

The Spurious Effect of ARCH Errors on Linearity Tests: A Theoretical Note and an Alternative Maximum Likelihood Approach

Abstract

Linearity tests against smooth transition nonlinearity are typically based on the standard least-squares (LS) covariance matrix estimator. We derive an expression for the bias of the LS estimator in the presence of ARCH errors. We show that the bias is downward, and increases dramatically with the persistence of the variance process. As a consequence, conventional tests spuriously indicate nonlinearity. Next, we examine an alternative maximum likelihood approach. Our findings suggest that this approach has substantially better size properties than tests based on least-squares and heteroskedasticity-consistent matrix estimators, and performs comparably to a bootstrap technique.

Keywords: linearity tests; smooth transition models; ARCH; spurious inference.

JEL Classification: C12, C22

1 Introduction

During the last few decades nonlinear models have become very popular in economic theory and in econometrics (Ruge-Murcia, 2012, Haldrup, Meitz, and Saikkonen, 2014). This popularity has led to the development of several tests that examine whether a time series exhibits nonlinear dynamics. At the same time, a vast empirical literature has emerged that suggests that many economic and financial variables exhibit volatility clustering.

A question that naturally arises is what are the properties of linearity tests in the presence of time-varying volatility. Granger and Teräsvirta (1993) highlight that ARCH errors may cause complications of linearity tests. Wong and Li (1997) show that a test for threshold autoregression can be heavily oversized in the presence of ARCH innovations. A similar finding is provided by Hurn and Becker (2009) for the neural network test of Teräsvirta, Lin, and Granger (1993), and by Pavlidis, Ivan, and Peel (2010) for the test of Escribano and Jordá (1999).

The findings of Wong and Li (1997), Hurn and Becker (2009), and Pavlidis et al. (2010) are based on Monte Carlo simulations. In this paper, we adopt a simple theoretical framework to investigate analytically the properties of three popular linearity tests against smooth transitional nonlinearity in the presence of ARCH dynamics. We show that because ARCH dynamics induce a downward bias in the LS covariance matrix estimator, all three tests frequently indicate the presence of nonlinearity in mean even when the process under examination is in fact linear. Having established the poor performance of linearity tests based on LS, we examine an alternative maximum likelihood approach, and compare its performance to that of four widely-used non-parametric methods: three heteroskedasticity-consistent covariance matrix estimators and the fixed design wild bootstrap.

2 Testing Linearity

Consider a basic Smooth Transition Autoregressive (STAR) model given by

$$y_t = \pi_0 + \pi_1 y_{t-1} + \pi_2 y_{t-1} F(y_{t-1}; \gamma, c) + \varepsilon_t, \quad t = 0, \dots, T, \quad (1)$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$. The transition function $F(\cdot)$ is at least fourth order continuously differentiable with respect to γ and bounded between 0 and 1. The selection of the transition function specifies the two common forms of the STAR model. For the exponential STAR, the transition function is given by

$$F(y_{t-1}; \gamma, c) = 1 - \exp\left(-\gamma(y_{t-1} - c)^2\right), \quad (2)$$

and for the logistic STAR, the transition function is

$$F(y_{t-1}; \gamma, c) = [1 + \exp(-\gamma(y_{t-1} - c))]^{-1}, \quad (3)$$

where c is a constant, y_{t-1} is the transition variable, and $\gamma \in (0, \infty)$ is the smoothness parameter that determines the speed of transition of $F(\cdot)$ towards the inner or outer regime of the process and, therefore, the degree of nonlinearity.

Several testing procedures have been proposed in the literature to examine whether a series exhibits STAR nonlinearity (see, e.g., Teräsvirta, 2006). Testing for the nonlinear part of Equation (1) is not trivial due to an unidentified parameter problem (Davies, 1987). Specifically, linearity corresponds to two null hypotheses, $H_0: \pi_2 = 0$ and $H_0: \gamma = 0$. Under the first, the parameters γ and c are not identified. Under the second, the unidentified parameters are π_2 and c . Consequently, classical test statistics do not follow standard distributions.

