Testing and selecting local proper scoring rules[*][†]

Emilio Zanetti Chini[‡]

July 2, 2014

Abstract

We provide a frequentist framework to elicit the forecaster's expected utility by introducing an LM-type test for the null of locality of scoring rule underlining the observed probabilistic forecast. The scoring rule is looked as an observed transition variable in a smooth transition autoregression in order to nest a known framework. The test behaves consistently with the requirement of the theoretical literature. Applications to US Business Cycle, several economic time series and Bank of Norway's Fan Charts reveal that scoring rules affect the dating algorithm of recessions events and the model-based forecast performances in favor of a nonlinear specification, advocating the importance of a correct score selection and that the locality of the scoring rule underlining the estimated predictive density is a strong assumption.

Keywords: Forecast Evaluation, Nonlinear Models, Statistical Decision Theory, Local Scoring Rules, Hypothesis Testing, Business Cycle.
JEL: C12, C22, C44, C53.

---

[*]This paper has been mainly developed when the author was visiting PhD student at CREATES - 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 constant and careful supervision. He also thanks Barbara Annicchiarico, Cristina Amado, Gianna Boero, Alessandro Giovannelli, Alessandra Luati, Gael Martin, Elisa Nicolato, Alessia Paccagnini, Francesco Ravazzolo, Francesco Violante and Andrey Vasnev for their useful feedback and discussions, as well as all seminar participants to the "Introduction to Risk Management, ALM and Derivative Prices" Conference held in Aarhus University (January 2013), CREATES Lunch Seminar (May 2014) in the same University and the 34^{th} International Symposium on Forecasting in Rotterdam (July 2014). This paper has been awarded of a Travel Grant by the International Institute of Forecaster, which is particularly acknowledged. The usual disclaimer applies.

[†]All the computations in this paper have been performed by using MATLAB R2009b. The codes for replication of the figures and tables are available from the author upon request. In particular, the results on the empirical applications in Section 5 have been obtained adapting the codes developed by James Engel, who is gratefully acknowledged.

[‡]University of Rome "Tor Vergata", Department of Economics and Institutions, Via Columbia 2, 00133, Rome, Italy - *E-mail*: Emilio.Zanetti.Chini@uniroma2.it - *Phone*: (+39) 06 72595715 - *Mobile*: (+39) 340 9641055 - *Fax*: (+39) 06 2020500.

1 Introduction

Making coherent forecasts of uncertain events is one of the most critical *desiderata* for any economic or statistical model. As proved by the seminal De Finetti (1937) contribution, uncertainty has intrinsically probabilistic nature and induces the forecaster to a strategic behavior as consequence of the presence of bad incentives (Vovk and Shafer, 2005). Consider a professional forecaster who is asked by a Central Bank to give a probability to the event "recession", conditional on the series of some proxy variable like the index of Industrial Production; then, he may well see the problem in terms of his own utility function, reflecting the danger of quoting a high value, so encouraging the Bank to implement an expansive monetary policy when the crisis is not still in course, and the lesser danger of quoting a low value when subsequent checks reveal the beginning of a recession. This example motivates the need of eliciting forecaster's utility when holding with his predictions.

In Statistics, one of the first tools created in order to evaluate the goodness-of-fit of forecasts of the whole predictive density function has been the probability integral transform (PIT) (Rosenblatt, 1952; Diebold, Gunther, and Tay, 1998), still today the standard result for contemporary literature in uniformity testing; see Corradi and Swanson (2006) for a survey. Unfortunately, PITs do not conveys any information about the utility underlining forecast. Consequently, they could be unable to recognize different forecaster (or model) types. These idiosyncrasies concretize in the Hamill (2001) puzzling result that different, misspecified forecasters produces a uniformly distributed PITs; figure 1 gives an example of this puzzle for different specifications of US. Industrial production. The Hamill's Paradox makes the disclosure of the forecaster's fair incentive via alternative methodology a primary objective for econometric research.

Scoring rules are the natural candidates for the solution to this problem. They originate from the independent, pioneering contributions by Brier (1950); De Finetti (1962); Savage (1971) and entered only recently in the Econometric literature Geweke and Amisano (2011); see Gneiting and Raftery (2007) (GR) for a survey. A scoring rule assigns a numerical score to several competing (model) density forecasts in order to have a quantitative assessment of them, corresponding to different forecaster's utility. The forecaster has no incentive to quote a forecast Q of an observation x drawn from the distribution Q, if and only if the reward for such quotation is larger or equal than the one deriving from the forecast under $P \neq Q$. A scoring rule with this property is called *proper*; it's *strictly proper* if the equality verify if and only if $P = Q$. It is *(m-)local* if it depends only on its value in x and first (m-)derivatives of the predictive distribution, that is, in an infinitesimal neighborhood of the observation that materializes. In this sense, proper scoring rules incentive the forecaster to be honest[1]. Gneiting, Balabdaoui, and Raftery (2007) (GBR) demonstrate that Hamill's paradox can be solved if appropriate scoring rules are used, advocating, in this sense, for a re-discover in applied analysis of these objects[2].

This paper moves from two observations: first, density forecasts, in particular the ones generated by nonlinear models, are often multimodal (Wei and Tanner, 1990; Teräsvirta, 2006), making them prone to be confused for a process driven by a mixture of distributions (*ergo*, of utilities); see Figure 2 as an example. Second, and related, there exists a large number of proper scoring rules (but not all strictly propers), a lot of which can be nested; Tables 1 - 2 summarize the best of our knowledge of this literature. Hence the need of a criterion to select the reward that effectively drives the quoted forecast. An important hint comes from the fundamental result that a scoring rule is proper if and only if it is logarithmic. This has important inferential and computational implications, in what the locality of a scoring rule coincides with the likelihood principle (Bernardo, 1979). This mathematical result

[1] "*The scoring rule is constructed according to the basic idea that the resulting device should oblige each participant to express his true feelings, because any departure from his own personal probability results in a diminution of his own average score as he sees it*" (De Finetti, 1962, p. 359).

[2] Mitchell and Wallis (2011) severely criticize this new approach in what the data generating process used by Hamill and GBR is not robust to some basic time-series feature. This implies that traditional diagnostic tools can still be successfully applied. We argument that this theory could be still useful for a number of theoretical and applied problems.

has still not a counterfactual in Econometrics, despite the fact that it is a testable hypothesis and that it is object of a renewed interest in Mathematical Statistics (Parry, Dawid, and Lauritzen, 2012; Dawid, Lauritzen, and Parry, 2012; Ehm and Gneiting, 2012). This motivates our research question: *do data allow for local, that is logarithmic, scoring rule? And, additionally, is it really the most likely in the set of well-known scorings?*

The applied literature assumes somehow that scoring rules to be known by forecast user when assessing the forecaster quotation, whilst this is exactly one of the source of uncertainty in the game. One of the first solution to circumvent this problem, introduced in Climatology, has been to standardize the average of the estimated scores (Murphy, 1974). In Econometrics, the most common strategy is comparing and combining average predictive densities in the spirit of Bates and Granger (1969) *via* Kullback-Liebler Information Criterion (Bao, Lee, and Saltoglu, 2007; Mitchell and Hall, 2005; Hall and Mitchell, 2007; Amisano and Giacomini, 2007) or bayesian methods (Kascha and Ravazzolo, 2010; Ravazzolo and Vahey, 2013). Two valuable exceptions are constituted by Patton and Timmermann (2007) and Boero, Smith, and Wallis (2011): in the first case, the loss function is assumed unknown with consequences for optimality of FED predictions of U.S. output gap, while, in the second, the loss function is selected via Giacomini and White (2006) test for equal predictive ability; however, the role of locality remains still not investigated or is treated as an assumption.

We fill this gap by treating the scoring rule as an endogenous transition variable in a Smooth Transition Autoregressive (STAR) model. This allows us to use a frequentist framework and rely on a well-known literature to build-up a test for the null hypothesis of locality of the scoring rule underlining the specified model's density forecast. The scoring rules and the tests are then used to evaluate the probability of recession events and to assess the density forecast from four economic time series. Our results confirm the feasibility of the locality test and the theoretical properties of the so treated score and underline the strength of locality assumption in many economic time series and the role of the scoring rule selection.

The rest of the paper is so organized: Section 2 discusses the theoretical framework; Section 3 builds-up the test; Sections 4 and 5 show the results of the simulation study and two illustrations, respectively; Section 6 concludes.

2 Theoretical framework

We consider a time series $Y_t = \{y_t\}_{t=1}^{T} = \{y_1, \ldots, y_t, \ldots, y_T\}^{\intercal}$, with $\intercal$ denoting the transposition, which is fully represented by a conditional density $P(Y_t)$ on the information set $\{\Omega = y_{t-1}, y_{t-2}, \ldots\}$ and $F(Y_t)$ the associated cumulative distribution function (cdf). Consequently the (1-step-ahead) probability forecast is $P(Y_{t+1})$. We denote the best forecaster's judgement of the distributional forecast of Y_t as $Q(Y_t)$ (we will omit Y_t for notational convenience) and x a draw of Q which materializes in $T+1$. Then, we denote the predictive cdf associated to the materialization of x as $F(x)$. To nest the theoretical framework by Vovk and Shafer (2005) and Dawid (2007), we introduce two agents: Forecaster and Nature; the first one is assumed to have an information set at most equal to the Nature one.

Let be: $\mathcal{X}$ being a set of the possible forecaster's outcomes, $\mathcal{P}$ the family of distributions on $\mathcal{X}$ in which P is belonged and $\mathcal{A}$ a σ-algebra of subset of $\mathcal{X}$ representing the set of actions. In particular, if the sample space is discrete (that is, dichotomous for events like the probability of a recession, categorical for, say, ranking position of a firm or State), $\mathcal{P} = \{\mathbf{p} \in \mathcal{A} : \sum_x p_x = 1\}$ is the set of all real vectors corresponding to strictly positive probability measures; if it is continuous (like the conditional mean of an economic time series), P is defined by $\mathcal{M}$, the set of all distributions on $\mathcal{X}$ which are absolutely continuous with respect to a $\sigma-$finite measure μ. The same is for $\mathbf{q}$.

Forecaster aims to solve a decision problem defined by the triple $(\mathcal{X}, \mathcal{A}, L(P, \mathrm{a}))$, where: $\mathcal{X}$ is previously defined; $\mathcal{A}$ is the action space; and $L(P, \mathrm{a}^*)$ is a real-valued loss function, suffered by Forecaster as effect of discrepancy on his own quotations, of the action $a^* \in \mathcal{A}$, which minimizes the expected loss computed using the density P believed the true DGP, with the expected loss denoted as $EL := \int L(P, \mathrm{a})P(Y_t)dY_t$. Let the functions $H(P) : \mathcal{P} \to \overline{\mathbb{R}}$ and $D(P,Q) : \mathcal{P} \times \mathcal{Q} \to \overline{\mathbb{R}}$ be associated to any $L(P, \cdot)$. The resulting system is defined as follows:

Definition 1 (Scoring rules, entropy functions, divergence functions)**.** We define:

i. *Scoring rule* the function $S(x, Q) := L(P, \mathrm{a}_Q)$

ii. *Entropy function* the function $H(P) := S(P, P) \equiv \sup_{Q \in \mathcal{P}} S(P, Q)$

iii. *Divergence function* the function $D(P, Q) := H(P) - S(P, Q)$

$S(\cdot, \cdot)$, $H(P)$ and $D(\cdot, \cdot)$ constitute an *unicum*, which, under proper conditions, is able to characterize coherent forecasts. Since any decision problem can be fully characterized by a (strictly) proper scoring rule, let formally define them as follows:

Definition 2 ((Strictly) Proper Scoring Rule)**.** The scoring rule $S(x, Q)$ is (strictly) proper relative to the class of probability measures $\mathcal{P}$ if

$$S(P, P) \leq S(P, Q) \quad \forall \quad P, Q \ \in \ \mathcal{P} \tag{1}$$

with equality if (and only if) $Q = P$

According to Parry, Dawid, and Lauritzen (2012), we introduce local proper scoring rules as follows:

Definition 3 (m-Local scoring rule)**.** A scoring rule $S : \mathcal{X} \times \mathcal{P} \to \mathbb{R}$ is said to be local of order m or m-local if it can be expressed in form of:

$$S(x, Q) = s(x, q(x), q'(x), q''(x), \ldots, q^{(m)}(x)) \tag{2}$$

where $s = \mathcal{X} \times \mathcal{Q}_m \to \mathbb{R}$, $\mathcal{Q}_m := \mathbb{R}^+ \times \mathbb{R}^m$ is a real-valued, infinitely differentiable function, $s()$ is called scoring function (or q-function) of $S(x, Q)$, $q()$ is the density function of Q, m is a finite integer, and the prime (') denote the differentiation with respect to x.

