Unit roots and smooth transitions: either, neither or both?*

Robinson Kruse and Philipp Sibbertsen

CREATES, University of Aarhus and Leibniz University Hannover

October 2008

Abstract

Without doubt unit roots and related tests play an important role in empirical economics and it is widely acknowledged that non-linear features like structural breaks or regime switches are present in time series data. Furthermore, there are numerous examples of time series which are under $I(0)/I(1)$ consideration and suspected to exhibit non-linearities at the same time. Therefore, we consider the model selection problem of discriminating between linear $I(0)$ and linear $I(1)$ as well as non-linear $I(0)$ and non-linear $I(1)$ models. However, as a specific type of nonlinearity we constrain ourselves to STAR processes which are quite frequently used in economic applications. We propose two-step testing-based model selection rules and compare their small-sample performance by means of a large-scale Monte Carlo study. The one-step procedure by Harvey and Leybourne (2007) serves as the benchmark. As none of the considered rules appears to be dominating, we suggest a unified two-step rule which is based on robust pre-testing for linearity. The unified procedure outperforms the benchmark especially for non-linear models. An empirical application to US interest rates underlines its usefulness in practice.

JEL-numbers: C22, C52

Keywords: Unit roots, nonlinearity, smooth transitions, model selection

*We would like to thank the participants of the CREATES research seminar for their helpful comments.

1 Introduction

Since the seminal works of Dickey and Fuller (1979) and Nelson and Plosser (1982) the question whether a time series contains a unit root and is therefore integrated of order 1, that is $I(1)$, or whether it is a stationary process, that is $I(0)$, attracted much attention in econometric research. As an $I(0)$ process can be interpreted as a process fluctuating around a stable equilibrium the question whether a given time series is $I(0)$ or $I(1)$ is equivalent to confirm or reject miscellaneous economic theories.

The literature on testing for a unit root against stationarity and vice versa was for a long time concentrated on linear processes. However, in the last decade it became more and more evident that many economic time series share properties of non-linear processes. Especially, structural breaks and regime switching are the most often considered types of non-linearity. Whereas early papers concentrated on the question whether a stationary process is linear or non-linear (see for example Luukkonen et al. (1988), Lee et al. (1993) or Brock et al. (1996)) the focus of econometric research recently is on testing for a unit root against stationary non-linear alternatives (see for example Kapetanios et al. (2003)). Choi and Moh (2007) provide an extensive simulation study on the usefulness of sevearl unit root tests against non-linear alternatives.

Moreover, there are numerous examples of time series which are under $I(0)/I(1)$ consideration and suspected to exhibit nonlinearities at the same time. Therefore, it is important to distinguish between linear and non-linear models in addition to stationary and non-stationary models. We consider in this work the model selection problem of discriminating between linear $I(0)$ and linear $I(1)$ as well as non-linear $I(0)$ and nonlinear $I(1)$ models. Each of these models has a different economic interpretation and other implications in terms of forecasting, economic modeling and analysis of impulse-response functions. The aim of this paper is to suggest testing-based model selection rules which allow a reliable classification of time series into one of these four model classes.

We study two different types of model selection rules. As a measure of performance we use the number of correct selections instead of size and power as we do not consider a test but testing-based model selection rules. First, we discuss a simultaneous procedure based on a linearity and a stationarity test which are independently computed. This procedure was proposed by Harvey and Leybourne (2007). They use a linearity test which is robust against non-stationarity of the time series. As this robustification induces significant power losses in small samples, we propose an alternative procedure which can be seen as a two-step or sequential procedure. In a first step, a unit root or a stationarity test is applied and the test result is used to choose the appropriate test regression for linearity. If the test suggests that the time series is $I(0)$, then the Wald-type linearity test is based on a test regression in levels and otherwise in first differences. As a stationarity test we use the test of Harris et al. (2003) or alternatively the non-linear unit root test of Kapetanios et al. (2003). However, it turns out that both procedures have better success rates if the Harris et al. (2003) test is used. More importantly, two-step procedures outperform simultaneous model selection rules in general, and especially the one proposed by Harvey and Leybourne (2007).

It should also be mentioned that we focus on STAR non-linearity which is one of the most popular non-linear models. As there are innumerable non-linear models around, it is impossible to create a unified procedure for all of them but our procedure can easily be generalized to Threshold AR or Markov Switching AR non-linearities. In addition, the linearity tests which are used here are based upon a Taylor approximation and the resulting test regressions may capture a lot of different types of non-linearities.

The paper is organized as follows. Section 2 introduces STAR models. Section 3 describes the linearity tests used and section 4 gives the unit root and stationarity tests. In section 5 several model selection rules are proposed and section 6 contains a Monte Carlo study showing the success rates of the various model selection rules. Section 7 includes empirical applications to US government bond yields, the one-month interbank rate and the spread between them. Section 8 concludes. All Tables and Figures can be found in the Appendix of this chapter.

2 Non-linear STAR model

In the following section we briefly discuss the often applied first-order stationary STAR process and a non-stationary variant of it that has been studied by Harvey and Leybourne (2007).

Non-linear stationary STAR model

Consider the non-linear data generating process (DGP) for y_t with constant μ and let time be $t = 1, 2, \ldots, T$,

$$\begin{aligned} y_t &= \mu + v_t & (1)\\ v_t &= \phi v_{t-1} + \delta f(v_{t-1}, \theta) v_{t-1} + \varepsilon_t & (2)\\ \varepsilon_t &\sim i.i.d.(0, \sigma^2)\ . & (3) \end{aligned}$$

The error term ε_t is assumed to be a white noise process with mean zero and variance σ^2. The autoregressive parameters in this model are ϕ and δ. Non-linearity arises due to the presence of the smooth transition function f which depends on the two-dimensional parameter vector $\theta = (\gamma, c)'$, where $\gamma > 0$ determines the shape and $c \in \mathbb{R}$ the location of f. Common specifications for f are the exponential (f_E) and the logistic (f_L) smooth transition function

$$f_E(v_{t-1}, \theta) = 1 - \exp\{-\gamma(v_{t-1} - c)^2\} \tag{4}$$

$$f_L(v_{t-1}, \theta) = \frac{2}{1 + \exp\{-\gamma(v_{t-1} - c)\}} - 1 \; . \tag{5}$$

The main difference between them is the symmetry of f_E and the asymmetry of f_L with respect to $v_{t-1} - c$. It is implicitly assumed that the DGP is self-exciting, which means that a lag of the process itself, the first lag (y_{t-1}) in our case, is the transition variable. A further assumption is that there are two regimes with a smooth transition between them. The first regime is characterized by $f = \{f_E, f_L\} = 0$ and the second by $f = 1$,

$$v_t = \phi v_{t-1} + \varepsilon_t \; , \quad f = 0 \tag{6}$$

$$v_t = (\phi + \delta) v_{t-1} + \varepsilon_t \; , \quad f = 1 \; . \tag{7}$$

Since this model comprises a linear AR process in each regime, we can measure local persistence by the sign and the magnitude of the autoregressive parameter in the respective regime, which is given by ϕ and $\phi + \delta$, respectively. In particular, local persistence changes smoothly from ϕ to $\phi + \delta$. Note that all other characteristics of the process, e.g. the variance of the error term ε_t, are not changing as a regime switch occurs.