A way to overcome this problem is to replace the transition function by a Taylor series approximation around $\gamma = 0$. This re-parameterization resolves the identification problem since it does not involve nuisance parameters. Luukkonen, Saikkonen, and Teräsvirta (1988) suggest replacing the transition function by a first order Taylor series approximation. This yields the following auxiliary regression

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1}^3 + u_t, \quad (4)$$

where $u_t = \varepsilon_t + R(\gamma, y_{t-1})$ and $R(\cdot)$ is the remainder term of the Taylor series. The null hypothesis of linearity takes the form $H_0: \beta_2 = \beta_3 = 0$. A drawback of the Luukkonen et al. (1988) test is that it may exhibit low power when y_t is an LSTAR process and only intercept changes are significant across regimes (Escribano and Jordá, 2001). To achieve better performance, Teräsvirta (1994) advocates the use of a third order Taylor series approximation of the logistic function. This gives rise to the auxiliary regression

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1}^3 + \beta_4 y_{t-1}^4 + u_t. \quad (5)$$

and the null hypothesis of linearity becomes $H_0: \beta_2 = \beta_3 = \beta_4 = 0$. While Teräsvirta (1994) uses a third order Taylor expansion of the logistic function and a first order expansion for the exponential function, Escribano and Jordá (1999) suggest augmenting the regression equation with a second order expansion of the exponential function. The idea of using a second order Taylor expansion is based on the fact that the logistic function has one inflection point but the exponential has two. The resulting auxiliary regression is

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1}^3 + \beta_4 y_{t-1}^4 + \beta_5 y_{t-1}^5 + u_t, \quad (6)$$

and the null hypothesis of linearity $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$.

A general representation that encompasses all auxiliary regressions is

$$y_t = \beta' x_t + u_t, \quad (7)$$

where $\beta = (\beta_0, \dots, \beta_n)'$ and $x_t = (1, y_{t-1}, y_{t-1}^2, \dots, y_{t-1}^n)'$. For the tests of Luukkonen et al. (1988), Teräsvirta (1994), and Escribano and Jordá (1999) n is equal to 3, 4, and 5, respectively. Using matrix notation, the null hypothesis of linearity can be written as $H_0: R\hat{\beta} = 0$, where R is the appropriate selector matrix. The corresponding Wald test statistic is given by

$$W = \left(R\hat{\beta} \right)' \left(R V_{\hat{\beta}} R' \right)^{-1} \left(R\hat{\beta} \right), \quad (8)$$

where $V_{\hat{\beta}} = (X'X)^{-1} X' \Omega X (X'X)^{-1}$ denotes the covariance matrix of the least squares estimator $\hat{\beta}$. Typically, residuals are assumed to be homoskedastic, $\Omega^0 = I \sigma_u^2$, which implies that W follows a χ^2 distribution with degrees of freedom equal to the number of restrictions $n - 1$, and $W/(n - 1)$ follows an F distribution with degrees of freedom $n - 1$ and $T - n - 1$.

2.1 Linearity Testing in the Presence of ARCH Errors

Suppose that the true DGP for y_t is the ARCH(1) process of Engle (1982)

$$y_t = \varepsilon_t \sigma_t, \quad (9)$$

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2, \quad (10)$$

with $\varepsilon_t \sim \mathcal{N}(0, 1)$ and $\alpha \in [0, 1)$, and consider the ratio of the true generalized variance to the generalized variance based on the assumption of homoskedastic errors

$$\frac{\det(V_{\hat{\beta}})}{\det(V_{\hat{\beta}}^0)} = \frac{\det\left(\frac{1}{T} X' \Omega X\right)}{\det\left(\frac{1}{T} X' \Omega^0 X\right)}. \quad (11)$$

In the presence of ARCH, Ω and Ω^0 are diagonal matrices with tt element equal to the conditional variance $\sigma_t^2 = \omega + \alpha y_{t-1}^2$ and the unconditional variance $\omega/(1 - \alpha)$, respectively. It follows that the ij element of $(1/T)X' \Omega X$ is

$$v_{ij} = 1/T \left(\omega \sum y_{t-1}^{i+j-2} + \alpha \sum y_{t-1}^{i+j} \right) \simeq \omega E y_t^{i+j-2} + \alpha E y_t^{i+j}, \quad (12)$$

where $i, j = 1, \dots, n+1$, and Ey_t^{i+j} is the $i+j$ moment of y_t . Similarly, the ij element of $(1/T)X'\Omega^0X$ is

$$h_{ij} = 1/T \frac{\omega}{1-\alpha} \sum y_{t-1}^{i+j-2} \simeq \frac{\omega}{1-\alpha} Ey_t^{i+j-2}. \quad (13)$$

