Generalizing smooth transition autoregressions

Emilio Zanetti Chini
(Università di Pavia)

# 114 (01-16)

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

January 2016
Generalizing smooth transition autoregressions

EMILIO ZANETTI CHINI

University of Pavia
Department of Economics and Management
Via San Felice 5 - 27100, Pavia (ITALY)
e-mail: ezc@eco.unipv.it

FIRST VERSION: October 2013
THIS VERSION: December 2015

Abstract

We introduce a new time series model capable to parametrize the joint asymmetry in duration and length of cycles - the dynamic asymmetry - by using a particular generalization of the logistic function. The modelling strategy is discussed in detail, with particular emphasis on two different tests for the null of symmetric adjustment and three diagnostic tests, whose power properties are explored via Monte Carlo experiments. Four case studies in classical economic and biological real datasets illustrate the versatility of the new model in different fields. In all the cases, the dynamic asymmetry in the cycle is efficiently detected and modelled. Finally, a rolling forecasting exercise is applied to the resulting estimates. Our model beats linear and conventional nonlinear competitors in point forecasting, while this superiority becomes less evident in density forecasting, specially when relying on robust measures.

Keywords: Dynamic asymmetry, Nonlinear time series, Econometric Modelling, Point forecasts, Density forecasts, Evaluating forecasts, Combining forecasts, Error measures.

JEL: C22, C51, C52
1 Introduction

Many of the economic and natural sciences time series show asymmetric fluctuations, see Tong (1990); Teräsvirta et al. (2010) *inter alia*. For example, in Business Cycle literature, Sichel (1993) gives a double definition of asymmetry: the first - the *steepness* - happens when contractions in the levels are steeper than expansions (symmetry in the level axis); the second - the *deepness* - when the series undergoes at an accelerating time until a minimum after which it starts to recover with high, decreasing acceleration, until to smoothly recover the peak (symmetry in time axis). When these two definitions are combined, we call this *dynamic asymmetry*.

Smooth transition autoregressions (STAR), originated by the pioneering contribution by Bacon & Watts (1971) in Biostatistics, then developed in time series framework by Haggan & Ozaki (1981); Chan & Tong (1986) and Teräsvirta (1994), are currently one of the most simple and successful tools to model the nonlinear dynamics in applied literature. In particular, a logistic transition is commonly postulated when the series under consideration is assumed having asymmetric oscillations from its conditional mean. We argue that, being the logistic function reflectively symmetric by construction, the resulting logistic STAR does not match the theoretical definition of dynamic asymmetry. In other words, the available models allow the econometrician, at the best, to answer to the question: *Does the series return to its original regime and when?* Here, our objective is to answer to another, more challenging question: *Is the rate of change (if any) in the left tail of the logistic transition different with respect of the right tail and how much?* As we will show, an appropriate solution to this methodological question, *per se* interesting for descriptive aims, improves the forecasting ability of STAR models family.

The econometric literature provides two strategies: the first, proposed by Sollis et al. (1999) (SLN1) and Lundbergh & Teräsvirta (2006) (LT) is to raise the STAR’s transition function to an exponent using an idea by Nelder (1961); the second, suggested by Sollis et al. (2002) (SLN2) is to add a parameter inside the transition function in
such a way to control for the asymmetry of both the tails of the transition function by simply using a Heaviside indicator.

Unfortunately, both of these solutions present some criticality: Figure 1, panel (a) clearly shows that in the SLN2 case, the transition function could be non-smooth; on the other hand, the SLN1 and LT parametrization, plotted in panel (b) conveys a smooth transition, but the effect of increasing of the asymmetry parameter could translate just in a shift effect, if not properly restricted as stated in the same article; moreover, this parametrization does not provide an immediate description of the behavior of each tail of the transition function (which is instead the beauty of SLN2). Thus, the detection and assessment of the dynamic asymmetry in a statistically well-specified time series model seem still an open issue.

In the next Section 2 we contribute to this strand of literature by applying to the autoregressive (AR) model a generalized version of the logistic transition function with two parameters governing the the two tails of the logistic sigmoid and a logarithmic/exponential rescaling able to preserve the smoothness of the transition without requiring any restriction in the parameters. The resulting Generalized STAR (GSTAR) model encloses the symmetric STAR, so we modify the general-to-specific modeling procedure following Granger & Teräsvirta (1993) (GT); this is done in Section 3, where estimation and forecasting methods are also discussed. Two different LM-type tests for the null hypothesis that the two tails of the transition function are reflexively symmetric - a situation which is called dynamic symmetry for what follows - are built-up in Section 4. Section 5 modifies three diagnostic tests originally introduced by Eitrheim & Teräsvirta (1996) (ET). Section 6 provides a simulation study according to which the SLT-type test seems less restrictive than the Score test. Four different case studies on U.S. industrial production and unemployment rate, International Sunspot Number and Canadian Lynx data are illustrated in Section 7, jointly with a rolling forecasting exercise where both point and density forecasting evaluation are investigated: in all these examples, the dynamic asymmetry is found...
to be a non negligible feature to deal with. Finally, Section 8 discusses the relevance of such a result and concludes.

2 The Model

**Definition 1.** Let be \( y_t \) a realization of a time series observed at \( t = 1 - p, 1 - (p - 1), \ldots, -1, 0, 1, \ldots, T - 1, T \). Then the univariate process \( \{y_t\}^T_t \) follows a GSTAR(p) model if

\[
y_t = \phi'z_t + \theta'z_tG(\gamma, h(c, s_t)) + \epsilon_t, \quad \epsilon_t \sim I.I.D.(0, \sigma^2),
\]

\[
G(\gamma, h(c, s_t)) = \left( 1 + \exp \left\{ -K \prod_{k=1}^{\gamma} h(c, s_t) \right\} \right)^{-1},
\]

\[
h(c, s_t) = \begin{cases} 
\gamma_1^{-1}\exp(\gamma_1 |s_t - c_k| - 1) & \text{if } \gamma_1 > 0, \\
 s_t - c_k & \text{if } \gamma_1 = 0, \\
 -\gamma_1^{-1}\log(1 - \gamma_1 |s_t - c_k|) & \text{if } \gamma_1 < 0,
\end{cases}
\]

for \((s_t - c_k) > 0\) (or, equivalently, \(h(c, s_t) > 1/2\)) and

\[
h(c, s_t) = \begin{cases} 
-\gamma_2^{-1}\exp(\gamma_2 |s_t - c_k| - 1) & \text{if } \gamma_2 > 0, \\
 s_t - c_k & \text{if } \gamma_2 = 0, \\
 \gamma_2^{-1}\log(1 - \gamma_2 |s_t - c_k|) & \text{if } \gamma_2 < 0,
\end{cases}
\]

for \((s_t - c_k) \leq 0\) (or, equivalently, \(h(c, s_t) < 1/2\)), where \( y_t \) is a dependent variable, \( z_t = (1, y_{t-1}, \ldots, y_{t-p})' \), \( \phi = (\phi_0, \phi_1, \ldots, \phi_p)' \), \( \theta = (\theta_0, \theta_1, \ldots, \theta_p)' \) are parameter vectors, the transition function \( G(\cdot, \cdot, \cdot) \) is a continuous function in the vector \( \gamma = (\gamma_1, \gamma_2) \) and in the function \( h(c, s_t) \), which is strictly increasing in the transition variable \( s_t = y_{t-d}, d > 0 \) is a delay parameter, and the \( K = \{1, 2\} \) location parameter(s) \( c_k \).

In what follows we simplify the notation by denoting the kernel of the model corresponding to the \( k \)-esim location with \( \eta_{k,t} \equiv s_t - c_k \) and by \( h(\eta_{k,t}) \) the associated
function, so that the general form of the transition function $G(\cdot)$ can be written as:

$$G(\gamma, h(\eta_{k,t})) = \left(1 + \exp\left\{-\prod_{k=1}^{K} \left[h(\eta_{k,t})I(\gamma_1 \leq 0, \gamma_2 \leq 0) + h(\eta_{k,t})I(\gamma_1 \leq 0, \gamma_2 > 0) + h(\eta_{k,t})I(\gamma_1 > 0, \gamma_2 \leq 0) + h(\eta_{k,t})I(\gamma_1 > 0, \gamma_2 > 0)\right]\right\}\right)^{-1}. \quad (5)$$

Equation (3) (equation (4)) models the higher (lower) tail of the probability function, so allowing for the asymmetric behavior introduced by the slope parameter $\gamma_1$ ($\gamma_2$) which controls the velocity of the transition. The case in which $h(\eta_{k,t}) = \eta_{k,t}$ implies that the function nests a one-parameter symmetric logistic STAR model with slope $\gamma_1 = \gamma_2 = \gamma$. When $\gamma_1, \gamma_2 > 0$ ($\gamma_1, \gamma_2 < 0$), $h(\eta_{k,t})$ is an exponential (logarithmic) rescaling which increases more quickly (more slowly) than a standard logistic function. Model (1) can be generalized to other distributions of exponential family. The Indicator functions in (5) stress that slope parameters are not constrained, as in the classical STAR model (whereas the positiveness of the slope parameter was an identifying condition). When $\gamma \rightarrow +\infty$, both the models nests an indicator function $I(s_t > c)$, in which case the model become a (Self Exciting) Threshold Autoregression (SETAR), see Tong (1983); on the other side, they nest a straight line around 1/2 for each $s_t$ when $\gamma \rightarrow -\infty$.

The Generalized Logistic is plotted in Figure 2: the resulting sigmoid is clearly consistent with the Sichel (1993) definition of dynamic asymmetry (see, e.g., the case in which $\gamma_1 = -2$ and $\gamma_2 = 4$) and maintains the global slope of the transition function unchanged with respect to the traditional LSTAR one, so that no additional identification restriction is needed with respect to the traditional STAR model.

Remark 1. The model described in this section is the time series variant of the original generalized logistic model proposed by Stukel (1988), which differs for the definition of $z = (x_1, \ldots, x_N)'$, for $\{x_i\}_{i=1}^{N}$ being N exogenous regressors, and conse-
Remark 2. The GLSTAR model described by equation (1)-(4) nests a linear AR model for $\gamma = 0$ if $h(\eta_t)$ is modified as follows:

$$h(\eta_t)^{EZC} = \begin{cases} \gamma_1^{-1} \exp(\gamma_1 |\eta_t| - 1) & \text{if } \gamma_1 > 0, \\ 0 & \text{if } \gamma_1 = 0, \\ -\gamma_1^{-1} \log(1 - \gamma_1 |\eta_t|) & \text{if } \gamma_1 < 0, \end{cases}$$

for $\eta_t \geq 0 (\mu > 1/2)$ and

$$h(\eta_t)^{EZC} = \begin{cases} -\gamma_2^{-1} \exp(\gamma_2 |\eta_t| - 1) & \text{if } \gamma_2 > 0, \\ 0 & \text{if } \gamma_2 = 0, \\ \gamma_2^{-1} \log(1 - \gamma_2 |\eta_t|) & \text{if } \gamma_2 < 0, \end{cases}$$

for $\eta_t < 0 (\mu < 1/2)$. The label "EZC" distinguishes this version from the original Stukel’s generalized logistic function for exposition matter. This special case is necessary in order to build a test for the null of linearity against of dynamic asymmetry, see next Section 4.

Remark 3. As in the traditional STAR, the process $\{\epsilon_t\}_t$ is assumed to be a martingale difference sequence with respect to the history of the time series up to time $t - 1$, denoted as $\Omega_{t-1} = [y_{t-1}, \ldots, y_{t-p}]$, i.e., $E[\epsilon_t|\Omega_{t-1}] = 0$. This is sufficient to build up tests based on artificial regressions as demonstrated in Davidson & McKinnon (1990) and has important consequence for applied aims, in what the "All-in-One" test discussed in Section 4 and the three diagnostic tests discussed in Section 5 can still be meaningful if the normality test reject this hypothesis. For expositional purposes, we restrict the conditional variance of the process $\{\epsilon_t\}_t$ to be constant, $E[\epsilon_t^2|\Omega_{t-1}] = \sigma^2$. Moreover the parameter vectors $\phi$ and $\theta$ are assumed to not change in time and the number of regimes is assumed to not exceed $K = 2$. However, these restriction can be relaxed¹ and tested as discussed in Section 5.

¹For advancements in this sense see Gonzáles-Rivera (1998); Lundbergh et al. (2003); McAleer
Remark 4. As in the traditional STAR, if process is characterized by $G(0, h(\eta_t)^{EZC})$, we assume $Q(z) = z^p - \phi_1 z^{p-1} - \cdots - \phi_p = 0$ has its roots inside the unit circle, since this implies that the model is stationary and ergodic under the null hypothesis of linearity. Also this assumption can be relaxed, as in Kapetanios et al. (2003); Vougas (2006).

We now discuss three relevant cases of GSTAR model.

