Economics Bulletin

Volume 29, Issue 3

Testing for a unit root against ESTAR nonlinearity with a delay parameter greater than one.

Stephen Norman
University of Washington, Tacoma

Abstract

In this paper, the tests of Kapetanios, Shin, and Snell (2003) and Bec, Ben Salem, and Carrasco (2004), which are designed to detect nonstationarity verses globally stationary exponential smooth transition autoregressive (ESTAR) nonlinearity, are extended to allow for a delay parameter, d, that is greater than one. Based on Monte Carlo simulations, it is shown that when the true delay parameter is greater than one, using the test with the correct value of d improves power almost uniformly compared to constraining the delay parameter to be unity. Using the tests when the delay parameter is not known and must be estimated is also addressed.

I thank, without implicating in any way, Tim Vogelsang for useful discussions regarding this paper. I also am indebted to the comments of an anonymous referee. This paper was written as part of my Ph.D. dissertation.
Citation: Stephen Norman, (2009) "Testing for a unit root against ESTAR nonlinearity with a delay parameter greater than one.", *Economics Bulletin*, Vol. 29 no.3 pp. 2152-2173.
Submitted: Sep 11 2008. **Published:** September 02, 2009.

1 Introduction

Using nonlinear time series models has become increasingly popular with data, such as real exchange rates and real interest rates, where the presence of a unit root cannot be rejected using conventional linear unit root tests.[1] This has made testing for the presence of nonlinear mean reversion against nonstationarity an important topic of research. Papers that develop such tests include Kapetanios, Shin, and Snell (2003), hereafter KSS, and Bec, Ben Salem, and Carrasco (2004), hereafter BBC. These two tests are based upon Taylor approximations of smooth transition autoregressive (STAR) models. STAR models which are used to describe nonlinear mean reversion are almost always specified as being "self-exciting." This implies that the transition variable, which controls the degree of transition from one regime to another, is a lagged dependent variable with lag or delay, d.

The delay parameter measures the time it takes for the adjustment process to begin that will tend to restore an economic variable to its long run equilibrium. The tests of KSS and BBC are limited by the fact that the delay parameter is assumed to be unity. To understand the practical importance of allowing for $d > 1$ it is instructive to focus on perhaps the most popular application of this type of nonlinear mean reversion, the study of real exchange rates. In the case of yearly real exchange rate data, a situation where the value of d is very large would be unlikely. On the other hand using monthly real exchange rate data, $d > 1$ is a distinct possibility. Baum et al. (2001), while studying nonlinear mean reversion in deviations from PPP, find evidence of delay parameters up to 12 months for many country pairs. Other examples of empirical evidence of delay parameters that are not one include Sarantis (1999) who finds evidence of $d > 1$ while using a STAR model to study real effective exchange rates. Also, Taylor, van Dijk, Franses, and Lucas (2000) find evidence of $d > 1$ when using a STAR model to study mispricing of futures and spot prices in a UK stock index where the data frequency is measured in minutes.

The main contribution of this paper is the generalization of the two tests of KSS and BBC to allow d to be any positive integer. It is shown that the asymptotic distributions of both test statistics is identical to the case when $d = 1$. Based on Monte Carlo simulations, it is also shown that when the true delay parameter is greater than one, using the test with the correct value of d improves power in almost all cases studied compared to using the constraint, $d = 1$. Following Hansen (1997), the theory developed in this paper is based upon the assumption that d is known with certainty. Suggestions for practical use of the test when d is unknown and must be estimated are also included.

The remainder of this paper is as follows: Section 2 describes the STAR model and associated tests of nonlinearity in the presence of possible nonstationarity. The asymptotic distribution of the generalized tests are also derived. Section 3 focuses on the small sample

[1]See Baum, Barkoulas, and Caglayan (2001), Kapetanios et al. (2003), Michael, Nobay, and Peel (1997), Sarno, Taylor, and Chowdhury (2004), Taylor, Peel, and Sarno (2001), Taylor and Peel (2000)

properties of the generalized tests. Section 4 contains suggestions for the use of the test when d is unknown. Section 5 concludes the paper.

2 Testing for the presence of a globally stationary ESTAR process against the null of a unit root

2.1 STAR model and Linearity tests

A general two regime smooth transition autoregressive (STAR) model can be expressed in the following manner:

$$y_t = (\alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p}) \\ + [(\beta_0 - \alpha_0) + (\beta_1 - \alpha_1)y_{t-1} + \ldots + (\beta_p - \alpha_p)y_p]R(z_t, \theta) + \varepsilon_t. \quad (1)$$

Whether the time series process follows an AR(p) model parameterized by $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_p)'$, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$, or some convex combination of the two is governed by the transition function $R(\cdot)$, which is itself a function of some transition variable z_t and a set of parameters θ. In the smooth transition model, $R(z_t, \theta)$ is a smooth function bounded between zero and one, $R : \mathbb{R} \rightarrow [0, 1]$. The value of the transition function determines the proportion of each regime present in the dynamics of the process given the value of the transition variable. There are two popular transition functions used in practice: the logistic function,

$$R(z_t; \theta) = \frac{1}{1 + \exp[-\gamma(z_t - c)]}, \quad (2)$$

and exponential function,

$$R(z_t; \theta) = 1 - \exp[-\gamma^2(z_t - c)^2]. \quad (3)$$

A STAR model with transition function (2) is known as an LSTAR model, and a STAR model with transition function (3) is known as an ESTAR model. In both cases, larger values of γ are associated with faster transitions. An important sub-class of STAR models are the self-exciting STAR processes. A transition model is self-exciting if the transition variable is a lagged dependent variable, $z_t = y_{t-d}$. The conditions for geometric ergodicity and associated asymptotic stationarity of the ESTAR model, which this paper will focus on, are discussed in Kapetanios, Shin, and Snell (2003).

The concept of testing linearity in the STAR framework is complicated by the fact that there are unidentified nuisance parameters under the null hypothesis of linearity. This can be seen in one of two ways. If the model is linear then either the parameter governing the rate of transition, γ, is zero or there is no difference between the autoregressive dynamics between regimes, $\alpha = \beta$. If the former is true then β is unidentified because the process always follows the autoregressive process parameterized by α. On the other hand, if there is no difference between regimes then different values γ will result in an identical model.