To get the system as defined in Definition 1 operational, we invoke the following assumptions:

A. 1. *$\mathcal{P}$ is assumed such that EL exists for all $a \in \mathcal{A}$, $P \in \mathcal{P}$.*

A. 2. *The set $\mathcal{A}$ is compact.*

A. 3. *The function $L()$ is strictly convex in* a.

A. 4. *The entropy $H(P)$ associated to $S()$ is (strictly) convex in P, integrable with respect to $P \in \mathcal{P}$ and quasi-integrable with respect to all $Q \in \mathcal{P}$ and such that H^* is a sub-tangent of H at point P.*

A. 5. *$S(P, Q)$ is affine, real-valued for all $P, Q \in \mathcal{P}$ and minimized in Q at $Q = P$.*

A. 6. *$D(P, Q) - D(P, Q_0)$ is affine in P, and $D(P, Q) \geq 0$, with equality achieved at $Q = P$*

Proposition 1. *The Forecaster's reward $S(P, Q)$ is a proper scoring rule if and only if A1 - A5 are satisfied.*

Proof. This is essentially the Theorem 1 in GR. □

Remark 1. A1 - A3 are necessary (but not sufficient) to define the Forecaster's reward as scoring rule. In particular, A1 is encloses the three "basic assumptions" discussed in (Dawid, 2007, p. 80)[3] and, in terms of GR treatment, implies that the reward is measurable with respect to $\mathcal{A}$ and quasi-integrable with respect to all $P \in \mathcal{P}$. A2 and A3 are convenience assumption which are necessary to having a

[3] That is, in our simplified notation: a) there exists exactly one $\boldsymbol{p} \in \mathcal{A}$ for any $P \in \mathcal{P}$; b) distinct distributions in P have distinct actions in $\mathcal{A}$; c) Every a $\in \mathcal{A}$ is a Bayes act for some $P \in \mathcal{P}$.

unique maximizing action. A4 characterizes the general representation of scoring rules; see Thm 1 in GR. A5 stresses the fact that Forecaster has no loss only if his DGP coincides with the Nature's one; see Thm 1 in Bernardo (1979) and GBR. A6 is fundamental to characterize a very general family of scoring rules for the case, important in applications, that every $Q \in \mathcal{P} = \mathcal{A} = \mathcal{M}$, has a density $q(x)$ with respect to $\mu \in \mathcal{X}$, the *Brègman score*:

$$S(x,Q)^B = \psi'\big[q(x)\big] + \int \Big\{\psi\big[q(x)\big] - q(x)\psi'\big[q(x)\big]\Big\}d\mu(x) \tag{3}$$

with associated *Brègman divergence* (Brègman, 1967):

$$d(P,Q)^B = \int \Big\{\Big(\psi\big[p(x)\big] + \big[p(x) - q(x)\big]\psi'\big[q(x)\big]\Big) - \psi[p(x)]\Big\}d\mu(x), \tag{4}$$

where ψ is a (strictly) concave function. This is a very general class of non-metric distance able to characterize most of the scoring rules described in Table 1. We are particularly interested in the special case $\psi = k - \lambda \log(\cdot)$, under which the system is a logarithmic scoring rule with Shannon Entropy and Kullback-Liebler distance, since this is the framework used by the literature mentioned in Section 1.

Remark 2. All the functions $S(\cdot,\cdot)$ in Definition 2 can be interpreted in terms of utility: $S(x,P)$ is the Forecaster's reward for the fact that the event x (truly) materializes. This is a function defined on the extended real line, that is $S(x,P) \in \overline{\mathbb{R}} = [-\infty, +\infty]$. Consequently the expected forecaster's reward, *conditionally to* Q can be denoted as: $S(P,Q) \equiv \int_{-\infty}^{+\infty} S(P,x)dQ(x)$. $H(P)$ can be interpreted as the maximum utility that Forecaster can achieve using Nature's true DGP to predict P. The divergence function is the difference between the maximum utility and that one achieved by predicting the quoted predictive distribution Q given the true distribution P. The necessary and sufficient conditions under which $D(P,Q)$ admits representation $S(P,P) - S(P,Q)$ are provided by Hendrickson and Buehler (1971).

The next result concerns about the representation of local proper scoring rules and constitutes the motivation for the rest of the analysis:

Proposition 2. *Let $S(x, Q)$ be a scoring rule, possibly of Brègman-type, with q-function s. Then, $S(x, Q)$ is local and strictly proper if and only if s satisfies the "key condition":*

$$\mathbb{L}s = 0, \qquad \mathbb{L} := \sum_{k \geq 0} (-1)^k \mathbb{D}^k q_0 \frac{\partial}{\partial q_k}, \quad \mathbb{D} := \frac{\partial}{\partial x} + \sum_{j>>0} q_{j+1} \frac{\partial}{q_j} \tag{5}$$

where $\mathbb{D}$ and $\mathbb{L}$ are total derivative and linear differential operators, respectively.

Proof. This is essentially the condition (i) in Theorem 6.4 in Parry, Dawid, and Lauritzen (2012). □

A local scoring rule fulfilling equation (5) is called *key score*. For theoretical reasons, other two conditions concerning the representation of s via Lagrange-Beltrami operators are needed. Nevertheless, we argue that, for applied aims, the key condition is sufficient in what it is the only empirically testable hypothesis to assess the logarithmic form of the Forecaster's reward. The next Section shows that (5) corresponds to a simple restriction in the score matrix of a general family of nonlinear models, making us able to built a LM-type test.

3 The Locality Test

This section deals with the problem to check if the "key equation" (5) is reasonable from an empirical point of view. In order to do this, we assume the q-function s generating $S(x, Q)$ as observed transition variable of a peculiar variant of the family of nonlinear models, introduced by Chan and Tong (1986); Teräsvirta (1994). This is necessary to set-up the null hypothesis and introduce an LM-type test *à la* Luukkonen, Saikkonen, and Teräsvirta (1988) (LST): the idea is to linearize the original nonlinear parametrization via Taylor expansion to arrive an auxiliary model with augmented regressors, the number of which depends on the type of non-linearity suspected; this artificial model can be investigated by standard χ^2 asymptotics. We stress the fact that the test does not impose any form for the scoring rule, because these are treated as unknown variables. The rest of the discussion is divided in two subsections: Subsection 3.1 briefly describes the model and the null hypothesis; Subsection 3.2 describes the tests statistics and selection procedure.

3.1 The Null Hypothesis

The process $\{y_t\}$ observed at $t = 1-p, 1-(p-1), \dots, -1, 0, 1, \dots, T-1, T$ is assumed to be driven by the following structure:

$$y_t = \boldsymbol{\phi}'\mathbf{z}_t + \boldsymbol{\theta}'\mathbf{z}_t G(\boldsymbol{\gamma}, h(\mathbf{w}_t, \mathbf{c}_k)) + \epsilon_t \tag{6}$$

$$G(\boldsymbol{\gamma}, h(\mathbf{w}_t, \mathbf{c}_k)) = \left(1 + \exp\left\{ -\prod_{k=1}^{K} h\big(\mathbf{w}_t, \mathbf{c}_k\big)\right\}\right)^{-1}, \tag{7}$$

where:

$$\mathbf{1}_{\{\mathbf{w}_t - \mathbf{c}_k > 0\}} h(\mathbf{w}_t, \mathbf{c}_k) = \begin{cases} \gamma_1^{-1}\exp(\gamma_1|\mathbf{w}_t - \mathbf{c}_k| - 1) & \text{if } \gamma_1 > 0, \\ 0 & \text{if } \gamma_1 = 0, \\ -\gamma_1^{-1}\log(1 - \gamma_1|\mathbf{w}_t - \mathbf{c}_k|) & \text{if } \gamma_1 < 0, \end{cases} \tag{8}$$

$$\mathbf{1}_{\{\mathbf{w}_t - \mathbf{c}_k \leq 0\}} h(\mathbf{w}_t, \mathbf{c}_k) = \begin{cases} -\gamma_2^{-1}\exp(\gamma_2|\mathbf{w}_t - \mathbf{c}| - 1) & \text{if } \gamma_2 > 0, \\ 0 & \text{if } \gamma_2 = 0, \\ \gamma_2^{-1}\log(1 - \gamma_2|\mathbf{w}_t - \mathbf{c}|) & \text{if } \gamma_2 < 0, \end{cases} \tag{9}$$

$$\mathbf{w}_t = a'\mathbf{z}_t \odot s,\ \ a = [a_1, \dots, a_p]',\ \ \mathbf{z}_t = (1, y_{t-1}, \dots, y_{t-p})',\ \ \boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_p)',$$
$$\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_p)',\ \ \mathbf{c}_k = (c_1, \dots, c_K),\ \ a_i = \begin{cases} 0 \text{ if } i = d \\ 1 \text{ if } i \neq d \end{cases}$$

with d denoting the delay parameter, $1 \leq d \leq p$, and s the score function in equation (2) of Definition 3 and $\mathbf{1}_{\{\cdot\}}$ the Heviside indicator. The vector a is introduced to stress the fact that d is unknown.

The transition function $G(\boldsymbol{\gamma}, h(\mathbf{w}_t, \mathbf{c}_k))$ is a continuous, twice differentiable function in the vector $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ and in the function $h(\mathbf{w}_t, \mathbf{c}_k)$, which in turn is strictly increasing in the transition variable $\mathbf{w}_t$ and the vector of location parameters, $\mathbf{c}_k$. Equations (6)-(9) define the Generalized Logistic STAR (GLSTAR) model. Equations (8)-(9) model the higher and lower tail of the probability function, respectively, so allowing for the asymmetric behavior introduced by the slope parameter γ_1 (γ_2) which controls the velocity of the transition. The case in which $h(\mathbf{w}_t, \mathbf{c}_k) = 0$ implies that for $\gamma_1 = \gamma_2 = 0$ ($\gamma_1 = \gamma_2 = \gamma \neq 0$), the function nests a linear AR (a symmetric logistic STAR with varying central slope) model. When $\gamma_1, \gamma_2 > 0$ ($\gamma_1, \gamma_2 < 0$), $h()$ is an exponential (logarithmic) rescaling which increases more quickly (more slowly) than a standard logistic function. The most common choices for K are $K = 1$, in which case the parameters $\boldsymbol{\phi} + \boldsymbol{\theta} G(\boldsymbol{\gamma}, h(\mathbf{w}_t, \mathbf{c}_k))$ change monotonically as a function of s_t from $\boldsymbol{\phi}$ to $\boldsymbol{\phi} + \boldsymbol{\theta}$ and $K = 2$, in which case the parameters $\boldsymbol{\phi} + \boldsymbol{\theta} G(\boldsymbol{\gamma}, h(\mathbf{w}_t, \mathbf{c}_k))$ change asymmetrically at some (undefined) point where the function reaches its own minimum. A peculiar form of this latter case is when $K = 2$ and $c_1 = c_2$ and the transition function defines the Generalized Exponential STAR (GESTAR) model. When $\gamma_1 = \gamma_2 \to \infty$, the model (6) nests a two-regimes Threshold Autoregressive model (Tong, 1983). To ease the notation, we will use $G(\boldsymbol{\gamma}, \mathbf{w}_t, \mathbf{c}_k)$, omitting the $h(\cdot, \cdot)$ function in what follows.

Let denote the log-likelihood function of the T observations by $\boldsymbol{\Lambda}_t(\mathbf{z_t}, \Xi)$ with $\Xi = [\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\gamma}, c]$ and the score vector by $\boldsymbol{\Sigma}_t(\mathbf{z_t}, \Xi)$ evaluated at $(\boldsymbol{\theta_0}, \boldsymbol{\phi_0}, \mathbf{0}, c_0)$. Then, standard results lead to the following log-likelihood function:

$$
\begin{aligned}
\boldsymbol{\Lambda}_t(\mathbf{z_t}, \Xi) &= const + \frac{T}{2} \ln \sigma^2 - \frac{1}{2} \sigma^2 \sum_t (y_t - \boldsymbol{\phi}' \mathbf{z_t} - \boldsymbol{\theta}' \mathbf{z}_t G)^2 \\
&= const + \frac{T}{2} \ln \sigma^2 - \frac{1}{2} \sigma^2 \sum_t u_t^2(\Xi),
\end{aligned}
\tag{10}
$$

with *const* and u_t denoting a constant and the model's residual, respectively, and to the score:

$$\mathbf{\Sigma}_t(\mathbf{z_t}, \Xi) = \nabla_{\Xi}\mathbf{\Lambda}_t(\mathbf{z_t}, \Xi) = \frac{1}{\sigma^2}\sum_t u_t(\Xi)\mathbf{d}_t, \tag{11}$$

$$\mathbf{d}_t = \nabla_{\Xi}u_t(\Xi) = [\mathbf{z}_t, \mathbf{z}_t G, \boldsymbol{\theta}'\mathbf{z}_t G_{\gamma_1}, \boldsymbol{\theta}'\mathbf{z}_t G_{\gamma_2}, \boldsymbol{\theta}'\mathbf{z}_t G_c]', \tag{12}$$

with $G_{\gamma_1} = \partial G/\partial\gamma_1$, $G_{\gamma_2} = \partial G/\partial\gamma_2$, and $G_c = \partial G/\partial\gamma_c$ denoting the first derivatives of G, see Appendix A1 in Zanetti Chini (2013) for analytics.

In terms of scoring rule locality, we wish to test that in (6):

$$H_{0i} : \boldsymbol{\theta} = \mathbf{0} \ \text{ vs } \ \boldsymbol{\theta} \neq \mathbf{0}. \tag{13}$$

This hypothesis system corresponds to the Key equation (5) (in particular to the null hypothesis that H_0^*: $\lambda = 0$ in the q-function s) and requires a simple LM-type test. Note that the parameters γ_i, a and c are not identified under H_0. In a similar way, we could choose $H_0' : \gamma = 0$ as our locality hypothesis, in which case neither c, a nor $\boldsymbol{\theta}$ would be identified under H_0'. This implies that the conventional maximum likelihood theory is not directly applicable to deriving test procedures for testing (13). In this case, it is still possible to proceed to built-up a null hypothesis by keeping fix the unspecified parameters, as suggested by Davies (1977).