Example 1. If $K = 1$, the parameters $\phi + \theta G(\gamma, c, s_t)$ change monotonically as a function of $s_t$ from $\phi$ to $\phi + \theta$. The corresponding transition function is:

$$G(\gamma, h(\eta_t)) = \left(1 + \exp \left\{ -\left[ h(\eta_{1,t})I(\gamma_1 \leq 0, \gamma_2 \leq 0) + h(\eta_{1,t})I(\gamma_1 \leq 0, \gamma_2 > 0) + h(\eta_{1,t})I(\gamma_1 > 0, \gamma_2 \leq 0) + h(\eta_{1,t})I(\gamma_1 > 0, \gamma_2 > 0) \right]\right\} \right)^{-1},$$

(8)

with $h(\eta_{1,t})$ corresponding to (3) and (4).

Example 2. When $K = 2$ and $c_1 \neq c_2 = c$, the model (1) nests the following STAR model with second order Generalized Logistic (GLSTAR2) function:

$$G(\gamma, h(\eta_t)) = 1 - \exp \left\{ -h(\eta_{2,t}) \right\},$$

(9)

where:

$$h(\eta_{2,t}) = \begin{cases} \gamma_1^{-1} \exp(\gamma_1 |(s_t - c_1)(s_t - c_2)| - 1) & \text{if } \gamma_1 > 0, \\ (s_t - c_1)(s_t - c_2) & \text{if } \gamma_1 = 0, \\ -\gamma_1^{-1} \log(1 - \gamma_1 |(s_t - c)(s_t - c_2)|) & \text{if } \gamma_1 < 0, \end{cases}$$

(10)

for $(s_t - c)^2 > 0$ (or, equivalently, $h(\eta_t) > 1/2$) and

$$h(\eta_{2,t}) = \begin{cases} -\gamma_2^{-1} \exp(\gamma_2 |(s_t - c)(s_t - c_2)| - 1) & \text{if } \gamma_2 > 0, \\ (s_t - c)(s_t - c_2) & \text{if } \gamma_2 = 0, \\ \gamma_2^{-1} \log(1 - \gamma_2 |(s_t - c_1)(s_t - c_2)|^2) & \text{if } \gamma_2 < 0, \end{cases}$$

(11)

& Medeiros (2008); Amado & Teräsvirta (2013); Silvennoinen & Teräsvirta (2013).
for \((s_t - c_1)(s_t - c_2) < 0\) (or, equivalently, \(h(\eta_{2,t}) < 1/2\)), where \(\eta_t \equiv \eta_t = (s_t - c_1)(s_t - c_2)\). Figure 3 shows the transition function for a set of different combinations of \(\gamma_1\) for fixed \(\gamma_2\) (upper panel) and vice versa (lower panel).

**Example 3.** A particular case of GLSTAR2 holds when \(K = 2\) and \(c_1 = c_2 = c\), in which case the model (1) nests an exponential generalized exponential autoregressive (GESTAR) model, which is defined as in (9) - (11), apart the fact that \(h(\eta_{2,t}) = (s_t - c)^2\) if \(\gamma_1 = 0\) for \((s_t - c)^2 > 0\) and \(\gamma_2 = 0\) for \((s_t - c)^2 \leq 0\). In this case, the parameters \(\phi + \theta G(\cdot)\) change asymmetrically at some (undefined) point where the function reaches its own minimum.

A simulated example of GLSTAR model (in both Stukel’s and EZC’s versions), jointly with its symmetric Teräsvirta (1994) counterpart, is shown in Figure 4. For each of these three models, we used two different specifications, which differ for the location parameter \(c\). As easy seen in panel (a), the Stukel and EZC model coincides; the associated transition functions versus time plotted in panel (b) and versus ordered \(\eta_t\) in panel (c) confirm this finding; on the other hand, the plot of \(h(\eta_t)\) versus ordered \(\eta_t\) in panel (d) is more informative with respect to the effect of the different kind of asymmetry in the process: \(h(\eta_t)^{EZC}\) is a 45° angle straight line, while the rescaling effect is visible in the Stukel’s \(h(\eta_t)\) parametrization.

## 3 Econometric Modelling and Methods

### 3.1 Modelling strategy

According to GT, the investigator should always be interested in testing whether a linear AR(p) representation is adequate when building a GSTAR model. If the answer is negative, then the second step will be the selection of a nonlinear symmetric model. Then, the issue of testing for dynamic symmetry hypothesis arises as further step, when finding good specifications of STAR models becomes too difficult.
or whenever suggested by the economic theory. The resulting General-to-Specific modelling strategy consists in the following 7 steps:

1. Specify a linear autoregressive model.

2. Test linearity for different values of $d$, and if rejected, determining $d$ in (2) or (9).

3. Choose between LSTAR, LSTAR2 or ESTAR by the Teräsvirta’s rule.

4. Test the symmetry of the tails transition function according to the result in Step 3.

5. If the hypothesis of symmetry is rejected, estimate the GSTAR model with the most appropriate transition function given by step 3.

6. Evaluate the new parametrization by some diagnostic tests.

7. Use the estimated GSTAR model for forecasting aims.

The autoregressive order $p$ is selected according to Bayesian Information Criterion (Schwarz, 1978), which is combined with the result with a portmanteau test for serial correlation in order to avoid a wrong rejection of symmetry hypothesis. This is due to the fact that the GSTAR model requires a lower autoregressive order with respect to its symmetric counterpart.

For what concerns the Step 4, the dynamic symmetry hypothesis is tested by two different LM-type tests. In the first test the series is assumed to follow a STAR model, so that testing for symmetry is a second step with respect to testing for linearity. Hence, we will refer to this test as "Two-Step test". In the second test we do not assume any prior of the nonlinearity of the series, so that it enclose all steps from 2) to 5) of the General-to-Specific modelling strategy above mentioned; hence, the use of the label "All-in-one" to distinguish it from the different null hypothesis of "Two-Step" test. The choice of what test to use depends on the
needs of the investigator. Our experience suggests to perform the "All-in-One" test should be used if the investigator wants to be conservative against evidence of asymmetric dynamics, while the "Two-Step" tends to not reject the null unless extreme situations (see Section 7 for details). Both the tests will be discussed in the next Section 4.

The choice of the delay parameter $d$ and the choice of the transition function can be done with the same procedure adopted in Tsay (1989) and Teräsvirta (1994).

3.2 Estimation

Following Leybourne et al. (1998), estimation is done by concentrating the Sum of Square Residuals function with respect to $\theta$ and $\phi$, that is minimizing:

$$SSR = \sum_{t=1}^{T} \left( y_t - \hat{\psi}' x_t' \right)^2,$$

(12)

where:

$$\hat{\psi} = [\hat{\phi}, \hat{\theta}] = \left( \sum_{t=1}^{T} x_t'(\gamma, c)x_t(\gamma, c) \right)^{-1} \left( \sum_{t=1}^{T} x_t'(\gamma, c)y_t \right),$$

(13)

and

$$x_t(\hat{\gamma}, \hat{c}) = \left[ z, z'_t G(\hat{\gamma}, h(\hat{c}, s_t)) \right].$$

(14)

This is possible because if $\gamma$ and $c$ are known and fixed, the GSTAR model is linear in $\theta$ and $\phi$, which can be easy computed via conditional OLS. In a such a way, the nonlinear least square minimization problem, otherwise necessary, more demanding in terms of parameters to estimate and not available in closed-form, reduces to a minimization on three (four) parameters, and is solved via a grid search over $\gamma_1, \gamma_2, c$ ($c_1, c_2$ in case of GLSTAR2).

In our applications, both $\gamma_1$ and $\gamma_2$ are chosen between a minimum value of -10 and a maximum of 10 with rate 0.5 in the first three examples (-150 and 150 with rate

\footnote{Our simulation study shows that the two tests behave differently in terms of empirical power. See Section 6 for details.}
15 in the fourth one); the grid for parameter \(c_1\) (\(c_2\)) is the set of values computed between the 10\(^{th}\) and 90\(^{th}\) percentile of \(s_t\) with rate computed as the difference of the two and divided for an arbitrarily high number (here, 200).

### 3.3 Forecasting

The literature on point forecasting and on evaluation of individual density forecasts under linear models is nowadays established, see Timmermann (2006) and Corradi & Swanson (2006). When the model is nonlinear, the one-step forecast is immediately available if knowing the nonlinear function in what, by least-square criterion, \(E(\epsilon_{t+1}|I_t) = 0, I_t = y_{t-i}, i \geq 1\) in (1). The multi-step ahead forecast is not available in closed form and requires numerical integration. Hence at \(t+1\), we generate, \(1, \ldots, m, \ldots, M\) draws conditionally on the estimated parameters and obtain the forecast \(y_{t+1} \sim f(y_{t+1}|I_t);\) in turn, this is collected to draw, at \(t + 2\), the forecast \(y_{t+2} \sim f(y_{t + 2}|I_t, y_{t+1}^{(m)}),\) and so on until, at \(t + h,\) the forecast \(y_{t+h|t} = f(t + h|I_t, y_{t+1}^{(m)}, \ldots, y_{t+h-1}^{(m)})\) is obtained and then evaluated as:

\[
\hat{y}_{t+h} = \frac{1}{M} \sum_{m=1}^{M} y_{t+h|t}^{(m)} \tag{15}
\]

The out-of-sample predictive properties of the estimated models are investigated via rolling forecast experiment, according to which the series \(y_t\) is divided in a "pre-forecast" period (going from time \(\{1 \ldots t\}\)) from which the model is estimated and the \(h\)-step-ahead forecast are computed and compared with the "test" period, going from time \(\{T^* \ldots T\}\) where \(T^* = t + h;\) this allows to measure \(T - T^* - h + 1\) out-of-sample forecasts. Let denote the corresponding realization of the series as \(y_t, y_T^*\) and \(y_T,\) as well as the corresponding density forecasts as \(f_t, f_T^*\) and \(f_T.\) Since our interest lies in short-run forecasting we consider \(h = \{1, 3, 6, 12\}.\) The point predictive performances of the model \(j\) are investigated by four different measures: the mean forecast error (MFE), the symmetric mean absolute percentage error (sMAPE), the
median relative absolute error (mRAE) and the root mean square forecast error (RMSFE)\textsuperscript{3}:

\[ MFE_{j,h} = \frac{1}{T-h-T^*} + \sum_{t=T^*}^{T-h} \left( y_{t+h} - \hat{y}_{t+h|t}^j \right) \] \quad (16)

\[ sMAPE_{j,h} = \frac{100|y_{t+h} - \hat{y}_{t+h}^j|}{0.5(y_{t+h} - \hat{y}_{t+h|t})} \] \quad (17)

\[ mRAE_{j,h} = \frac{|y_{t+h} - \hat{y}_{t+h}^j|}{|y_{t+h} - \hat{y}_{t+h|t}|} \quad \text{with (1) indexing the benchmark model; (18)} \]

\[ RMSFE_{j,h} = \frac{1}{T-h-T^*} + \sum_{t=T^*}^{T-h} \left( y_{t+h} - \hat{y}_{t+h|t}^j \right)^2 \] \quad (19)

For completeness of analysis, we check the statistical superiority of the dynamic asymmetric forecasts with respect to other specifications via the Giacomini & White (2006) test.

Differently, literature on aggregation of density forecasts is instead in a development phase, and focuses on the so called scoring rules (or opinion pools), peculiar functions enabling the forecaster to properly aggregate the set of conditional predictive density as well as more common measures as Mean Square Forecast Error et similia do for point forecasts. Despite their dated origins in statistics, as documented by Gneiting & Raftery (2007), scoring rules are becoming increasingly applied by contemporaneous econometric literature only recently; see, inter alia, Geweke & Amisano (2011). In a similar fashion, concerning about density forecasting, four different scoring rules are used for aggregate the \( T - T^* - h + 1 \) predictive densities produced by the same forecasting exercise\textsuperscript{4}:

\textsuperscript{3}In particular, sMAPE and mRAE are recommended when the series is known to present volatility effects or skewness, two features often associated to nonlinearity; see the discussion in Tashman (2000).

\textsuperscript{4}The scoring rules here considered are just a fraction of the many nowadays available. The choice of the scores has been done for easy of treatment and does not imply any preference.
• the logarithmic score (LogS):

\[
LogS_{j,h} = \frac{1}{T-h-T_s+1} \sum_{t=T_s}^{T-h} \log \hat{f}_{t+h|t}^j
\]  

(20)

corresponding to a Kullback-Liebler distance from the true density; models with higher LogS are preferred.

• The quadratic score, somehow the equivalent of the MSFE in point forecasting, is defined as:

\[
QSR_{j,h} = \frac{1}{T-h-T_s+1} \sum_{t=T_s}^{T-h} K \sum_{k=i}^{K} (f_{t+h|t}^j - d_{kt})^2
\]  

(21)

where \(d_{kt} = 1\) if \(k = t\) and 0 otherwise; models with lower QSR are preferred.

• The (aggregate) continuous-ranked probability (CRPS) score, equivalent to the sMAPE, is defined as:

\[
CRPS_{j,h} = \frac{1}{T-h-T_s+1} \sum_{t=T_s}^{T-h} \left( |f_{t+h} - \hat{f}_{t+h|t}^j| - 0.5 |f_{t+h} - f'_{t+h}| \right),
\]  

(22)

where \(f\) and \(f'\) are independent random draws from the predictive density and \(f_{t+h|t}\) the observed value; models with lower CRPS are preferred.