Because all the odd moments of an ARCH process are equal to zero, both matrices have a chessboard structure (i.e., $v_{ij} = h_{ij} = 0$ for $i+j$ odd). Further, we can see from (12) and (13) that the true generalized variance is a function of the first $n+1$ even moments of y_t while the generalized variance based on the assumption of homoskedasticity is a function of just the first n moments.

The regularity condition for the existence of the $2m$ th moment of an ARCH(1) process (Engle, 1982, Theorem 1) is

$$\xi_m \alpha^m < 1, \quad (14)$$

where

$$\xi_1 = 1, \quad \xi_j = \prod_{i=1}^j (2i-1), \quad j = 1, 2, \dots$$

For the tests of Luukkonen et al. (1988), Teräsvirta (1994), and Escribano and Jordá (1999) $n = 3, 4$, and 5 and, hence, the moments required are the 8th, 10th and 12th, respectively. From (14), these moments exist if α is smaller than 0.31, 0.25, and 0.21. In general, as higher powers of the lag dependent variable enter in the auxiliary regression (i.e., as n increases), the region for the ARCH coefficient for which all the required moments exist becomes smaller. Unfortunately, even for relatively small values of n , as for example for the test of Luukkonen et al. (1988), there is no guarantee that the regularity condition (14) will be satisfied in empirical applications.

Bollerslev (1986, Theorem 2) provides a recursive expression for the $2m$ th moment of an ARCH process

$$E(y_t^{2m}) = \frac{\xi_m}{1 - \xi_m \alpha^m} \left[\sum_{k=0}^{m-1} \xi_k^{-1} E(y_t^{2k}) \omega^{m-k} \binom{m}{m-k} \xi_k \alpha^k \right]. \quad (15)$$

Given the moments of y_t , the ratio of generalized variances (11) can be calculated using the Leibniz formula for determinants. For example, for the Luukkonen et al. test, we obtain

$$\frac{\det(V_{\hat{\beta}})}{\det(V_{\hat{\beta}}^0)} = \frac{v_{11}v_{22}v_{33}v_{44} + v_{22}^2v_{33}^2 - v_{11}v_{33}^3 - v_{22}^3v_{44}}{h_{11}h_{22}h_{33}h_{44} + h_{22}^2h_{33}^2 - h_{11}h_{33}^3 - h_{22}^3h_{44}}. \quad (16)$$

The expressions for the Teräsvirta (1994) and the Escribano and Jordá (1999) tests are lengthy and we omit them to save space. Figure 1 plots the log of $\det(V_{\hat{\beta}})/\det(V_{\hat{\beta}}^0)$ against the ARCH coefficient for the three linearity tests. We observe that the ratio of generalized variances increases rapidly with the value for the ARCH coefficient, and with the number of regressors n in the auxiliary regression. Strikingly, for values close to the upper bound implied by the moment condition, the true generalized variance is about 10^4 times larger than the generalized variance based on the assumption of homoskedasticity. The extreme bias of $V_{\hat{\beta}}^0$ highlights why caution should be taken when testing a series for nonlinearity and there are suspicions of conditional heteroskedasticity.