If $f = f_E$, then y_t is globally stationary if $|\phi + \delta| < 1$, while $|\phi \pm \delta| < 1$ must be fulfilled in order to achieve global stationarity under $f = f_L$, see Harvey and Leybourne (2007). Therefore, the stationarity condition for logistic STAR models are far more restrictive. Note, that a local unit root ($\phi = 1$) or even local explosiveness ($\phi > 1$) is permitted in the case of an exponential smooth transition function ($f = f_E$) while maintaining global stationarity of y_t. If $f = f_L$ is specified, such behavior is ruled out due to stronger restrictions for stationarity.

Non-linear non-stationary STAR model

Analogously to the nonlinear $I(0)$ DGP, Harvey and Leybourne (2007) consider an $I(1)$ version of it where non-linearity enters through first differences, i.e.

$$y_t = \mu + v_t \tag{8}$$

$$\Delta v_t = \phi \Delta v_{t-1} + \lambda f(\Delta v_{t-1}, \theta) \Delta v_{t-1} + \varepsilon_t \; . \tag{9}$$

In contrast to the previously discussed DGP, this one has an autoregressive lag structure of two and is globally non-stationary, which becomes more obvious after some rearrangements,

$$y_t = \mu + v_t \tag{10}$$

$$v_t = [1 + \phi + \lambda f(\Delta v_{t-1}, \theta)]\, v_{t-1} - [\phi + \lambda f(\Delta v_{t-1}, \theta)]\, v_{t-2} + \varepsilon_t \; . \tag{11}$$

The autoregressive parameters sum up to one, implying at least one unit root, regardless of the value of the smooth transition function f. This property still holds if the two extremes of f are considered,

$$v_t = (1+\phi) v_{t-1} - \phi v_{t-2} + \varepsilon_t \; , \quad f = 0 \tag{12}$$

$$v_t = (1+\phi+\lambda) v_{t-1} - (\phi+\lambda) v_{t-2} + \varepsilon_t \; , \quad f = 1 \; . \tag{13}$$

Suppose that $f = f_E$. If $\phi = 0$, then the lag structure changes from one to two as a regime shift occurs. Furthermore, if $f = 0$, then the process exhibits a unit root and if $f = 1$, then one root lies on the unit circle while the other one lies outside of it as long as $0 < \lambda < 1$ holds. Moreover, if $\phi = 1$ and $-1 < \lambda < 0$ then two unit roots are present under $f = 0$, while for $f = 1$ one root lies again on the unit circle and the second lies outside. A third case that is studied in the following is given by the setting $\phi = 1.5$ and $-1.5 < \lambda < -0.5$ which implies one unit root and one root inside the unit circle for $f = 0$, and one unit root and one root outside the unit circle for $f = 1$. Hence, the non-linear and non-stationary exponential STAR model can have very different local persistence properties while it is globally non-stationary. Similarly to the stationary non-linear STAR model, such rich dynamics are not permitted if the logistic transition function $(f = f_L)$ is assumed.

3 Testing time series linearity

The non-linear DGPs that were presented in the previous section become linear under the constraint that the smoothness parameter equals zero, i.e. $\gamma = 0$. This holds true for the stationary as well as for the non-stationary DGP. Additionally, linearity can be achieved by setting δ (for the stationary DGP) or λ (for the non-stationary DGP) equal to zero. This means that if $H_0 : \gamma = 0$ is tested against $H_1 : \gamma > 0$, δ or λ appears to be a nuisance parameter under the null hypothesis. This circumstance is often referred to as the Davies problem, see Davies (1987). Luukkonen et al. (1988) suggested to overcome this problem of unidentified parameters under H_0 by applying a Taylor approximation to the smooth transition function f around $\gamma = 0$. This approach has been widely adopted and also applied by Harvey and Leybourne (2007) who employ a second-order expansion. In particular, when such an approximation is applied to the stationary and non-stationary DGP respectively, we get

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1}^3 + \varepsilon_t & (14) \\ \Delta y_t &= \beta_0 + \beta_4 \Delta y_{t-1} + \beta_5 (\Delta y_{t-1})^2 + \beta_6 (\Delta y_{t-1})^3 + \varepsilon_t \, . & (15) \end{aligned}$$

The first equation is the Taylor approximation for the stationary $I(0)$ process and the second equation is the one for the corresponding non-stationary $I(1)$ process defined above. These auxiliary regression serve as the basis for testing linearity which is done by testing $H_0' : \beta_2 = \beta_3 = 0$ (for the stationary DGP) or $H_0' : \beta_5 = \beta_6 = 0$ (for the non-stationary DGP). Allowing for both degrees of integration simultaneously, $I(0)$ as well as $I(1)$, Harvey and Leybourne (2007) propose a hybrid test regression that incorporates terms from both individual test regressions, i.e.

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1}^3 \\ &\quad + \beta_4 \Delta y_{t-1} + \beta_5 (\Delta y_{t-1})^2 + \beta_6 (\Delta y_{t-1})^3 + \varepsilon_t \, . \qquad (16) \end{aligned}$$

Now, the null hypothesis of linearity corresponds to four restrictions formulated as $H_0' : \beta_2 = \beta_3 = \beta_5 = \beta_6 = 0$, while the alternative of non-linearity can be written as H_1' : at least one of $\beta_2, \beta_3, \beta_5, \beta_6 \neq 0$. Harvey and Leybourne (2007) suggest the Wald statistic

$$W_T = \frac{RSS_0 - RSS_1}{RSS_1/T} , \tag{17}$$

where RSS_i denotes the sum of squared residuals under H_i and T is the number of observations used in the test regression. Let us denote the Wald statistic (17) computed via (14) by W_T^0 and the one computed via (15) by W_T^1, where the exponent indicates the (implicitly) assumed degree of integration. Standard results imply that the limiting distribution of W_T^d is $\chi^2(2)$ if y_t is $I(d)$ with $d = \{0, 1\}$. Harvey and Leybourne (2007) derive the non-standard distribution of W_T^0 if y_t is a linear random walk. Following the notation of Harvey and Leybourne (2007), $W_0 (= \chi^2(4))$ denotes the limiting distribution of W_T^0 under $y_t \sim I(0)$ and W_1 denotes its limiting distribution under $y_t \sim I(1)$. In order to achieve the same limiting distribution under both degrees of integration, Harvey and Leybourne (2007) make use of Vogelsang's (1998) approach. Consider the transformed Wald test statistic

$$W_T^* = \exp\{-bH_T\} W_T , \tag{18}$$

where b is a non-zero constant and H_T is a statistic for testing $I(1)$ versus $I(0)$ with a pivotal limiting distribution under the null hypothesis. In addition, it is necessary that it converges to zero in probability under the alternative. Harvey and Leybourne (2007) set $H_T = |DF_T|^{-1}$, with DF_T being the Dickey–Fuller t–statistic obtained from

$$y_t = \pi_0 + \pi_1 y_{t-1} + \kappa \Delta y_{t-1} + \varepsilon_t .$$

In order to have the same critical values under both degrees of integration, that is $P(W_0 > c_\alpha) = P(\exp\{-bH\} W_1 > c_\alpha) = \alpha$, the constant b, which depends on the significance level α, has to be chosen accordingly. Harvey and Leybourne

(2007) provide a response surface by fitting a seventh–order polynomial. Therefore, asymptotic critical values can be computed easily for any desired significance level α. This approaches, however, makes it impossible to use p-values because the test statistic W_T^* depends on the significance level.