• Finally, the quantile score (qS) can be obtained if \(f_{t+h|t}^j\) is replaced in (20) by a predictive \(\alpha\)-level quantile \(q_{\alpha_{t+h|t}}\) (and the logarithmic function removed); this score is used in risk analysis because provide information about deviations from the true tail of the distribution.

4 Testing for Dynamic Symmetry

In this section we discuss two LM-type tests for the null of dynamic symmetry according to the General-to-Specific strategy stated in the previous section 3. The
"Two-Step" test, illustrated in Subsection 4.1, is a classical Score test on the two slope parameters and is a simple adaptation for time series of the original Stukel (1988)'s parametrization. On the other side, the "All-in-One" test, derived in Subsection 4.2, is modified version of the Taylor-expansion-based test by Luukkonen et al. (1988) (LST), where the $G(\gamma, h(\eta^{EZC}))$ is linearized by third-order Taylor expansion in $G(\cdot)$; this leads to an augmented artificial model which in turn can be investigated by a classical $\chi^2$ or $F$-test. This is due to the fact that the Information matrix is the same as in GT$^5$.

4.1 "Two-Step" Test

Consider the general formulation (1)-(2). Then, the null hypothesis of no logarithmic (exponential) deviations from the logistic transition in systems (1)-(8) or (1)-(11) can be tested by setting the following hypotheses testing system:

$$H_{0i} : (\gamma_1, \gamma_2) = (0, 0) \text{ vs } H_{1i} : (\gamma_1, \gamma_2) \neq (0, 0), \ i = 1, 2, 3 \ , \ (23)$$

with subscript $i$ indicating the type of underlining transition function, namely $i = 1$ for generalized logistic (eq. (5)), $i = 2$ for generalized second order logistic (eq. (9)) and $i = 3$ for the generalized exponential one.

This hypothesis system requires a simple score test. Let denote by $\Xi = [\phi, \theta, \gamma, c]$ the hyper-parameter vector of the model, so that the log-likelihood function of the $T$ observations can be denoted by $L_t(z_t, \Xi)$ and the score vector by $q_t(z_t, \Xi) = \sum_t q(z_t, \Xi) = \partial L_t(z_t, \Xi)/\partial \Xi$ evaluated at $(\theta_0, \phi_0, 0, c_0)$. Then, standard results

$^5$See GT, pp. 64-5, adjust the notation for an autoregressive framework and notice that we only modify the definition of nonlinear part $f_t = f(w_t; \psi)$, which does not vary the general result.
lead to the following log-likelihood function:

\[
L_t(z_t, \Xi) = \text{const} + \frac{T}{2} \ln \sigma^2 - \frac{1}{2} \sigma^2 \sum_t (y_t - \phi'z_t - \theta'G_t)^2 \\
= \text{const} + \frac{T}{2} \ln \sigma^2 - \frac{1}{2} \sigma^2 \sum_t u_t^2(\Xi),
\]  

(24)

with \(\text{const}\) denoting a constant and \(u_t\) the model’s residual, and to the score:

\[
q_t(z_t, \Xi) = \sum_t q(z_t, \Xi) = \frac{\partial L_t(z_t, \Xi)}{\partial \Xi} = \frac{1}{\sigma^2} \sum_t w_t(\Xi)k_t,
\]  

(25)

where

\[
k_t = \frac{\partial u_t(\Xi)}{\partial \Xi} = (z_t, z_tG, \theta'z_tG_{\gamma_1}, \theta'z_tG_{\gamma_2}, \theta'z_tG_{t,c}),
\]  

(26)

with \(G_{t,\gamma_1} = \partial G/\partial \gamma_1, G_{t,\gamma_2} = \partial G/\partial \gamma_2,\) and \(G_c = \partial G/\partial \gamma_c\) being defined in Appendix A.1.

Under \(H_0\), the test statistic is:

\[
S_1(\Xi)_{LM} = \frac{1}{\sigma^2} \hat{u}'H'H - H'Z(Z'Z)^{-1}Z'H^{-1}H'\hat{u},
\]  

(27)

where \(\hat{u} = [\hat{u}_1, \ldots, \hat{u}_T]'\), \(Z = (z_1', \ldots, z_T')'\), \(H = [(h^0_1)', \ldots, (h^0_T)']'\), with \((h^0_t)' = k_t^G\), \(k_t^G\) denoting the sub-vector \([\theta'z_tG_{\gamma_1}, \theta'z_tG_{\gamma_2}, \theta'z_tG_{t,c}]\)' and \(n = \text{dim}(k_t^G)\). Under \(H_0\), statistic \(S_1\) is asymptotically distributed as a \(\chi^2_n\). Just minor modifications are needed in notation of \(k_t\) and \(q_t\) in case of GLSTAR2 model due to an additional \(c\) parameter with respect to the GLSTAR.

### 4.2 "All-in-One" Test

Consider (2) with \(G(\gamma, h(\eta^{E2Z}))|_{\gamma=0}\) and define \(\tau = (\tau_1, \tau_2)'\), where \(\tau_1 = (\phi_0, \phi')'\), \(\tau_2 = \gamma\). Let \(\hat{\tau}_1\) the LS estimator of \(\tau_1\) under \(H_0 : \gamma = 0\), \(\hat{\tau} = (\tau_1', \theta')'\). Moreover, let \(z_t(\tau) = \frac{\partial q_t}{\partial \tau}\) and \(\tilde{z}_t = z_t(\hat{\tau}) = (\tilde{z}_{1,t}, \tilde{z}_{2,t})\), where the partition conforms to that of
Then the general form of LM statistic is:

\[
S_2(\Xi)^{LM} = \frac{1}{\hat{\sigma}^2} \hat{u}' \hat{Z}_2 (\hat{Z}_2' \hat{Z}_2 - \hat{Z}_1' \hat{Z}_1) \hat{Z}_1' \hat{Z}_2 \hat{u},
\]

(28)

where \( \hat{u} \) is previously defined, \( \hat{\sigma}^2 = \frac{1}{T} \sum_1^T \hat{u}_t^2 \) and \( \hat{u}_t = y_t - \hat{\tau}' \hat{z}_t \), \( \hat{\tau} = (\hat{\tau}_{i1}, \ldots, \hat{\tau}_{iT})' \), \( i = \{1, 2\} \), \( t = 1, \ldots, T \). When the model is an GLSTAR, \( \hat{z}_{1,t} = -z_t = -(1, y_{t-1}, \ldots, y_{t-p})' \) while \( \hat{z}_{2t} \equiv \frac{\partial^2 y_t}{\partial \gamma \partial \gamma'} \bigg|_{\gamma=0} = -\frac{1}{2} \{ \theta_20[y_t(y_{t-d})] - cy_t \theta' z_t + \theta' z_t y_{t-d} \} \), where \( d \) is the delay parameter. The change in the definition of \( z_{2t} \) is not significant in terms of LM statistic build-up. This implies that no change of treatment with respect to the original parametrization is needed. In particular, in order to circumvent the Davies (1977)’s problem of unidentification of nuisance parameters \( \theta_0 \) and \( \hat{\theta} = [\theta_1, \ldots, \theta_p]' \) under the null hypothesis, the same LST approach can be used. The linearized GLSTAR model

\[
y_t = \phi' z_t + \theta' z_t T_3 \left[ h(\eta_{k,t}) I_{(\gamma_1 \leq 0, \gamma_2 \leq 0)} + h(\eta_{k,t}) I_{(\gamma_1 \leq 0, \gamma_2 > 0)} + h(\eta_{k,t}) I_{(\gamma_1 > 0, \gamma_2 \leq 0)} + h(\eta_{k,t}) I_{(\gamma_1 > 0, \gamma_2 > 0)} \right] + \epsilon_t,
\]

(29)

leads to the following auxiliary regression for testing linearity and symmetry:

\[
\hat{u}_t = \hat{z}_{11}' \hat{\beta}_1 + \sum_{j=1}^p \beta_{2j} y_{t-j} y_{t-d} + \sum_{j=1}^p \beta_{3j} y_{t-j} y_{t-d}^2 + \sum_{j=1}^p \beta_{4j} y_{t-j} y_{t-d}^3 + v_t,
\]

(30)

where \( v_t \) is a \( N.I.D.(0, \sigma^2) \) process, \( \hat{\beta}_1 = (\beta_{10}, \beta'_1)' \), \( \beta_{10} = \phi_0 - (c/4) \theta_0 \), \( \beta_1 = \phi - (c/4) \theta + (1/4) \theta_0 e_d \), \( e_d = (0, 0, \ldots, 0, 1, 0, \ldots, 0)' \) with the \( d \)-th element equal to unit and \( T_3(G) = f_1 G + f_3 G^3 \) is the third-order Taylor expansion of \( G(\Xi) \) at \( \gamma = 0 \), \( f_1 = \partial G(\Xi)/\partial \Xi \bigg|_{\gamma=0} \) and \( f_3 = (1/6) \partial^3 G(\Xi)/\partial \Xi \bigg|_{\gamma=0} \), \( G(\Xi) \) being defined in previous section. The null hypothesis is

\[
H_0 : \beta_{2j} = \beta_{3j} = \beta_{4j} = 0 \quad j = 1, \ldots, p,
\]

(31)
The test statistic:

\[ LM_1 = (SSR_0 - SSR)/\hat{\sigma}_v^2, \]  

with \( SSR_0 \) and \( SSR \) denoting the sum of squared estimated residuals from the estimated auxiliary regression \( (30) \) and under the null and alternative, respectively and \( \sigma_v^2 = (1/T)SSR \), has an asymptotic \( \chi^2_3p \) distribution under \( H_0 \).

If the model is an GESTAR(p), then \( \hat{z}_1 = -z_t \) as in the generalized logistic case, while \( \hat{z}_{2,t} = -2\theta_2'z_{\hat{}}_ty_{\hat{}}_t^2 - 2\theta_2'y_{\hat{}}_t - 2c\theta_2'w_{\hat{}}_ty_{\hat{}}_t + 4\theta_2'z_{\hat{}}_ty_{\hat{}}_t^2 - 2c\theta_{20} = 2\hat{z}_{2,t}^{ESTAR} \), where \( ESTAR \) denotes the vector \( \hat{z}_{2,t} \) for the ESTAR model. That is, the vector \( \hat{z}_{2,t} \) of the generalized ESTAR model is found to be two times the symmetric one. The corresponding auxiliary regression is

\[ \hat{\epsilon}_t = \hat{\beta}_1'\hat{z}_1 + \beta_2'y_{\hat{}}_t - d + \beta_3'y_{\hat{}}_t^2 + v'_t, \]  

where \( v'_t \) is a \( N.I.D.(0, \sigma^2) \) error term and \( \hat{\beta}_1 = (\beta_{10}, \beta_1')' \), with \( \beta_{10} = \phi_0 - c^2\theta_0 \) and \( \beta_1 = \phi - c^2\theta + 2c\theta_0e_d \); moreover \( \beta_2 = 2c\theta - \theta_0e_d \) and \( \beta_3 = -\theta \). Thus the null hypothesis of linearity is

\[ H'_0: \beta_2 = \beta_3 = 0, \]  

which can be tested by the test statistic:

\[ LM_2 = (SSR_0 - SSR)/\hat{\sigma}_{v1}^2, \]  

where \( SSR_0 \) and \( SSR \) are the sum of squared residuals from \( (33) \) under the null and the alternative respectively, \( \hat{\sigma}_{v1}^2 = (1/T)SSR \). When the null is true, the statistic \( (35) \) is asymptotically \( \chi^2_p \) distributed.
5 Diagnostics Tests

For what concerns the diagnostics, the new parametrization can be applied directly to the three tests developed by ET, which will be discussed in detail.