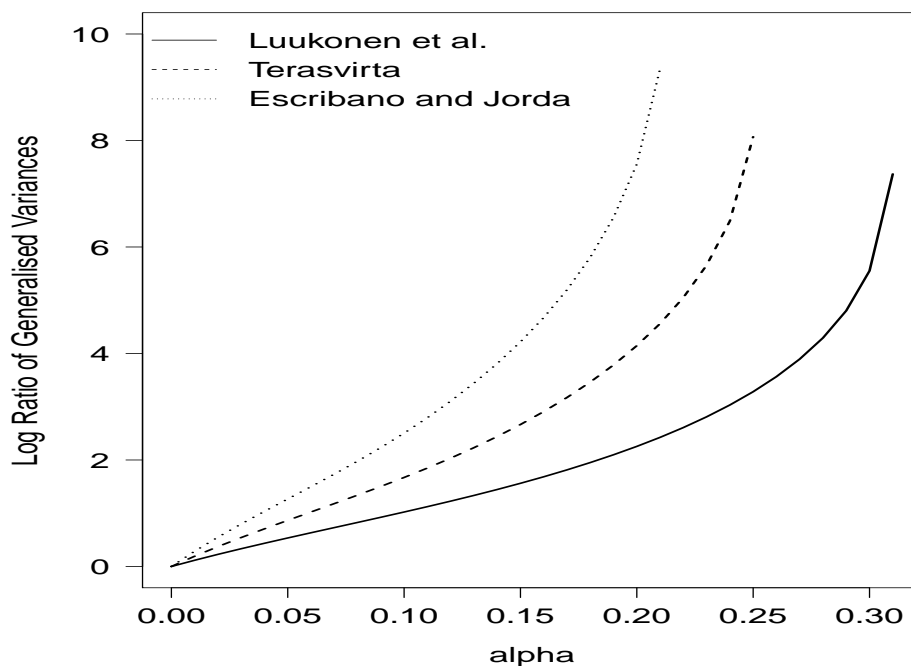


Figure 1: The figure plots the log of the true generalized variance to the generalized variance based on the assumption of homoskedastic errors, $\det(V_{\hat{\beta}})/\det(V_{\hat{\beta}}^0)$, against the value of the ARCH coefficient for the linearity tests of Luukkonen et al. (1988) (solid line), Teräsvirta (1994) (dashed line), and Escribano and Jordá (1999) (dotted line).

To illustrate the effect of the bias of the least squares variance estimator on the performance of linearity tests, we estimate the empirical size of the tests (i.e., the number of times each test rejects the null of linearity when the null is true) when y is an ARCH process. Table 1 reports results for $\alpha \in \{0.00, 0.05, \dots, 0.40\}$ and sample size $T = 100, 500$ and 5000 . The nominal significance level is set equal to 5 percent and the number of Monte Carlo simulations equal to 5000. In line with the theoretical results, the degree of oversizing increases with the ARCH coefficient and with n . Furthermore, it increases with the sample size. As T and α take larger values, it becomes more likely to reject the null of linearity than not, even though the process under consideration is linear. For instance, for $n = 4$, $\alpha = 0.4$, and $T = 5000$, the tests of Luukkonen et al. (1988), Teräsvirta (1994), and Escribano and Jordá (1999) falsely reject linearity in 61, 73 and 80 percent of the cases, respectively.

α	T=100			T=500			T=5000		
	LST	T	EJ	LST	T	EJ	LST	T	EJ
0.00	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05
0.05	0.07	0.07	0.07	0.08	0.09	0.09	0.10	0.12	0.12
0.10	0.08	0.09	0.09	0.11	0.13	0.14	0.14	0.18	0.21
0.15	0.11	0.12	0.13	0.16	0.18	0.20	0.21	0.28	0.32
0.20	0.14	0.15	0.17	0.20	0.24	0.27	0.27	0.36	0.42
0.25	0.15	0.17	0.19	0.25	0.29	0.33	0.36	0.46	0.54
0.30	0.18	0.21	0.23	0.30	0.36	0.41	0.45	0.57	0.66
0.35	0.20	0.24	0.27	0.36	0.43	0.48	0.52	0.63	0.71
0.40	0.24	0.27	0.31	0.42	0.49	0.55	0.61	0.73	0.80

Table 1: The table reports the empirical size of the linearity tests of Luukkonen et al. (1988) (LST), Teräsvirta (1994) (T), and Escribano and Jordá (1999) (EJ).