The test regressions (14) and (15) can be used instead of (16) if the degree of integration is known, which is hardly the case in practice. Harvey and Leybourne (2007) show that the robust test has a good overall performance, but the price paid for robustification against non-stationarity can be high in terms of power. For example, if $T = 300$, the power loss that results from using W_T^* instead of W_T^1 can be up to twenty or nearly thirty percent for exponential or logistic processes, respectively.

4 Testing for and against unit roots

Unit root test

The unit root test we consider in this paper is the one proposed by Kapetanios et al. (2003). It builds upon a first-order Taylor approximation of a stationary exponential STAR model. The resulting test regression reads

$$\Delta y_t = \psi y_{t-1}^3 + \sum_{i=1}^{p-1} \rho_i \Delta y_{t-i} + \varepsilon_t$$

where the error term ε_t contains the Taylor approximation remainder that equals zero under the null hypothesis $H_0 : \psi = 0$. The alternative hypothesis is given by $H_1 : \psi < 0$ which ensures global stationary. The authors suggest a Dickey-Fuller-type t-statistic given by

$$t_T = \frac{\hat{\psi}}{\sqrt{\hat{\text{var}}(\hat{\psi})}} = \frac{\sum_{t=1}^T y_{t-1}^3 \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^T y_{t-1}^6}}, \tag{19}$$

where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T (\Delta y_t - \hat{\psi} y_{t-1}^3)^2$ is the usual estimator of the error variance. For reasons of comparability with the stationarity test described in the next paragraph and because of empirical relevance we use de-meaned data $\tilde{y}_t \equiv y_t - \bar{y}$ where $\bar{y}$ denotes the mean of y_t.

Following the proof of consistency given in Kapetanios et al. (2003), one can verify that the test is consistent against stationary linear autoregressive processes as well. However, little is known about the small sample performance of this test if the true data generating process is actually non-stationary but non-linear as well. In section 6, we conduct the empirical size and power of the Kapetanios et al. (2003) test if data is generated by (non-stationary) exponential and logistic STAR models. In the following, the lag length p is set equal to two because the non-stationary STAR process is of order two.

Stationarity test

Harris et al. (2003) propose a test for stationarity against a unit root that is based on sample autocovariances. Define $a_{t,k} = \tilde{y}_t \tilde{y}_{t-k}$, where $\tilde{y}_t$ denotes the deviation of y_t from its mean $\bar{y} \equiv \frac{1}{T}\sum_{t=1}^{T} y_t$. The test statistic is given by

$$S_T = \frac{1}{T^{1/2}} \frac{\sum_{t=k+1}^{T} a_{t,k}}{\hat{\omega}(a_{t,k})} \xrightarrow{d} N(0,1) \tag{20}$$

where $\hat{\omega}(a_{t,k})^2$ is the Bartlett kernel-based long run variance estimator of $a_{t,k}$. More specifically,

$$\hat{\omega}(a_{t,k})^2 = \hat{\gamma}_0(a_{t,k}) + 2\sum_{j=1}^{l}\left(1 - \frac{j}{l}\right)\hat{\gamma}_j(a_{t,k}) \tag{21}$$

$$\hat{\gamma}_j(a_{t,k}) = \frac{1}{T}\sum_{t=j+k+1}^{T} a_{t,k} a_{t-j,k} \tag{22}$$

The test rejects the null hypothesis of stationarity for large values of S_T. Since the simulation study in Harris et al. (2003) is somewhat limited, we extend their simulation analysis by considering the empirical power of S_T if the data generating process is (non-)stationary and non-linear, see section 6.

5 Testing-based model selection rules

The benchmark testing-based model selection rule is the one used in Harvey and Leybourne (2007) and it can be classified as a simultaneous procedure since it consists of two independently computed test statistics. These two statistics are the Harris et al. (2003) stationarity statistic S_T and the robust linearity statistic W_T^*. This model selection rule is referred to as R_1. If, for example, both tests lead to a rejection (R) of their respective null hypotheses, we conclude that the process is non-linear $I(1)$. This procedure is depicted as follows, where NR stands for a non-rejection:

$$S_T, W_T^* \xrightarrow{\text{NR, NR}} \text{L–}I(0)$$

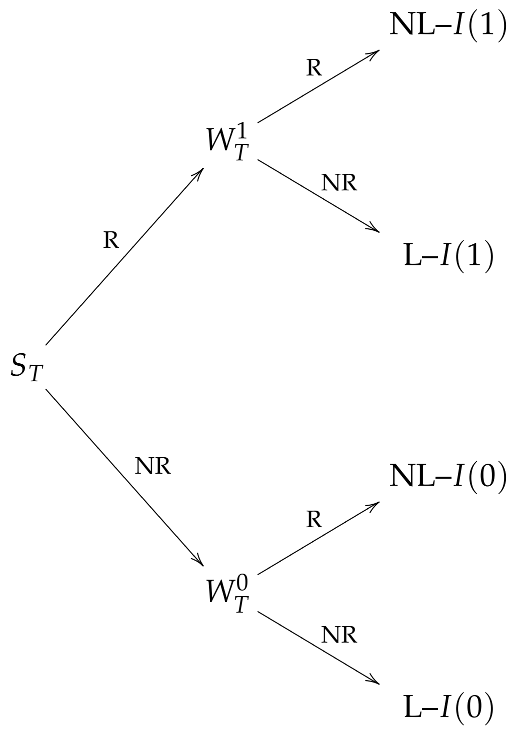
$$S_T, W_T^* \xrightarrow{\text{R, NR}} \text{L–}I(1)$$

$$S_T, W_T^* \xrightarrow{\text{NR, R}} \text{NL–}I(0)$$

$$S_T, W_T^* \xrightarrow{\text{R, R}} \text{NL–}I(1)$$

The success rate of a model selection rule is measured as the the relative frequency of correct selections. If, for example, the true DGP is linear $I(0)$, the success rate is simply the percentage of correct decisions for the category L–$I(0)$.