To deal with the problem of nuisance parameters in the STAR framework, Saikkonen and Luukkonen (1988) propose replacing the transition function with a second order Taylor series approximation around $\gamma = 0$. van Dijk, Teräsvirta, and Franses (2002) explain that, given this approximation, the error in the Taylor series approximation is then a part of the regression error term. Under the null, the Taylor approximation error would be zero, and as a result the properties of the error term would not be affected. The test of nonlinearity would then simply be a test that the coefficients on the variable affected by the Taylor approximation are zero. Specifically, let

$$R(z_t;\theta) = \delta_0 + \delta_1 z_t + \delta_2 z_t^2 + T(z_t;\theta) \tag{4}$$

be the Taylor approximation with $T(z_t;\theta)$ as the Taylor remainder term. Testing linearity is then based upon the auxiliary regression

$$y_t = \phi_0' x_t + \phi_1' x_t z_t + \phi_2' x_t z_t^2 + \varepsilon_t^*, \tag{5}$$

where $x_t = (1, y_{t-1}, \ldots, y_{t-p})$ and $\varepsilon_t^* = \varepsilon_t + T(y_{t-1};\theta)(\beta - \alpha)x_t$. The linearity test can then be stated as $H_0 : \phi_1 = \phi_2 = 0$ against the alternative that $H_1 : \phi_1 \neq 0$ or $\phi_2 \neq 0$.

2.2 Testing for ESTAR nonlinearity versus a unit root

The tests of KSS and BBC, which test for the presence of nonlinear mean reversion against the null hypothesis of a unit root, are based upon the Dickey-Fuller representation of the Taylor expansion used in the previously discussed linearity test proposed by Saikkonen and Luukkonen (1988) and discussed further in van Dijk et al. (2002). Specifically, the model employed is given by the following:

$$\Delta y_t = \sum_{r=r_1}^{r_2} \delta_r y_{t-1} y_{t-d}^r + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} + \varepsilon_t, \tag{6}$$

where $\varepsilon_t \sim$ i.i.d$(0, \sigma^2)$. The tests are based upon the statistical significance of the parameters $(\delta_{r_1}, \ldots, \delta_{r_2})$. KSS set $r_1 = r_2 = 2$ and derive the distribution of the t-statistic testing $\delta_2 = 0$ against the null hypothesis of $\delta_2 < 0$ when the true data generating process (DGP) follows a unit root and $d = 1$:

$$t_{\text{NL}} = \hat{\delta}_2 / \text{s.e.}(\hat{\delta}_2). \tag{7}$$

BBC set $r_1 = 1, r_2 = 2$ and derive the distribution of the Wald statistic, F_{NL}, testing $\delta_1 = \delta_2 = 0$ against the null of $\delta_1 \neq 0$ or $\delta_1 \neq 0$ when the actual DGP is a unit root and $d = 1$. The distributions of both test statistics are free of nuisance parameters. The proofs of the following theorems are given in the appendix.

Theorem 1. *Under the null of a unit root, for $d \in \{2, 3, 4, \ldots\}$ and $p \in \{0, 1, 2, \ldots\}$, the t_{NL} statistic has the following asymptotic distribution,*

$$t_{NL} \xrightarrow{D} \frac{\int_0^1 W(r)^3 dW(r)}{\sqrt{\int_0^1 W(r)^6 dr}}$$

where $W(r)$ is the standard Brownian motion.

Theorem 2. *Under the null of a unit root, for $d \in \{2, 3, 4, \ldots\}$ and $p \in \{0, 1, 2, \ldots\}$, the F_{NL} statistic has the following asymptotic distribution,*

$$F_{NL} \xrightarrow{D} h'Q^{-1}h$$

where $Q = \begin{pmatrix} \int_0^1 W(r)^4 dr & \int_0^1 W(r)^5 dr \\ \int_0^1 W(r)^5 dr & \int_0^1 W(r)^6 dr \end{pmatrix}$, $h = \begin{pmatrix} \int_0^1 W(r)^2 dW(r) \\ \int_0^1 W(r)^3 dW(r) \end{pmatrix}$ *and $W(r)$ is the standard Brownian motion.*

The distributions of both t_{NL} and F_{NL} for $d > 1$ are the same as the case when $d = 1$ (see KSS and BBC). This is convenient because the same asymptotical critical values can be used regardless of the value of d. KSS note that to accommodate processes with non-zero means and/or time trends, it may be necessary to de-mean and/or de-trend the data prior to calculating the appropriate test statistic. In each case the asymptotic distribution of both test statistics will be the same except that standard Brownian motion, $W(r)$, will be replaced by de-meaned and/or de-trended standard Brownian motion. For brevity and because the non-zero mean case is perhaps the most common case encountered empirically when using an ESTAR model, only the de-meaned case will be studied in this paper.

3 Small sample properties

Following KSS, the size properties of the t_{NL} and F_{NL} statistics are studied by using the following DGP:

$$y_t = y_{t-1} + \varepsilon_t \text{ with } \varepsilon_t = \rho\varepsilon_{t-1} + u_t, \tag{8}$$

where $\rho = \{0, 0.5\}$ and u_t follows the standard normal distribution. Each statistic was calculated based on a sample size of $T = \{50, 100, 200\}$ in each of 50,000 iterations. The size properties were studied for $d = \{1, 2, 6, 12\}$. When $\rho = 0.5$, the model in (6) was estimated with $p = 1$. The nominal size was set to 5%.[2] The results in Table 1 show that the size for both tests is close to the nominal level for all values of d.

[2]Asymptotical critical values were not supplied in BBC. The critical value was calculated using stochastic simulations with T=1000 and 50,000 replications. The 5% critical value produced was 10.13.

Also following KSS, the small sample power properties are based upon the following ESTAR DGP:

$$y_t = y_{t-1} + \gamma[1 - \exp(-\theta y_{t-d}^2)]y_{t-1} + \varepsilon_t \tag{9}$$

with $\varepsilon_t \sim \mathrm{N}(0,1)$, $\gamma = \{-1.5, -1, -0.5, -0.1\}$, and $\theta = \{0.01, 0.05, 0.1, 1\}$. In particular, to gauge the importance of not restricting $d = 1$, we set $d = \{1, 2, 6, 12\}$ and calculate the power for each test statistic with the correct value of d and with $d = 1$. The power simulations are based upon 50,000 iterations.

The results of the power simulations are given in Tables 2 and 3. As expected, in almost all of the cases presented in the Monte Carlo simulations, when the correct value of d was used, the tests were more powerful. The few exceptions, which are in bold, occurred mainly when the amount of nonlinearity was extreme, $\theta = 1$, and the difference between regimes was small, $\gamma = -0.1$. Simulations appear to indicate that the cause for the power loss in these cases stems from the fact that when there is a large amount of nonlinearity and there is little mean reversion in the outer regime, the Taylor approximation does a relatively poor job at approximating the relevant portions of the transition function transition function.

With less mean reversion the variance of the entire process is larger. This causes the data to be spread out over a larger portion of the transition function. Given that there is already a large degree of nonlinearity, this means that the parabolic Taylor approximation is trying to fit a transition function which has a value of one for a large portion the data and then drops sharply to zero near the mean of the process. This implies that the value of $T(z_t;\theta)$ in (4) is large compared with a situation when the transition function is more parabolic in shape.