3 A Maximum Likelihood Approach

The results of the previous section demonstrate in a clear manner how ARCH errors can induce substantial downward bias in the LS covariance matrix estimator. A direct implication of this bias is that Wald statistics do not follow standard distributions, and F and χ^2 tests can frequently lead to spurious results. A number

of alternative methods have been put forth in the literature for drawing statistical inference in this context. The most popular, due to their easy implementation and low computational cost, are methods based on heteroskedasticity-consistent (HC) covariance matrix estimators. These estimators include, among others, the HC0: $\widehat{\Omega} = \text{diag}(\widehat{u}_t^2)$ of White (1980), and two estimators proposed by MacKinnon and White (1985), HC2: $\widehat{\Omega} = \text{diag}(\widehat{u}_t^2/(1 - h_{tt}))$ and HC3: $\widehat{\Omega} = \text{diag}(\widehat{u}_t^2/(1 - h_{tt})^2)$, where h_{tt} denotes the t th diagonal element of the hat matrix.

In addition to HC estimators, researchers have also employed alternatives based on bootstrap techniques (Hurn and Becker, 2009). The basic idea behind these techniques is to approximate the unknown distribution of the test statistic by simulating under the null. Although computationally intensive, these techniques appear to display good size and power properties. A bootstrap technique that appears to perform particularly well in the context of linearity testing is the fixed-design wild bootstrap (WB) of Wu (1986) and Mammen (1993). For a detailed description of the WB, see Pavlidis et al. (2010).

The HC and WB methods are non-parametric, in that they do not require the specification of the functional form of the variance process. It may well be the case, however, that a parametric approach which involves modelling both the mean of y_t and the error variance displays superior performance. Perhaps surprisingly, there is little, if any, work on this topic. We contribute to the existing literature by examining the performance of a simple approach that consists of estimating the general auxiliary regression (7)

$$y_t = \beta' x_t + u_t,$$

together with an ARCH process for the error variance

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2,$$

by maximum likelihood (ML), and testing the null of linearity $H_0: R\widehat{\beta} = 0$.¹ The conditional log-likelihood function of (7) given $u_t, t = 1, \dots, T$, is

$$l(\theta) = -\frac{1}{2} \sum_{t=1}^T (\ln \sigma_t + u_t^2 / \sigma_t),$$

where $\theta = (\beta, \omega, \alpha)'$, and the corresponding ML estimator is $\widehat{\theta} = \text{argmax}_{\theta \in \Theta} l(\theta)$, with Θ being a compact subset of \mathbb{R} . As shown by Engle (1982), the information

¹We are grateful to a referee for motivating the use of a ML approach, and for bringing the reference of Pantula (1988) to our attention.

matrix associated with $\hat{\theta}$ is block-diagonal. Therefore, the estimation of parameters in the conditional mean and conditional variance functions can be implemented in two stages while maintaining asymptotic efficiency. In related work, Weiss (1984) and Pantula (1988) illustrate that ML estimates are strongly consistent and asymptotically normal for stationary autoregressive models with ARCH errors, and Chan and McAleer (2002) derive similar results for stationary STAR-GARCH models.

3.1 Monte Carlo Simulations

To compare the properties of the LS, HC, ML, and WB tests, we employ two Monte Carlo experiments. For both experiments, we set, as before, the nominal significance level to 5 percent, the number of simulations to 1000, the value of the ARCH coefficient to $\alpha = \{0.0, 0.1, 0.2, 0.3, 0.4\}$, $\omega = 1 - \alpha$, and the sample size to $T = 100, 500$, and 5000.