Define $\boldsymbol{\tau} = (\boldsymbol{\tau_1}, \boldsymbol{\tau_2})'$, where $\boldsymbol{\tau_1} = (\phi_0, \boldsymbol{\phi}')'$, $\boldsymbol{\tau_2} = \boldsymbol{\gamma}$. Let $\hat{\boldsymbol{\tau}}_\mathbf{1}$ the LS estimator of $\boldsymbol{\tau_1}$ under $H_0 : \boldsymbol{\gamma} = \mathbf{0}$, $\hat{\boldsymbol{\tau}} = (\boldsymbol{\tau}_\mathbf{1}', \mathbf{0}')'$. Moreover, let $\hat{\mathbf{d}}_\mathbf{t} = \mathbf{d_t}(\hat{\boldsymbol{\tau}}) = (\hat{\mathbf{d}}_{\mathbf{1,t}}, \hat{\mathbf{d}}_{\mathbf{2,t}})$, where the partition conforms to that of $\boldsymbol{\tau}$. Under H_0, the test statistic is:

$$S(\Xi)^{LM} = \frac{1}{\hat{\sigma}^2}\hat{\mathbf{U}}'\hat{\mathbf{D}}_\mathbf{2}(\hat{\mathbf{D}}_\mathbf{2}'\hat{\mathbf{D}}_\mathbf{2} - \hat{\mathbf{D}}_\mathbf{2}'\hat{\mathbf{D}}_\mathbf{1}(\hat{\mathbf{D}}_\mathbf{1}'\hat{\mathbf{D}}_\mathbf{1})^{-1}\hat{\mathbf{D}}_\mathbf{1}'\hat{\mathbf{D}}_\mathbf{2})^{-1}\hat{\mathbf{D}}_\mathbf{2}'\hat{\mathbf{U}}, \tag{14}$$

$\mathbf{D}_i = [\mathbf{d}_{i1}, \dots, \mathbf{d}_{iT}]'$, $\hat{\mathbf{D}}_\mathbf{i} = [\hat{\mathbf{d}}_\mathbf{i1}, \dots, \hat{\mathbf{d}}_\mathbf{it}, \dots, \hat{\mathbf{d}}_\mathbf{iT}]'$, $i = \{1, 2\}$, $t = 1, \dots, T$, $\hat{\sigma}^2 = \frac{1}{T}\sum_1^T \hat{u}_t^2$ and $\hat{u}_t = y_t - \hat{\boldsymbol{\tau}}_\mathbf{1}'\mathbf{z_t}$. When the model is an GLSTAR, $\hat{\mathbf{d}}_{\mathbf{1,t}} = -\mathbf{z_t} = -(1, y_{t-1}, \dots, y_{t-p})'$ while $\hat{\mathbf{z}}_\mathbf{2t} \equiv \frac{\partial^2 u_t}{\partial\gamma\partial\gamma'}\big|_{\boldsymbol{\gamma}=\mathbf{0}} = -\frac{1}{2}\big\{\theta_{20}[y_t(y_{t-d})] - cy_t\boldsymbol{\theta}'\mathbf{z_t} + \boldsymbol{\theta}_\mathbf{2}'\mathbf{z_t}y_t y_{t-d}\big\}$

Just minor modifications are needed in notation of $\hat{\mathbf{d}}_t$ and $\mathbf{s}_t^L$ in case of LSTAR2 model due to an additional c parameter with respect to the LSTAR. Note the similarity with the original parametrization in Teräsvirta (1994) and Zanetti Chini (2013) with respect to whom we differ for the definition of the transition function and the presence of the score as multiplicative constant of the transition variable.

As explained in (Teräsvirta, 1994, p. 209), the proposed test statistic depends on θ and is still unidentified unless $\theta_2 = 0$. LST prove that this problem can be circumvented by linearizing the nonlinear model via (third order) Taylor expansion. The same argument is used here. We argue that, in this context, this linearization have a double importance in what it is a way to link the null hypothesis to a regression framework.

3.2 The Test Statistics

The proposed parametrization is not sensitive to $\boldsymbol{\theta}$ parameters, that is the model is not specified under the null hypothesis. The linearized GLSTAR model

$$\begin{aligned} y_t = \phi'\mathbf{z_t} + \theta'\mathbf{z_t}T_3\Big[&\mathbf{1}_{\{\gamma_1\leq 0,\gamma_2\leq 0\}}G\big(\boldsymbol{\gamma},\mathbf{w}_t,\mathbf{c}_k\big) + \mathbf{1}_{\{\gamma_1\leq 0,\gamma_2>0\}}G\big(\boldsymbol{\gamma},\mathbf{w}_t,\mathbf{c}_k\big)+ \\ &+\mathbf{1}_{\{\gamma_1>0,\gamma_2\leq 0\}}G\big(\boldsymbol{\gamma},\mathbf{w}_t,\mathbf{c}_k\big) + \mathbf{1}_{\{\gamma_1>0,\gamma_2>0\}}G\big(\boldsymbol{\gamma},\mathbf{w}_t,\mathbf{c}_k\big)\Big] + \epsilon_t' \ , \end{aligned} \tag{15}$$

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

$$\hat{\epsilon}_t = \hat{\mathbf{z}}_{\mathbf{1t}}'\tilde{\boldsymbol{\beta}}_{\mathbf{1}} + \sum_{j=1}^{p}\beta_{2j}sy_{t-j}y_{t-d} + \sum_{j=1}^{p}\beta_{3j}sy_{t-j}y_{t-d}^2 + \sum_{j=1}^{p}\beta_{4j}sy_{t-j}y_{t-d}^3 + v_t \ , \tag{16}$$

where v_t is a $NIID(0,\sigma^2)$ process, $\tilde{\boldsymbol{\beta}}_{\mathbf{1}} = (\beta_{10},\boldsymbol{\beta}_{\mathbf{1}}')'$, $\beta_{10} = \phi_0 - (c/4)\theta_0$, $\boldsymbol{\beta}_{\mathbf{1}} = \boldsymbol{\phi} - (c/4)\boldsymbol{\theta} + (1/4)\theta_0\mathbf{e}_d$, $\mathbf{e}_d = (0,0,\ldots,0,1,0,\ldots,0)'$ with the d-th element equal to unit and $T_3(G) = f_1G + f_3G^3$ is the third-order Taylor expansion of $G(\Xi)$, $f_1 = \partial G(\Xi)/\partial\Xi\big|_{\boldsymbol{\gamma}=\mathbf{0}}$ and $f_3 = (1/6)\partial^3G(\Xi)/\partial\Xi\big|_{\boldsymbol{\gamma}=\mathbf{0}}$, $G(\Xi)$ being defined in previous section. The null hypothesis is

$$H_0 : \beta_{2j} = \beta_{3j} = \beta_{4j} = 0 \;\; j = 1, \ldots, p, \tag{17}$$

The test statistic:

$$LM_1 = (SSR_0 - SSR)/\hat{\sigma_v}^2 \,, \tag{18}$$

with SSR_0 and SSR denoting the sum of squared estimated residuals from the estimated auxiliary regression (3.2) and under the null and alternative, respectively and $\sigma_v^2 = (1/T)SSR$, has an asymptotic χ^2_{3p} distribution under H_0.

If the model is a GESTAR(p), then it is possible to show that the corresponding auxiliary regression is

$$\hat{\epsilon}_t = \tilde{\boldsymbol{\beta}}_{\mathbf{1}}' \hat{\mathbf{z}}_{\mathbf{1}} + \boldsymbol{\beta}_{\mathbf{2}}' \mathbf{z_t} s y_{t-d} + \boldsymbol{\beta}_{\mathbf{3}}' \mathbf{z_t} s y_{t-d}^2 + v_t' \,, \tag{19}$$

where v_t' is a $NIID(0, \sigma^2)$ error term and $\tilde{\boldsymbol{\beta}}_{\mathbf{1}} = (\beta_{10}, \boldsymbol{\beta}_{\mathbf{1}}')'$, with $\beta_{10} = \phi_0 - c^2\theta_0$ and $\boldsymbol{\beta}_{\mathbf{1}} = \boldsymbol{\phi} - c^2\boldsymbol{\theta} + 2c\theta_0\boldsymbol{e}_d$; moreover $\boldsymbol{\beta}_{\mathbf{2}} = 2c\boldsymbol{\theta} - \theta_0\mathbf{e}_d$ and $\boldsymbol{\beta}_{\mathbf{3}} = -\boldsymbol{\theta}$. Thus the null hypothesis of linearity is

$$H_0' : \boldsymbol{\beta}_{\mathbf{2}} = \boldsymbol{\beta}_{\mathbf{3}} = 0, \tag{20}$$

which can be tested by the test statistic:

$$LM_2 = (SSR_0 - SSR)/\hat{\sigma}^2_{v1}, \tag{21}$$

where SSR_0 and SSR are the sum of squared residuals from (19) under the null and the alternative respectively, $\hat{\sigma}^2_{v1} = (1/T)SSR$. When the null is true, the statistic (23) is asymptotically χ^2_{2p} distributed. A peculiar case of (23) is when $\boldsymbol{\beta}_2 = 0$ as $\theta_0 = c = 0$, in which case, under the null

$$H_0' : \boldsymbol{\beta}_{\mathbf{3}} = 0, \tag{22}$$

the test has a statistic

$$LM_3 = (SSR_0 - SSR)/\hat{\sigma}^2_{v2} \tag{23}$$

distributed as a χ^2_p, with SSR_0, SSR and σ_{v2} defined in a similar way with respect the LM_2 case.

The F-version of LM_1, LM_2 and LM_3, denoted as F_1, F_2 and F_3, may be preferable when testing (18) or (23) in order to preserve power in low samples. In practice the form of G is not known by the investigator. Teräsvirta (1994) proposes a battery of F-tests on the auxiliary model :

$$H_{01} : \beta_4 = 0 \quad \text{vs} \quad H_{11} : \beta_4 \neq 0 \tag{24}$$

$$H_{02} : \beta_3 = 0|\beta_4 = 0 \quad \text{vs} \quad H_{12} : \beta_3 \neq 0|\beta_4 = 0 \tag{25}$$

$$H_{03} : \beta_2 = 0|\beta_3 = 0 \ \text{ and } \ \beta_4 = 0 \quad \text{vs} \quad H_{22} : \beta_2 \neq 0|\beta_3 = 0 \ \text{ and } \ \beta_4 = 0. \tag{26}$$

The next Section 4 shows some results from this LST-derived test.

4 Simulation Study

This section investigates the empirical properties of the proposed locality test by a Monte Carlo experiment. We organize this section as follows: Sub-section 4.1 describes the design of the experiment; Sub-section 4.2 reports the results; Sub-section 4.3 provides a brief discussion.

4.1 Simulation Design

We consider two different data generating processes (DGP):

$$y^{(n)}_{1,t} = 0.4y^{(n)}_{1,t-1} - 0.25y^{(n)}_{1,t-2} + (0.01 - 0.9y^{(n)}_{1,t-1} + 0.795y_{1,t-2})^{(s)}G^{(n)}(\Xi) + \epsilon^{(n)}_{1,t} \ , \tag{27}$$

and

$$y_{2,t}^{(n)} = 0.8y_{2,t-1}^{(n)} - 0.7y_{2,t-2}^{(n)} + (0.01 - 0.9y_{2,t-1}^{(n)} + 0.795y_{2,t-2}^{(i)})G^{(n)}(\Xi) + \epsilon_{2,t}^{(n)} \ , \quad (28)$$

where $G^{(i)}(\Xi)$ has the same step form as in (15), $\epsilon_t^{(n)} \sim N(0,1)$, $n = \{1, \ldots, N\}$ denoting the n-esim drow of the process $\{y_t\}_{t=1}^T$ with $s = y_{t-1}$, $c = \frac{1}{T}y_t^{(i)}$, $N = 1,000$. $y_{2,t}^{(n)}$ (henceforth "DGP 1") is an additive nonlinear model with accentuated nonlinear behavior, due to the high autoregressive parameters driving $G(\Xi)$ that gave a high sensitivity to the size of the slope parameters; this can be the case of a macroeconomic indicator affected by an unexpected shocks affecting the whole dynamics. On the other hand, $y_{2,t}^{(n)}$ (henceforth "DGP 2") describes a mixed scenario.

In order to simulate the function $h(\mathbf{w}_t, \mathbf{c})$ we use a set of values of vector $\boldsymbol{\gamma}$. These combinations allow us to investigate: i) the different cases of null, small, medium, extreme asymmetry respectively; ii) the effect of having different kinds of asymmetry, due to the different signs in the two γ-s. Moreover, we consider three different hypotheses for T and the size α, namely $T = \{100, 300, 1000\}$ and $\alpha = \{0.01, 0.05, 0.10\}$. The first 100 simulations have been discarded in order to avoid the initialization effect.

4.2 Results

The results of the simulation on locality test for all the F-statistics from the nested hypothesis system 24 discussed in Section (3) are reported in Table 3. Several findings can be easily noticed: first, the two tests tends to be well-behaving for what concerns the empirical size. Second, and conversely, the empirical power is poor if an almost linear model is used, and in general for DGP 2. Second, the power is sensitive to the signs of the slope parameters: the F_1-statistic in DGP 1 with T=300 and 1000 evidence that a highly negative γ kills all the power that one could acheive if both parameters are positive, see columns 1-3, raws 22-24 and compare with rows

34-36 at the same columns. However, under low-asymmetric case, the sign effect is limited (for example, the cases of $\boldsymbol{\gamma} = (0.5, 1)$ and $\boldsymbol{\gamma} = (-0.5, 1)$). This last finding degenerates in the case of Statistics F_3, which is not sensitive in any exercise.