In the following section 6 we compare R_1 with a little modification of it, labeled as R_4, where we employ the Kapetanios et al. (2003) test instead of the Harris et al. (2003) test, by means of a Monte Carlo study. Moreover, we propose two versions of a two-step procedure that consists of a unit root or a stationarity statistic in a first step and a linearity test in the second step. The first-stage result is used to select the appropriate test regression: if the test in the first stage gives evidence for $I(0)$ we run the linearity test regression in levels, see equation (14), and otherwise in first differences, see equation (15).[1] The model selection rule R_2 applies the Kapetanios et al. (2003) test in the first step, while R_3 uses the Harris et al. (2003) test instead. Note, that the second step of R_2 and R_3 is identical. For example, the R_3 procedure can be depicted as:



The intuition behind the two-step procedures is as follows: the robustification of the linearity test that is achieved by applying W_T^* instead of W_T^0 or W_T^1 induces a substantial power loss in small samples which is not surprising at all. We analyze whether this power loss can be reduced by first selecting the appropriate linearity test regression given by equation (14) or (15). Note that the linearity test can be improved if the degree of integration is known, see Table 3 in Harvey and Leybourne

[1]This approach is also mentioned in Sandberg (2008).

(2007). In practice, the degree of integration is generally unknown, but if we can exploit the information of a powerful test for $I(0)$ or $I(1)$, then the assumption of an unknown degree of integration becomes superfluous. Table 1 gives a short overview over different data generating processes and model selection rules.

6 Monte Carlo study

This section reports results of a variety of simulation experiments that shed light on the empirical small sample properties of the four model selection rules. In a first step, we study the performance of the Kapetanios et al. (2003) unit root test and the Harris et al. (2003) stationarity test under non-linear $I(0)$ and $I(1)$ processes. On the one hand, it is not clear whether the Kapetanios et al. (2003) test is correctly sized under non-linear $I(1)$ DGPs and on the other hand there are no simulation results available yet that allow to draw conclusions about the behavior of the Harris et al. (2003) stationarity test under non-linear $I(0)$ processes. Furthermore, we are interested in the power of the Kapetanios et al. (2003) and the Harris et al. (2003) test if the DGP is of non-linear STAR-type. The interpretation of outcomes of different model selection rules will be easier when we have these results in mind.

The parameter settings we consider are a large subset of those used in Harvey and Leybourne (2007) in order to achieve comparability. The nominal significance levels are one, five and ten percent for the size and power analysis of the Kapetanios et al. (2003) and the Harris et al. (2003) test, while we use a nominal five percent level of significance for the study of model selection rules. The sample size is chosen as $T = 300$ which is also used in Harvey and Leybourne (2007). Following Harvey and Leybourne (2007), we set $k = [2T^{1/2}]$ and $l = [12(T/100)^{1/4}]$, where $[x]$ denotes the nearest integer of x, for the Harris et al. (2003) test. Moreover, the lag length (p) for the Kapetanios et al. (2003) test is set equal to one.

Table 2 reports the empirical sizes and powers of the Kapetanios et al. (2003) test and the Harris et al. (2003) test under ESTAR $I(0)/I(1)$ and LSTAR $I(0)/I(1)$ processes. We observe that the Kapetanios et al. (2003) test is a bit undersized if the the true DGP is non-linear $I(1)$. The only exception can be found in the last row. The same conclusions hold true for the Harris et al. (2003) test as well, where we also observe an exception. Although we do not provide an analytic proof, we conjecture that the distributions of t_T and S_T depend on the parameters of the non-linear DGP because their performance varies with these parameters. In order to cope with this problem a suitable bootstrap algorithm could be used to obtain more accurate critical values which is beyond the scope of this paper. Regarding the power properties both tests show quite good performance. The Kapetanios et al. (2003) test often reaches hundred percent of power, while the power of the Harris et al. (2003) test is also quite high. In particular, both tests appear to be very powerful against logistic STAR models.

When analyzing the performance of the four model selection rules we choose a nominal significance level of five percent. We start with linear first-order autoregressive processes, labeled as L–$I(0)$, see Table 3. The autoregressive parameter ϕ takes the values $0.00, 0.30, \ldots, 0.99, 1.00$. The local-to-unity values are chosen because it gets very difficult for unit root and stationarity tests to distinguish $I(0)$ from $I(1)$ processes in this region. If ϕ lies between zero and 0.7 we cannot observe big differences between the four model selection rules and all of them show very good performance. In the local-to-unity region, the performance of R_2 and R_4, which are both based on the Kapetanios et al. (2003) test, worsens dramatically. On the contrary, the model selection rules R_1 and R_3, which both use the Harris et al. (2003) test, are performing relatively good although the frequency of correct decisions is not extremely high at $\phi = 0.99$, but this could not be expected anyway. Wrong decision are made clearly in the direction of linear $I(1)$ processes which is due to the behavior of the unit root and the stationarity test.

Next, the accuracy of model selection rules R_1 to R_4 is analyzed for stationary non-linear exponential STAR processes. Results are reported in Table 4. Both simultaneous rules (R_1 and R_4) are clearly outperformed by both two-step rules and in particular by R_2. However, in the experiments where $\phi = 0$, R_2 is dominated by R_3. Furthermore, the results for local unit root ($\phi = 1.0$) and local explosiveness ($\phi = 1.5$) suggest that the differences between both two-step methods are not big but R_2 dominates R_3. One exception to this is the case of $\phi = 1.5, \lambda = -1.0, \gamma = 0.1$, which is due to the fact that the Harris et al. (2003) test is definitely oversized in this case, see Table 2. The overall performance of R_2 and R_3 is satisfying and the gains with respect to R_1 range from four to nineteen percent. Additionally, we observe only little differences between the two simultaneous rules R_1 and R_4. Moreover, wrong decisions are often made in the direction of linear $I(0)$ processes because of the type II-error of linearity tests. It is rarely the case that the process is misclassified as non-stationary regardless of which model selection rule is applied.

Turning to non-stationary exponential STAR processes, the results in Table 5 suggest that R_2 is best performing, followed by R_3. The gains from using a two-step procedure are evident as they range from four to twenty-six percent and they are higher on average than for ESTAR $I(0)$ processes. Again, R_1 and R_4 show quite similar performance but R_4 is preferable to R_1. Nonetheless, their success rates are far below those of R_2 and R_3. Our conclusions do not change a lot when interpreting the outcomes for logistic stationary (upper part of Table 6) and non-stationary (lower part of Table 6) STAR processes. Nonetheless, the differences between simultaneous and two-step rules are less pronounced for stationary processes. Furthermore, the frequency of correct decisions increases with the smoothness parameter γ in the case of logistic STAR processes because it does not become linear in limit ($\gamma \to \infty$). On the contrary, ESTAR processes become linear in the limit.

A unified two-step procedure

Recall that the two-step procedure based on the Kapetanios et al. (2003) test (R_2) performs relatively poor in the case of linear processes but very good for non-linear processes. Further note that the two-step procedure using the Harris et al. (2003) test (R_3) shows relatively good performance for linear processes as well. Hence, it is worthwhile to think of a procedure that takes the best out of both. One approach we suggest is to pre-test for linearity using the robust Wald statistic W_T^* and to proceed with R_2 in the case of a rejection and with R_3 in the case of a non-rejection. More formally, the unified model selection rule, labeled as R_5 in the following, is defined by

$$\begin{aligned} \text{If } W_T^* &\geq \chi^2_{1-\alpha}(4), \text{ then } R_5 = R_2 \\ \text{If } W_T^* &< \chi^2_{1-\alpha}(4), \text{ then } R_5 = R_3 \ , \end{aligned}$$

where $\chi^2_{1-\alpha}(4)$ denotes the $(1-\alpha)\%$ asymptotic critical value for the W_T^* statistic. This unification is somehow in the spirit of the methodology used in Harvey et al. (2008b) which is based on the comments of Breitung to Harvey et al. (2008a).