Figures 1 and 2 plot two transition functions along with an estimate of the respective distribution of the data produced by the ESTAR model. The data distribution was estimated by simulating data (100,000 observations) using model (9) with a standard normal error term and then using kernel density estimation with a bandwidth that is optimal for estimation normal distributions. Figure 1 illustrates a case with a high degree of nonlinearity and little mean reversion. It is clear that over the portion of the transition function where the data is found, the parabolic Taylor approximation would result in a large amount of error. Figure 2 is parameterized by a low degree of nonlinearity and a greater degree of mean reversion. In this case, over the spread of the data, the Taylor approximation would result in a good fit.

The fact that the error in the transition function approximation enters the error term, as in (5), implies that this issue can be viewed as a kind of omitted variable bias. The potential correlation between the excluded Taylor approximation error, $T(y_{t-d};\theta)$, and the included variable $y_{t-1}y_{t-d}^2$ will tend to bias the estimate of δ_2 in model (6). With $d > 1$ the correlation will tend to be greater when the correct term, $y_{t-1}y_{t-d}^2$, is included compared with the case when the delay parameter is constraint to be unity and y_{t-1}^3 is included in the regression model. This greater correlation, when the correct delay parameter is used, appears to cause the estimate of δ_2 to become less significant which in turn lowers the power of the test in these cases. As a result, in practice, care should be taken if it appears that the Taylor approximation will do a poor job at approximating the true transition function.

It is also important to note that there appears to be only small differences between the power of the t_{NL} and the F_{NL} statistic. In other words, it appears the addition of the extra term in the Taylor expansion adds little to the ability to reject the presence of nonstationarity when the true DGP is ESTAR. Among all of the cases in the Monte Carlo experiments in the paper, when using the true value of d, the average difference in power between t_{NL} and F_{NL} was less than 0.01.

4 Use of the tests when d is unknown

In most practical situations the value of d will not be known and must be estimated. To understand the implications of using an estimated value of the delay parameter while testing for nonlinearity, Monte Carl simulations were constructed to assess the impact on the size and power of the tests. The delay parameter was chosen from $d = (1, ..., 15)$ such that sum of squared errors from the test regression in (6) is minimized. Given that maximizing the fit of the regression is equivalent to maximizing the t or F stat, size distortions should be expected.

Table 4 contains the results from 50,000 simulations following similar procedures as those outlined in the previous section. Both tests are heavily over-sized. The table also contains the critical value that yields the correct size of 5%. These critical values were used to calculate the size adjusted power of the two tests in Tables 2 - 3. For $d = 1, 2$ there is a meaningful decrease in the power of both tests when the estimated value of the delay parameter is used. This result is to be expected given that the test should be more powerful in these cases if the correct delay parameter is used or if d is constrained to be unity. On the other hand for the cases when $d = 6, 12$ the power of the test using the estimated delay parameter is close to the power when the correct delay is employed.

These results suggest that in the ideal case, where d is known, then the correct value should be used so that the power if the test is maximized. In the case where the value of the delay parameter is unknown, there are two choices that encompass virtually all of empirical use of the STAR model. This first choice is to constrain the the value of d to be one and the second is to estimate the delay parameter using some optimization criterion. The results of these simulations suggest that there will be a decrease in power if either the delay parameter is constrained to be one and it is actually large or if the delay parameter is estimated and it is close to one. Given this, the practitioner's best option is to be guided by economic theory as to whether the true value of the delay parameter is large or small. If there is strong evidence that the delay parameter is large then the best option is to estimate the value of the delay parameter and use a size-adjusted test statistic to test for nonlinearity. If it is very unlikely that the delay parameter is larger than one, then the delay parameter should be constrained to be one and the regular test statistic should be used.

5 Conclusion

The paper generalizes two previously developed tests for detecting nonlinear mean reversion in a STAR framework against the presence of nonstationarity by allowing a delay parameter greater than one. It was shown that the asymptotic distribution of the test statistics remains unchanged. Also, it was also shown that when the delay parameter of the data generating process is greater than one, using the test with the correct value of d improves power in almost all cases studied compared to constraining the delay parameter to unity. Practical suggestions for using the test in the case when the delay parameter must be estimated were also given.

References

Balke, N. S. and T. B. Fomby (1997): "Threshold cointegration," *International Economic Review*, 38, 627–645.

Baum, C. F., J. T. Barkoulas, and M. Caglayan (2001): "Nonlinear adjustment to purchasing power parity in the post-bretton woods era," *Journal of International Money and Finance*, 20, 379–399.

Bec, F., M. Ben Salem, and M. Carrasco (2004): "Detecting mean reversion in real exchange rates from a multiple regime star model," RCER Working Papers 509, University of Rochester - Center for Economic Research (RCER).

Eitrheim, O. and T. Teräsvirta (1996): "Testing the adequacy of smooth transition autoregressive models," *Journal of Econometrics*, 74, 59–75.

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

Hansen, B. (1997): "Inference in TAR models," *Studies in Nonlinear dynamics and Econometrics*, 2, 1–14.

Kapetanios, G., Y. Shin, and A. Snell (2003): "Testing for a unit root in the nonlinear star famework," *Jornal of Econometrics*, 112, 359–370.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

Michael, P., R. A. Nobay, and D. A. Peel (1997): "Transaction costs and nonlinear adjustment in real exchange rates: An empirical investigation," *Journal of Political Economy*, 105, 862–879.

Park, J. Y. and P. C. B. Phillips (2001): "Nonlinear regressions with integrated time series," *Econometrica*, 69, 117.

Saikkonen, P. and R. Luukkonen (1988): "Lagrange multiplier tests for testing non-linearities in time series models," *Scandinavian Journal of Statistics*, 15, 55–68.

Sarantis, N. (1999): "Modeling non-linearities in real effective exchange rates," *Journal of International Money and Finance*, 18, 27–45.

Sarno, L., M. P. Taylor, and I. Chowdhury (2004): "Nonlinear dynamics in deviations from the law of one price: a broad-based empirical study," *Journal of International Money and Finance*, 23, 1.

Taylor, M., D. Peel, and L. Sarno (2001): "Nonlinear mean-reversion in real exchange rates: Towards a solution to the purchasing power parity puzzles," *International Economic Review*, 42, 1015–42.

Taylor, M. P. and D. A. Peel (2000): "Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals," *Journal of International Money and Finance*, 19, 33–53.

Taylor, N., D. van Dijk, P. H. Franses, and A. Lucas (2000): "Sets, arbitrage activity, and stock price dynamics," *Journal of Banking and Finance*, 24, 1289–1306.

Teräsvirta, T. (1994): "Specification, estimation and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association*, 89, 208–218.

Teräsvirta, T. (1998): "Modelling economic relationships with smooth transition regressions," in D. E. A. Giles and A. Ullah, eds., *Handbook of Applied Economic Statistics*, New York: Dekker.

van Dijk, D., T. Teräsvirta, and P. H. Franses (2002): "Smooth transition autoregressive models - a survey of recent developments," *Econometric Reviews*, 21, 1–47.

White, H. (2000): *Asymptotic Theory for Econometricians*, New York: Academic Press.

A Appendix: Proofs

These derivations and notation closely follow those found in KSS and BBC.