Table 4 reports the results of a similar exercise with the same scoring rules defined in Table 1, apart the previously investigated logarithmic score. We focus on the extreme nonlinearity case, by setting the slopes to $\gamma_1 = 200$ and $\gamma_2 = 100$, respectively. Two facts stand out immediately: first, that the power of the test is not affected by different scoring rules. Second, that the result is more extreme than the logarithmic case for many aspect: i) the disparity of the power for different sample sizes; see, for example Table 3, column 3 and compare it with the same one of Table 4: we pass from a range of [0.12 - 0.28] of the logarithmic case to a 0.14 for all the scores for the case of medium sample of $T = 300$, while for the long sample case, we pass from a (longer) range [0.27 - 0.81] to a full power of 1.000; ii) the F_3 statistics annihilates completely; iii) on the other side, an horizontal comparisons of the F_1 vs F_2 statistics suggests a tendence to a more uniform behavior: if one compares columns 1-3 with columns 4-6 for both the tables, would see that the range of the powers for similar nominal and sample size decays ([0.19 - 0.28] against [0.06 - 0.14] for medium sample and [0.74 - 0.81] vs [0.99 - 1.00] for long sample.

4.3 Discussion

These results convey a non-trivial picture of the role of the locality of the scoring rules in applied forecasting. In terms of statistical theory, we hold with two problems:

Model dependance: Estimating a strongly nonlinear model underlining the predictive density is the only way to capture such an information, despite on the form of the scoring rule.

Score Invariance: In any case, Statistician is still apparently unsure on the form of the Forcaster's expected utility reward because of the uniformity of the results for different scores.

Heuristically, we can reply to these idiosyncrasies as follows. For the model dependance problem, the scoring rule, although treated as transition variable, is per se exogenous to the (G)STAR model parametrization, because it is just a functional of its distribution. In addition, we used the implicit assumption that scoring rule has an observable variable, when the converse is considerably more appropriate from a theoretical point of view. It is not so unrealistic to think that such a direct application of the LST test to this object is responsible to add some nonlinearity not captured by the AR parameters, hence leading to a spurious rejection of the null of locality symmetry when it is true. Moreover, the generalized logistic function is known to having a near-to-linear behavior under extreme negative γ parameters, see Zanetti Chini (2013).

The Score Invariance is one of the critical assumptions of the Lindley (1982) generalized theory on the admissibility of the Forecaster' reward and clearly stated as a condition for treating the scores as finitely-additive probability-behaving objects. In addition, the Lemma 4 of the same work proves the equivalence between two different scores on the same variable conditional on some event, with different quotations, enhancing in a such a way the status of probability transform of the obtained value x. In this sense our simulations are fully consistent with the theory.

5 Illustrations

5.1 Characterizing the U.S. Business Cycle

In this application, we evaluate the US business cycle according to the "Bry-Boshan Quarterly" (BBQ) algorithm, introduced by Harding and Pagan (2002); Engel, Haugh, and Pagan (2005) (EHP), which in turn is an advancement of the pre-existing Bry-Boshan Algorithm for the detection of turning points. Here, the focus is twofold: the evaluation of the asymmetry of the business cycle by following the literature started with Burns and Mitchell (1946) and the assessment via scoring rule of the probability of recession in line with Diebold and Ruderbusch (1989) (DR). We use the US Index of Industrial Production in quarterly data as proxy variable in logarithmic transformation. The source is Federal Reserve Bank of St. Louis, Research Division[4]. The sample goes from 1947Q1 to 2013Q1. A couple of *caveat* should be noted: first, our choice of the EHP methodology does not imply a preference for it and should be only considered as a first step for a more complete treatment of elicitation of business cycle, which can be done via more elaborated methods like Markov-Switching approach by Artis, Marcellino, and Proietti (2004). Second, differently to DR, we do not assume a bayesian strategy for the replication of probabilities of downturn. The results of the test should be looked in light of Teräsvirta's modelling strategy.

Understanding the business cycle means to realize how the transition probabilities are likely to be affected by the nature of Δy_t, which is the scope of the EHP dating algorithm, in which the phases are defined as $S_t = 1(\Delta y_t > 0)$, i.e. when there is a positive growth rate at time t, the economy is in a state of expansion (E), while a negative one refers to a contraction (C). Under random walk hypothesis, the probabilities of a change in phase at time t is

$$\begin{aligned} Pr(EC) : Pr(S_{t+1} = 0|S_t = 1) = Pr(\Delta y_{t+1} < 0|\Delta y_t > 0) \\ Pr(CE) : Pr(S_{t+1} = 1|S_t = 0) = Pr(\Delta y_{t+1} > 0|\Delta y_t < 0) \end{aligned} \tag{29}$$

A set of restrictions have to be imposed in order to define properly what movement can be defined as cycle and what is contrarily treated as outlier. The literature suggests: $K = 2$ for definition of peak, which produces the expression (30); the minimal duration to have a complete cycle is 5 quarters; the turning phase (the number of quarters before the peak) has been set at 1. The results of original EHP algorithm are shown in Figure 3. Our innovation consists in understanding how

[4]Data can be downloaded from: http://research.stlouisfed.org/fred2.

the transition probabilities are likely to be affected by the nature of Δy_t *given* the estimated scoring rule(s).

Let the turning point dates produce K phases of expansions and contractions, with an index $(i = 1, ..., K)$ denoting the i-th expansion or contractions. Then the definition of peaks and troughs can be stated follows:

$$\begin{aligned} \text{peak at } t &= \{(\Delta_2 y_t, \Delta y_t) > 0, (\Delta y_{t+1}, \Delta_2 y_{t+2}) < 0\} \\ \text{trough at } t &= \{(\Delta_2 y_t, \Delta y_t) < 0, (\Delta y_{t+1}, \Delta_2 y_{t+2}) > 0\} \end{aligned} \tag{30}$$

with $\Delta_2 y_t = y_t - y_{t-2}$. Additionally we are interested the following indicators of cycle:

- Duration: D_i
- Amplitude: A_i
- Cumulative Movements: $0.5 \cdot (D_i \cdot A_i)$
- Excess cumulated movements: $E_i = (C_{Ti} - C_i + 0.5 \cdot A_i)/D_i$
- Coefficient Variation: $CV_i = \frac{\sqrt{(1/K \sum_{i=1}^{K} (D_j^i - \bar{D}^i))}}{1/K \sum_{j=1}^{K} D_j^i}$

The indicators stated above can be thought as two states of a Markov process, whose transition is the object of investigation. Thus, a peak at time t demarcates a state of expansion at time t, $S_t = 1$, from a state of contraction at time t $+1$, $S_{t+1} = 0$. The states S_t are then a binary Markov process, which might be summarized with the transition probabilities $Pr(S_{t+1}|S_t)$.

Model selection has been conducted via Bayesian Information Criterion (BIC); this suggests an autoregressive order 5. We first estimated a linear AR(5) model; the estimated model are the basis for computing the h-step ahead density forecast via Monte Carlo simulation. A set of 26 different scoring rules is then achieved and used to proceed with our modified version of EHP algorithm, which we call Scored Bry-Boshan (SBB) algorithm. Figure 4, panel (a) plots the series and the estimated recession indicator via SBB , while panel (b) the corresponding fan chart. We notice how reciprocally similar are the pattern of the scored and non scored algorithm, with the noticeable difference of the last Great Recession, which is already recovered only if the SBB is used. Table 5 shows the effects of adopting the scoring rules in BBQ algorithm. Two main findings immediately emerges: first, the high duration of the expansion phase with respect to the contraction, in line with all the literature. Secondly, the computed value of many indicators seems to be unaffected by different scoring rules. This seems the most counter-intuitive result, since, being each scoring rule different (consistently, in some cases) each from others, we would expect different values for each of them. On the other side, this seems coherent with the Score Invariance principle empirically proved in the previous section. This result could be the effect of the fact that many of the selected scoring rules are nested.

We then repeat the analysis by estimating a Multiple Regime-STAR model of order 5 in both the linear and nonlinear part and a delay of 4; the data suggest the presence of three states and consequently two transitions. Figure 5 summarizes the main findings: panel (a) shows the estimated series and the corresponding SBB algorithm: the final part of the data are considered still in a different way with respect to the AR case (expansion just for the last observation); panel (b) shows the fan chart for 60 step-ahead density forecasts: the slope parameter ($\gamma_1 = 0.493$) leads to a really smooth transition between the phases of recessions (G=0) and expansions (G=1), while the second transition is clearly nonlinear ($\gamma_2 = 2.046$); panel (d) plots the same transition function against time: the second transition recalls almost perfectly the observed data. Table 6 shows the effects of adopting the scoring rules in BBQ algorithm under the estimated MR-STAR model. All the indicators shows that the asymmetry of the cycle is increased, and in some cases, exacerbated (see the Excess of % triangle area).

The effects of SBB algorithm are graphically shown in Figure 6. In panel (a) a Weighted Power Score is introduced in the EHP algorithm on the estimated AR(4) process; panel (b) shows the effect of Logarithmic Score on the estimated STAR model; this last scoring rule seems clearly over-estimating the recession phases; the density forecasts of that processes are shown in panel (c) and panel (d) respectively; notice how the median of the predictive densities is slightly increasing with respect to that of the (non scored) processes. This is an interesting feature, since it seems to detect an increasing trend structure after the Great Recession.

5.2 Assessing univariate forecasts

In this application we compute the locality test statistic in four univariate time series which are of particular interest in applied economics, namely the U.S.index of industrial production (IIP) unemployment rate (UN), the long-term spread of the interest rate between Italy and Germany (SPR) and Norway's Inflation rate (NIRF). We first apply the locality test to U.S. industrial production in growth rates, at the light of the Business Cycle analysis in previous Section 5.1. The series go from 1948:01 to 2013:01. The BIC suggest an order of 15; the smooth nonlinear model conveys a bimodal estimated density forecast, see Figure 7, panel (a). The monthly U.S. rate of unemployment contains 783 observations from 1948:01 to 2013:03. The BIC suggest an order of 15. The highly nonlinear model conveys an estimated density forecast with skewness effect (Figure 7, panel (b); still, the null of locality is rejected. The series of the long-term interest spread between Italy and Germany contains 266 observations from 1991:03 to 2013:05 and presents similar characteristics in terms of predictive distribution, as shown 6, panel (c). Our results show that a GSTAR(3) is a good representation of the data. Inflation rate is one of the most important variable in applied macro and is constantly monitored by Central Banks. The case of Norway is an interesting example of due to its nonlinearity, as mentioned in Kascha and Ravazzolo (2010). The series for Norway inflation rate goes from 1985:01 to 2013:12 and has been downloaded by Norges Bank website. Still BIC suggests $p = 3$.

Table 7 shows the results of the Locality test introduced in Section 3: the null hypothesis of logarithmic structure of the forecaster reward is strongly rejected in all the cases.

This unexpected results improve the need to investigate better the properties of the scoring rules of the estimated GSTAR(p) models. To this end we used a linear AR(p) as benchmark; the results are shown in Table 8. Several findings emerges: first, some of the estimated scoring rules are not able to discriminate between the two models, in particular the quadratic and the PseudoSpherical family, the PseudoSpectrum and Logarithmic scores. Second, the general finding for the preference for linear specifications seems generally confirmed; however, some important remarks have to be noticed: in two cases (Weighted Power Scores with $\alpha = 1/2$ and $\alpha = 2$) the score computed by the GSTAR model is less than the AR; in three cases (Energy, Generalized Energy and Hyvaarinen and PseudoSpherical Scores with $\alpha = 1$) is the AR forecast is only weakly preferable to the nonlinear one. Finally, we notice that this last case does not coincide with its limiting case, the Logarithmic Score.

5.3 Bank of Norway's Fan Chart

The output gap (OG) measures the percentage deviation between GDP and projected potential GDP and is one of the most important variables measured in applied macro. Assessing the predictive ability of the out-of-sample estimates of OG is thus currently done in several Central Banks. We nest our analysis by using the t-test by Diebold and Mariano (1995) and extended by Amisano and Giacomini (2007) and Gneiting and Ranjan (2011). Let be $\bar{S}_n^f = \frac{n-k-1}{\sum}{}^{m+n-k}_{t=m} S(\hat{f}_{t+k}, y_{t+k})$ and $\bar{S}_n^g = \frac{n-k-1}{\sum}{}^{m+n-k}_{t=m} S(\hat{g}_{t+k}, y_{t+k})$ the average scores of two density forecast, f and g, respectively; then, the test statistic for the null hypothesis that $\Delta^* = \bar{S}_n^f - \bar{S}_n^g = 0$ is

$$t_n = \sqrt{n}\frac{\Delta^*}{\hat{\sigma}_n}, \quad \hat{\sigma}_n^2 = \frac{1}{n-k+1} \sum_{-(k-1)}^{++1} \sum_{m}^{m+n-k-|j|} \Delta_{t,k}\Delta_{t+|j|,k} \tag{31}$$

and $\Delta_{t,k} = S_n^f - S_n^g$. In this kind of analysis, f is preferable to g if and only if $S^f < S^g$.

The Bank of Norway's Monetary Policy Report (MPR) has issued probabilistic forecasts of OG since March 2008 to December 2017, by using fan charts to visualize the deciles of the predictive distributions. The quarterly Bank of Norway OG is available online at http://www.norges-bank.no/en/about/published/publications/monetary-policy-report/ . The rates are percentage changes over 12 months. The first quarter extends from March 31st to May 30, the second quarter from July 1st to September 30th, and so on. Following the example of Gneiting and Ranjan (2011), we compare the Bank of Norway's density forecasts of OG to those derived from a simplistic autoregressive time series model. The simplistic competitor is a Gaussian autoregression of order 1 that uses a rolling estimation window of length $m = 6$ quarters. This method results in Gaussian density forecasts. Table 9 compares the two methods at a prediction horizon of $k = 1$ quarters ahead, for a test period ranging from the first quarter of 2008 to the first quarter of 2017, for a total of $n = 34$ density forecast cases. The superiority of the Bank of Norway is not unambiguously clear, how shown by the different values τ and P-values. Moreover, in several occasion, the benchmark model is superior, as demonstrated by some of the WPowerS cases and CRPS.