Tables 7 and 8 report the results for the model selection rule R_5 which clearly show that the proposed unification works very well. On the one hand, R_5 has the satisfying properties of R_3 when the true DGP is linear and on the other hand it shares the qualities of R_2 if the non-linearities are present. The gains are obvious in the case of non-linear (non-)stationary STAR processes when compared to the performance of R_1, see Tables 4, 5, 6 and 8.

7 Empirical application

The unified two-step procedure R_5 is applied to the US government bond yield, the one-month interbank rate and the spread between them. Data is taken from Datastream.[2] Our sample spans from 1986, February to May, 2008 and consists of 268 monthly observations. Figure 1 depicts the three time series. No clear trend can be detected in the spread by visual inspection and economic theory does not suggest that there are deterministic trends in interest rates, too. Therefore, we include only constants in the test regressions or we use de-meaned data.

As a by-product of this application we test the expectation hypothesis of the term structure (EHT) that requires the term spread to be stationary. However, the main aim of this empirical application is to classify the time series as (non-)linear and/or (non-)stationary. Such classification is of big importance for model building, the analysis of monetary shocks and for forecasting.

Results are reported in Table 9. As in Harvey and Leybourne (2007), we select the lag length for the test regressions by using a general-to-specific methodology at the ten percent level of significance. The maximum lag order is set equal to four and the minimum equal to two. In a first step, we use the W_T^* test statistic in order to choose the appropriate model selection rule which is R_2 for all time series, because W_T^* is significant in all cases. We conclude that the government bond yield and the interbank rate are non-stationary. In both cases, the Kapetanios et al. (2003) test does not reject the unit root hypothesis at the employed ten percent level of significance. On the contrary, the term spread appears to be stationary which supports the EHT and hints at cointegration between the government bond yield and the interbank rate. Although not reported in Table 7, the Harris et al.

[2] The relevant codes are USGBOND. and BBUSD1M for the government bond yield and the interbank rate, respectively.

(2003) test confirms the conclusions drawn by the Kapetanios et al. (2003) test results, since S_T equals 2.721 (government bond yield), 1.702 (interbank rate) and 0.360 (term spread). The asymptotic critical value equals 1.282 at the nominal ten percent level of significance. Hence, the linearity test is carried out using first differences, see equation (15), in each case. In two cases we have to reject the null hypothesis of linearity in favor of STAR-type non-linearity. We conclude that non-linearities are more important for the shorter maturity and that the type of cointegration between the government bond yield and the interbank rate is in fact non-linear.

8 Conclusions

This paper deals with a model selection problem that appears when (non-)stationary and (non-)linear models are considered at once. The first testing-based approach, which serves as our benchmark in the following, was put forward by Harvey and Leybourne (2007) who suggested to use two robust test statistics independently. The applied stationarity test by Harris et al. (2003) is robust with respect to non-linearity while the linearity test proposed by Harvey and Leybourne (2007) is robust with respect to the integer degree of integration. Based on the outcomes of both individual tests, one of the four possible models is selected: linear $I(0)$, linear $I(1)$, non-linear $I(0)$ or non-linear $I(1)$.

We propose the usage of two-step methods which apply a test for discriminating between stationary and non-stationary time series in the first step. Depending on this pre-test decision, the linearity test is carried out in levels or first differences. If the degree of integration is correctly determined, the linearity tests in the second stage are more powerful than the robust test in small samples. Therefore, two-step methods may be more accurate in selecting the true model.

In a large-scale Monte Carlo study we investigate the empirical performance of various testing-based model selection rules. The usage of two-step methods lead to more precise model selection than one-step methods. However, as none of the considered rules appears to be dominating, we suggest to use a unified procedure that takes the best out of the two best performing rules. The unified rule is based on robust pre-testing for linearity and it outperforms the benchmark especially for non-linear models. In an empirical application to US interest rate data we find evidence for the expectation hypothesis of the term structure and nonlinearities in the spread between long- and short-term interest rates.

References

Brock, W.A., J.A. Scheinkman, W.D. Dechert and B. LeBaron (1996): "A Test for Independence Based on the Correlation Dimension." *Econometric Reviews* 15, 197–235.

Choi, C.Y. and Y.K. Moh (2007): "How Useful are Tests for Unit-Root in Distinguishing Unit-Root Processes from Stationary but Non-Linear Processes?" *Econometrics Journal* 10, 82–112.

Davies, R.B. (1987): "Hypothesis Testing When a Nuisance Parameter is Present Under the Alternative." *Biometrika* 74, 33–43.

Dickey, D.A. and W.A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series With a Unit Root." *Journal of the American Statistical Association* 74, 427–431.

Harris, D., B.P. McCabe and S.J. Leybourne (2003): "Some limit theory for autocovariances whose order depends on sample size." *Econometric Theory* 19, 829–864.

Harvey, D.I. and S.J. Leybourne (2007): "Testing for Time Series Linearity." *Econometrics Journal* 10, 149–165.

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor (2008a): "Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition (with commentaries and rejoinder)." *Econometric Theory*, forthcoming.

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor (2008b): "Testing for unit roots and the impact of quadratic trends, with an application to relative primary commodity prices." *Granger Centre Discussion Paper Series* 08/04.

Kapetanios, G., Y. Shin and A. Snell (2003): "Testing for a unit root in the non-linear STAR framework." *Journal of Econometrics* 112, 359–379.

Lee, T.H., H. White and C.W.J. Granger (1993): "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests." *Journal of Econometrics* 56, 269–290.

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

Nelson, C.R. and C.I. Plosser (1982): "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics* 10, 139–162.

Sandberg, R. (2008): "Critical values for linearity tests in time-varying smooth transition autoregressive models when data are highly persistent." *Econometrics Journal*, forthcoming.

Vogelsang, T.J. (1998): "Trend Function Hypothesis Testing in the Presence of Serial Correlation." *Econometrica* 66, 123–148.