A.1 Distribution of $t_{\mathbf{NL}}$ for $d = \{2, 3, 4, \ldots\}$

The relevant statistic for $r_1 = r_2 = 2$ in (6), as in KSS, is the t-statistic for $\delta_2 = 0$ against $\delta_2 < 0$,

$$t_{\text{NL}} = \hat{\delta}_2 / \text{s.e.}(\hat{\delta}_2). \tag{10}$$

A.1.1 Case 1: $p = 0$

Under the null, the process follows a unit root, so $\Delta y_t = \varepsilon_t$. Thus

$$t_{\text{NL}} = \frac{\sum_{t=1}^{T} y_{t-1} y_{t-d}^2 \varepsilon_t}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^{T} (y_{t-1} y_{t-d}^2)^2}} \tag{11}$$

Substituting $y_{t-1} = y_{t-d} + \sum_{i=1}^{d-1} \varepsilon_{t-i}$ for $d > 1$ yields

$$\begin{aligned}
t_{\text{NL}} &= \frac{\sum_{t=1}^{T} (y_{t-d} + \sum_{i=1}^{d-1} \varepsilon_{t-i}) y_{t-d}^2 \varepsilon_t}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^{T} ((y_{t-d} + \sum_{i=1}^{d-1} \varepsilon_{t-i}) y_{t-d}^2)^2}} && (12)\\
&= \frac{\sum_{t=1}^{T} (y_{t-d}^3 \varepsilon_t + \sum_{i=1}^{d-1} \varepsilon_{t-i} \varepsilon_t y_{t-d}^2)}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^{T} (y_{t-d}^3 + \sum_{i=1}^{d-1} \varepsilon_{t-i} y_{t-d}^2)^2}} && (13)\\
&= \frac{\sum_{t=1}^{T} (y_{t-d}^3 \varepsilon_t + \sum_{i=1}^{d-1} \varepsilon_{t-i} \varepsilon_t y_{t-d}^2)}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^{T} (y_{t-d}^6 + 2 \sum_{i=1}^{d-1} \varepsilon_{t-i} y_{t-d}^5 + y_{t-d}^4 (\sum_{i=1}^{d-1} \varepsilon_{t-i})^2)}} && (14)
\end{aligned}$$

Focusing on the simplest case, set $d = 2$

$$t_{\text{NL}} = \frac{\sum_{t=1}^{T} (\varepsilon_t y_{t-2}^3 + \varepsilon_{t-1} \varepsilon_t y_{t-2}^2)}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^{T} (y_{t-2}^6 + 2\varepsilon_{t-1} y_{t-2}^5 + \varepsilon_{t-1}^2 y_{t-2}^4)}} \tag{15}$$

It is straightforward to show that $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$. Noting that the asymptotics of $\sum_{t=1}^{T} \varepsilon_t y_{t-d}^k$, for some positive integer k, will be the same as $\sum_{t=1}^{T} \varepsilon_t y_{t-1}^k$ by the i.i.d. nature of ε_t, Theorem 3.1 of Park and Phillips (2001) can be applied to get the following results:

$$T^{-2} \sum_{t=1}^{T} \varepsilon_t y_{t-2}^3 \xrightarrow{D} \sigma^4 \int_0^1 W(r)^3 \mathrm{d}W(r) \tag{16}$$

$$T^{-4} \sum_{t=1}^{T} y_{t-2}^6 \xrightarrow{D} \sigma^6 \int_0^1 W(r)^6 \mathrm{d}r \tag{17}$$

$$T^{-3} \sum_{t=1}^{T} \varepsilon_{t-1} y_{t-2}^5 \xrightarrow{D} \sigma^6 \int_0^1 W(r)^5 \mathrm{d}W(r) \tag{18}$$

The remaining terms to consider are $\sum_{t=1}^{T} \varepsilon_{t-1}\varepsilon_t y_{t-2}^2$ and $\sum_{t=1}^{T} \varepsilon_{t-1}^2 y_{t-2}^4$. First note that $u_t = \varepsilon_{t-1}\varepsilon_t$ is a martingale difference sequence and is uncorrelated with y_{t-2}. Again using Theorem 3.1 of Park and Phillips (2001), we get the following:

$$T^{-3/2} \sum_{t=1}^{T} u_t y_{t-2}^2 \xrightarrow{D} \sigma^4 \int_0^1 W(r)^2 \mathrm{d}U(r) \tag{19}$$

The remaining term can be expressed as

$$\sum_{t=1}^{T} \varepsilon_{t-1}^2 y_{t-2}^4 = \sum_{t=1}^{T} (\varepsilon_{t-1}^2 - \sigma^2) y_{t-2}^4 + \sum_{t=1}^{T} \sigma^2 y_{t-2}^4 \tag{20}$$

Note that $v_t = \varepsilon_{t-1}^2 - \sigma^2$ is a martingale difference sequence. Thus using the same theorem we get the following:

$$T^{-5/2} \sum_{t=1}^{T} v_t y_{t-2}^4 \xrightarrow{D} \sigma^6 \int_0^1 W(r)^4 \mathrm{d}V(r) \tag{21}$$

$$T^{-3} \sum_{t=1}^{T} \sigma^2 y_{t-2}^4 \xrightarrow{D} \sigma^6 \int_0^1 W(r)^4 \mathrm{d}r \tag{22}$$

Thus with $d = 2$

$$t_{\mathrm{NL}} \xrightarrow{D} \frac{\int_0^1 W(r)^3 \mathrm{d}W(r)}{\sqrt{\int_0^1 W(r)^6 \mathrm{d}r}} \tag{23}$$

For $d > 2$ the preceding arguments are sufficient to show that

$$t_{\mathrm{NL}} \xrightarrow{D} \frac{\int_0^1 W(r)^3 \mathrm{d}W(r)}{\sqrt{\int_0^1 W(r)^6 \mathrm{d}r}} \tag{24}$$

This is because there are only two terms in (6) who asymptotics will change with $d > 2$. They are

$$\sum_{t=1}^{T} \sum_{i=1}^{d-1} \varepsilon_{t-i} \varepsilon_t y_{t-d}^2 \tag{25}$$

in the numerator, and

$$\sum_{t=1}^{T} y_{t-d}^2 (\sum_{i=1}^{d-1} \varepsilon_{t-i})^2 \tag{26}$$

in the denominator.

Each term in (25) and the cross product terms in (26), $\sum_{t=1}^{T} y_{t-d}^2 (\sum_{i=1}^{d-1} \varepsilon_{t-i} (\sum_{j=1, j \neq i}^{d-1} \varepsilon_{t-j}))$, behave the same as (19). The rest of the $(d-1)$ terms in (26), $\sum_{t=1}^{T} y_{t-d}^2 \sum_{i=1}^{d-1} \varepsilon_{t-i}^2$, will behave the same as (20). Note that the distribution of the t_{NL} statistic does not depend on σ or any other nuisance parameters and has the same distribution for d=1. ■

A.1.2 Case 2: $p > 0$

Follows directly from the Proof of Theorem 2.2 of KSS.