6 Conclusions

We introduced a LM-type test to verify the hypothesis of locality of the reward of forecaster's expected utility. The nonlinearity of the time series process underlining the predictive density forecast allows us to see the scoring rule as transition variables, and consequently to nest tools of a very well-known and established literature. This finding impacts on the power properties of the test, which depends on the length of the two slope parameters. Anyway, the power properties of the test are consistent with the statistical theory.

An application to Business Cycle and to assessment of univariate several time series density forecast is provided. The results can be so summarized: (a) the asymmetry in the cycle is exacerbated; (b) Recessions tend to be over-evaluated; (c) the measures of cycle tend to be not sensitive to different scores; (d) this last effect is considerably weaken if the model underlining the predictive density and its score is generated by a nonlinear model; (e) the density forecasts based on scored models seem to be more reliable than simple model-based ones; (f) the hypothesis of logarithmic structure of the forecaster reward, necessary for likelihood principle to hold is a too strong assumption in four cases on four. This last point seems to us the most astonishing one because, if confirmed, it would imply that the systematic failure of Forecast User to assess the Forecaster's honesty.

From a methodological point of view, this explains the importance of the use of scoring rule as way to help the econometrician, in particular if holding with nonlinear forecasting. On the other side, a careful selection of the scoring rule on the many known in statistical literature is a primary need for applied aims.

The so presented analysis is quite still a stylized way to see the more general problem of the elicitation of the true Forecaster's incentive and can be generalized in many ways. For example, the scoring rule has been here assumed unknown but observed to exploit the properties of STAR family of models. Relaxing this last implicit assumption via unobserved component models could be an interesting development. Finally, the counterintuitive results of our applications makes a deeper theoretical investigation on the links between Locality test and other Equal Predictive ability tests necessary.

References

Amisano, G., and R. Giacomini (2007): "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," *Journal of Business and Economic Statistics*, 25(2), 177–190.

Artis, M., M. Marcellino, and T. Proietti (2004): "Dating Business Cycles: A Methodological Contribution with an Application to the Euro Area," *Oxford Bullettin of Economics and Statistics*, 66(4), 537–565.

Bao, Y., T. Lee, and B. Saltoglu (2007): "Comparing Density Forecast Models," *Journal of Forecasting*, 26(3), 203–225.

Bates, J., and C. Granger (1969): "The combination of forecasts," *Operations Research Quarterly*, 20, 451–468.

Bernardo, J. (1979): "Expected Information as Expected Utility," *The Annals of Statistics*, 7, 686–690.

Boero, G., J. Smith, and K. Wallis (2011): "Scoring rules and survey density forecasts," *International Journal of Forecasting*, 27, 379–393.

Brègman, L. (1967): "The Relaxation Method of Finding the Common Point of Convex Sets and Its Application to the Solution of Problems in Convex Programming," *USSR Computational Mathematics and Mathematical Physics*, 7, 200–217.

Brier, G. (1950): "Verification of the forecasts Expressed in Terms of Probability ," *Monthly Weather Review*, 78, 1–3.

Burns, A., and W. Mitchell (1946): "Measuring Business Cycles," National Bureau of Economic Research.

Cervera, J., and J. Munoz (1996): "Proper Scoring Rules for Fractiles," in *Bayesian Statistics 5*, ed. by J. Bernardo, J. Berger, A. Dawid, and A. Smith, pp. 513–519. Oxford University Press, Oxford, UK.

Chan, K., and H. Tong (1986): “On estimating thresholds in autoregressive models,” *Journal of Time Series Analysis*, 7(3), 178–190.

Corradi, V., and N. S. Swanson (2006): “Predictive Density Evaluation,” in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann. North Holland.

Davies, R. (1977): “Hypothesis Testing When a Nuisance Parameter is Present only Under the Alternative,” *Biometrika*, 64(1), 247–254.

Dawid, A. (2007): “The geometry of proper scoring rules,” *The Annals of the Institute of Statistica Mathematics*, 59, 77–93.

Dawid, A., S. Lauritzen, and M. Parry (2012): “Proper Local Scoring Rules on Discrete Sample Space,” *The Annals of Statistics*, 40(1), 593–608.

Dawid, P., and P. Sebastiani (1999): “Coherent Dispersion Criteria for Optimal Experimental Design,” *The Annals of Statistics*, 27, 65–81.

De Finetti, B. (1937): “La prévision : ses lois logiques, ses sources subjectives,” *Annales de l'institut Henri Poincarè*, 7, 1–68.

——— (1962): “Does it make sense to speak of "good probability appraisers"?,” in *The Scientist Speculates*, ed. by I. Good. Wiley, New York.

Diebold, F., and R. Mariano (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 20(1), 134–144.

Diebold, F., and G. Ruderbusch (1989): “Scoring the Leading Indicators,” *The Journal of Business*, 62, 369–391.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998): “Evaluating Density Forecasts With Applications to Financial Risk Management,” *International Economic Review*, 39(4), 863–883.

Diks, C., V. Panchenko, and D. van Dijk (2011): "Likelihood-based scoring rules for comparing density forecasts in tails," *Journal of Econometrics*, 163(2), 215–230.

Eaton, M., A. Giovagnoli, and P. Sebastiani (1996): "A Predictive Approach to the Bayesian Design Problem with Application to Normal Regression Models," *Biometrika*, 83, 111–125.

Ehm, W., and T. Gneiting (2012): "Proper Local Scoring Rules on Discrete Sample Space," *The Annals of Statistics*, 40(1), 609–637.

Engel, J., D. Haugh, and A. Pagan (2005): "Some methods for assessing the need for non-linear models in business cycle analysis," *International Journal of Forecasting*, 21, 651–662.

Epstein, E. (1969): "A scoring system for probability forecasts of ranked categories," *Journal of Applied Meteorology*, 8, 985–987.

Geweke, J., and G. Amisano (2011): "Optimal prediction pools," *Journal of Econometrics*, 164, 130–141.

Giacomini, A., and H. White (2006): "Tests for Conditional Predictive Ability," *Econometrica*, 74(6), 1545–1578.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007): "Probabilistic forecasts, calibration and sharpness," *Journal of Royal Statistical Society*, 69(2), 243–268.

Gneiting, T., and A. Raftery (2007): "Strictly Proper Scoring Rules, Prediction and Estimation," *Journal of the American Statistical Association*, 102(477), 359–378.

Gneiting, T., and R. Ranjan (2011): "Comparing Density Forecasts Using

Threshold- and Quantile-Weighted Scoring Rules," *Journal of Business & Economic Statistics*, 29(3), 411–422.

Good, I. (1952): "Rational Decisions ," *Journal of Royal Statistical Society, Ser. B*, 14, 107–114.

——— (1971): "Comment on "Measuring Information and Uncertainty" by R. J. Buheler," in *Foundations of Statistical Inference*, ed. by V. Godambe, and A. Sprott, pp. 337–339. Holt, Rinehart and Winston, Toronto.

Hall, S., and J. Mitchell (2007): "Combining Density Forecasts," *International Journal of Forecasting*, 23(1), 1–13.

Hamill, T. (2001): "Interpretation of rank histograms for verifying ensemble forecasts," *Monthly Weather Review*, 125, 550–560.

Harding, D., and A. Pagan (2002): "Dissecting the cycle: a methodological investigation," *Journal of Monetary Economics*, 49, 365–381.

Hendrickson, A., and R. Buehler (1971): "Proper Scores for Probability Forecasters," *The Annals of Mathematical Statistics*, 42(6), 1916–1921.

Hyvärinen, A. (2005): "Estimation of non-normalized statistical models using score matching," *Journal of Machine Learning Research*, 6, 695–709.

Jose, V., R. Nau, and R. Winkler (2008): "Scoring Rules, Generalized Entropy, and Utility Maximization," *Operation Research*, 56, 1146–1157.

Kascha, C., and F. Ravazzolo (2010): "Combining Inflation Density Forecasts," *Journal of Forecasting*, 29(1–2), 231–250.

Lindley, D. (1982): "Scoring Rules and the Inevitability of Probability," *Revue Internationale de Statistique*, 50, 1–11.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988): “Testing linearity against smooth transition autoregressive models,” *Biometrika*, 75(3), 491–499.

Matheson, J., and R. Winkler (1972): “Scoring Rules for Continuous Probability Distributions,” *Management Science*, 22, 1087–1096.

Mitchell, J., and S. Hall (2005): “Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR "Fan" Charts of Inflation,” *Oxford Bulletin of Economics and Statistics*, 67(S1), 995–1033.

Mitchell, J., and K. Wallis (2011): “Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness,” *Journal of the Applied Econometrics*, 26(6), 1023–1040.

Murphy, A. (1974): “A Sample Skill Score for Probability Forecasts,” *Monthly Weather Review*, 102, 48–55.

Parry, M., A. Dawid, and S. Lauritzen (2012): “Proper Local Scoring Rules,” *The Annals of Statistics*, 40(1), 561–592.

Patton, A., and A. Timmermann (2007): “Testing Forecast Optimality Under Unknown Loss,” *Journal of American Statistical Association*, 102, 1172–1184.

Ravazzolo, F., and S. Vahey (2013): “Forecast densities for economic aggregates from disaggregate ensembles,” *Studies of Nonlinear Dynamics and Econometrics*, forthcoming.

Rosenblatt, M. (1952): “Remarks on a Multivariate Transformation,” *Annals of Mathematical Statistics*, 23(3), 470–472.

Savage, L. (1971): “Elicitation of Personal Probabilities and Expectations,” *Journal of American Statistical Association*, 66, 783–801.

Teräsvirta, T. (1994): "Specification, estimation and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association*, 89(425), 208–218.

——— (2006): "Forecasting Economic Variables with Nonlinear Models," in *Handbook of Econometric Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann. Elsevier, New York.

Tong, H. (1983): *Threshold Models in Non-Linear Time Series Analysis*, no. 21 in Lecture Notes in Statistics. Springer-Verlag, New York.

Tsallis, C. (1988): "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, 52, 479–487.

Vovk, V., and G. Shafer (2005): "Good randomized sequential probability forecasting is always possible," *Journal of Royal Statistical Society, ser. B*, 67(747-763), 491–499.

Wei, G., and M. Tanner (1990): "Calculating the Content and Boundary of the Highest Posterior Density Region via DataAugmentation," *Biometrika*, 77(3), 649–652.

Winkler, R. (1972): "A Decision-Theoretic Approach to Interval Estimation," *Journal of American Statistical Association*, 67, 187–191.

Zanetti Chini, E. (2013): "Generalizing smooth transition autoregressions," CREATES Research Paper 2013-32.