5.1 Serial independence

Consider the general additive model (1), where:

\[ \epsilon_t = a' v_t + u_t = \sum_{j=1}^{q} a_j L^j \epsilon_t + u_t, \quad u_t \sim I.I.D.(0, \sigma^2), \quad (36) \]

with \( L^j \) denoting the lag operator, \( v_t = (u_{t-1}, \ldots, u_{t-q})' \), \( a = (a_1, \ldots, a_q)' \), \( a_q \neq 0 \). Under the assumption of stationarity and ergodicity (see Section 2), the null hypothesis of serial independence is \( H_0 : a = 0 \). By pre-multiplying eq. (2) by \( 1 - \sum_{j=1}^{q} a_j L^j \) we get:

\[ y_t = \sum_{j} a_j L^j y_t + \phi' z_t - \sum_{j} a_j L^j \phi' z_t + \theta' G(z_t) - \sum_{j} a_j \theta' G(z_t) + \epsilon_t, \quad (37) \]

hence, assuming the necessary initial values \( y_0, y_{-1}, \ldots, y_{-(p+q)+1} \) fixed, the pseudo normal loglikelihood for \( t = 1, \ldots, T \) is:

\[ L_t = \text{constant} + \frac{1}{2} \ln \sigma^2 - \frac{\epsilon_t^2}{2\sigma^2}, \]

\[ \epsilon_t = y_t - \sum_{j} a_j L^j y_t - \phi' z_t + \sum_{j} a_j L^j \phi' z_t - \theta' G(z_{t-j}, \Xi) + \sum_{j} a_j \theta' G(z_{t-j}, \Xi). \quad (38) \]
Consistently with the model initial assumptions, the information matrix is block diagonal, hence we can consider \( \sigma^2 \) fixed for the rest of the derivations. So we have:

\[
\frac{\partial \mathcal{L}_t}{\partial a_j} = \frac{\epsilon_t}{\sigma^2} [y_{t-j} - \phi'z_{t-j} - \theta'G(z_{t-j}, \hat{\Xi})] \tag{39}
\]

\[
\frac{\partial \mathcal{L}_t}{\partial \Xi} = \frac{\epsilon_t}{\sigma^2} \left[ \theta'z_t \frac{\partial G(z_{t-j}, \hat{\Xi})}{\partial \Xi} - \sum_j a_j \theta' \frac{\partial G(z_{t-j}, \hat{\Xi})}{\partial \Xi} \right]. \tag{40}
\]

Under \( H_0 \), consistent estimators of (39) - (40) are:

\[
\frac{\partial \hat{\mathcal{L}}_t}{\partial a_t} \bigg|_{H_0} = \frac{1}{\sigma^2} \hat{u}_t \hat{v}_t \quad \frac{\partial \hat{\mathcal{L}}_t}{\partial \hat{\Xi}_t} \bigg|_{H_0} = -\frac{1}{\sigma^2} \hat{u}_t \hat{z}_t, \tag{41}
\]

where \( \hat{u}_t = (\hat{v}_{t-1}, \ldots, \hat{v}_{t-q})' \), \( \hat{v}_{t-j} = y_{t-j} - \phi'z_{t-j} - \theta'G(z_{t-j}, \hat{\Xi}), j = 1, \ldots, q \), \( \hat{\Xi} \) is the QMLE of \( \Xi \) and \( \hat{z}_t = \frac{\partial G(z_t, \hat{\Xi})}{\partial \hat{\Xi}} = k_t^G = [\theta'z_tG_{\gamma_1}, \theta'z_tG_{\gamma_2}, \theta'z_tG_{\gamma_3}] \). The resulting LM statistic is:

\[
LM = \frac{1}{\hat{\sigma}^2} \left( \hat{u}_t' \hat{v}_t \right) \left( \hat{v}_t' \hat{v}_t - \hat{v}_t' \hat{z}_t \left( \hat{z}_t \hat{z}_t' \right)^{-1} \hat{z}_t' \hat{v}_t \right)^{-1} \left( \hat{v}_t' \hat{u}_t \right), \tag{42}
\]

with \( \hat{\sigma}^2 = \frac{1}{T} \sum_t u_t^2 \). Under the null hypothesis, statistics (42) is asymptotically \( \chi^2_q \) distributed. The partial derivatives of \( G(\cdot) \) are shown in Appendix A.1. Another possibility is to use the same three-step procedure for carrying an \( F \)-test:

1. Estimate the GSTR model under the assumption of uncorrelated errors and compute the residual sum of squares \( SSR_0 = \sum_{t=1}^T \hat{u}_t^2 \).

2. Regress \( \hat{u}_t \) on \( \hat{v}_t, z_t, \hat{G}(z_{t-d}), \hat{G}_{\gamma_1}, \hat{G}_{\gamma_2}, \hat{G}_c \) (eventually \( \hat{G}_{c2} \) in case of GLSTR2) and compute SSR;

3. Compute the test statistic \( F_{LM} = \frac{SSR_0 - SSR}{q} / \frac{SSR}{T - n - q} \), where \( n = \dim(\hat{z}_t) \).

The \( F \)-statistics is preferable to the \( \chi^2 \) statistics which may suffer from size problems when the number of lags is high and time series is short, so that the estimated residuals can be non-orthogonal to the gradient vector \( \hat{z}_t \). In this case ET suggests to add an extra-step to the step (i), consisting in regressing the estimated
errors to $z_t$, $z_t \hat{G}(z_{t-1})$, $\hat{G}_{\gamma_1}$, $\hat{G}_{\gamma_2}$, $\hat{G}_c$; the resulting errors $\hat{u}_t$ is used to derive the $SSR_1 = \sum_{t=1}^T \hat{u}_t^2$.

5.2 No remaining asymmetry

As in the symmetric STAR model, we are interested to detect possible misspecification. In this case there are two plausible issues to investigate: neglected (additive) nonlinearity and, in our case, neglected asymmetry. Consider the additive GSTAR model:

$$y_t = \phi'z_t + \theta'z_t G_1(\gamma, h(\eta_t)) + \pi'z_t G_2(\gamma, h(\eta_t))^{EZW} + u_t, \quad (43)$$

with $u_t \sim I.I.D. (0, \sigma^2)$. The null of neglected asymmetry is:

$$H_0 : h(\eta_t)^{(2),EZW} = 0 \quad \text{vs} \quad H_0 : h(\eta_t)^{(2),EZW} \neq 0. \quad (44)$$

If $\gamma$ is found being not null, the investigator can easily check if the additive nonlinear part is significant. The EZC version of $h(\eta_t)$ is necessary in order to nest the discussion to ET framework. We assume that, under $H_0$, $\Xi$ can be consistently estimated by QML. Similarly to the symmetric case, it should be noticed that the model is not identified under $H_0$, so that the Taylor expansion of the $G(\cdot)$ suggested by LST can be used in order to circumvent this problem. In this case, we assume $G_2(\cdot)$ as generalized logistic and replace it with its third-order Taylor expansion about $h(\gamma)^{(2)} = 0$. This implies:

$$T_2 = g_{20} + g_{21} y_{t-1} + g_{22} y_{t-1}^2 + g_{23} y_{t-1}^3,$$

where $g_{2j}$, $j = 0, 1, 2, 3, 4$ are functions of $\gamma^{(2)}$ such that $g_{20} = g_{21} = g_{22} = g_{23} = 0$ for $\gamma^{(2)} = 0$, consistently with the definition of $h_\gamma(s_t)$. By re-parametrizing, the
model (43) became:

\[ y_t = \beta_0' z_t + \theta' z_t G_1(\cdot) + \beta_1' \tilde{z}_t y_{t-1} + \beta_2' \tilde{z}_t y_{t-2}^2 + \beta_3' \tilde{z}_t y_{t-3}^3 + r_t, \tag{46} \]

where \( \tilde{z}_t = (y_{t-1}, \ldots, y_{t-p})' \). The null hypothesis of no additive nonlinearity is \( H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \), and, as in the symmetric case, under \( H_0 \), \( r_t = u_t \). The LM statistics distributes as a \( \chi^2(3p) \). As in the symmetric case, the test preserves power also against generalized exponential transition. Since there are no modifications in the statistical assumptions concerning the errors distribution, the asymptotic theory is the same of the symmetric STAR case. The test statistic is (28) with \( \tilde{z}_t = (z_t, z_t \hat{G}(\cdot), \hat{G}_{\gamma_1}, \hat{G}_{\gamma_2}, \hat{G}_{c})' \) (or \( \hat{G}_{c1}, \hat{G}_{c2} \) in case of GLSTR2), whereas \( v_t = (\tilde{z}_t y_{t-1}, \tilde{z}_t y_{t-2}^2, \tilde{z}_t y_{t-3}^3)' \). As in the symmetric STAR model, the test is implemented with the same procedure for serial correlation, the F-test has \((3p)\) and \((T - n - 3p)\) degrees of freedom and the Teräsvirta rule can be applied to (46) in order to select the form of the transition. If this selection is not desirable, a polynomial expansion of (43) can be performed to build up an omnibus test, but in this case, a rejection of the null of no additive nonlinearity will not give any qualitative information, that is why we do not take in consideration this scenario.

### 5.3 Parameter constancy

Consider the model:

\[ y_t = \phi(t)' \tilde{z}_t + \theta(t)' \tilde{z}_t G(\gamma, h(\eta_t)) + u_t, \quad u_t \sim I.I.D. (0, \sigma^2), \tag{47} \]

with \( \tilde{z}_t \) denoting the \( k \leq p + 1 \) element of \( z_t \) for which the corresponding element of \( \phi \) is not assumed zero a priori, \( \tilde{z}_t \) is the same \((l \times 1)'\) for the element of \( \theta \). Let \( \tilde{\phi} \) and \( \tilde{\theta} \) denote the equivalent \((k + 1)\) and \((l + 1)\) parameter vectors, \( \phi(t) = \tilde{\phi} + \lambda_1 G_j(t; \gamma, h(\eta_t))^{(1)} \) and \( \theta(t) = \tilde{\theta} + \lambda_2 G_j(t; \gamma, h(\eta_t))^{(2)} \) with \( \lambda_1 \) and \( \lambda_2 \) being a \((k \times 1)\) and \((l \times 1)\) vectors respectively. Then the null of parameter constancy in
Three forms for $G_j$ can be considered:

$$G_1(t; \gamma, h(c, s_t)) = (1 + \exp(-h(\eta_t^{GL})))^{-1} \text{ with } \eta_t^{GL} \equiv t - c,$$

$$G_2(t; \gamma, h(c, s_t)) = 1 + \exp(-h(\eta_t^{GE})) \text{ with } \eta_t^{GE} \equiv (t - c)^2,$$

$$G_3(t; \gamma, h(c, s_t)) = (1 + \exp(-h(\eta_t^C)))^{-1} \text{ with } \eta_t^C \equiv (t^3 + c_{12}t^2 + c_{11}t + c_{10}).$$

The null of parameter constancy is $H_0 : \gamma = 0$. Notice that in this case the model is identified also in case of $\gamma < 0$, so that the only identifying restriction is that $\gamma \neq 0$. $G_1$ and $G_2$ are the Generalized Logistic and Exponential smooth transition of the change in parameters, while $G_3$ is a cubic function which allows for both monotonically and non-monotonically changing parameters and can be seen as a general case of $G_1$ and $G_2$ when building up a test. As suggested by the literature, we use a third-order Taylor expansion of $G_3$ about $\gamma = 0$:

$$T_3(t; \gamma, h(\eta_t)) = \frac{1}{4} h(\gamma)(t^3 + c_{12}t^2 + c_{11}t + c_{10}) + R(t, \gamma, h(\eta_t)).$$

in order to approximate $\phi(t)$ and $\theta(t)$ in (47) using (50). This yields to:

$$y_t = \beta_0'(\bar{z}_t) + \beta_1'(t\bar{z}_t) + \beta_2'(t^2\bar{z}_t) + \beta_3'(t^3\bar{z}_t) +$$

$$+ \{ \beta_4'(\bar{z}_t) + \beta_5'(t\bar{z}_t) + \beta_6'(t^2\bar{z}_t) + \beta_7'(t^3\bar{z}_t) \} G(t; \gamma, h(\eta_t)) + r_t^*,$$

where $r_t^* = u_t + R(t; \gamma, h(\eta_t))$. Under $H_0$, $r_t^* = u_t$. In (51), $\beta_j = h(\eta_t)\tilde{\beta}$, $j = 1, \ldots, 7$, hence the null hypothesis in terms of (51) becomes $H_0 : \beta_j = 0$, $j = 1, \ldots, 7$. Consequently, the locally approximated pseudo normal log-likelihood under
\( H_0 \) (ignoring \( R \)) is

\[ \mathcal{L}_t = \text{const} - \frac{1}{2} \ln \sigma^2 - \frac{1}{2} \sigma^2 [y_t - \beta'_0 \mathbf{w}_t - \beta'_1 (t\bar{\mathbf{w}}_t) - \beta'_2 (t^2\bar{\mathbf{w}}_t) - \beta'_3 (t^3\bar{\mathbf{w}}_t) - \{ \beta'_4 (\bar{\mathbf{w}}_t) + \beta'_5 (t\bar{\mathbf{w}}_t) + \beta'_6 (t^2\bar{\mathbf{w}}_t) + \beta'_7 (t^3\bar{\mathbf{w}}_t) \} G(y_{t-d}; \gamma, h(\eta_t))^2] \]  

(52)

The partial derivatives are:

\[
\begin{align*}
\frac{\partial \mathcal{L}_t}{\partial \beta_j} &= \frac{1}{\sigma^2} u_t (t^j\bar{\mathbf{w}}_t), \quad j=0, \ldots, 3, \\
\frac{\partial \mathcal{L}_t}{\partial \beta_j} &= \frac{1}{\sigma^2} u_t (t^j\bar{\mathbf{w}}_t) G(y_{t-d}; \gamma, h(\eta_t)), \quad j=4, \ldots, 7, \\
\frac{\partial \mathcal{L}_t}{\partial \gamma_1} &= \frac{1}{\sigma^2} u_t \{ \beta'_4 (\bar{\mathbf{w}}_t) + \beta'_5 (t\bar{\mathbf{w}}_t) + \beta'_6 (t^2\bar{\mathbf{w}}_t) + \beta'_7 (t^3\bar{\mathbf{w}}_t) \} G_{\gamma_1}, \\
\frac{\partial \mathcal{L}_t}{\partial \gamma_2} &= \frac{1}{\sigma^2} u_t \{ \beta'_4 (\bar{\mathbf{w}}_t) + \beta'_5 (t\bar{\mathbf{w}}_t) + \beta'_6 (t^2\bar{\mathbf{w}}_t) + \beta'_7 (t^3\bar{\mathbf{w}}_t) \} G_{\gamma_2}, \\
\frac{\partial \mathcal{L}_t}{\partial c} &= \frac{1}{\sigma^2} u_t \{ \beta'_4 (\bar{\mathbf{w}}_t) + \beta'_5 (t\bar{\mathbf{w}}_t) + \beta'_6 (t^2\bar{\mathbf{w}}_t) + \beta'_7 (t^3\bar{\mathbf{w}}_t) \} G_c,
\end{align*}
\]