A.2 Distribution of F_{NL} for $d = \{2, 3, 4, \ldots\}$

For $r_1 = 1, r_2 = 2$ in (6), as in BBC, the relevant test statistic is the Wald test statistic for $\delta_1 = \delta_2 = 0$ against $\delta_1 \neq 0$ or $\delta_2 \neq 0$. Adopting the same reparametrization as BBC and KSS, let $\mathbf{y}_d^j = (y_0 y_{-d}^j, \ldots, y_{T-1} y_{T-d}^j)'$, $\Delta \mathbf{y}_{-j} = (\Delta y_{1-j}, \ldots, \Delta y_{T-j})'$, $\mathbf{Z} = (\Delta \mathbf{y}_{-1}, \ldots, \Delta \mathbf{y}_{-1})$. Define the $T \times T$ idempotent matrix $\mathbf{Z}_T = \mathbf{I}_T$ - $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$. Model (6) can be rewritten as

$$\Delta \mathbf{y} = \sum_{j-1}^{p-1} \alpha_j \Delta \mathbf{y}_{-j} + \delta_1 \mathbf{y}_d^1 + \delta_2 \mathbf{y}_d^2 + \varepsilon,$$

which can also be expressed as

$$\mathbf{M}_T\Delta\mathbf{y} = \delta_1\mathbf{M}_T\mathbf{y}_d^1 + \delta_2\mathbf{M}_T\mathbf{y}_d^2 + \mathbf{M}_T\varepsilon.$$

Let $\hat{\theta} = (\hat{\delta}_{r_1}, \hat{\delta}_{r_2})'$, $X = [\mathbf{y}_d^1\ \mathbf{y}_d^2]$, and

$$\Gamma_T = \begin{bmatrix} T^{3/2} & 0 \\ 0 & T^2 \end{bmatrix}.$$

The Wald test statistic we are concerned with is

$$F_{\text{NL}} = \hat{\theta}' \left[\hat{\sigma^2}(X'\mathbf{M}_T X)^{-1}\right]^{-1} \hat{\theta}. \tag{27}$$

To establish the limiting distribution, note the following:

$$\Gamma_T\hat{\theta} = \Gamma_T^{-1}\theta + (\Gamma_T^{-1}(X'\mathbf{M}_T X)\Gamma_T^{-1})^{-1}\Gamma_T^{-1}X'\mathbf{M}_T\varepsilon,$$

$$\begin{aligned} \Gamma_T^{-1}X'\mathbf{M}_T\varepsilon &= \begin{pmatrix} \frac{\mathbf{y}_d^{1'}\mathbf{M}_T\varepsilon}{T^{-3/2}} \\ \frac{\mathbf{y}_d^{2'}\mathbf{M}_T\varepsilon}{T^{-2}} \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{y}_d^{1'}\varepsilon}{T^{-3/2}} \\ \frac{\mathbf{y}_d^{2'}\varepsilon}{T^{-2}} \end{pmatrix} + o_p(1) = \begin{pmatrix} \frac{\sum_{t=1}^T y_{t-1}y_{t-d}\varepsilon_t}{T^{3/2}} \\ \frac{\sum_{t=1}^T y_{t-1}y_{t-d}^2\varepsilon_t}{T^2} \end{pmatrix} + o_p(1) \\ &\xrightarrow{D} \sigma \begin{pmatrix} \eta^2\int_0^1 W(r)^2\mathrm{d}W(r) \\ \eta^3\int_0^1 W(r)^3\mathrm{d}W(r) \end{pmatrix} \equiv \sigma \begin{bmatrix} \eta^2 & 0 \\ 0 & \eta^3 \end{bmatrix} h, \end{aligned} \tag{28}$$

where $\eta = \sigma/(1-\alpha_1-\ldots-\alpha_p)$. The limiting distribution for the sums in (28) can be determined using the same arguments to derive the distributions for the numerator in (11).

Also,

$$\begin{aligned} \Gamma_T^{-1}(X'\mathbf{M}_T X)^{-1}\Gamma_T^{-1} &= \begin{pmatrix} \frac{\mathbf{y}_d^{1'}\mathbf{M}_T\mathbf{y}_d^{1'}}{T^{3/2}} & \frac{\mathbf{y}_d^{1'}\mathbf{M}_T\mathbf{y}_d^{2'}}{T^{7/2}} \\ \frac{\mathbf{y}_d^{2'}\mathbf{M}_T\mathbf{y}_d^{1'}}{T^{7/2}} & \frac{\mathbf{y}_d^{2'}\mathbf{M}_T\mathbf{y}_d^{2'}}{T^4} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\mathbf{y}_d^{1'}\mathbf{y}_d^{1'}}{T^{3/2}} & \frac{\mathbf{y}_d^{1'}\mathbf{y}_d^{2'}}{T^{7/2}} \\ \frac{\mathbf{y}_d^{2'}\mathbf{y}_d^{1'}}{T^{7/2}} & \frac{\mathbf{y}_d^{2'}\mathbf{y}_d^{2'}}{T^4} \end{pmatrix} + o_p(1) \\ &= \begin{pmatrix} \frac{\sum_{t=1}^T y_{t-1}^2y_{t-d}^2}{T^3} & \frac{\sum_{t=1}^T y_{t-1}^2y_{t-d}^3}{T^{7/2}} \\ \frac{\sum_{t=1}^T y_{t-1}^2y_{t-d}^3}{T^{7/2}} & \frac{\sum_{t=1}^T y_{t-1}^2y_{t-d}^4}{T^4} \end{pmatrix} + o_p(1) \\ &\xrightarrow{D} \begin{pmatrix} \eta^4\int_0^1 W(r)^4\mathrm{d}r & \eta^5\int_0^1 W(r)^5\mathrm{d}r \\ \eta^5\int_0^1 W(r)^5\mathrm{d}r & \eta^6\int_0^1 W(r)^6\mathrm{d}r \end{pmatrix} \\ &\equiv \begin{bmatrix} \eta^2 & 0 \\ 0 & \eta^3 \end{bmatrix} Q \begin{bmatrix} \eta^2 & 0 \\ 0 & \eta^3 \end{bmatrix} \end{aligned} \tag{29}$$

The limiting distribution for the sums in (29) can be determined using the same arguments to derive the distributions for the denominator in (11).