A Appendix

A.1 Tables and Graphs

Table 1: Scoring Rules for continuous variables and their features

Score	S(P,x)	H(p,x)	Measure	$d(P,Q)$	Brègman type	Reference
QS	$2p(x) - \|\|p\|\|_2^2$	$\|\|p(x)\|\|_2^2$	L_2	$\|\|p-q\|\|_2^2$	Yes	Brier (1950)
LogS	$k \log p(x)$	$\sum_{j=1}^m p \log p$	L_2	$\sum_j q_j \ln(\frac{q}{p})$	Yes	Good (1952)
RPS	$\int \{\|Q(A_t) - 1_{A_t}(x))\}^2 d\mu(t)$	$\int P(A_t)\{1-P(A_t)\}d\mu(t)$	μ	$\int \{P(A_t)Q(A_t)\}^2 d\mu(t)$	No	Epstein (1969)
PseudoSph	$\frac{p(x)^{\alpha-1}}{\|\|p\|\|_\alpha^{1-\alpha}}$	$\|\|p\|\|_\alpha$	L_α	$\|\|p\|\|_\alpha$	No	Good (1971)
IntS	$(u-l) + \frac{2}{\alpha}(l-x)I_{(x<l)} +$ $+\frac{2}{\alpha}(x-u)I_{(x>u)}$	$\int S_\alpha^{int} dp(x)$	$\mathcal{P}$	$\|\|p\|\|_\alpha$	No	Winkler (1972)
CRPS	$\frac{1}{2}E_F\|\|X-X'\|\| - E_F\|\|X-x\|\|$	$\frac{1}{2}E_F\|\|X-X'\|\|$	$\mathcal{P}_1$	$\int_{-\infty}^{+\infty}(F(x)-G(x))$	No	Matheson and Winkler (1972)
TsallisS	$\frac{k}{d(x)-1}\sum_{t=1}^W p_i(x)(1-p_i(x)^{d-1})$	$-\sum p(x)^d$	L	$\sum p(x)q(x)^{(d-1)} -$ $-(d-1)H(Q) - H(P)$	Yes	Tsallis (1988)
PseudoSpectrum	$-\|\phi_P(\mathbf{y}) - e^{i<\mathbf{x},\mathbf{y}>}\|^2$	$-\|\phi_P(\mathbf{y})\|$	$\mathcal{P}$	$\int_u \|\|\alpha - \beta\|\|^2$	No	Eaton, Giovagnoli, and Sebastiani (1996)
DispersionS	$K(Q_V) + tr\{\mathbf{V}_P - \mathbf{V}_Q\mathbf{\Gamma}_Q\} -$ $-(\mathbf{x}-\boldsymbol{\mu}_P)'\mathbf{\Gamma}_P^{-1}(\mathbf{x}-\boldsymbol{\mu}_P)$	$-\log\det\mathbf{\Gamma}_P - mK$	$\mathcal{P}$	$tr(\mathbf{\Gamma}_P^{-1}\mathbf{\Gamma}_Q) - \log\det(\mathbf{\Gamma}_P - 1\mathbf{\Gamma}_Q) +$ $+ (\boldsymbol{\mu}_P - \boldsymbol{\mu}_Q)'\mathbf{\Gamma}_P^{-1}(\boldsymbol{\mu}_P - \boldsymbol{\mu}_Q) - K$	Yes	Dawid and Sebastiani (1999)
Hyvaärinen	$((\ln q)'(x)^2 + 2(\ln q)''(y)$	$E_P p(x) \bigtriangledown \ln p(x)$	L	$\frac{1}{2}\int p(x)\|\bigtriangledown \ln p(x) - \bigtriangledown \ln q(x)\|dx$	Yes	Hyvärinen (2005)
ES	$\frac{1}{2}E_F\|\|\mathbf{X}-\mathbf{X}'\|\|^\beta - E_F\|\|\mathbf{X}-\mathbf{x}\|\|^\beta$	$\frac{1}{2}E_F\|\|\mathbf{X}-\mathbf{X}'\|\|$	$\mathcal{P}_\beta$	$\int_{-\infty}^{+\infty}(F(\mathbf{x})-G(\mathbf{x}))$	No	Gneiting and Raftery (2007)
GES	$\frac{1}{2}E_F\|\|\mathbf{X}-\mathbf{X}'\|\|_\alpha^\beta - E_F\|\|\mathbf{X}-\mathbf{x}\|\|_\alpha^\beta$	$\frac{1}{2}E_F\|\|\mathbf{X}-\mathbf{X}'\|\|_\alpha^\beta$	$\mathcal{P}$	$\int_{-\infty}^{+\infty}(F(\mathbf{x})-G(\mathbf{x}))$	No	Gneiting and Raftery (2007)
WPower	$\frac{(p_i/q_i)^{\beta-1}-1}{\beta-1} - \frac{E_\mathbf{p}[(\mathbf{p}/\mathbf{q})^{\beta-1}]-1}{\beta}$	$\frac{E_\mathbf{p}[(\mathbf{p}/\mathbf{q})^{\beta-1}]-1}{\beta}$	L_β	$\frac{(E_\mathbf{p}[(\mathbf{p}/\mathbf{q})^{\beta-1}])^{1/\beta}-1}{\beta-1}$	No	Jose, Nau, and Winkler (2008)
WPseudoSph	$\frac{1}{\beta-1}((\frac{p_i/q_i}{(E_\mathbf{p}[\mathbf{p}/\mathbf{q}^{\beta-1}])^{1/\beta}} - 1)$	$\frac{p_i/q_i}{(E_\mathbf{p}[\mathbf{p}/\mathbf{q}^{\beta-1}])^{1/\beta}}$	L_β	$\frac{E_\mathbf{p}[(\mathbf{p}/\mathbf{q})^{\beta-1}]-1}{\beta(\beta-1)}$	No	Jose, Nau, and Winkler (2008)
QuantS	$2(I_{[x\leq F^{-1}(\alpha)]} - \alpha)(F^{-1}(\alpha) - y)$	$\int S(\alpha;x)dp(x)$	$\mathcal{P}$	$\|\|p-q\|\|_2^2$	No	Cervera and Munoz (1996)
CLS	$I_{(y_t+1\in A_t)}\log(\frac{\hat{f}_t(y_{t+1})}{\int_{A_t}\hat{f}_t(s)ds})$	$\int_A p\log p$	L_2	$\int_t p_t(x)\ln(\frac{q(x)}{p(x)})dx$	No	Diks, Panchenko, and van Dijk (2011)
CsLS	$I_{(y_t+1\in A_t)}\log\hat{f}_t(x_{t+1}) +$ $+I_{(y_{t+1}\in A_t^c)}\log(\int_{A_t^c}\hat{f}_t(s)ds)$	-	L_2	-	No	Diks, Panchenko, and van Dijk (2011)
TW-CRPS	$\frac{1}{2}w(z)E_F\|\|X-X'\|\| - E_F\|\|X-x\|\|$	$\frac{1}{2}E_F\|\|X-X'\|\|$	$\mathcal{P}_1$	$\int_{-\infty}^{+\infty}(F(x)-G(x))$	No	Gneiting and Ranjan (2011)
QW-CRPS	$2(I_{[x\leq F^{-1}(\alpha)]} - \alpha)(F^{-1}(\alpha) - y)w(\alpha)d\alpha$	$\frac{1}{2}E_F\|\|X-X'\|\|$	$\mathcal{P}_1$	$\int_{-\infty}^{+\infty}(F(x)-G(x))$	No	Gneiting and Ranjan (2011)
Log-coshS	$-\ln\cosh\frac{q'(x)}{q(x)} + \frac{q'(x)}{q(x)}\tanh\frac{q'(x)}{q(x)} +$ $+(\frac{q''(x)}{q(x)} - \frac{q(x)'}{q(x)})(1-\tanh\frac{q'(x)}{q(x)})$				Yes	Ehm and Gneiting (2012)

NOTE: $A_t \subseteq \mathcal{X}$, $t \in \mathcal{T} = \mathcal{X}$ so that $\{A_t\} \equiv \{t\}$; $\mathcal{P}$: Borel probability measure; L: Lebesgue probability measure, μ: σ-finite measure

Table 2: Weighted power and pseudospherical scoring rules: special cases

	$S^P(p,q)$	$S^S(p,q)$
$\beta=-1$	$-\frac{1}{2}(1+(q_i/p_i)^2)+E_q(q/p)$	$\frac{1}{2}(1-((q_i/p_i)/E_q[q/p])^2)$
$\beta=0$	$1-(q_i/p_i)+E_q[ln(q/p)])$	$1-(q_i/p_i)\exp(-E_q[ln(q/p)])$
$\beta=\frac{1}{2}$	$2(2-\sqrt{q_i/p_i}-E_p[\sqrt{q/p}])$	$2(1-\sqrt{q_i/p_i}-E_p[\sqrt{q/p}])$
$\beta=1$	$\ln(p_i/q_i)$	$\ln(p_i/q_i)$
$\beta=2$	$((p_i/q_i)-1)-\frac{1}{2}(E_p[q/p]-1)$	$((p_i/q_i)/\sqrt{E_p[q/p]})-1$

Source: Jose, Nau, and Winkler (2008)

Table 3: Empirical Size and Power of LM test for Locality for different slopes parameters

			Empirical Size																	
			DGP 1									DGP 2								
T	γ_1	γ_2	F_1			F_2			F_3			F_1			F_2			F_3		
			$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
100			0.0001	0.0026	0.0162	0.0001	0.0036	0.0156	0.0000	0.0001	0.0058	0.0024	0.0164	0.0185	0.0039	0.0039	0.0123	0.0024	0.0027	0.0082
300			0.0017	0.0071	0.0188	0.0009	0.0045	0.0136	0.0009	0.0060	0.0073	0.0001	0.0081	0.0272	0.0001	0.0017	0.0221	0.0005	0.0019	0.0032
1000			0.0015	0.0121	0.0326	0.0004	0.0092	0.0285	0.0001	0.0017	0.0082	0.0036	0.0460	0.1018	0.0032	0.0441	0.0811	0.0001	0.0040	0.0069
			Empirical Power																	
T	γ_1	γ_2	F_1			F_2			F_3			F_1			F_2			F_3		
			$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
	5	2	0.0202	0.0824	0.1824	0.0098	0.0286	0.0728	0.0000	0.0014	0.0029	0.0185	0.0878	0.1801	0.0039	0.0217	0.0632	0.0024	0.0034	0.0105
	-5	2	0.0184	0.0679	0.1437	0.0000	0.0256	0.0480	0.0000	0.0001	0.0023	0.0185	0.0990	0.1811	0.0039	0.0181	0.0705	0.0024	0.0025	0.0082
	5	-2	0.0157	0.0767	0.1508	0.0041	0.0250	0.0515	0.0000	0.0017	0.0017	0.0212	0.0878	0.1765	0.0039	0.0217	0.0660	0.0024	0.0034	0.0084
	50	20	0.0316	0.1472	0.2502	0.0101	0.0736	0.1137	0.0001	0.0014	0.0043	0.0212	0.0878	0.1792	0.0039	0.0217	0.0660	0.0024	0.0034	0.0084
	-50	20	0.0151	0.0729	0.1664	0.0038	0.0241	0.0601	0.0001	0.0014	0.0017	0.0278	0.0927	0.1636	0.0039	0.0186	0.0583	0.0024	0.0051	0.0113
	50	-20	0.0162	0.0932	0.1875	0.0059	0.0308	0.0811	0.0001	0.0014	0.0017	0.0151	0.0799	0.1552	0.0039	0.0148	0.0400	0.0024	0.0034	0.0036
	50	0	0.0170	0.0836	0.1601	0.0038	0.0308	0.0653	0.0001	0.0014	0.0024	0.0174	0.0810	0.1579	0.0039	0.0141	0.0429	0.0024	0.0034	0.0037
100	-50	0	0.0094	0.0801	0.1288	0.0001	0.0197	0.0392	0.0001	0.0028	0.0058	0.0233	0.0986	0.1662	0.0039	0.0254	0.0689	0.0024	0.0037	0.0082
	0	50	0.0237	0.0921	0.1732	0.0060	0.0283	0.0716	0.0000	0.0001	0.0046	0.0216	0.0776	0.1490	0.0039	0.0198	0.0508	0.0024	0.0051	0.0117
	200	100	0.0289	0.1546	0.2557	0.0130	0.0813	0.1398	0.0001	0.0014	0.0065	0.0116	0.0504	0.1198	0.0021	0.0129	0.0351	0.0015	0.0015	0.0077
	-200	100	0.0173	0.0890	0.1647	0.0059	0.0268	0.0642	0.0001	0.0014	0.0017	0.0177	0.0737	0.1444	0.0039	0.0193	0.0507	0.0015	0.0034	0.0086
	200	-100	0.0195	0.0936	0.1689	0.0059	0.0324	0.0701	0.0001	0.0014	0.0017	0.0147	0.0674	0.1533	0.0039	0.0175	0.0365	0.0015	0.0035	0.0054
	200	0	0.0166	0.0855	0.1707	0.0038	0.0346	0.0658	0.0001	0.0014	0.0017	0.0180	0.0683	0.1369	0.0039	0.0154	0.0394	0.0015	0.0035	0.0054
	-200	0	0.0094	0.0782	0.1325	0.0001	0.0214	0.0392	0.0001	0.0028	0.0069	0.0233	0.0990	0.1670	0.0039	0.0273	0.0689	0.0024	0.0037	0.0082
	0	200	0.0230	0.0807	0.1514	0.0045	0.0225	0.0497	0.0000	0.0001	0.0079	0.0160	0.0810	0.1454	0.0039	0.0143	0.0472	0.0024	0.0024	0.0084
	5	2	0.0530	0.1534	0.2503	0.0150	0.0927	0.1340	0.0009	0.0031	0.0061	0.0239	0.1005	0.1960	0.0014	0.0278	0.0575	0.0001	0.0019	0.0031
	-5	2	0.0178	0.1033	0.1773	0.0054	0.0193	0.0665	0.0009	0.0029	0.0054	0.0174	0.0894	0.1837	0.0012	0.0282	0.0638	0.0005	0.0019	0.0032
	5	-2	0.0257	0.1221	0.1998	0.0071	0.0314	0.1064	0.0009	0.0029	0.0115	0.0254	0.1016	0.1919	0.0006	0.0276	0.0549	0.0000	0.0019	0.0031
	50	20	0.1252	0.2900	0.3935	0.0747	0.2005	0.2803	0.0009	0.0014	0.0034	0.0154	0.0762	0.1701	0.0001	0.0264	0.0511	0.0000	0.0001	0.0015
	-50	20	0.0262	0.1312	0.2043	0.0074	0.0466	0.1093	0.0009	0.0014	0.0057	0.0132	0.1156	0.2064	0.0014	0.0357	0.0720	0.0001	0.0009	0.0015
	50	-20	0.0348	0.1433	0.2218	0.0085	0.0581	0.1299	0.0009	0.0014	0.0072	0.0384	0.1225	0.2005	0.0001	0.0218	0.0399	0.0001	0.0006	0.0031
	50	0	0.0353	0.1339	0.2059	0.0065	0.0482	0.1012	0.0009	0.0014	0.0071	0.0341	0.1068	0.1847	0.0001	0.0245	0.0420	0.0000	0.0012	0.0031
300	-50	0	0.0190	0.1033	0.1476	0.0009	0.0209	0.0497	0.0009	0.0051	0.0078	0.0031	0.0950	0.2043	0.0014	0.0279	0.0607	0.0006	0.0026	0.0032
	0	50	0.0467	0.1606	0.2284	0.0077	0.0424	0.1040	0.0012	0.0012	0.0063	0.0098	0.0664	0.1682	0.0004	0.0205	0.0605	0.0001	0.0004	0.0032
	200	100	0.1493	0.3309	0.4407	0.0986	0.2437	0.3328	0.0009	0.0010	0.0022	0.0074	0.0347	0.1010	0.0001	0.0063	0.0335	0.0000	0.0001	0.0029
	-200	100	0.0360	0.1500	0.2150	0.0074	0.0548	0.1216	0.0009	0.0014	0.0057	0.0090	0.0652	0.1623	0.0001	0.0142	0.0534	0.0001	0.0009	0.0042
	200	-100	0.0413	0.1513	0.2269	0.0085	0.0597	0.1356	0.0009	0.0014	0.0062	0.0155	0.0728	0.1579	0.0001	0.0082	0.0345	0.0000	0.0006	0.0031
	200	0	0.0380	0.1416	0.2281	0.0078	0.0513	0.1152	0.0009	0.0014	0.0079	0.0159	0.0710	0.1527	0.0001	0.0082	0.0332	0.0000	0.0016	0.0031
	-200	0	0.0190	0.1033	0.1464	0.0009	0.0209	0.0500	0.0009	0.0048	0.0078	0.0216	0.0970	0.2043	0.0014	0.0279	0.0621	0.0006	0.0026	0.0032
	0	200	0.0475	0.1614	0.2349	0.0080	0.0513	0.1197	0.0012	0.0027	0.0063	0.0192	0.0924	0.1933	0.0012	0.0258	0.0545	0.0006	0.0019	0.0032
	5	2	0.0475	0.1614	0.2349	0.0080	0.0513	0.1197	0.0012	0.0027	0.0063	0.1146	0.3375	0.4697	0.0322	0.1438	0.2601	0.0001	0.0024	0.0080
	-5	2	0.0491	0.1537	0.2253	0.0224	0.0713	0.1195	0.0001	0.0047	0.0089	0.0923	0.2783	0.4189	0.0294	0.1269	0.2507	0.0001	0.0024	0.0024
	5	-2	0.1004	0.2120	0.3370	0.0380	0.1245	0.1845	0.0007	0.0053	0.0140	0.1215	0.3439	0.4709	0.0306	0.1252	0.2518	0.0001	0.0024	0.0104
	50	20	0.5337	0.7927	0.8732	0.4438	0.6777	0.7814	0.0015	0.0076	0.0135	0.0708	0.2574	0.4093	0.0171	0.1181	0.2175	0.0000	0.0001	0.0081
	-50	20	0.1150	0.2338	0.3527	0.0545	0.1566	0.2257	0.0000	0.0040	0.0113	0.1601	0.3856	0.5310	0.0383	0.1432	0.2711	0.0000	0.0001	0.0074
	50	-20	0.1334	0.2772	0.4222	0.0748	0.1721	0.2688	0.0006	0.0040	0.0090	0.1543	0.3604	0.5036	0.0082	0.0764	0.1708	0.0001	0.0050	0.0125
	50	0	0.1344	0.3275	0.4455	0.0479	0.1467	0.2243	0.0006	0.0031	0.0084	0.1316	0.3275	0.4669	0.0068	0.0786	0.1628	0.0001	0.0024	0.0130
1000	-50	0	0.0446	0.1193	0.2482	0.0140	0.0375	0.0606	0.0001	0.0029	0.0089	0.1001	0.2840	0.4272	0.0275	0.1310	0.2342	0.0001	0.0040	0.0063
	0	50	0.1636	0.3471	0.4766	0.0633	0.1604	0.2469	0.0024	0.0050	0.0254	0.0666	0.2232	0.3532	0.0101	0.0908	0.1828	0.0000	0.0001	0.0122
	200	100	0.6660	0.8601	0.9307	0.5699	0.7839	0.8639	0.0045	0.0072	0.0195	0.0093	0.0720	0.1555	0.0018	0.0018	0.0447	0.0001	0.0004	0.0024
	-200	100	0.1290	0.2733	0.4298	0.0696	0.1810	0.2699	0.0006	0.0040	0.0105	0.0437	0.1961	0.3173	0.0068	0.0712	0.1348	0.0001	0.0024	0.0104
	200	-100	0.1344	0.2871	0.4381	0.0760	0.1883	0.2863	0.0006	0.0040	0.0109	0.0422	0.2042	0.3123	0.0042	0.0403	0.1083	0.0000	0.0001	0.0101
	200	0	0.1412	0.3459	0.4640	0.0605	0.1636	0.2435	0.0006	0.0044	0.0100	0.0413	0.1844	0.3092	0.0030	0.0393	0.1099	0.0001	0.0001	0.0067
	-200	0	0.0462	0.1223	0.2390	0.0140	0.0377	0.0598	0.0001	0.0029	0.0089	0.1016	0.2828	0.4262	0.0279	0.1306	0.2355	0.0001	0.0040	0.0063
	0	200	0.1692	0.3616	0.5052	0.0727	0.1825	0.2619	0.0021	0.0050	0.0242	0.0214	0.1439	0.2526	0.0058	0.0485	0.1058	0.0000	0.0001	0.0098