	T=100			T=500			T=5000		
	LST	T	EJ	LST	T	EJ	LST	T	EJ
$\alpha = 0.0$									
LS	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05
HC0	0.17	0.29	0.40	0.11	0.19	0.31	0.06	0.08	0.12
HC2	0.11	0.17	0.20	0.09	0.13	0.18	0.06	0.07	0.10
HC3	0.06	0.08	0.08	0.07	0.09	0.11	0.06	0.06	0.08
ML	0.06	0.03	0.03	0.05	0.02	0.03	0.06	0.04	0.03
WB	0.04	0.05	0.05	0.06	0.05	0.06	0.06	0.05	0.05
$\alpha = 0.1$									
LS	0.08	0.09	0.09	0.11	0.13	0.14	0.14	0.18	0.21
HC0	0.18	0.33	0.46	0.12	0.23	0.35	0.06	0.11	0.18
HC2	0.12	0.19	0.22	0.09	0.16	0.20	0.05	0.10	0.13
HC3	0.07	0.09	0.09	0.07	0.11	0.10	0.05	0.08	0.09
ML	0.08	0.04	0.03	0.05	0.03	0.04	0.05	0.05	0.04
WB	0.04	0.05	0.06	0.05	0.05	0.04	0.04	0.06	0.06
$\alpha = 0.2$									
LS	0.14	0.15	0.17	0.20	0.24	0.27	0.27	0.36	0.42
HC0	0.21	0.39	0.52	0.16	0.28	0.41	0.10	0.17	0.28
HC2	0.14	0.22	0.29	0.12	0.19	0.26	0.09	0.13	0.17
HC3	0.09	0.10	0.11	0.09	0.12	0.13	0.08	0.10	0.11
ML	0.07	0.03	0.06	0.05	0.04	0.05	0.05	0.06	0.04

WB	0.07	0.06	0.06	0.07	0.07	0.06	0.06	0.06	0.05
$\alpha = 0.3$									
LS	0.18	0.21	0.23	0.30	0.36	0.41	0.45	0.57	0.66
HC0	0.26	0.42	0.54	0.23	0.37	0.52	0.12	0.24	0.42
HC2	0.16	0.24	0.30	0.17	0.23	0.32	0.10	0.17	0.25
HC3	0.10	0.10	0.12	0.12	0.13	0.14	0.08	0.12	0.14
ML	0.09	0.04	0.06	0.05	0.06	0.06	0.05	0.05	0.04
WB	0.06	0.06	0.06	0.06	0.07	0.07	0.06	0.06	0.07
$\alpha = 0.4$									
LS	0.24	0.27	0.31	0.42	0.49	0.55	0.61	0.73	0.80
HC0	0.31	0.47	0.61	0.27	0.44	0.59	0.15	0.31	0.48
HC2	0.20	0.28	0.32	0.19	0.26	0.35	0.12	0.20	0.28
HC3	0.12	0.12	0.10	0.12	0.15	0.16	0.09	0.11	0.13
ML	0.08	0.05	0.06	0.07	0.05	0.06	0.07	0.06	0.06
WB	0.07	0.09	0.08	0.08	0.08	0.08	0.06	0.05	0.05

Table 2: The table reports the empirical size of the LS, HC0, HC2, HC3, ML and WB versions of the linearity tests of Luukkonen et al. (1988) (LST), Teräsvirta (1994) (T), and Escribano and Jordá (1999) (EJ).

We first examine the empirical size of the various tests by generating from a white noise DGP, $y_t \sim \mathcal{N}(0, \sigma_t^2)$. Table 2 reports the corresponding rejection rates. As is evident from the table, with the exception of HC3, tests based on heteroskedasticity-consistent covariance matrix estimators can display severe size distortions when the sample size is small and the ARCH coefficient takes large values. As the sample size increases, the degree of oversizing tends to decrease but it does not disappear even for $T = 5000$. Turning to HC3, we observe that it displays substantially better size properties than HC0 and HC2. However, the best performing methods are the ML and the WB. The rejection rates for these two methods are close to the 5 percent nominal significance level irrespective of the sample size and the value of the ARCH coefficient.

	T=100			T=500			T=5000		
	LST	T	EJ	LST	T	EJ	LST	T	EJ
$\alpha = 0.0$									
LS	0.19	0.17	0.24	0.74	0.74	0.91	1.00	1.00	1.00
HC0	0.16	0.06	0.08	0.61	0.44	0.34	1.00	1.00	1.00
HC2	0.16	0.07	0.10	0.56	0.41	0.64	1.00	1.00	1.00