Thus the Wald test statistic in (27) is

$$\begin{aligned} F_{\mathrm{NL}} &= \hat{\theta}' \left[\hat{\sigma^2} (X' \mathbf{M}_T X)^{-1} \right]^{-1} \hat{\theta} \\ &= (\Gamma_T^{-1} \hat{\theta})' \left[\Gamma_T^{-1} (X' \mathbf{M}_T X)^{-1} \Gamma_T^{-1} \right]^{-1} \Gamma_T^{-1} \hat{\theta} / \hat{\sigma^2} \\ &\xrightarrow{D} (Q^{-1} h)' Q^{-1} (Q^{-1} h) \\ &= h' Q^{-1} h \end{aligned}$$

Note that the limiting distribution does not depend on σ or any other nuisance parameters and is the same as the distribution for d=1. ■

Table 1: Size properties of t_{NL} and F_{NL}

	$d=1$		$d=2$		$d=6$		$d=12$	
	t_{NL}	F_{NL}	t_{NL}	F_{NL}	t_{NL}	F_{NL}	t_{NL}	F_{NL}
$\rho=0$								
$T=50$	0.050	0.045	0.048	0.048	0.055	0.058	0.050	0.064
$T=100$	0.050	0.048	0.048	0.045	0.047	0.051	0.051	0.053
$T=200$	0.050	0.049	0.047	0.045	0.043	0.046	0.045	0.048
$\rho=0.5$								
$T=50$	0.061	0.063	0.051	0.049	0.059	0.060	0.053	0.070
$T=100$	0.055	0.055	0.047	0.045	0.048	0.050	0.053	0.056
$T=200$	0.054	0.052	0.048	0.047	0.042	0.044	0.047	0.049

Table 2: Power properties of t_{NL}

$\theta =$	0.01			0.05			0.10			1.00		
$\bar{d} =$	d	1	$\hat{d}$	d	1	$\hat{d}$	d	1	$\hat{d}$	d	1	$\hat{d}$
$d = 1$												
$\gamma = -1.5$												
$T = 50$	0.256	-	0.129	0.825	-	0.385	0.967	-	0.611	1.000	-	0.965
$T = 100$	0.690	-	0.263	0.998	-	0.820	1.000	-	0.966	1.000	-	1.000
$T = 200$	0.992	-	0.739	1.000	-	0.999	1.000	-	1.000	1.000	-	1.000
$\gamma = -1.0$												
$T = 50$	0.178	-	0.108	0.622	-	0.257	0.855	-	0.407	0.996	-	0.853
$T = 100$	0.488	-	0.194	0.980	-	0.620	0.999	-	0.852	1.000	-	0.998
$T = 200$	0.955	-	0.570	1.000	-	0.987	1.000	-	0.999	1.000	-	1.000
$\gamma = -0.5$												
$T = 50$	0.114	-	0.089	0.305	-	0.147	0.473	-	0.205	0.798	-	0.465
$T = 100$	0.242	-	0.130	0.778	-	0.321	0.934	-	0.490	0.992	-	0.872
$T = 200$	0.725	-	0.342	0.997	-	0.831	1.000	-	0.957	1.000	-	0.999
$\gamma = -0.1$												
$T = 50$	0.068	-	0.065	0.084	-	0.079	0.090	-	0.085	0.115	-	0.101
$T = 100$	0.089	-	0.080	0.148	-	0.107	0.189	-	0.128	0.257	-	0.179
$T = 200$	0.159	-	0.135	0.428	-	0.256	0.549	-	0.336	0.609	-	0.461
$d = 2$												
$\gamma = -1.5$												
$T = 50$	0.290	0.131	0.125	0.775	0.492	0.349	0.911	0.757	0.525	0.993	0.993	0.899
$T = 100$	0.685	0.466	0.286	0.991	0.980	0.792	0.999	0.999	0.936	1.000	1.000	0.998
$T = 200$	0.987	0.977	0.765	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000
$\gamma = -1.0$												
$T = 50$	0.206	0.105	0.104	0.598	0.304	0.236	0.780	0.522	0.352	0.948	0.945	0.717
$T = 100$	0.504	0.301	0.204	0.955	0.899	0.606	0.993	0.986	0.802	**0.999**	**1.000**	0.983
$T = 200$	0.944	0.903	0.603	1.000	1.000	0.981	1.000	1.000	0.998	1.000	1.000	1.000
$\gamma = -0.5$												
$T = 50$	0.124	0.080	0.080	0.313	0.150	0.137	0.450	0.234	0.180	0.629	0.599	0.355
$T = 100$	0.263	0.162	0.135	0.743	0.565	0.321	0.886	0.791	0.458	**0.958**	**0.975**	0.766
$T = 200$	0.714	0.600	0.373	0.994	0.989	0.819	0.999	0.999	0.938	1.000	1.000	0.996
$\gamma = -0.1$												
$T = 50$	0.066	0.062	0.063	0.086	0.072	0.072	0.092	0.080	0.079	0.104	0.103	0.092
$T = 100$	0.087	0.078	0.081	0.145	0.120	0.107	0.181	0.150	0.121	**0.218**	**0.236**	0.167
$T = 200$	0.163	0.140	0.142	0.413	0.355	0.262	0.509	0.474	0.331	**0.552**	**0.606**	0.447

$\theta =$	0.01			0.05			0.10			1.00		
$\bar{d} =$	d	1	$\hat{d}$	d	1	$\hat{d}$	d	1	$\hat{d}$	d	1	$\hat{d}$
$d = 6$												
$\gamma = -1.5$												
$T = 50$	0.253	0.082	0.129	0.538	0.189	0.312	0.696	0.317	0.447	0.962	0.885	0.849
$T = 100$	0.616	0.142	0.384	0.934	0.410	0.755	0.980	0.613	0.876	0.999	0.982	0.994
$T = 200$	0.959	0.482	0.841	1.000	0.874	0.991	1.000	0.951	0.998	1.000	0.999	1.000
$\gamma = -1.0$												
$T = 50$	0.197	0.074	0.104	0.416	0.133	0.220	0.532	0.215	0.305	0.824	0.725	0.636
$T = 100$	0.491	0.116	0.291	0.851	0.296	0.621	0.930	0.465	0.745	0.988	0.940	0.950
$T = 200$	0.899	0.375	0.738	0.998	0.780	0.967	1.000	0.894	0.991	1.000	0.996	1.000
$\gamma = -0.5$												
$T = 50$	0.136	0.066	0.081	0.253	0.094	0.133	0.312	0.124	0.163	0.423	0.389	0.303
$T = 100$	0.297	0.090	0.173	0.620	0.177	0.389	0.730	0.267	0.479	0.828	0.731	0.673
$T = 200$	0.687	0.233	0.512	0.965	0.572	0.852	0.989	0.725	0.915	0.992	0.966	0.975
$\gamma = -0.1$												
$T = 50$	0.072	0.062	0.059	0.087	0.065	0.066	0.088	0.073	0.071	**0.089**	**0.098**	0.085
$T = 100$	0.094	0.066	0.083	0.144	0.089	0.109	0.160	0.102	0.124	**0.159**	**0.182**	0.155
$T = 200$	0.180	0.099	0.159	0.361	0.184	0.295	0.416	0.260	0.345	**0.396**	**0.482**	0.426
$d = 12$												
$\gamma = -1.5$												
$T = 50$	0.133	0.091	0.111	0.413	0.222	0.328	0.614	0.350	0.517	0.953	0.878	0.924
$T = 100$	0.450	0.128	0.380	0.855	0.357	0.800	0.954	0.553	0.931	0.999	0.979	0.998
$T = 200$	0.894	0.254	0.883	0.997	0.613	0.997	1.000	0.811	1.000	1.000	0.998	1.000
$\gamma = -1.0$												
$T = 50$	0.102	0.082	0.089	0.274	0.164	0.210	0.417	0.249	0.329	0.811	0.734	0.746
$T = 100$	0.351	0.107	0.288	0.707	0.268	0.640	0.847	0.414	0.802	0.986	0.934	0.980
$T = 200$	0.806	0.198	0.790	0.982	0.491	0.983	0.996	0.688	0.996	1.000	0.991	1.000
$\gamma = -0.5$												
$T = 50$	0.073	0.070	0.071	0.140	0.106	0.119	0.196	0.147	0.157	**0.408**	**0.423**	0.355
$T = 100$	0.234	0.083	0.188	0.449	0.160	0.379	0.558	0.247	0.488	0.787	0.725	0.762
$T = 200$	0.610	0.138	0.583	0.893	0.327	0.892	0.946	0.470	0.947	0.984	0.939	0.989
$\gamma = -0.1$												
$T = 50$	**0.056**	**0.062**	0.056	**0.061**	**0.071**	0.064	**0.064**	**0.078**	0.068	**0.071**	**0.115**	0.084
$T = 100$	0.093	0.064	0.079	0.125	0.082	0.110	0.133	0.099	0.120	**0.131**	**0.182**	0.152
$T = 200$	0.186	0.084	0.181	0.319	0.134	0.310	0.344	0.177	0.356	**0.313**	**0.407**	0.425