Table 4: Empirical Power of LM test for Locality for different Scoring functions and $(\gamma_1, \gamma_2) = (200, 100)$.

S(p, x)	DGP 1									DGP 2								
	F_1			F_2			F_3			F_1			F_2			F_3		
$T = 100$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
QSR	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
WPowerS	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
PSph	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
WPSph	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
IntS	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
TsallisS	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
ES	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
GES	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
PSpctr	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
CRPS	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
QuantS	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
CLS	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
CsLS	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
HS	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
LCS	0.0000	0.0013	0.0211	0.0000	0.0000	0.0053	0.0000	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
$T = 300$																		
QSR	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
PSph	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
WPowerS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
WPSph	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
IntS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
TsallisS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
ES	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
GES	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
PSpctr	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
CRPS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
QuantS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
CLS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
CsLS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
HS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
LCS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	0.0000	0.0000	0.0000	0.0000	0.0336	0.1270	0.0000	0.0002	0.0180	0.0000	0.0000	0.0000
$T = 1000$																		
QSR	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
WPowerS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
PSph	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
WPSph	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
IntS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
TsallisS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
ES	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
GES	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
PSpctr	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
CRPS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
QuantS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
CLS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
CsLS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
HS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000
LCS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	0.0000	0.0000	0.0000	0.0000	0.0223	0.0613	0.0000	0.0040	0.0168	0.0000	0.0000	0.0000

Table 5: BC effects via SRs estimated from AR(5)

S(Q, y)	$S(\cdot,\cdot)^{STAR}$	Duration		Amplitude		Cumulated Value		Excess Mov. as % of Triangular Area		Cum.Val. of Duration		Cum.Val. of amplitde		Cum.Value of Excess	
		C	E	C	E	C	E	C	E	C	E	C	E	C	E
No SR	0.0000	4.0769	15.0000	-0.0402	0.1806	-0.1879	2.2817	79.7578	60.0237	0.4063	0.6700	-1.2077	0.7509	2.1099	3.3342
QSR	0.8127	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
WPowerS	-2.6311	-	-	-	-	-	-	-	-	-	-	-	-	-	-
" ($\alpha = -1$)	32.9164	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
" ($\alpha = 0$)	34.3658	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
" ($\alpha = 1/2$)	-63.1365	-	-	-	-	-	-	-	-	-	-	-	-	-	-
" ($\alpha = 1$)	-2.2902	-	-	-	-	-	-	-	-	-	-	-	-	-	-
" ($\alpha = 2$)	-17.6323	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PseudoSph	1.7631	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
" ($\alpha = -1$)	1.9145	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
" ($\alpha = 0$)	0.5000	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
" ($\alpha = 1/2$)	1.0000	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
" ($\alpha = 1$)	-19.2982	-	-	-	-	-	-	-	-	-	-	-	-	-	-
" ($\alpha = 2$)	-0.9825	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LogS	0.0021	4.0769	15.0000	-0.0402	0.1806	-0.1879	2.2817	79.7578	60.0237	0.4063	0.6700	-1.2077	0.7509	2.1099	3.3342
IntS	3.5000	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
TsallisS	1.2455	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
ES	-1.1195	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GES	-0.8426	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PseudoSpectrumS	-16.1862	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRPS	-8.1440	-	-	-	-	-	-	-	-	-	-	-	-	-	-
QuantS	1.2735	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
CLS	2.6982	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059
CsLS	0.0066	4.2500	16.1538	-0.0433	0.1908	-0.2028	2.4879	68.9083	11.1129	0.3771	0.6277	-1.1379	0.6588	2.4808	7.0334
HS	0.0381	3.6667	56.0000	-0.0047	0.5954	-0.0235	25.2830	190.4135	19.7733	0.1575	0.6819	-1.6418	0.7247	1.1166	0.0357
LCS	0.1529	4.4545	17.6667	-0.0475	0.2070	-0.2212	3.0766	93.1130	10.2991	0.3384	0.6446	-1.0399	0.7133	1.6780	7.9059

Table 6: BC effects via SRs estimated from MR-STAR(5,5,4)

S(Q, y)	$S(\cdot,\cdot)^{STAR}$	Duration		Amplitude		Cumulated Value		Excess Mov. as % of Triangular Area		Cum.Val. of Duration		Cum.Val. of amplitde		Cum.Value of Excess	
		C	E	C	E	C	E	C	E	C	E	C	E	C	E
No SR	0.0000	4.2308	16.7500	-0.0604	0.2250	-0.1824	2.9900	31.5354	81.6785	0.4223	0.6531	-0.8133	0.7162	2.2606	2.3361
QSR	0.8127	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
WPowerS	1.4812	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
" ($\alpha=-1$)	54.7040	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
" ($\alpha=0$)	-106.5745	-	-	-	-	-	-	-	-	-	-	-	-	-	-
" ($\alpha=1/2$)	-0.2337	-	-	-	-	-	-	-	-	-	-	-	-	-	-
" ($\alpha=1$)	-0.8191	9.2727	14.3000	-1.7576	1.5429	-7.4182	37.0634	-181.0128	131.2204	1.0861	1.0388	-1.2176	1.1698	-4.1934	1.6266
" ($\alpha=2$)	-27.4562	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PseudoSph	1.7631	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
" ($\alpha=-1$)	1.9689	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
" ($\alpha=0$)	0.4999	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
" ($\alpha=1/2$)	1.0000	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
" ($\alpha=1$)	-94.9719	-	-	-	-	-	-	-	-	-	-	-	-	-	-
" ($\alpha=2$)	-0.8928	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LogS	0.0021	4.2308	16.7500	-0.0604	0.2250	-0.1824	2.9900	31.5354	81.6785	0.4223	0.6531	-0.8133	0.7162	2.2606	2.3361
IntS	3.5000	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
TsallisS	1.2455	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
ES	-1.1383	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GES	-16.7098	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PseudoSpectrumS	-8.1440	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRPS	0.9012	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
QuantS	0.0076	4.2308	16.7500	-0.0604	0.2250	-0.1824	2.9900	31.5354	81.6785	0.4223	0.6531	-0.8133	0.7162	2.2606	2.3361
CLS	0.7807	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529
CsLS	0.0072	4.2308	16.7500	-0.0604	0.2250	-0.1824	2.9900	31.5354	81.6785	0.4223	0.6531	-0.8133	0.7162	2.2606	2.3361
HS	1.7303	5.9524	6.7000	-2.7365	2.8518	-11.3085	13.7579	56.5168	77.8627	0.6297	0.8032	-0.7702	1.1889	3.2995	2.0767
LCS	0.1529	4.4167	18.4545	-0.0652	0.2452	-0.1968	3.4011	16.6674	29.2511	0.3917	0.6075	-0.7358	0.6208	2.9447	1.4529

Table 7: Results of Locality Test from real data

Series	LM_1		LM_2		LM_3	
	F-statistic	P-value	F-statistic	P-value	F-statistic	P-value
US IIP	220.5935	0.0000	220.0645	0.0000	623.5069	0.0000
US UN	1.0e^{21} 3.1651	0.0000	1.0e^{21} 3.1651	0.0000	1.0e^{21} 3.1651	0.0000
SPR	1.0e^{22} 1.4019	0.0000	1.0e^{22} 1.4019	0.0000	1.0e^{21} 1.4019	0.0000
NINF	1.0e^{7} 1.9633	0.0000	1.0e^{7} 0.03041	0.0000	1.0e^{7} 0.3925	0.0000

Table 8: Scoring rules for GSTAR and AR-based density forecasts on 1-step ahead for four time series

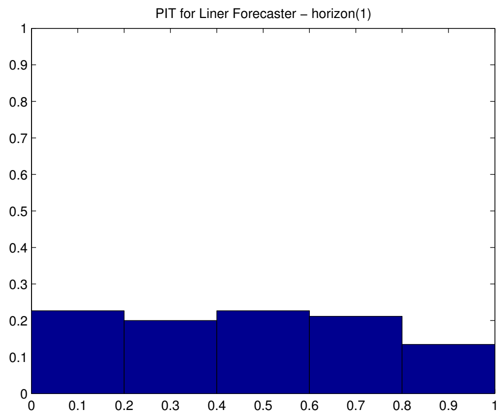
	US IIP		US UN		SPR		NINF	
S(Q, y)	GSTAR	AR	GSTAR	AR	GSTAR	AR	GSTAR	AR
QSR	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125
WPowerS	3.8186	3.8152	3.4514	3.5162	3.4610	3.3874	3.8600	3.9244
" ($\alpha = -1$)	4,189.7251	3.8152	1,179.0102	3.5162	2150.4860	3.3874	16,358.9195	3.9244
" ($\alpha = 0$)	4,103.3601	4,024.9000	1,172.2002	1143.4955	2143.8558	1,964.6571	16,231.2396	23,519.985
" ($\alpha = 1/2$)	-8,903.6397	-8,719.2999	-2,404.8807	-2,369.4963	-3,972.9757	-224.9477	-33,489.9179	-51,084.1416
" ($\alpha = 2$)	-1,964.3152	-1,928.5125	-578.2900	-561.1846	-4,346.7567	-3,972.9757	5.5544	-11,254.2211
PseudoSph	158.3423	158.3423	158.3423	158.3423	2.7888	2.5189	158.3423	0.0000
" ($\alpha = -1$)	0.5000	0.5000	578.2900	0.4999	0.5000	0.5000	0.5000	0.5000
" ($\alpha = 0$)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
" ($\alpha = 1/2$)	-11,3421.4023	1.0000	-9,682.9521	1.0000	-1,7411.4046	1.0000	-53,0059.2982	1.0000
" ($\alpha = 1$)	5.1689	5.1200	1.0000	1.0000	2.7888	2.5189	5.5544	6.9202
" ($\alpha = 2$)	1.6867	1.6080	-9,682.952	1.0000	-0.6500	-0.7207	1.0120	5.4647
LogS	0.0064	0.0064	0.0065	0.0065	0.0190	0.0190	0.0145	0.0145
IntS	3.5000	3.5000	3.5000	3.5000	3.5000	3.5000	3.5000	3.5000
TsallisS	1.0063	1.0063	1.0063	1.0063	1.0063	1.0063	1.0063	1.0063
ES	-1.2919	-1.2916	-2.3400	-2.3365	-1.9413	-1.9407	-0.3538	-0.2049
GES	-31.0931	-31.0489	-49.0791	-48.7288	-26.8488	-26.6362	-8.5122	-7.1544
PseudoSpectrumS	-9.9754	-9.9754	-9.9754	-9.9754	-9.9754	-9.9754	-9.9754	-9.9754
CRPS	0.14765	0.14765	1.5539	1.1139	2.4889	2.6810	0.0141	0.0150
QuantS	-11.8036	-11.8036	-11.9461	-11.9984	-11.6053	-11.5867	-11.8762	-11.8690
CLS	1.1508e-05	7.1093e-06	-0.0009	-0.0008	-0.0030	-0.0034	0.0002	6.7182e-05
CsLS	-0.0006	-0.0006	-0.0020	-0.0020	-0.0119	-0.0119	-0.0005	-0.0003
HS	6.9858	6.9819	26.7302	26.6323	84.8447	84.7913	3.5312	2.2900
LCS	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001