HC3	0.14	0.12	0.14	0.52	0.40	0.75	1.00	1.00	1.00
ML	0.15	0.13	0.26	0.70	0.74	0.93	1.00	1.00	1.00
WB	0.22	0.22	0.30	0.76	0.78	0.93	1.00	1.00	1.00
$\alpha = 0.1$									
LS	0.12	0.12	0.18	0.53	0.50	0.79	1.00	1.00	1.00
HC0	0.09	0.05	0.08	0.44	0.30	0.19	0.99	0.99	1.00
HC2	0.10	0.09	0.09	0.41	0.31	0.42	0.99	0.98	1.00
HC3	0.09	0.10	0.13	0.36	0.32	0.59	0.99	0.97	1.00
ML	0.12	0.12	0.20	0.70	0.68	0.91	1.00	1.00	1.00
WB	0.17	0.18	0.27	0.61	0.64	0.86	1.00	1.00	1.00
$\alpha = 0.2$									
LS	0.09	0.09	0.14	0.29	0.27	0.56	0.99	0.99	1.00
HC0	0.08	0.06	0.09	0.30	0.14	0.07	0.91	0.85	0.97
HC2	0.09	0.06	0.09	0.30	0.23	0.22	0.89	0.82	1.00
HC3	0.11	0.09	0.12	0.30	0.26	0.45	0.86	0.81	1.00
ML	0.09	0.13	0.17	0.68	0.68	0.88	1.00	1.00	1.00
WB	0.18	0.19	0.27	0.52	0.56	0.82	0.99	0.99	1.00
$\alpha = 0.3$									
LS	0.05	0.04	0.07	0.15	0.15	0.28	0.83	0.78	0.99
HC0	0.06	0.07	0.07	0.14	0.05	0.05	0.69	0.36	0.18
HC2	0.07	0.06	0.07	0.18	0.12	0.08	0.70	0.43	0.86
HC3	0.09	0.08	0.10	0.18	0.16	0.26	0.67	0.46	0.94
ML	0.09	0.12	0.16	0.66	0.59	0.86	1.00	1.00	1.00
WB	0.16	0.18	0.26	0.42	0.48	0.73	0.90	0.94	1.00
$\alpha = 0.4$									
LS	0.04	0.05	0.07	0.06	0.06	0.08	0.41	0.36	0.75
HC0	0.04	0.05	0.06	0.12	0.06	0.06	0.43	0.11	0.05
HC2	0.06	0.05	0.06	0.13	0.07	0.05	0.46	0.23	0.41
HC3	0.09	0.05	0.12	0.14	0.10	0.17	0.44	0.31	0.79
ML	0.08	0.11	0.19	0.63	0.67	0.88	1.00	1.00	1.00
WB	0.15	0.16	0.26	0.31	0.38	0.62	0.72	0.81	0.96

Table 3: The table reports the empirical size-adjusted power of the LS, HC0, HC2, HC3, and ML versions of the linearity tests of Luukkonen et al. (1988) (LST), Teräsvirta (1994) (T), and Escribano and Jordá (1999) (EJ), and the empirical power of the WB.

For the power experiment, we consider the following ESTAR DGP

$$y_t = 0.3y_{t-1} - 0.9y_{t-1}(1 - \exp(-y_{t-1}^2)) + \varepsilon_t$$

which is similar to a STAR process employed by Escribano and Jordá (1999). Table 3 reports the empirical size-adjusted power of LS, HC and ML tests, as well as the empirical power of the wild bootstrap. At least two interesting conclusions emerge from the results. The first is that the test proposed by Escribano and Jordá (1999) is more powerful than those of Luukkonen et al. (1988) and Teräsvirta (1994). This result is expected since Escribano and Jordá (1999) augment their auxiliary regression with a second order expansion of the exponential function. Because the exponential transition function (2) possesses two inflection points, a second-order Taylor expansion leads to power improvements for ESTAR processes. The second conclusion is that, though the WB test is superior to all other tests for small samples, the ML test is the one that performs best for samples greater than 100 observations. Specifically, the ML displays similar power to the LS and WB methods in the case of homoskedasticity, and higher (in many cases substantially) power when the error term is heteroskedastic. Most notably, the difference in size-adjusted power between ML and LS (WB) tests reaches up to 80 (32) percentage points for $T = 500$ and $\alpha = 0.4$. The fact that the ML method exhibits good size and power properties, and its relatively low computational cost, make it particularly attractive for high-frequency financial time series.