(53) \hspace{1cm} (54) \hspace{1cm} (55) \hspace{1cm} (56) \hspace{1cm} (57)

where \( G_{\gamma_1}, G_{\gamma_2}, G_c \) are the derivatives of \( G(y_{t-d}, \gamma, h(\eta_t)) \) with respect to \( \gamma_1, \gamma_2 \) and \( c \). With this notation, the estimators of \( \frac{\partial \hat{G}_{\gamma_1}}{\partial \gamma_1}, \frac{\partial \hat{G}_{\gamma_2}}{\partial \gamma_2} \) and \( \frac{\partial \hat{G}_c}{\partial c} \) are \( \frac{\partial \hat{G}_{\gamma_1}}{\partial \gamma_1} = \frac{1}{\sigma^2} u_t \hat{G}_{\gamma_1}, \frac{\partial \hat{G}_{\gamma_2}}{\partial \gamma_2} = \frac{1}{\sigma^2} u_t \hat{G}_{\gamma_2}, \frac{\partial \hat{G}_c}{\partial c} = \frac{1}{\sigma^2} u_t \hat{G}_c \) respectively, so that: \( \hat{\mathbf{z}}_t = (1, \hat{\mathbf{z}}'_1, \hat{\mathbf{z}}'_2, \hat{\mathbf{z}}'_3, \hat{\mathbf{z}}'_4, \hat{\mathbf{z}}'_5, \hat{\mathbf{z}}'_6, \hat{\mathbf{z}}'_7, \hat{\mathbf{z}}'_8) \) and \( \hat{\mathbf{u}}_t = (t\hat{\mathbf{z}}'_1, t^2\hat{\mathbf{z}}'_1, t^3\hat{\mathbf{z}}'_1, t\hat{\mathbf{z}}'_2, t^2\hat{\mathbf{z}}'_2, t^3\hat{\mathbf{z}}'_2, t\hat{\mathbf{z}}'_3, t^2\hat{\mathbf{z}}'_3, t^3\hat{\mathbf{z}}'_3) \). Like in the symmetric scenario, under \( H_0 \), the statistic (42) has a \( \chi^2 \) distribution with \( 3(k+l) \) degrees of freedom and the equivalent \( F \)-distribution has \( 3(k+l) \) and \( T-4(k+l)-2 \) degrees of freedom (the statistic is denoted \( LM_3 \)). The following rule is used: if \( H_1 \) is (47) with transition function \( G_3 \), then (42) is based on (51) assuming \( \beta_3 = 0 \) and \( \beta_7 = 0 \) (statistic \( LM_2 \)) and, if the same alternative hypothesis has the transition function \( G_2 \), the test is based on (51), assuming \( \beta_2 = 0 \) and \( \beta_6 = 0 \) (statistic \( LM_2 \)).
6 Simulation Study

6.1 Simulation design

A Monte Carlo simulation experiment is settled in order to investigate the empirical properties of the proposed asymmetry tests. We consider two different data generating processes (DGP):

\[ y_{1,t}^{(i)} = 0.4y_{1,t-1}^{(i)} - 0.25y_{1,t-2}^{(i)} + (0.02 - 0.9y_{1,t-1}^{(i)} + 0.795y_{1,t-2}^{(i)})G^{(i)}(\Xi) + \epsilon_{1,t}^{(s)}, \quad (58) \]

and

\[ y_{2,t}^{(i)} = 0.8y_{2,t-1}^{(i)} - 0.7y_{2,t-2}^{(i)} + (0.01 - 0.9y_{2,t-1}^{(i)} + 0.795y_{2,t-2}^{(i)})G^{(i)}(\Xi) + \epsilon_{2,t}^{(s)}, \quad (59) \]

where

\[ G^{(i)}(\Xi) = \left(1 + \exp \left\{ -h(\eta_t)^{(i)}I(\gamma_1<0,\gamma_2<0) + h(\eta_t)^{(i)}I(\gamma_1>0,\gamma_2<0) + h(\eta_t)^{(i)}I(\gamma_1<0,\gamma_2>0) + h(\eta_t)^{(i)}I(\gamma_1>0,\gamma_2>0) \right\} \right)^{-1}, \quad (60) \]

with \( \epsilon_t^{(i)} \sim N(0,1) \), \( i = \{1, \ldots, I\} \) denoting the \( i \)-esim simulation of the process \( \{y_t\}_{t=1}^T \) with \( s = y_{t-1}, \ c = \frac{1}{T}y_t^{(i)} \), \( I = 1,000 \).

\( y_{2,t}^{(i)} \) (henceforth "DGP 1") is an additive nonlinear model with accentuated nonlinear behavior, due to the high autoregressive parameters driving \( G(\Xi) \) and the low ones driving the linear part; this can be the case of a macroeconomic indicator affected by an unexpected shocks affecting the whole dynamics. On the other hand, \( y_{2,t}^{(i)} \) (henceforth "DGP 2") describes a more balanced scenario.

In order to simulate the function \( h(\eta_t) \) we use a set of values of vector \( \gamma \). The same different combinations of \((\gamma_1, \gamma_2)\) of the two symmetry tests has been used to investigate the empirical size and the empirical power of the three diagnostic test
described in Section 5. These combinations allow us to understand the behavior of the test in the different cases of null, medium and extreme asymmetry, respectively, as well as the effect of having different kinds of asymmetry, due to the different signs in the two $\gamma$-s. Moreover, we consider three different hypotheses for $T$ and the size $\alpha$, namely $T = \{100, 300, 1000\}$ and $\alpha = \{1\%, 5\%, 10\%\}$. For each DGP we explore the possibility of both types of different functional form of asymmetry in $G(\cdot)$ and compute the corresponding statistics (32) - (35), jointly to the "Two-Step" test hypothesis corresponding to statistics (27). In this experiment, the first 100 simulations have been discarded in order to avoid the initialization effect.

For what concerns the three diagnostic tests, in the error autocorrelation test we assumed the errors of the generating process followed an AR(1) process $u_t = \rho u_{t-1} + \epsilon$, $\epsilon \sim NID(0,1)$ and $\rho = \{0.2, 0.4\}$. In the test for no additive asymmetry we added to the previously described DGP a generalized logistic function $G_2(\gamma^{(2)}, h(c, y_{t-1}))$ with coefficients $\pi_0 = 0.01, \pi_2 = -1.8, \pi_3 = 1.6, \gamma^{(2)} = (\gamma_3, \gamma_4) = \{(5, 2), (50, 20), (500, 200)\}$ and $\gamma^{(1)}$ fixed at $(120, 70)$; this ensures that the behavior of the additive component remains isolated from the second; our experience shows that if higher parameter are set, the inversion become problematic. For the test for parameter constancy, the coefficients has been simulated according to a generalized logistic smooth change with $\lambda_1 = (0, 0.4, -0.25)'$ and $\lambda_2 = (0.2, -0.9, -0.795)'$. All these devices should make our simulation exercise comparable to the ET results.

6.2 Results

The results of the "All-in-one" and "Two-step" tests discussed in Section 4 for single DGP 1 and DGP 2 are reported in Table 1 and Table 2, respectively. Several findings can be easily noticed:

- The two tests have good and similar size properties. Only in large samples the two models behave in a slightly different way because of the statistics $LM_2$, being its size in for DGP 1 (0.0735 at a nominal size 5%) slightly oversized
with respect to DGP 2 (0.0391), while statistics $LM_1$ and the "Two-Step" test are more consistent.

- The two tests react similarly to different DGPs: the statistics $LM_1$ is more powerful of $LM_2$, regardless to the DGP 2 as sample size grows, although the empirical power is similar for moderate asymmetry and T=100. The "Two-Step" test makes an exception: under $DGP_1$, the power of $S_1$ is very similar to $LM_1$ and $LM_2$, while, under DGP 2, the $S_1$ power is full when one of the slopes is 0 and the other is positive (see rows 4, 6 or 10 and 12 in Table 2). An important difference between the two scenarios is the change is scale of the empirical power under DGP 2; for example, when T=300, $LM_1$ statistic passes from 0.99 to 0.13 for $\alpha = 5\%$. This implies that the detection of a dynamic asymmetric movement of the series when the underlining process is not unambiguously nonlinear remains critical.

- Both the tests are quite sensitive to different couples of $(\gamma_1, \gamma_2)$ with respect to signs and scale: the empirical power of both tests tends to decline while $\gamma$ has opposite signs. In particular, for $\gamma_1 < 0$, the power decays up to one third (see the case of $\gamma = (50, 10)$ for $\alpha = 5\%$ in statistic $LM_2$ at DGP1). In any case, all the statistics requires high slopes (500, 100 and similar) to get power in low sample. Heuristically, this is justified with the fact that the Stukel’ s function approximates a near-to-linear function for extreme negative slopes.

- The results for the three diagnostic tests deliver a similar picture, see Table 3 and Table 4. Some caveat are required to interpret the empirical power properties: under DGP 1, all the tests have good power, in particular for serial correlation test; the test of no additive asymmetry and parameter constancy are characterized by a duality: when the two slopes are high, that is $\gamma = (500, 200)$, the power is extreme, while it decays for low-medium asymmetry (0.21 vs 1.00 at $\alpha = 5\%$, T=100 in no additive asymmetry test, 0.44 vs 0.87 for
\( LM_2 \) statistic at the same nominal and sample size for parameter constancy).

On the other hand, under DGP 2, the change in scale of the power is evident only for the parameter constancy test; interestingly, the test for no serial correlation is more powerful.

7 Illustrations

7.1 Data

In this section the GSTAR model is applied to four time series, namely: the U.S. index of industrial production and unemployment (IP and UN, respectively); the yearly average of daily International Sunspot Number (YSSN), and the Canadian Lynx data (LYNX). Further informations on the dataset can be found in Table 5. We consider also the monthly average of Sunspot Number from January 1850 to December 2013 (1962 observations) for which three different kinds of data transformations are compared to link our model to the existing literature: the logarithmic (logMSSN), square root (sqrtMSSN) and the growth rate (DLMSSN); in this case, the Kalman-smoothed version of the series is available and necessary to avoid inversion problems due to the high noise. The series and the resulting (eventually, multiple) transition function(s), plotted versus time are reported in Figure 5, while the same transitions plotted versus the selected transition variable are shown in Figure 6.

7.2 Results

Tables 6 and 7 report the results of the modelling strategy discussed in Section 3, and specifically: the descriptive analysis of the series, using basic statistics and a battery of test for normality, ARCH-effect, serial uncorrelation, identical distribution and the \( t \)-statistics of Dikey-Fueller test augmented for two lags in the first panel; the result of the LST linearity test, the selected model according to the Teräsvirta rule.
and two symmetry tests introduced in Section 4 (second panel); parameter estimates and HAC standard errors of the selected GSTAR model with its equivalent symmetric specification, for which the possibility of multiple regimes has been taken in consideration (MR-STAR, third panel); the diagnostic tests (fourth panel). Informations on rolling forecasting exercise are shown in Table 8.

Several facts arises:

• The dynamic asymmetry here introduced is never rejected if a GSTAR model is assumed, as the "All-in-One" test suggests; however, the "Two-Step" approach, which strictly follows the original Stukel’s methodology changes this result; this seems reasonable at least for case of UN. This is not the case of monthly sunspots series, for which the "Two-Step" test, although still not able to reject the null for the selected model, starts to reject if different data and models are used\(^6\).

• The GSTAR transition function differs from its symmetric equivalent and this holds also when the two estimated slopes are very similar. In particular, Figure 5, panels (a), (c) and (d), shows that the estimated asymmetric \(G(\gamma, s_t, c)\) functions tends to concentrate in the upper part of the space of continuum of states (between 0.5 and 1); on the other side, in panel (b) of the same figure, the estimated GLSTAR transition for UN (where \(\gamma_1 = \gamma_2 = 0.001\)) reduces of 30% from the full [0, 1] range (from 1950 to 1980) to a [0.2, 0.9] range (after first 1980s), allowing to reproduce the cyclical movements of the data and their change in scale better than the traditional parametrization. The differences between transition functions are still more evident in Figure 6.

• The GSTAR model is sensitive to changes in scale, so that a further transformation tends to over-smoothing (exacerbate) the nonlinear dynamics of

\(^6\)The results, not shown here, can be provided upon request.
the process if further transformations are applied. In this sense, Figure 7 on monthly Sunspot series is almost self-explanatory: under square root-transformation, (panels (c) and (d)), the GSTAR transition reproduces most of the asymmetric cycle fluctuations of the data in a very precise way and its symmetric version seems almost equivalent. On the other side, if growth rates are considered (same figure, panels (e) and (f)), the GSTAR model reproduces the asymmetric dynamics of the data, but the states oscillates in a windows [0.45-0.6], leading to a quasi-linear behavior, as confirmed by the low slopes coefficients; on the other side, the MR-STAR transitions are well-behaving.