Appendix

Table 1: Data generating processes and model selection rules

Type	Expression
L–$I(0)$	$(1-\phi L)y_t = \varepsilon_t$
L–$I(1)$	$(1-L)y_t = \varepsilon_t$
NL–$I(0)$	$(1-\phi L)y_t = \delta f(y_{t-1}, \gamma) y_{t-1} + \varepsilon_t$
NL–$I(1)$	$(1-\phi L)\Delta y_t = \lambda f(\Delta y_{t-1}, \gamma)\Delta y_{t-1} + \varepsilon_t$
R_1	Simultaneous, Harris et al. (2003) Test & W^*
R_2	Two-step, Kapetanios et al. (2003) Test & W_0/W_1
R_3	Two-step, Harris et al. (2003) Test & W_0/W_1
R_4	Simultaneous, Kapetanios et al. (2003) Test & W^*

Table 2: Empirical size and power of Kapetanios et al. (2003) and Harris et al. (2003) test

			ESTAR–$I(1)$						ESTAR–$I(0)$					
			t_T			S_T			t_T			S_T		
ϕ	λ	γ	1.0	5.0	10.0	1.0	5.0	10.0	1.0	5.0	10.0	1.0	5.0	10.0
0.0	0.7	0.1	1.0	2.2	4.8	63.9	87.3	92.2	99.0	99.7	99.9	0.4	3.4	8.0
		0.5	1.4	3.5	4.9	63.2	87.0	92.2	84.1	95.2	97.8	0.4	3.4	8.8
		0.9	1.0	4.6	6.2	63.4	87.3	93.3	90.6	97.8	99.1	0.3	4.4	8.6
	0.9	0.1	1.0	5.0	9.1	63.7	86.7	92.4	94.6	97.5	98.2	0.4	3.6	8.5
		0.5	1.1	3.8	7.2	65.0	89.2	93.7	15.2	36.9	52.7	0.2	3.3	8.0
		0.9	1.0	4.2	8.0	68.6	91.0	94.9	26.2	55.0	70.2	0.4	4.8	10.3
1.0	-0.7	0.1	0.5	3.7	8.0	66.7	89.0	93.7	100.0	100.0	100.0	0.4	3.0	7.0
		0.5	0.6	3.9	8.7	63.2	87.3	92.8	100.0	100.0	100.0	0.5	4.1	9.0
		0.9	1.0	4.7	8.8	63.4	87.5	92.5	100.0	100.0	100.0	0.7	4.6	9.3
	-0.9	0.1	0.4	3.8	8.0	64.6	88.7	93.2	100.0	100.0	100.0	0.7	4.0	8.2
		0.5	0.6	4.4	9.1	62.9	88.1	92.8	100.0	100.0	100.0	0.6	4.1	8.1
		0.9	1.0	4.5	9.2	64.4	87.0	92.7	100.0	100.0	100.0	0.5	3.5	8.2
1.5	-1.0	0.1	0.0	0.4	1.1	87.6	98.4	99.1	90.6	96.0	97.7	9.2	27.6	40.0
		0.5	0.4	3.2	7.2	67.0	89.7	94.8	100.0	100.0	100.0	0.5	3.9	8.2
		0.9	0.8	4.4	9.2	64.5	87.6	92.5	100.0	100.0	100.0	0.6	3.6	8.2
	-1.4	0.1	0.1	1.0	3.2	76.6	94.4	96.8	100.0	100.0	100.0	1.0	6.1	12.6
		0.5	0.7	4.0	8.7	64.4	87.4	92.7	100.0	100.0	100.0	0.6	3.6	8.6
		0.9	1.2	5.0	9.6	65.7	88.6	93.3	100.0	100.0	100.0	0.4	4.0	9.4
			LSTAR–$I(1)$						LSTAR–$I(0)$					
0.0	0.7	0.1	0.7	4.0	8.4	65.1	88.5	93.4	100.0	100.0	100.0	0.2	3.6	8.2
		0.5	1.0	3.8	6.5	95.1	99.3	99.6	99.7	99.8	99.9	0.7	4.5	9.0
		0.9	2.3	5.3	7.9	99.6	99.9	100.0	99.3	99.8	99.9	0.5	4.1	8.8
	0.9	0.1	0.9	5.2	9.4	69.0	89.8	94.1	99.9	99.9	100.0	0.6	3.8	8.5
		0.5	1.8	4.8	7.3	98.6	99.9	100.0	98.5	99.2	99.6	0.4	3.7	9.1
		0.9	7.1	11.4	13.6	99.8	100.0	100.0	84.2	90.2	93.0	0.3	3.0	7.2

Table 3: Success rates for linear AR

	L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
ϕ	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.00	89.6	94.5	93.7	94.9	4.8	0.0	1.6	0.0	5.3	5.5	4.6	5.1	0.3	0.0	0.1	0.0
0.30	92.5	95.4	91.4	96.5	3.3	0.0	3.7	0.0	3.9	4.6	4.9	3.5	0.3	0.0	0.0	0.0
0.50	92.8	96.9	91.8	96.1	3.8	0.0	3.9	0.0	3.3	3.1	4.0	3.9	0.1	0.0	0.3	0.0
0.70	92.7	94.4	90.2	94.5	3.6	0.3	4.8	0.1	3.5	5.3	4.8	5.4	0.2	0.0	0.2	0.0
0.90	88.6	74.8	87.2	75.2	6.7	18.6	7.5	19.3	3.9	5.4	4.8	4.7	0.8	1.2	0.5	0.8
0.95	72.1	37.6	71.4	33.0	23.0	55.6	22.5	59.3	3.3	5.0	5.5	5.0	1.6	1.8	0.6	2.7
0.99	21.0	5.5	23.2	6.9	73.3	87.6	70.9	87.1	0.3	2.0	1.8	1.2	5.4	4.9	4.1	4.8
1.00	12.7	3.1	9.1	3.6	82.6	91.4	85.8	90.4	0.9	0.6	1.6	1.2	3.8	4.9	3.5	4.8