4 Conclusion

In this paper, we adopted a simple framework to examine the properties of three linearity tests based on the LS covariance estimator in the presence of ARCH errors. We showed that the LS estimator is biased for all three tests, with the bias increasing with the persistence of the ARCH process and the order of the Taylor series expansion used to derive the auxiliary test regression. As a consequence, all tests heavily over-reject the null hypothesis of linearity for large values of the ARCH coefficient. We also examined the performance of an alternative maximum-likelihood approach to test for linearity. We showed through Monte Carlo experiments that this approach, contrary to LS tests, displays good size properties and, for relatively large samples, higher power than tests based on heteroskedasticity-consistent covariance matrix estimators and the fixed-design wild bootstrap. As such, it is particularly attractive for high-frequency financial data.

References

- Bollerslev, T. (1986): “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.
- Chan, F. and M. McAleer (2002): “Maximum likelihood estimation of STAR and STAR-GARCH models: Theory and monte carlo evidence,” *Journal of Applied Econometrics*, 17, 509–534.
- Davies, R. B. (1987): “Hypothesis testing when a nuisance parameter is present only under the alternative,” *Biometrika*, 74, 33–43.
- Engle, R. F. (1982): “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation,” *Econometrica*, 50, 987–1007.
- Escribano, A. and O. Jordá (1999): “Improved testing and specification of smooth transition regression models,” in P. Rothman, ed., *Nonlinear Time Series Analysis of Economic and Financial Data*, Dordrecht: Kluwer Academic Publishers, 289–320.
- Escribano, A. and O. Jordá (2001): “Testing nonlinearity: Decision rules for selecting between logistic and exponential star models,” *Spanish Economic Review*, 3, 193–209.
- Granger, C. W. J. and T. Teräsvirta (1993): *Modelling Nonlinear Economic Relationships*, Oxford University Press.
- Haldrup, N., M. Meitz, and P. Saikkonen (2014): *Essays in nonlinear time series econometrics*, Oxford University Press.
- Hurn, S. and R. Becker (2009): “Testing for Nonlinearity in Mean in the Presence of Heteroskedasticity,” *Economic Analysis and Policy*, 39, 311–326.
- Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988): “Testing linearity against smooth transition autoregressive model,” *Biometrika*, 75, 491–499.
- MacKinnon, J. G. and H. White (1985): “Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties,” *Journal of Econometrics*, 29, 305–325.
- Mammen, E. (1993): “Bootstrap and wild bootstrap for high dimensional linear models,” *Annals of Statistics*, 21, 255–285.

- Pantula, S. G. (1988): “Estimation of autoregressive models with arch errors,” *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, 50, 119–138.
- Pavlidis, E. G., P. Ivan, and D. A. Peel (2010): “Specifying Smooth Transition Regression Models in the Presence of Conditional Heteroskedasticity of Unknown Form,” *Studies in Nonlinear Dynamics & Econometrics*, 14, 1–40.
- Ruge-Murcia, F. (2012): “Estimating nonlinear dsge models by the simulated method of moments: With an application to business cycles,” *Journal of Economic Dynamics and Control*, 36, 914–938.
- Teräsvirta, T. (1994): “Specification, estimation, and evaluation of smooth transition autoregressive models,” *Journal of the American Statistical Association*, 89, 208–218.
- Teräsvirta, T. (2006): “Univariate nonlinear time series models,” *Palgrave Handbook of Econometrics*, 1, 396–424.
- Teräsvirta, T., C.-F. Lin, and C. W. J. Granger (1993): “Power of the neural network linearity test,” *Journal of Time Series Analysis*, 14, 209–220.
- Weiss, A. A. (1984): “Arma models with arch errors,” *Journal of Time Series Analysis*, 5, 129–143.
- White, H. (1980): “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica*, 48, 817–38.
- Wong, C.-S. and W.-K. Li (1997): “Testing for threshold autoregression with conditional heteroscedasticity,” *Biometrika*, 84, 407–418.
- Wu, C. F. J. (1986): “Jackknife, bootstrap and other resampling methods in regression analysis (with discussion),” *Annals of Statistics*, 14, 1261–1350.