's test for $T \gg n^2$ and small n , and then switch to the index test as $n^2 > Tk$, where k is some threshold value such as 0.25. Furthermore, while parsimony delivers a higher power, such that selection prior to implementing the test would appear to be beneficial, this may be a hazardous strategy if the linear term is irrelevant, but 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 via a non-linear transformation, resulting in a lower power to detect non-linearity if such a variable is excluded.

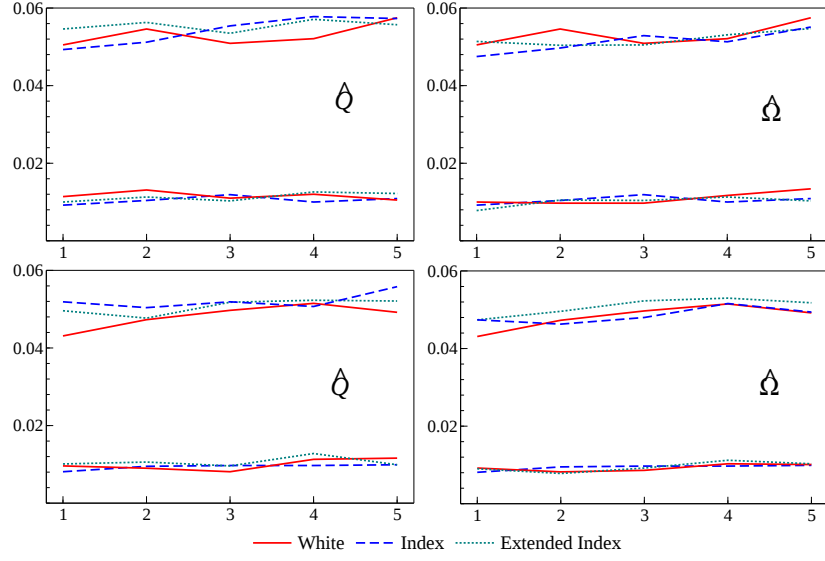


Figure 11: Sizes for ADL(1,1) model with diagonal and non-diagonal strong exogeneity, comparing $\hat{\Omega}$ and $\hat{Q}$

When the functional form and the set of relevant variables are both unknown, but nest the local DGP, the extended index test using quadratic and cubic terms has power to reject a false null in a wide range of circumstances: pure quadratic, pure cubic, pure quartic, and these in combinations, even for highly collinear data. We surmise it will have some power to detect quintics, but doubt the relevance of that case. The index test outperforms White's test in most of these situations, and can even be close to the optimal infeasible test. For larger departures from linearity, where several non-linear terms occur for several variables, its power will dominate that in the experiments illustrated here. Furthermore, simulation experiments suggest the test has the appropriate size and power properties for dynamic models. Thus, it promises to be a useful mis-specification test, providing a basis for examining the functional form of the initially-specified general model.

References

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.

- Castle, J. L., and Hendry, D. F. (2005). Extending the boundaries of PcGets: Non-linear models. Mimeo, Economics Department, University of Oxford.
- Chan, K. S., and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, **7**, 179–194.
- Doornik, J. A. (1995). Econometric Computing. Ph.D Thesis, University of Oxford, Oxford.
- Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.
- Hamilton, J. D. (1989). A new approach to economic analysis of nonstationary time series. *Econometrica*, **57**, 357–384.
- 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.
- 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.
- Lorenz, H.-W. (1989). *Nonlinear Dynamic Economics and Chaotic Motion*. New York: Springer-Verlag.
- Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity in univariate time series models. *Scandinavian Journal of Statistics*, **15**, 161–175.
- 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.
- Perez-Amaral, T., Gallo, G. M., and White, H. (2003). A flexible tool for model building: the relevant transformation of the inputs network approach (RETINA). *Oxford Bulletin of Economics and*

- Statistics*, **65**, 821–838.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*. New York: Academic Press.
- Puu, T. (1989). *Nonlinear Economic Dynamics*. New York: Springer-Verlag.
- Quandt, R. E. (1983). Computational problems and methods.. pp. 699–746: North Holland.
- 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.
- Subba Rao, T. (1985). On the theory of bilinear time series models. *Journal of the Royal Statistical Society, B*, **43**, 244–255.
- Subba Rao, T., and Gabr, M. M. (1984). *Introduction to Bispectral Analysis and Bilinear Time Series Models*. New York: Springer-Verlag.
- Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208–218.
- Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.
- Tsay, R. S. (1986). Non-linearity tests for time series. *Biometrika*, **73**, 461–466.
- White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.
- White, H. (1989). Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of the American Statistical Association*, **84**, 1003–1013.
- White, H. (1992). *Artificial Neural Networks: Approximation and Learning Theory*. Oxford: Basil Blackwell.