• Concerning point forecasting properties, the GSTAR model beats almost always its symmetric counterpart according to mRAE, while, in terms of RMSFE, the GSTAR wins in many forecast horizons of YSSN and LYNX and at longest horizons of UN; similar evidence is provided by MFE criterion: the new model prevails in two cases (UN and YSSN) whereas at very short term, the AR is still a good model for IP and LYNX. This superiority is less evident if considering sMAE: the two nonlinear specification almost equivalent for IP, while, for other three cases the MR-STAR prevails with a factor of less than 0.1%.

• In terms of density forecasting, the GSTAR wins for only UN according to LogS, while AR does for IP and short horizons of YSSN and MR-STAR for LYNX. The GSTAR returns to outperform if QSR is used, beating its competitors for YSSN and LYNX and at long horizons of IP and UN. Differently, the CRPS conveys a clear superiority of MR-STAR, which win in almost all cases, with the exception of YSSN and short-run horizons of IP (where AR better). The $qS^α$ enforces this result by providing evidence in favor linear specifications with the only exception of IP, and STAR being still the second best for IP and UN.

• Finally, there is evidence of significant improvement in prediction until $h=6$ if
a GSTAR(p) is considered with respect to a linear AR specification.

8 Discussion and Conclusions

The Generalized Logistic function is applied to STAR family of models as simple, statistically feasible way to capture the dynamic asymmetry in the conditional mean of a time series. The resulting GSTAR model ensures the smoothness of the transition function by construction without requiring further efforts for what concerns identification and estimation.

Two test for the null of dynamic symmetry and three diagnostic tests are proposed and investigated. Our simulation results make us confident in the use of the "All-in-One", as well as in the use of all diagnostic tests, while the "Two-Step" test seems us excessively conservative. In any case, such a feature remains not easy to detect if DGP is not properly imposed.

The GSTAR specification, due to its logarithmic (exponential) rescaling imposed by $h$-functions (3) - (4), is able to characterize some of the most prominent examples of nonlinear time series in applied sciences. Moreover, it allows the modeller to gain in terms of parsimony with respect to other symmetric counterparts. This is immediately evident in the LYNX example, where an autoregressive order 7 is sufficient to pass all the diagnostic tests, whereas more lags was required by previous literature using SETAR and STAR models (Tong, 1977; Teräsvirta, 1994). The monthly sunspot data enforces this finding.

The dynamic asymmetry is an important feature to take in account for point forecasting aims. The density forecasting exercise confirm - and possibly, because of the variety of datasets used, enforce - the Kascha & Ravazzolo (2010) evidence that the relation between highest LogS and lower RMSFE is not one-to-one. In addiction to this, we find that such a relation breaks under CRPS and reverts under $qS^a$. This means that dynamic asymmetric models are not superior to traditional STARs if
robust measures are used. In any case, nonlinear specifications remains preferable to linear ones.

Finally, the flexibility of the GSTAR model makes it suitable of several developments and applications in many important applied fields, like Finance and Climatology.

Acknowledgements

This paper has been mainly developed when the author was visiting PhD student at CREATE - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. The hospitality and the stimulating research environment provided by Niels Haldrup are gratefully acknowledged. The author is particularly grateful to Tommaso Proietti and Timo Teräsvirta for their supervision. He also thanks Barbara Annicchiarico, Gianna Boero, Alessandra Canepa, Jan G. De Gooijer, Menelaos Karanasos, Alessandra Luati, James Morley, Alessia Paccagnini, Phil Rothman and Howell Tong for helpful discussions. He is also grateful to seminar participants at "ECTS2011" Conference held in Villa Modragone, CFE 2012 in Oviedo, ICEEE-5th in Genoa, 8th BMRC-QASS Conference on Macro and Financial Economics in Brunel University and the 2013 Annual Conference of Royal Statistical Society in Newcastle. This paper has been awarded of the “James B. Ramsey” Prize for the best paper in Econometrics presented by a PhD student at the 21st Annual Symposium of the Society for Nonlinear Dynamics and Econometrics held in University of Milan-Bicocca. This work is in memory of Giancarlo Marini for his great support and encouragement. The usual disclaimer applies.

References


Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika, 64*, 247–254.


A Appendix

A.1 Mathematical derivations

A.1.1 Preliminary notation

Let denote $G_t = G(\Xi)$, $\Xi = [\gamma_1, \gamma_2, c]$ or $[\gamma_1, \gamma_2, c_1, c_2]$ in case of GLSTR1 and GESTAR (GESTR). Then we can re-define $G(\Xi)$ as:

\[
G^{(i)}(\Xi) = [1 + g(f^{(i)}(\Xi))]^j,
\]

\[
f^{\text{GLSTR}}(\Xi) = -[h(\eta^L)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^L)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^L)I_{(\gamma_1,0,\gamma_2,0)}],
\]

\[
f^{\text{GLSTR2}}(\Xi) = -[h(\eta^2L)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^2L)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^2L)I_{(\gamma_1,0,\gamma_2,0)}],
\]

\[
f^{\text{GESTR}}(\Xi) = -[h(\eta^F)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^F)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^F)I_{(\gamma_1,0,\gamma_2,0)}],
\]

with $i = \{L, 2L, E\}$, denoting the Logistic, Double Logistic and Exponential parametrization, $j = \{1; -1\}$, with $j = 1$ only if $f(\Xi) = f^{\text{GESTR}}(\Xi)$, $\eta^L = s_t - c$, $\eta^2L = (s_t - c_1)(s_t - c_2)$, $\eta^F = (s_t - c)^2$. Moreover, let $f'(\Xi) = -[h'(\eta)I_{(\gamma_1,0,\gamma_2,0)} + h'(\eta)I_{(\gamma_1,0,\gamma_2,0)} + h'(\eta)I_{(\gamma_1,0,\gamma_2,0)}]$ define the first derivative of $f(\Xi)$ and $D = 1 + g(\Xi)$ denote the denominator of the fraction which is the result of the computation of the second derivatives so that:

\[
D^2 = 1 + g(\Xi)^2 + 2g(\Xi),
\]

\[
g^{(i)}(\Xi)^2 = 1 + \exp\left\{-2(h(\eta^i)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^i)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^i)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^i)I_{(\gamma_1,0,\gamma_2,0)})\right\} + 2\exp\left\{h(\eta^i)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^i)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^i)I_{(\gamma_1,0,\gamma_2,0)}\right\} + h(\eta^i)I_{(\gamma_1,0,\gamma_2,0)} + h(\eta^i)I_{(\gamma_1,0,\gamma_2,0)}).
\]

A.1.2 LSTR1 case

When the transition equation is a Generalized Logistic, we have the following derivatives:
(i) 

\[ G_{\gamma_1}(\Xi) = -\frac{g'(f(\Xi)) \cdot f'(\Xi)}{D^2} \]  

(61)

where:

\[
h'(\eta_t) I(\eta_t > 0) = \frac{\partial}{\partial \gamma_1} h(\gamma) = \begin{cases} 
-\frac{1}{\gamma_1} \cdot \exp(|\eta_t| - 1)(|\eta_t| - 1) & \text{if } \gamma_1 > 0 \\
0 & \text{if } \gamma_1 = 0 \\
-\frac{1}{\gamma_1} \cdot \ln(1 - \gamma_1 |\eta_t| + \frac{|\eta_t|}{\gamma_1 |\eta_t|}) & \text{if } \gamma_1 < 0 
\end{cases}
\]  

and

\[
h'(\eta_t) I(\eta_t \leq 0) = \begin{cases} 
0 & \text{if } \gamma_2 > 0 \\
0 & \text{if } \gamma_2 = 0 \\
0 & \text{if } \gamma_2 < 0 
\end{cases}
\]  

(ii) \( G_{\gamma_2}(\cdot) \) : equal to (61) but with

\[
h'(\eta_t) I(\eta_t > 0) = \frac{\partial}{\partial \gamma_2} h(\gamma) = \begin{cases} 
0 & \text{if } \gamma_1 > 0 \\
0 & \text{if } \gamma_1 = 0 \\
0 & \text{if } \gamma_1 < 0 
\end{cases}
\]  

and

\[
h'(\eta_t) I(\eta_t \leq 0) = \begin{cases} 
\frac{1}{\gamma_2} \exp(1 - \gamma_2 |\eta_t|) \cdot (\frac{1}{\gamma_2} + |\eta_t|) & \text{if } \gamma_2 > 0 \\
0 & \text{if } \gamma_2 = 0 \\
-\frac{1}{\gamma_2} \left[ \frac{1}{\gamma_2} \ln(\gamma_2 |\eta_t| - 1) + \frac{|\eta_t|}{\gamma_2 |\eta_t|} \right] & \text{if } \gamma_2 < 0 
\end{cases}
\]  

(iii) \( G_{c}(\cdot) \) : equal to (61) but with

\[
f'(\Xi) = h(\eta_t) I(\eta_t \leq 0) + h(\eta_t) I(\eta_t > 0) 
\]  

(66)

A.1.3 LSTR2 case

When the transition equation is a (Generalized) Double Logistic as in model (??), we have the following derivatives:
(i) \(G_\gamma(\cdot)\) equal to equation (61) with: \(f^{(i)}(\Xi) = f^{(GLSTR^2)}(\Xi) = -[h(\eta_t^E)I_{(\gamma_1 \leq 0, \gamma_2 \leq 0) + h(\eta_t^2)I_{(\gamma_1 \leq 0, \gamma_2 > 0) + h(\eta_t^2)I_{(\gamma_1 > 0, \gamma_2 \leq 0) + h(\eta_t^2)I_{(\gamma_1 > 0, \gamma_2 > 0)}]}, g^{(GLSTR^2)} = \exp\{f^{(GLSTR^2)}(\Xi)\},

\(h'(\eta_t^2)I_{(\eta_t^2 \leq 0)}\) and \(h'(\eta_t^2)I_{(\eta_t^2 > 0)}\) equal to systems (62) and (63).

(ii) \(G_{\gamma_2}(\cdot)\) : equal to equation (61) with: \(f^{(GLSTR^2)}(\Xi)\) and \(g^{(GLSTR^2)}(\Xi)\) above defined as in case (i) and \(h'(\eta_t^2)I_{(\eta_t^2 \geq 0)}\) and \(h'(\eta_t^2)I_{(\eta_t^2 < 0)}\) equal to systems (64) and (65) respectively.

(iii) \(G_{\gamma_1}(\cdot)\): equal to equation (61) with: \(f^{(GLSTR^2)}(\Xi)\) and \(g^{(GLSTR^2)}(\Xi)\) defined as in case (i) and

\[
f'(\Xi) = h'(\eta_t^2)I_{(\eta_t^2 \leq 0)}(s_t - c_2) + h'(\eta_t^2)I_{(\eta_t^2 > 0)}(s_t - c_2) \quad (67)
\]

(iv) \(G_{\gamma_2}(\cdot)\): equal to equation (61) with: \(f^{(GLSTR^2)}(\Xi)\) and \(g^{(GLSTR^2)}(\Xi)\) defined as in case i) and

\[
f'(\Xi) = h'(\eta_t^2)I_{(\eta_t^2 \leq 0)}(s_t - c_1) + h'(\eta_t^2)I_{(\eta_t^2 > 0)}(s_t - c_1) \quad (68)
\]

### A.1.4 ESTR case

When the transition equation is an exponential as in model (??), we have: \(f^{(ESTR)}(\Xi) = -[h(\eta_t^E)I_{(\gamma_1 \leq 0, \gamma_2 \leq 0) + h(\eta_t^E)I_{(\gamma_1 \leq 0, \gamma_2 > 0) + h(\eta_t^E)I_{(\gamma_1 > 0, \gamma_2 \leq 0) + h(\eta_t^E)I_{(\gamma_1 > 0, \gamma_2 > 0)}]}\), \(g^{(ESTR)}(\Xi) = -\exp\{f^{(E)}\}\), hence the following derivatives:

(i) \(G_\gamma(\cdot) = f^{(ESTR)*}(\Xi)\) with: \(f'(\Xi) = -[h(\eta_t^E)I_{(\eta_t^E \leq 0)}(s_t - c)^2 + h'(\eta_t^E)I_{(\eta_t^E \leq 0)}(s_t - c)^2), h'(\eta_t^E)I_{(\eta_t^E > 0)}\) and \(h'(\eta_t^E)I_{(\eta_t^E < 0)}\) being the same of systems (62) and (63).

(ii) \(G_{\gamma_2}(\cdot)\): same as \(G_\gamma(\cdot)\) with \(h'(\eta_t^E)I_{(\eta_t^E > 0)}\) and \(h'(\eta_t^E)I_{(\eta_t^E \leq 0)}\) being the same of systems (64) and (65).

(iii) \(G_{\gamma_1}(\cdot) = f^{(ESTR)*}(\Xi)\) with \(f'(\Xi) = h'(\eta_t^E)I_{(\eta_t^E \leq 0)}(2c) + h'(\eta_t^E)I_{(\eta_t^E > 0)}(2c), h'(\eta_t^E)I_{(\eta_t^E > 0)}\) and \(h'(\eta_t^E)I_{(\eta_t^E \leq 0)}\) being the same of systems (64) and (65).
A.2 Tables and Graphs

**Figure 1**: Transition function for different parametrizations of Asymmetric STAR.

(a) Sollis et al. (2002)

(b) Sollis et al. (1999) - Lundbergh and Terasvirta (2006)
Figure 2: The Generalized Logistic function.