Table 4: Success rates for non-linear ESTAR $I(0)$

			L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
ϕ	λ	γ	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.0	0.7	0.1	68.2	61.6	60.7	72.7	2.8	0.2	3.1	0.0	28.0	38.2	36.1	26.8	1.0	0.0	0.1	0.5
		0.5	52.5	35.8	39.0	49.0	1.9	4.5	3.9	0.2	44.2	59.7	56.8	46.5	1.4	0.0	0.3	4.3
		0.9	77.5	76.2	73.4	81.9	2.9	1.7	2.5	0.5	18.6	22.1	24.0	15.7	1.0	0.0	0.1	1.9
	0.9	0.1	49.7	38.7	38.1	52.6	1.5	2.6	2.1	0.0	47.1	58.5	59.8	45.4	1.7	0.2	0.0	2.0
		0.5	45.8	17.2	29.7	24.4	1.9	62.8	3.3	27.8	50.9	17.1	66.8	11.3	1.4	2.9	0.2	36.5
		0.9	86.8	54.4	82.6	51.8	3.9	39.6	4.2	39.5	8.8	3.9	12.9	3.1	0.5	2.1	0.3	5.6
1.0	-0.7	0.1	20.1	9.2	9.1	17.5	0.3	0.0	3.7	0.0	76.7	90.8	86.8	82.5	2.9	0.0	0.4	0.0
		0.5	36.1	25.0	22.9	37.8	1.2	0.0	4.2	0.0	60.9	75.0	72.7	62.2	1.8	0.0	0.2	0.0
		0.9	64.3	54.6	52.2	66.9	1.8	0.0	3.1	0.0	32.1	45.4	44.4	33.1	1.8	0.0	0.3	0.0
	-0.9	0.1	7.0	3.1	3.3	9.6	0.1	0.0	4.6	0.0	89.9	96.9	90.9	90.4	3.0	0.0	1.2	0.0
		0.5	16.0	10.0	9.4	17.5	0.7	0.0	2.6	0.0	80.1	90.0	87.4	82.5	3.2	0.0	0.6	0.0
		0.9	43.0	33.2	32.0	44.6	1.6	0.0	2.8	0.0	52.7	66.8	65.0	55.4	2.7	0.0	0.2	0.0
1.5	-1.0	0.1	0.0	0.0	0.0	0.0	0.0	3.2	22.1	0.0	73.0	94.5	75.4	96.1	27.0	2.3	2.5	3.9
		0.5	6.5	2.8	2.0	6.8	0.2	0.0	4.8	0.0	90.1	97.2	92.3	93.2	3.2	0.0	0.9	0.0
		0.9	37.7	26.4	26.3	41.4	1.8	0.0	5.0	0.0	58.7	73.6	68.3	58.6	1.8	0.0	0.4	0.0
	-1.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	5.8	0.0	92.9	100.0	92.2	100.0	7.1	0.0	2.0	0.0
		0.5	0.2	0.1	0.0	0.1	0.0	0.0	3.1	0.0	96.0	100.0	95.4	99.9	3.8	0.0	1.5	0.0
		0.9	9.2	4.7	4.1	7.4	0.3	0.0	2.9	0.0	87.2	95.3	92.5	92.6	3.3	0.0	0.5	0.0

Table 5: Success rates for non-linear ESTAR $I(1)$

			L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
ϕ	λ	γ	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.0	0.7	0.1	10.0	4.1	13.0	2.7	65.2	60.7	52.2	70.5	3.1	0.4	1.0	2.4	21.7	34.8	33.8	24.4
		0.5	7.5	4.2	10.9	1.3	49.3	35.2	34.2	55.7	5.0	0.6	0.6	2.5	38.2	60.0	54.3	40.5
		0.9	10.4	3.3	11.3	3.9	70.9	70.3	65.9	79.0	2.7	1.0	1.1	0.8	16.0	25.4	21.7	16.3
	0.9	0.1	6.5	4.9	12.7	2.4	50.8	39.4	35.5	53.1	7.2	0.4	0.9	3.1	35.5	55.3	50.9	41.4
		0.5	7.2	2.0	10.2	1.3	46.3	29.9	29.0	49.9	5.7	1.5	2.0	2.3	40.8	66.6	58.8	46.5
		0.9	7.2	2.6	4.7	3.5	81.4	81.4	81.6	85.1	0.7	1.1	1.8	0.6	10.7	14.9	11.9	10.8
1.0	-0.7	0.1	2.2	4.7	9.0	0.5	20.4	8.5	7.6	25.0	8.6	0.1	1.9	3.4	68.8	86.7	81.5	71.1
		0.5	4.0	3.8	12.7	1.7	38.8	23.8	22.5	43.9	7.9	0.1	0.3	3.4	49.3	72.3	64.4	51.0
		0.9	7.2	4.5	11.1	2.7	63.5	50.7	47.8	65.5	3.4	0.2	0.3	2.3	25.9	44.6	40.8	29.5
	-0.9	0.1	1.4	3.7	8.3	0.3	12.8	2.7	3.2	14.7	11.6	0.0	0.7	4.2	74.2	93.5	87.8	80.8
		0.5	3.4	4.2	12.6	0.8	20.2	9.2	7.6	24.4	11.1	0.0	0.5	5.2	65.3	86.6	79.3	69.6
		0.9	6.6	3.9	13.2	1.4	45.9	31.6	27.8	50.1	6.2	0.1	0.5	3.1	41.3	64.4	58.5	45.4
1.5	-1.0	0.1	0.0	0.5	1.3	0.0	4.2	0.0	0.0	4.1	0.8	0.0	0.9	0.3	95.0	99.5	97.8	95.6
		0.5	0.4	3.2	10.7	0.1	15.6	2.7	2.2	13.6	9.2	0.1	0.6	2.4	74.8	94.0	86.5	83.9
		0.9	4.6	3.7	10.7	1.8	40.6	31.8	23.7	45.7	7.0	0.1	0.8	2.5	47.8	64.4	64.8	50.0
	-1.4	0.1	0.0	0.5	5.5	0.0	3.4	0.0	0.0	2.3	5.5	0.0	1.1	0.9	91.1	99.5	93.4	96.8
		0.5	0.0	3.6	12.2	0.0	4.1	0.2	0.0	3.4	13.0	0.1	0.1	4.0	82.9	96.1	87.7	92.6
		0.9	1.0	4.4	11.2	0.3	13.6	3.5	3.2	15.7	11.9	0.1	0.5	4.5	73.5	92.0	85.1	79.5

Table 6: Success rates for non-linear LSTAR $I(0)$

			L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
ϕ	λ	γ	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.0	0.7	0.1	88.9	89.5	88.2	90.9	3.6	0.0	2.6	0.0	7.2	10.5	8.7	9.1	0.3	0.0	0.5	0.0
		0.5	10.0	7.0	7.7	11.5	0.2	0.1	0.5	0.0	85.7	92.9	89.2	88.5	4.1	0.0	2.6	0.0
		0.9	0.2	0.0	0.1	0.6	0.0	0.1	0.6	0.0	95.6	99.4	96.7	99.4	4.2	0.5	2.6	0.0
	0.9	0.1	85.2	85.5	82.1	89.0	2.8	0.0	2.8	0.0	11.7	14.5	14.6	11.0	0.3	0.0	0.5	0.0
		0.5	1.4	0.5	1.0	2.5	0.1	0.1	1.3	0.0	94.8	99.0	95.8	97.2	3.7	0.4	1.9	0.3
		0.9	0.0	0.0	0.0	0.0	0.0	0.5	0.2	0.0	97.4	90.1	96.9	90.8	2.6	9.4	2.9	9.2

Table 6.6: Success rates for non-linear LSTAR $I(1)$

			L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
ϕ	λ	γ	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.0	0.7	0.1	8.3	2.8	9.0	2.9	82.9	86.1	79.4	88.2	0.8	0.8	1.3	0.7	8.0	10.3	10.3	8.2
		0.5	0.2	3.0	0.3	0.1	29.5	7.4	6.1	31.6	0.7	0.3	0.2	2.3	69.6	89.3	93.4	66.0
		0.9	0.0	2.1	0.0	0.1	12.2	0.0	0.1	14.0	0.0	0.6	0.0	2.6	87.8	97.2	99.9	83.3
	0.9	0.1	8.1	3.1	7.9	3.6	81.7	80.9	81.6	86.3	1.1	0.7	1.4	0.5	9.1	14.6	13.4	9.6
		0.5	0.0	2.3	0.2	0.0	19.4	0.4	0.2	16.8	0.2	0.4	0.0	3.3	80.4	97.1	99.0	79.9
		0.9	0.0	0.9	0.0	0.0	7.8	0.0	0.0	6.7	0.0	1.8	0.0	2.5	92.2	97.3	100.0	90.8