NOTES: $\bar{d}$ is the value of the delay parameter used to calculated t_{NL}, and d is the true value of the delay parameter used in the DGP. The cases where the test is more powerful when $\bar{d} = 1$ verses $\bar{d} = d$ are in bold. $\hat{d}$ refers to the case where the delay parameter is chosen from $d = 1, ..., 15$ through the estimation process.

Table 3: Power properties of F_{NL}

$\theta =$	0.01			0.05			0.10			1.00		
$\bar{d} =$	d	1	$\hat{d}$	d	1	$\hat{d}$	d	1	$\hat{d}$	d	1	$\hat{d}$
$d = 1$												
$\gamma = -1.5$												
$T = 50$	0.250	-	0.378	0.821	-	0.684	0.966	-	0.843	1.000	-	0.992
$T = 100$	0.715	-	0.511	0.999	-	0.934	1.000	-	0.991	1.000	-	1.000
$T = 200$	0.998	-	0.863	1.000	-	1.000	1.000	-	1.000	1.000	-	1.000
$\gamma = -1.0$												
$T = 50$	0.166	-	0.342	0.614	-	0.558	0.849	-	0.710	0.996	-	0.963
$T = 100$	0.501	-	0.423	0.984	-	0.827	0.999	-	0.949	1.000	-	1.000
$T = 200$	0.979	-	0.737	1.000	-	0.995	1.000	-	1.000	1.000	-	1.000
$\gamma = -0.5$												
$T = 50$	0.102	-	0.295	0.288	-	0.410	0.452	-	0.500	0.759	-	0.770
$T = 100$	0.236	-	0.328	0.796	-	0.579	0.941	-	0.736	0.993	-	0.963
$T = 200$	0.767	-	0.526	0.999	-	0.916	1.000	-	0.982	1.000	-	1.000
$\gamma = -0.1$												
$T = 50$	0.059	-	0.252	0.071	-	0.280	0.080	-	0.296	0.098	-	0.333
$T = 100$	0.081	-	0.236	0.139	-	0.296	0.174	-	0.332	0.234	-	0.409
$T = 200$	0.155	-	0.268	0.423	-	0.420	0.541	-	0.509	0.600	-	0.636
$d = 2$												
$\gamma = -1.5$												
$T = 50$	0.296	0.110	0.370	0.790	0.422	0.642	0.919	0.694	0.782	0.993	0.991	0.973
$T = 100$	0.712	0.420	0.524	0.993	0.978	0.913	0.999	0.999	0.979	1.000	1.000	1.000
$T = 200$	0.993	0.986	0.873	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
$\gamma = -1.0$												
$T = 50$	0.205	0.088	0.335	0.610	0.256	0.528	0.791	0.459	0.653	0.949	0.927	0.909
$T = 100$	0.528	0.273	0.438	0.964	0.889	0.800	0.994	0.984	0.923	1.000	1.000	0.997
$T = 200$	0.963	0.918	0.758	1.000	1.000	0.993	1.000	1.000	1.000	1.000	1.000	1.000
$\gamma = -0.5$												
$T = 50$	0.120	0.071	0.293	0.319	0.127	0.390	0.451	0.192	0.460	0.613	0.540	0.675
$T = 100$	0.270	0.145	0.338	0.760	0.521	0.570	0.894	0.757	0.697	**0.953**	**0.972**	0.918
$T = 200$	0.754	0.597	0.555	0.997	0.994	0.909	1.000	1.000	0.973	1.000	1.000	0.999
$\gamma = -0.1$												
$T = 50$	0.065	0.056	0.248	0.080	0.062	0.272	0.089	0.069	0.282	0.099	0.089	0.314
$T = 100$	0.082	0.071	0.240	0.141	0.106	0.293	0.174	0.133	0.317	**0.205**	**0.210**	0.392
$T = 200$	0.159	0.127	0.275	0.408	0.332	0.435	0.510	0.456	0.505	**0.540**	**0.594**	0.619