Figure 3: The Generalized Second-Order Logistic function.
**Figure 4:** An example of GLSTAR model.

(a) Simulated process

(b) Transition function

(c) Transition functions vs $\eta_t$

(d) The rescaling effect of $h(\eta_t)$

Simulation performed with following parameters: $\phi_0 = 0.05; \phi_1 = 0.4; \phi_2 = 0.25; \theta_0 = 0.2; \theta_1 = 0.4; \theta_2 = 0.25; \gamma_1 = 0.25; \gamma_2 = -1.0; c = 0; c_1 = 3; c_2 = 5; T = 100$
Figure 5: Estimated transition function for (MR)STAR and GSTAR specifications.

(a) U.S. Industrial Production

(b) U.S. Unemployment

(c) Yearly Sunspot Number

(d) Canadian Lynx

NOTE: The data are plotted in blue line.
Figure 6: Estimated transition functions vs transition variable.

(a) U.S. Industrial Production

(b) U.S. Unemployment

(c) Yearly Sunspot Number

(d) Canadian Lynx
Figure 7: Monthly SSN: estimated transition functions for different data transformations.

(a) logMSSN: \( G(Ξ) \) vs \( t \)

(b) logMSSN: \( G(Ξ) \) vs \( s_t \)

(c) sqrtMSSN: \( G(Ξ) \) vs \( t \)

(d) sqrtMSSN: \( G(Ξ) \) vs \( s \)

(e) DLMSSN: \( G(Ξ) \) vs \( t \)

(f) DLMSSN: \( G(Ξ) \) vs \( s \)

NOTE: The data are plotted in blue line of left-hand subfigures.
Table 1: Empirical Size and Power of "All-in-One" and "Two-Step" test for dynamic asymmetry under DGP 1.

<table>
<thead>
<tr>
<th>T</th>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
<th>( LM_1 )</th>
<th>( LM_2 )</th>
<th>( S_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \alpha = 0.01 )</td>
<td>( \alpha = 0.05 )</td>
<td>( \alpha = 0.10 )</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0.0030</td>
<td>0.0221</td>
<td>0.0458</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0058</td>
<td>0.0255</td>
<td>0.0565</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0073</td>
<td>0.0384</td>
<td>0.0798</td>
</tr>
<tr>
<td>300</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
<td>0.0230</td>
<td>0.0467</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0070</td>
<td>0.0288</td>
<td>0.0653</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0208</td>
<td>0.0735</td>
<td>0.1297</td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>0.0025</td>
<td>0.0135</td>
<td>0.0430</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0039</td>
<td>0.0168</td>
<td>0.0433</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0088</td>
<td>0.0527</td>
<td>0.0882</td>
</tr>
</tbody>
</table>

Empirical Power

<table>
<thead>
<tr>
<th>T</th>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
<th>( LM_1 )</th>
<th>( LM_2 )</th>
<th>( S_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \alpha = 0.01 )</td>
<td>( \alpha = 0.05 )</td>
<td>( \alpha = 0.10 )</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0</td>
<td>0.0315</td>
<td>0.1382</td>
<td>0.2366</td>
</tr>
<tr>
<td></td>
<td>-50</td>
<td>10</td>
<td>0.0197</td>
<td>0.0737</td>
<td>0.1502</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-10</td>
<td>0.0180</td>
<td>0.0939</td>
<td>0.1904</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0</td>
<td>0.0170</td>
<td>0.0836</td>
<td>0.1601</td>
</tr>
<tr>
<td></td>
<td>-50</td>
<td>50</td>
<td>0.0094</td>
<td>0.0301</td>
<td>0.1288</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>50</td>
<td>0.0128</td>
<td>0.0732</td>
<td>0.1486</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.7930</td>
<td>0.8165</td>
<td>0.8335</td>
</tr>
<tr>
<td></td>
<td>-50</td>
<td>100</td>
<td>0.0173</td>
<td>0.0890</td>
<td>0.1645</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-100</td>
<td>0.7931</td>
<td>0.8073</td>
<td>0.8241</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0</td>
<td>0.7954</td>
<td>0.8059</td>
<td>0.8215</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>500</td>
<td>0.8037</td>
<td>0.8083</td>
<td>0.8303</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>100</td>
<td>0.0198</td>
<td>0.0270</td>
<td>0.0379</td>
</tr>
<tr>
<td></td>
<td>-50</td>
<td>10</td>
<td>0.0207</td>
<td>0.1266</td>
<td>0.1941</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-10</td>
<td>0.0420</td>
<td>0.1458</td>
<td>0.2253</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0</td>
<td>0.0353</td>
<td>0.1339</td>
<td>0.2059</td>
</tr>
<tr>
<td></td>
<td>-50</td>
<td>50</td>
<td>0.0190</td>
<td>0.1033</td>
<td>0.1476</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.0176</td>
<td>0.0964</td>
<td>0.1898</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>100</td>
<td>0.9958</td>
<td>0.9961</td>
<td>0.9980</td>
</tr>
<tr>
<td></td>
<td>-500</td>
<td>100</td>
<td>0.0300</td>
<td>0.1485</td>
<td>0.2147</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-100</td>
<td>0.9958</td>
<td>0.9961</td>
<td>0.9961</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>500</td>
<td>0.9958</td>
<td>0.9961</td>
<td>0.9961</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0190</td>
<td>0.1039</td>
<td>0.1434</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>500</td>
<td>0.9917</td>
<td>0.9966</td>
<td>0.9972</td>
</tr>
<tr>
<td>300</td>
<td>10</td>
<td>0</td>
<td>0.1173</td>
<td>0.2498</td>
<td>0.3634</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0</td>
<td>0.0830</td>
<td>0.2032</td>
<td>0.2996</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-50</td>
<td>0.1111</td>
<td>0.2863</td>
<td>0.4373</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>50</td>
<td>0.0446</td>
<td>0.1192</td>
<td>0.2352</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0</td>
<td>0.0654</td>
<td>0.1799</td>
<td>0.2643</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>-50</td>
<td>100</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-100</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>500</td>
<td>0.9958</td>
<td>0.9961</td>
<td>0.9961</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0190</td>
<td>0.1039</td>
<td>0.1434</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>500</td>
<td>0.9917</td>
<td>0.9966</td>
<td>0.9972</td>
</tr>
<tr>
<td>1000</td>
<td>10</td>
<td>0</td>
<td>0.2139</td>
<td>0.3545</td>
<td>0.4518</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0</td>
<td>0.0580</td>
<td>0.1589</td>
<td>0.2205</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-50</td>
<td>0.0866</td>
<td>0.2054</td>
<td>0.2599</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>50</td>
<td>0.0107</td>
<td>0.0635</td>
<td>0.1281</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0</td>
<td>0.0110</td>
<td>0.0471</td>
<td>0.0893</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>-50</td>
<td>100</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-100</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Table 2: Empirical Size and Power of "All-in-One" and "Two-Step" test for dynamic asymmetry under DGP 2.

<table>
<thead>
<tr>
<th>T</th>
<th>γ1</th>
<th>γ2</th>
<th>$LM_1$</th>
<th>$LM_2$</th>
<th>$S_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.10$</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0.0024</td>
<td>0.0164</td>
<td>0.0185</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
<td>0.0081</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0036</td>
<td>0.0460</td>
<td>0.1018</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T</th>
<th>γ1</th>
<th>γ2</th>
<th>$LM_1$</th>
<th>$LM_2$</th>
<th>$S_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>0</td>
<td>0</td>
<td>0.0024</td>
<td>0.0164</td>
<td>0.0185</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
<td>0.0081</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0036</td>
<td>0.0460</td>
<td>0.1018</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T</th>
<th>γ1</th>
<th>γ2</th>
<th>$LM_1$</th>
<th>$LM_2$</th>
<th>$S_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>0.0024</td>
<td>0.0164</td>
<td>0.0185</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
<td>0.0081</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0036</td>
<td>0.0460</td>
<td>0.1018</td>
</tr>
</tbody>
</table>
Table 3: Empirical Size and Empirical Power of test for serial correlation, no additive asymmetry and parameter constancy under DGP 1.

### Empirical Size

<table>
<thead>
<tr>
<th>T</th>
<th>γ1</th>
<th>γ2</th>
<th>γ3</th>
<th>γ4</th>
<th>Nominal size</th>
<th>No error autocorrelation ( \rho = 0 )</th>
<th>No additional asymmetry</th>
<th>Parameter constancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>q = 1 q = 2 q = 4 q = 10</td>
<td>H0</td>
<td>LM1</td>
<td>LM2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0.0060 0.0041 0.0087 0.0053</td>
<td>0.0020</td>
<td>0.0000 0.0000 0.0005</td>
<td>0.0000 0.0000 0.0005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0628 0.0482 0.0354 0.0435</td>
<td>0.0034</td>
<td>0.0000 0.0000 0.0005</td>
<td>0.0000 0.0000 0.0005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0951 0.1008 0.0931 0.0955</td>
<td>0.0063</td>
<td>0.0000 0.0000 0.0029</td>
<td>0.0000 0.0000 0.0029</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>10</td>
<td></td>
<td></td>
<td>0.0455 0.1334 0.0700 0.0440</td>
<td>0.0000</td>
<td>0.0000 0.0000 0.0000</td>
<td>0.0000 0.0000 0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0526 0.0392 0.0186 0.0167</td>
<td>0.0000</td>
<td>0.0000 0.0000 0.0000</td>
<td>0.0000 0.0000 0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0797 0.0953 0.0515 0.0393</td>
<td>0.0000</td>
<td>0.0000 0.0000 0.0092</td>
<td>0.0000 0.0000 0.0092</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td>0.0102 0.0149 0.0098 0.0047</td>
<td>0.0000</td>
<td>0.0000 0.0000 0.0000</td>
<td>0.0000 0.0000 0.0000</td>
</tr>
</tbody>
</table>

### Empirical Power

<table>
<thead>
<tr>
<th>T</th>
<th>γ1</th>
<th>γ2</th>
<th>γ3</th>
<th>γ4</th>
<th>Nominal size</th>
<th>No error autocorrelation ( \rho = 0.2 )</th>
<th>No error autocorrelation ( \rho = 0.4 )</th>
<th>No additional asymmetry</th>
<th>Parameter constancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>q = 1 q = 2 q = 4 q = 10</td>
<td>q = 1 q = 2 q = 4 q = 10</td>
<td>H0</td>
<td>LM1</td>
<td>LM2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td></td>
<td>0.0136 0.0114 0.0112 0.0057</td>
<td>0.0124</td>
<td>0.0000 0.0000 0.0000</td>
<td>0.0000 0.0000 0.0000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0162 0.0272 0.0616 0.0436</td>
<td>0.0122</td>
<td>0.0000 0.0000 0.0081</td>
<td>0.0000 0.0000 0.0081</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>10</td>
<td>50</td>
<td>20</td>
<td>0.0087 0.0084 0.0043 0.0014</td>
<td>0.0018</td>
<td>0.0000 0.0000 0.0041</td>
<td>0.0000 0.0000 0.0041</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0622 0.0613 0.0287 0.0112</td>
<td>0.0018</td>
<td>0.0000 0.0000 0.0297</td>
<td>0.0000 0.0000 0.0297</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1245 0.1070 0.0723 0.0231</td>
<td>0.0097</td>
<td>0.0000 0.0000 0.0607</td>
<td>0.0000 0.0000 0.0607</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td>0.0250 0.0337 0.0149 0.0000</td>
<td>0.0142</td>
<td>0.0000 0.0000 0.0045</td>
<td>0.0000 0.0000 0.0045</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0621 0.0443 0.0618 0.0208</td>
<td>0.0657</td>
<td>0.0000 0.0023 0.0344</td>
<td>0.0000 0.0023 0.0344</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1221 0.1242 0.1195 0.0356</td>
<td>0.1083</td>
<td>0.0000 0.0557 0.0836</td>
<td>0.0000 0.0557 0.0836</td>
<td></td>
</tr>
<tr>
<td>Table 4: Empirical Size and Empirical Power of test for serial correlation, no additive asymmetry and parameter constancy under DGP 2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Empirical Size</strong></td>
<td><strong>Empirical Power</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.01$</td>
<td>$\rho = 0.05$</td>
<td>$\rho = 0.10$</td>
<td>$\rho = 0.01$</td>
<td>$\rho = 0.05$</td>
<td>$\rho = 0.10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>0.0042</td>
<td>0.0048</td>
<td>0.0056</td>
<td>0.0049</td>
<td>0.0055</td>
<td>0.0062</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>0.0058</td>
<td>0.0063</td>
<td>0.0071</td>
<td>0.0064</td>
<td>0.0070</td>
<td>0.0077</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.10$</td>
<td>0.0074</td>
<td>0.0080</td>
<td>0.0088</td>
<td>0.0081</td>
<td>0.0087</td>
<td>0.0094</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>0.0046</td>
<td>0.0052</td>
<td>0.0060</td>
<td>0.0055</td>
<td>0.0061</td>
<td>0.0068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>0.0052</td>
<td>0.0058</td>
<td>0.0066</td>
<td>0.0061</td>
<td>0.0067</td>
<td>0.0074</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.10$</td>
<td>0.0068</td>
<td>0.0074</td>
<td>0.0082</td>
<td>0.0077</td>
<td>0.0083</td>
<td>0.0090</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>0.0050</td>
<td>0.0056</td>
<td>0.0064</td>
<td>0.0059</td>
<td>0.0065</td>
<td>0.0072</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>0.0056</td>
<td>0.0062</td>
<td>0.0070</td>
<td>0.0065</td>
<td>0.0071</td>
<td>0.0078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.10$</td>
<td>0.0072</td>
<td>0.0078</td>
<td>0.0086</td>
<td>0.0081</td>
<td>0.0087</td>
<td>0.0094</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>0.0054</td>
<td>0.0060</td>
<td>0.0068</td>
<td>0.0063</td>
<td>0.0069</td>
<td>0.0076</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>0.0060</td>
<td>0.0066</td>
<td>0.0074</td>
<td>0.0069</td>
<td>0.0075</td>
<td>0.0082</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.10$</td>
<td>0.0076</td>
<td>0.0082</td>
<td>0.0090</td>
<td>0.0085</td>
<td>0.0091</td>
<td>0.0098</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
### Table 5: Datasets