Table 7: Success rates of R_5 for linear AR

ϕ	L–$I(0)$	L–$I(1)$	NL–$I(0)$	NL–$I(1)$
0.00	90.6	4.5	4.9	0.0
0.30	92.4	2.6	5.0	0.0
0.50	92.0	3.5	4.5	0.0
0.70	92.0	3.6	4.4	0.0
0.90	86.7	8.6	4.3	0.4
0.95	70.1	24.2	4.6	1.1
0.99	21.8	72.6	1.4	4.2
1.00	11.2	83.2	0.9	4.7

Table 8: Success rates of R_5 for non-linear ESTAR $I(0)$ and $I(1)$

ϕ	λ	γ	L–$I(0)$	L–$I(1)$	NL–$I(0)$	NL–$I(1)$	L–$I(0)$	L–$I(1)$	NL–$I(0)$	NL–$I(1)$
0.0	0.7	0.1	60.0	2.5	37.5	0.0	11.0	54.6	0.5	33.9
		0.5	35.8	5.4	58.3	0.5	9.0	33.8	0.2	57.0
		0.9	73.0	5.2	21.6	0.2	12.3	65.7	0.0	22.0
	0.9	0.1	38.2	4.0	57.8	0.0	10.1	37.0	0.5	52.4
		0.5	30.6	47.7	19.5	2.2	5.8	29.0	2.1	63.1
		0.9	80.8	15.0	3.1	1.1	6.7	81.7	0.3	11.3
1.0	-0.7	0.1	8.7	0.4	90.9	0.0	4.3	8.7	0.0	87.0
		0.5	23.0	1.4	75.6	0.0	8.1	22.5	0.1	69.3
		0.9	50.8	2.1	47.1	0.0	10.9	47.0	0.0	42.1
	-0.9	0.1	3.9	0.3	95.8	0.0	4.1	3.3	0.0	92.6
		0.5	8.5	0.4	91.1	0.0	6.0	8.0	0.4	85.6
		0.9	31.1	1.6	67.3	0.0	7.9	31.5	0.5	60.1
1.5	-1.0	0.1	0.0	2.9	95.6	1.5	0.4	0.0	0.0	99.6
		0.5	2.4	0.3	97.3	0.0	4.0	2.4	0.2	93.4
		0.9	27.1	1.5	71.4	0.0	9.2	23.0	0.1	67.7
	-1.4	0.1	0.0	0.0	100.0	0.0	0.4	0.0	0.1	99.5
		0.5	0.0	0.0	100.0	0.0	4.0	0.1	0.2	95.7
		0.9	3.5	0.3	96.2	0.0	5.5	3.7	0.3	90.5
Success rates of R_5 for non-linear LSTAR $I(0)$ and $I(1)$										
0.0	0.7	0.1	88.5	3.2	8.3	0.0	10.0	78.8	0.6	10.6
		0.5	5.8	0.2	94.0	0.0	1.9	6.0	0.2	91.9
		0.9	0.1	0.1	99.6	0.2	1.8	0.0	0.9	97.3
	0.9	0.1	80.9	2.3	16.8	0.0	9.4	76.3	0.8	13.5
		0.5	0.6	0.1	98.8	0.5	3.2	0.6	0.6	95.6
		0.9	0.0	0.9	91.8	7.3	1.3	0.0	1.8	96.9

Table 9: Empirical application to US interest rates

Time Series	W_T^*	t_T	S_T	W_T^0	W_T^1	Model selection rule R_5
Bond yield	8.695*	−2.073	2.721*	—	3.458	Linear-$I(1)$
Interbank rate	26.231*	−1.621	1.702*	—	31.740*	Non-linear-$I(1)$
Spread	27.181*	−3.027*	0.360	5.396*	—	Non-linear-$I(0)$

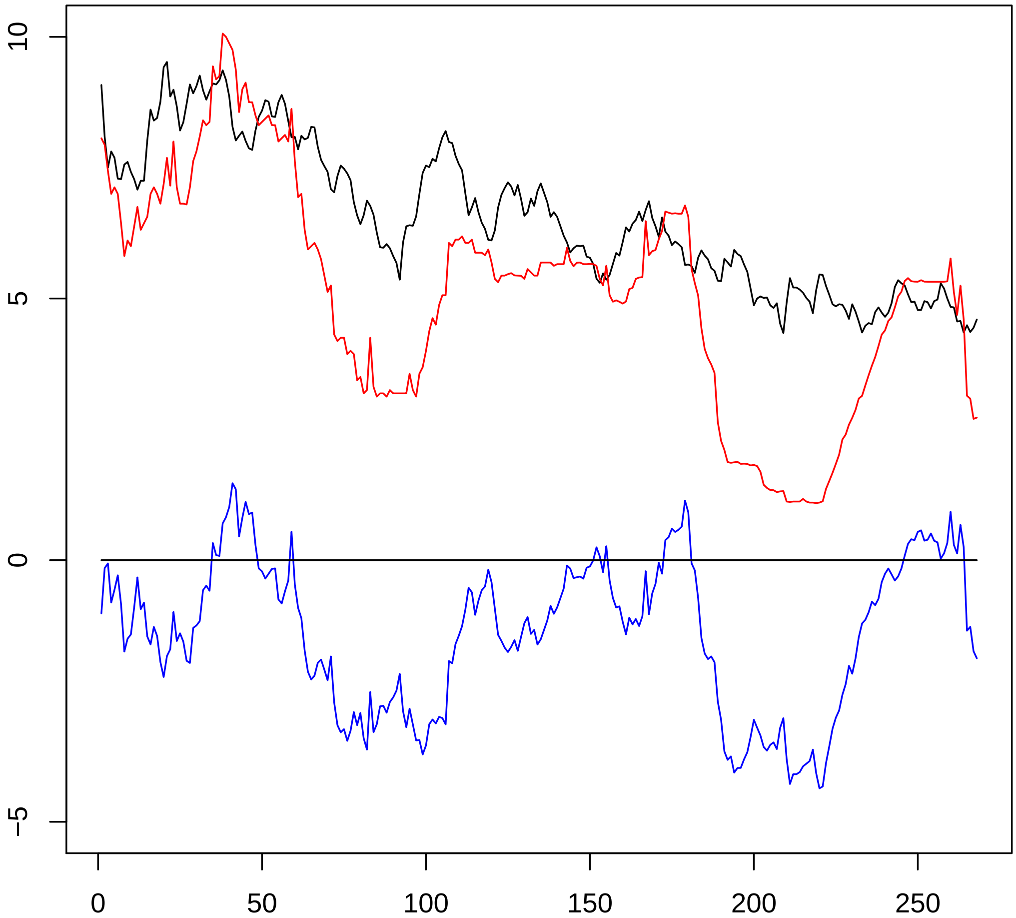


Figure 1: US bond yield (black), interbank rate (red) and spread (blue).