$\theta =$	0.01			0.05			0.10			1.00		
$\bar{d} =$	d	1	$\hat{d}$	d	1	$\hat{d}$	d	1	$\hat{d}$	d	1	$\hat{d}$
$d = 6$												
$\gamma = -1.5$												
$T = 50$	0.264	0.084	0.356	0.552	0.194	0.561	0.716	0.328	0.686	0.964	0.904	0.956
$T = 100$	0.638	0.132	0.601	0.941	0.396	0.872	0.983	0.610	0.947	0.999	0.991	0.998
$T = 200$	0.972	0.431	0.916	1.000	0.858	0.997	1.000	0.954	1.000	1.000	1.000	1.000
$\gamma = -1.0$												
$T = 50$	0.196	0.078	0.320	0.425	0.137	0.475	0.543	0.215	0.571	0.832	0.737	0.862
$T = 100$	0.515	0.106	0.518	0.863	0.280	0.779	0.937	0.446	0.869	0.988	0.956	0.986
$T = 200$	0.924	0.330	0.847	0.999	0.746	0.987	1.000	0.888	0.997	1.000	0.999	1.000
$\gamma = -0.5$												
$T = 50$	0.133	0.070	0.282	0.258	0.095	0.367	0.319	0.125	0.416	0.417	0.374	0.613
$T = 100$	0.311	0.087	0.393	0.646	0.165	0.603	0.745	0.245	0.683	0.813	0.736	0.856
$T = 200$	0.721	0.199	0.683	0.975	0.526	0.917	0.991	0.693	0.957	0.992	0.976	0.990
$\gamma = -0.1$												
$T = 50$	0.071	0.068	0.238	0.082	0.073	0.260	0.088	0.078	0.266	**0.087**	**0.097**	0.297
$T = 100$	0.094	0.067	0.247	0.148	0.081	0.301	0.166	0.098	0.322	**0.158**	**0.170**	0.368
$T = 200$	0.187	0.090	0.312	0.375	0.169	0.473	0.421	0.232	0.526	**0.379**	**0.451**	0.595
$d = 12$												
$\gamma = -1.5$												
$T = 50$	0.159	0.120	0.344	0.434	0.268	0.611	0.640	0.406	0.767	0.959	0.907	0.979
$T = 100$	0.464	0.143	0.633	0.858	0.404	0.921	0.957	0.604	0.977	0.999	0.991	1.000
$T = 200$	0.906	0.253	0.951	0.998	0.648	0.999	1.000	0.853	1.000	1.000	1.000	1.000
$\gamma = -1.0$												
$T = 50$	0.126	0.111	0.307	0.296	0.196	0.492	0.439	0.287	0.622	0.826	0.761	0.917
$T = 100$	0.370	0.115	0.547	0.715	0.294	0.830	0.852	0.451	0.920	0.986	0.957	0.995
$T = 200$	0.828	0.200	0.905	0.986	0.508	0.993	0.996	0.721	0.999	1.000	0.998	1.000
$\gamma = -0.5$												
$T = 50$	**0.097**	**0.104**	0.270	0.168	0.136	0.362	0.223	0.174	0.431	**0.428**	**0.436**	0.674
$T = 100$	0.240	0.092	0.428	0.457	0.174	0.646	0.563	0.263	0.743	0.782	0.745	0.916
$T = 200$	0.639	0.138	0.762	0.902	0.333	0.955	0.947	0.482	0.977	0.984	0.959	0.996
$\gamma = -0.1$												
$T = 50$	**0.070**	**0.105**	0.231	**0.076**	**0.108**	0.254	**0.081**	**0.111**	0.262	**0.084**	**0.141**	0.301
$T = 100$	0.092	0.075	0.243	0.128	0.092	0.307	0.138	0.109	0.330	**0.128**	**0.185**	0.382
$T = 200$	0.195	0.084	0.343	0.326	0.131	0.514	0.352	0.173	0.557	**0.300**	**0.395**	0.608

NOTES: $\bar{d}$ is the value of the delay parameter used to calculated F_{NL}, and d is the true value of the delay parameter used in the DGP. The cases where the test is more powerful when $\bar{d} = 1$ verses $\bar{d} = d$ are in bold. $\hat{d}$ refers to the case where the delay parameter is chosen from $d = 1, \ldots, 15$ through the estimation process.

Table 4: Size properties of t_{NL} and F_{NL} for $\hat{d} = 1, ..., 15$

	t_{NL}		F_{NL}	
	Size	$CV_{0.05}$	Size	$CV_{0.05}$
$\rho = 0$				
$T = 50$	0.164	-3.64	0.143	16.91
$T = 100$	0.152	-3.50	0.116	14.52
$T = 200$	0.105	-3.28	0.096	13.18
$\rho = 0.5$				
$T = 50$	0.142	-3.56	0.126	16.15
$T = 100$	0.121	-3.38	0.094	13.77
$T = 200$	0.085	-3.18	0.076	12.35

NOTES: $CV_{0.05}$ refers to the critical value of the test which yields no size distortions.

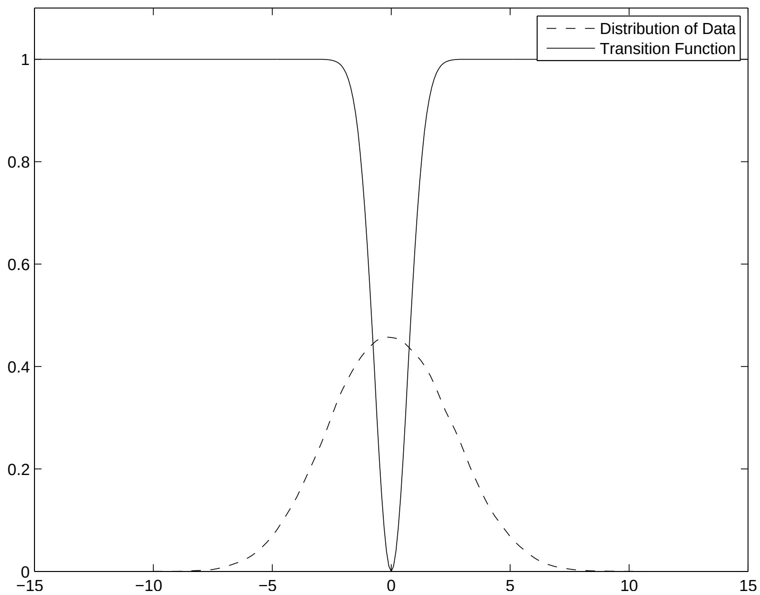

Figure 1: Transition Function and Data Distribution for $\theta = 1, \gamma = -0.1$

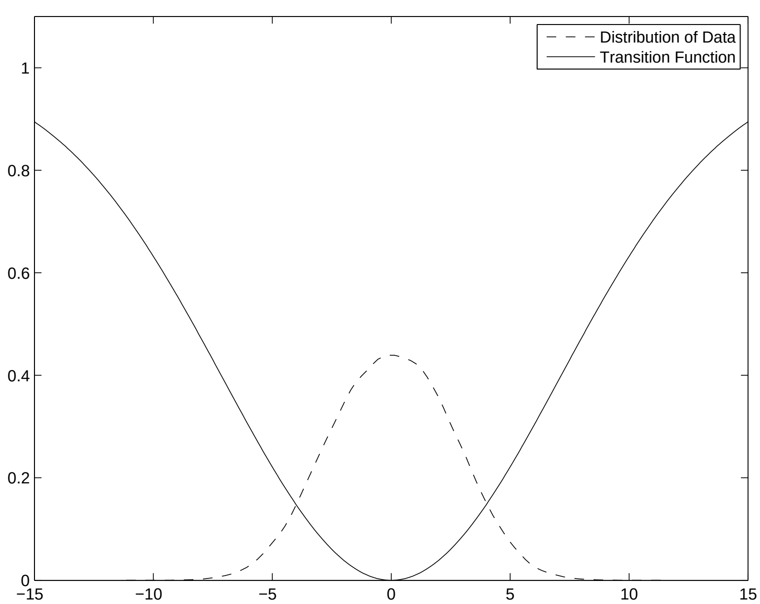

Figure 2: Transition Function and Data Distribution for $\theta = 0.01, \gamma = -1$