<table>
<thead>
<tr>
<th>Series</th>
<th>Sample</th>
<th>( T^* )</th>
<th>Testing Period</th>
<th>Source</th>
<th>Previous Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>1947 M1 — 2013 M3</td>
<td>500</td>
<td>1982 M9+h — 2013M3-12+h</td>
<td>OECD Main Economic Indicators</td>
<td>Anderson &amp; Teräsvirta (1992)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Proietti (1998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Teräsvirta et al. (2005)</td>
</tr>
<tr>
<td>UN</td>
<td>1948 M1 — 2013 M3</td>
<td>500</td>
<td>1983 M9+h — 2013M3-12+h</td>
<td>OECD Main Economic Indicators</td>
<td>Rothman (1991)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rothman (1998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Montgomery et al. (1998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Skalin &amp; Teräsvirta (2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Proietti (2003)</td>
</tr>
<tr>
<td>YSSN</td>
<td>1700 — 2013</td>
<td>150</td>
<td>1851+h — 2013-12+h</td>
<td>Solar Influences Data Analysis Center</td>
<td>Tong &amp; Lim (1980)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tsay (1989)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hansen (1999)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strikholm &amp; Teräsvirta (2005)</td>
</tr>
<tr>
<td>LYNX</td>
<td>1820 — 1934</td>
<td>50</td>
<td>1871+h — 1934-12+h</td>
<td>The Encyclopedia of Mathematics wiki</td>
<td>Elton &amp; Nicholson (1942)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran (1953)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tong (1977)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tsay (1989)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Teräsvirta (1994)</td>
</tr>
</tbody>
</table>
Table 6: Empirical application of the (G)STAR model to four real data samples

### Descriptive statistics

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean</th>
<th>Median</th>
<th>MAD</th>
<th>Std. Dev.</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>JB*</th>
<th>ARCH-effects b</th>
<th>DW*</th>
<th>Ks d</th>
<th>ADF(1)(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>0.2490</td>
<td>0.2790</td>
<td>0.3744</td>
<td>0.4789</td>
<td>-0.2815</td>
<td>4.5233</td>
<td>0.0010</td>
<td>0.0000</td>
<td>0.0293</td>
<td>0.0000</td>
<td>-11.1962**</td>
</tr>
<tr>
<td>UN</td>
<td>0.0974</td>
<td>-0.3764</td>
<td>1.3182</td>
<td>1.8054</td>
<td>0.9802</td>
<td>4.9645</td>
<td>0.0010</td>
<td>0.0000</td>
<td>0.0341</td>
<td>0.0001</td>
<td>-10.7775**</td>
</tr>
<tr>
<td>YSSN</td>
<td>6.3981</td>
<td>6.3236</td>
<td>2.4412</td>
<td>2.9486</td>
<td>0.1790</td>
<td>2.2904</td>
<td>0.0226</td>
<td>0.0000</td>
<td>0.0278</td>
<td>0.0000</td>
<td>-5.2805**</td>
</tr>
<tr>
<td>LYNX</td>
<td>2.9037</td>
<td>2.9870</td>
<td>0.4721</td>
<td>0.5584</td>
<td>-0.3620</td>
<td>2.4664</td>
<td>0.0082</td>
<td>0.0000</td>
<td>0.0376</td>
<td>0.0000</td>
<td>-5.8245**</td>
</tr>
</tbody>
</table>

### ARCH-effects

<table>
<thead>
<tr>
<th>Series</th>
<th>ARCH-effects b</th>
<th>DW*</th>
<th>Ks d</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>-</td>
<td>0.0010</td>
<td>0.0293</td>
</tr>
<tr>
<td>UN</td>
<td>-</td>
<td>0.0000</td>
<td>0.0341</td>
</tr>
<tr>
<td>YSSN</td>
<td>-</td>
<td>0.0000</td>
<td>0.0278</td>
</tr>
<tr>
<td>LYNX</td>
<td>-</td>
<td>0.0000</td>
<td>0.0376</td>
</tr>
</tbody>
</table>

### Diagnostics (p-values)

<table>
<thead>
<tr>
<th>Diagnostic Test</th>
<th>STAR</th>
<th>GSTAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>No error autocorrelation</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>q=1</td>
<td>0.1785</td>
<td>0.3341</td>
</tr>
<tr>
<td>q=2</td>
<td>0.4028</td>
<td>0.6278</td>
</tr>
<tr>
<td>q=4</td>
<td>0.0001</td>
<td>0.9206</td>
</tr>
<tr>
<td>q=10</td>
<td>0.0000</td>
<td>0.9993</td>
</tr>
</tbody>
</table>

### Notes:
- Jarque-Bera (p-value)
- Engle’s test (p-value)
- Durbin-Watson test (p-value)
- Kolmogorov-Smirnov test (p-value)
- Values expressed in test-statistics, significance denoted by '∗' (5%), '∗∗' (1%)
- Results acquired by JMulTi
- Results acquired by RATS
Table 7: Empirical application of the (G)STAR model to Monthly Smoothed Sunspot Number from 1850 to 2013.

### Descriptive statistics

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean (logSSN)</th>
<th>Median (logSSN)</th>
<th>MAD (logSSN)</th>
<th>Std. Dev. (logSSN)</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>JB</th>
<th>ARCH-effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>logSSN</td>
<td>3.5843</td>
<td>3.8597</td>
<td>0.8641</td>
<td>1.0848</td>
<td>-0.7149</td>
<td>2.7842</td>
<td>0.0010</td>
<td>0.0000</td>
</tr>
<tr>
<td>sqrtSSN</td>
<td>6.7691</td>
<td>6.8884</td>
<td>2.5196</td>
<td>3.0051</td>
<td>0.1714</td>
<td>2.1951</td>
<td>0.0010</td>
<td>0.0000</td>
</tr>
<tr>
<td>DLSSN</td>
<td>0.0021</td>
<td>-0.0076</td>
<td>0.0563</td>
<td>0.1224</td>
<td>22.6304</td>
<td>798.4752</td>
<td>0.0010</td>
<td>0.9901</td>
</tr>
</tbody>
</table>

### ARCH-effects

<table>
<thead>
<tr>
<th>Series</th>
<th>DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>logSSN</td>
<td>0.0012</td>
</tr>
<tr>
<td>sqrtSSN</td>
<td>0.0780</td>
</tr>
<tr>
<td>DLSSN</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

### Linearity and Asymmetry tests (p-values)

#### Linearity Test

<table>
<thead>
<tr>
<th>Parameter</th>
<th>logSSN</th>
<th>sqrtSSN</th>
<th>DLSSN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>SE</td>
<td>Value</td>
<td>SE</td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>-3.495100</td>
<td>0.3420</td>
<td>-21.8270</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-464.4000</td>
<td>2.1995</td>
<td>-32.574</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-67.1761</td>
<td>0.7843</td>
<td>-1.412</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>-16.3030</td>
<td>0.3175</td>
<td>-0.3175</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>-328.3510</td>
<td>0.2302</td>
<td>7.618000</td>
</tr>
<tr>
<td>$\phi_5$</td>
<td>-160.2940</td>
<td>0.9647</td>
<td>-1.196000</td>
</tr>
<tr>
<td>$\phi_6$</td>
<td>-336.7630</td>
<td>0.6516</td>
<td>-17.7000</td>
</tr>
<tr>
<td>$\phi_7$</td>
<td>107.1330</td>
<td>1.5158</td>
<td>-57.3000</td>
</tr>
<tr>
<td>$\phi_8$</td>
<td>-98.2139</td>
<td>2.6406</td>
<td>-147.2963</td>
</tr>
<tr>
<td>$\phi_9$</td>
<td>26.5310</td>
<td>0.7790</td>
<td>96.2139</td>
</tr>
<tr>
<td>$\phi_{10}$</td>
<td>-15.5313</td>
<td>1.5158</td>
<td>-11.960000</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-130.2640</td>
<td>1.5158</td>
<td>90.6000</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>-336.7630</td>
<td>0.6516</td>
<td>-17.7000</td>
</tr>
<tr>
<td>$\phi_{13}$</td>
<td>107.1330</td>
<td>1.5158</td>
<td>-57.3000</td>
</tr>
<tr>
<td>$\phi_{14}$</td>
<td>-98.2139</td>
<td>2.6406</td>
<td>-147.2963</td>
</tr>
<tr>
<td>$\phi_{15}$</td>
<td>26.5310</td>
<td>0.7790</td>
<td>96.2139</td>
</tr>
<tr>
<td>$\phi_{16}$</td>
<td>-15.5313</td>
<td>1.5158</td>
<td>-11.960000</td>
</tr>
<tr>
<td>$\phi_{17}$</td>
<td>-130.2640</td>
<td>1.5158</td>
<td>90.6000</td>
</tr>
<tr>
<td>$\phi_{18}$</td>
<td>-336.7630</td>
<td>0.6516</td>
<td>-17.7000</td>
</tr>
<tr>
<td>$\phi_{19}$</td>
<td>107.1330</td>
<td>1.5158</td>
<td>-57.3000</td>
</tr>
<tr>
<td>$\phi_{20}$</td>
<td>-98.2139</td>
<td>2.6406</td>
<td>-147.2963</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>26.5310</td>
<td>0.7790</td>
<td>96.2139</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-15.5313</td>
<td>1.5158</td>
<td>-11.960000</td>
</tr>
<tr>
<td>$\phi_{23}$</td>
<td>-130.2640</td>
<td>1.5158</td>
<td>90.6000</td>
</tr>
<tr>
<td>$\phi_{24}$</td>
<td>-336.7630</td>
<td>0.6516</td>
<td>-17.7000</td>
</tr>
<tr>
<td>$\phi_{25}$</td>
<td>107.1330</td>
<td>1.5158</td>
<td>-57.3000</td>
</tr>
<tr>
<td>$\phi_{26}$</td>
<td>-98.2139</td>
<td>2.6406</td>
<td>-147.2963</td>
</tr>
</tbody>
</table>

### Diagnostics (p-values)

#### No error autocorrelation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (q=1)</th>
<th>Value (q=2)</th>
<th>Value (q=4)</th>
<th>Value (q=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_1$</td>
<td>0.0091</td>
<td>0.0010</td>
<td>0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>$q_2$</td>
<td>0.0125</td>
<td>0.0010</td>
<td>0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>$q_{10}$</td>
<td>0.5558</td>
<td>0.1089</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

### Parameter constancy

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (H1)</th>
<th>Value (H2)</th>
<th>Value (H3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>$c_2$</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>$c_3$</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

### NOTE:

(a) Jarque-Bera ($p$-value) (b) Engle’s test ($p$-value); (c) Durbin-Watson test ($p$-value); (d) Kolmogorov-Smirnov test ($p$-value); (1) Values expressed in test-statistics, significance denoted by '∗' (5%), '∗∗' (1%); (2) Results acquired by JMulTi; (3) Results acquired by RATS
<table>
<thead>
<tr>
<th>Forecast horizon</th>
<th>Giacomini-White Test (p-values)</th>
<th>Scoring Rule</th>
<th>Density predictive performances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MPE</td>
<td>MRAE</td>
<td>RMSFE</td>
</tr>
<tr>
<td>1</td>
<td>0.0020</td>
<td>1.1568</td>
<td>0.2120</td>
</tr>
<tr>
<td>3</td>
<td>0.0020</td>
<td>1.1644</td>
<td>0.2181</td>
</tr>
<tr>
<td>6</td>
<td>0.0020</td>
<td>1.1665</td>
<td>0.2246</td>
</tr>
<tr>
<td>12</td>
<td>0.0020</td>
<td>1.2024</td>
<td>0.2310</td>
</tr>
</tbody>
</table>

**NOTE:** Forecast windows size is set to 500 for IP and UN, 50 for LYNX and 150 for YSSN. For the Giacomini-Wight test, an AR(1) model is used as benchmark.