Table 9: Amisano and Giacomini (2007) test for Norges Bank's Fan Chart of Output Growth at a prediction horizon of 12 month for different scoring rules and Bank of Norway taking the role of $\bar{S}^f$ and the benchmark AR density forecast the role of $\bar{S}^g$

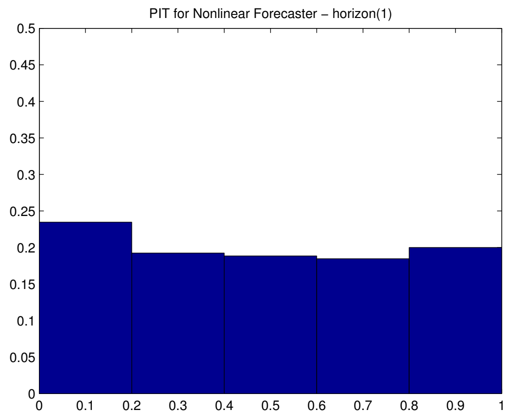
S(Q, y)	$\bar{S}^f$	$\bar{S}^g$	σ	t	P-value
QSR	0.55827	0.55827	0	-	0
WPowerS	2.8147	2.9534	0.1011	-3.3575	0.9991
" $(\alpha = -1)$	527.6771	2.9534	108,813.5611	-0.0032	0.5012
" $(\alpha = 0)$	527.5193	670.8714	108,103.8173	-0.0032	0.5012
" $(\alpha = 1/2)$	-1,062.301	-1,352.7636	443,826.9044	0.0016	0.4993
" $(\alpha = 1)$	1.1986	1.4481	0.3275	-1.8662	0.9653
" $(\alpha = 2)$	-262.6019	-333.8081	26,672.8241	0.006539	0.4974
PseudoSph	2.9817	2.9817	0	-	0
WPseudoSph	1.9916	1.9931	1.2709e-05	-299.5681	1
" $(\alpha = -1)$	0.4999	0.5000	7.8768e-12	-380,525.5137	1
" $(\alpha = 0)$	1	1	0	-	0
" $(\alpha = 1/2)$	-1,927.5277	1.0000	3,815,729.1164	0.0005	0.4997
" $(\alpha = 1)$	1.1986	1.4481	0.3275	-1.8662	0.9653
" $(\alpha = 2)$	-0.8559	-0.8361	0.0020	-23.4613	1.0000
LogS	0.0273	0.02730	0.0000	-	0.0000
IntS	3.5000	3.5000	0.0000	-	0.0000
TsallisS	1.2229	1.2229	0	-	0.0000
ES	-0.1237	-0.0834	0.0085	-11.5709	1.0000
GES	1.1626	1.2485	0.0388	-5.4196	1.0000
PseudoSpectrumS	-7.8530	-7.8530	0.0000	-	0.0000
CRPS	0.013201	0.012025	7.2746e-06	395.9622	0.0000
QuantS	-0.1909	-0.18357	0.0002827	-63.5174	1.0000
CLS	-0.1467	-0.4232	0.4021	1.6841	0.0499
CsLS	0.0088	0.0076	8.0784e-06	375.7488	0.0000
LCS	0.0552	0.0552	0.0000	-	0.0000

Figure 1: The PIT of four different forecasts

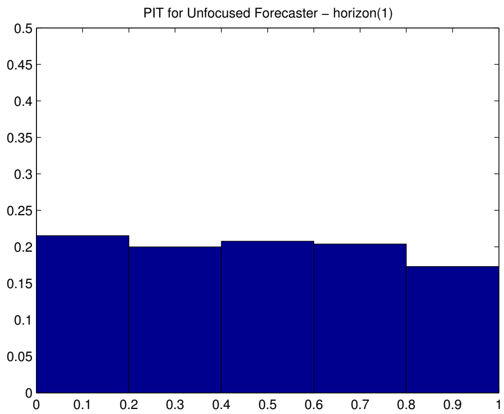
(a) PIT under linear forecast



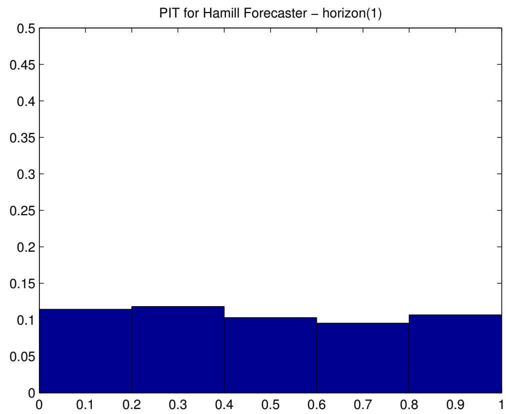
(b) PIT under GSTAR forecast



(c) PIT under unfocused forecast

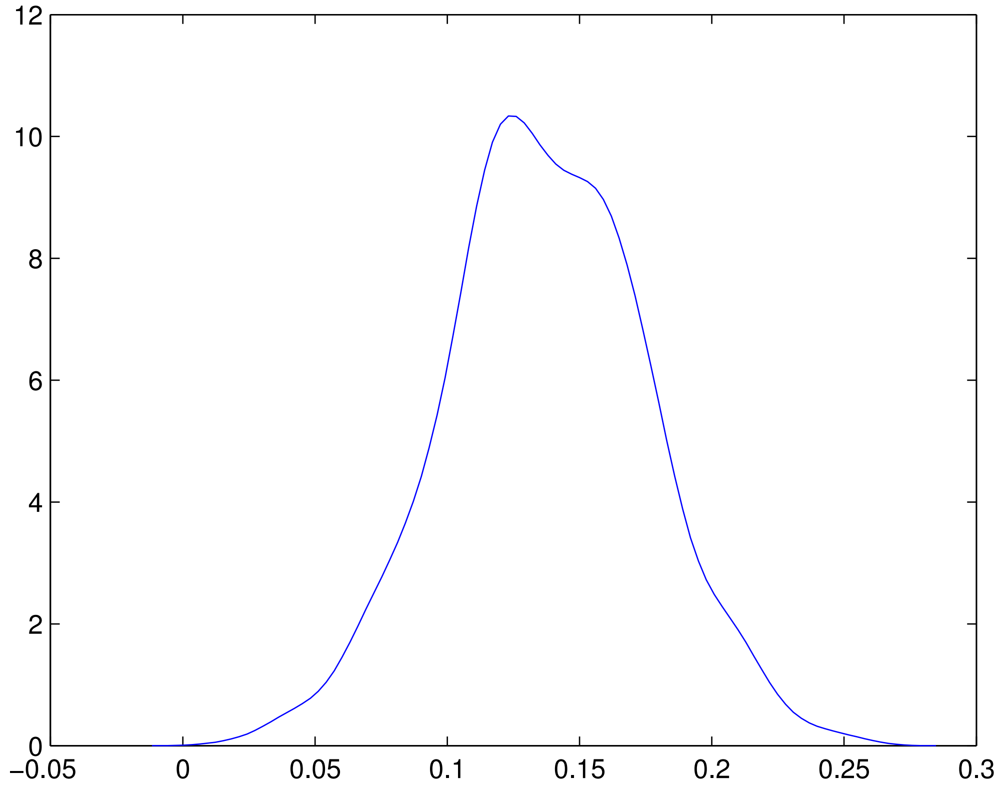


(d) PIT under Hamill's forecast



NOTE: At times $t = 1, 2, \ldots, T$, Nature picks a distribution G_t , and Forecaster chooses a probabilistic forecast F_t . The observations are independent random numbers y_t with distribution G_t. We write $N(\mu, \sigma^2)$ for the normal distribution with mean μ and variance σ^2. AR and GSTAR forecasts are drown from a $N(0, 1)$. The Unfocused forecast is drawn from a $\frac{1}{2}[N(\mu_t, 1) + N(\mu_t + \tau_t, 1)]$; Hamill's forecaster is drawn by $N(\mu_t + \delta_t, \sigma_t^2)$, $(\delta_t, \sigma_t^2) = 0.33 \cdot (0.5, 1) + 0.33 \cdot (-0.5, 1) + 0.33 \cdot (0, 167/100)$. The sequences $(\mu)_t$, $(\tau)_t$ and $(\delta)_t$, $(\sigma^2)_t$ are independent identically distributed and independent of each other

Figure 2: Bootstrapped density forecast from a simulated realization of GSTAR model



NOTE: Parameter values: $\phi_0 = 0.0$, $\phi_1 = 0.4$, $\phi_2 = -0.25$, $\theta_0 = 0.01$, $\theta_1 = -0.9$, $\theta_2 = 0.795$, $\gamma_1 = -200$, $\gamma_2 = 100$, $c = \bar{y}_t$, $S = 1,000$, $B = 10.000$

Figure 3: The QBB Algorithm on IIP

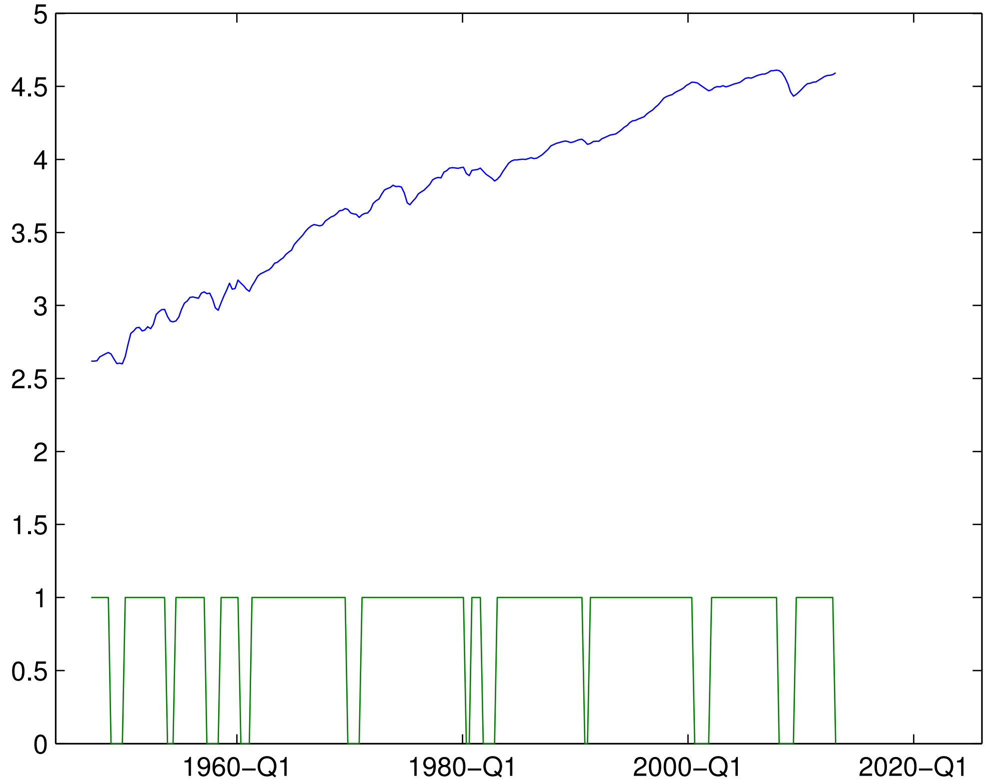
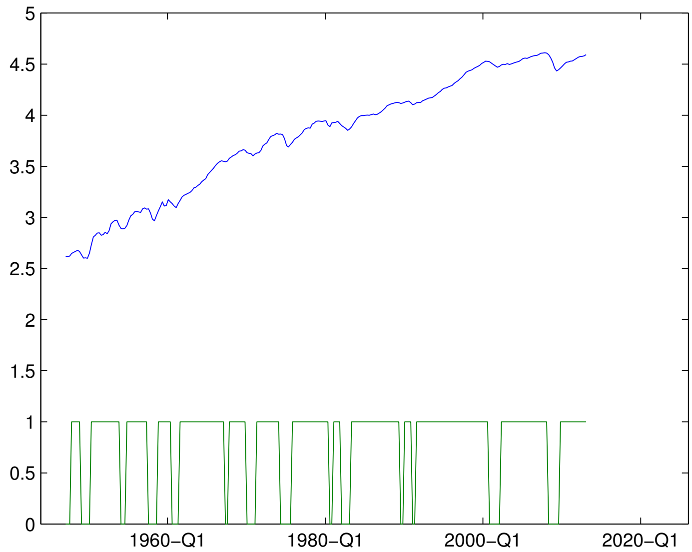


Figure 4: The SBBQ algorithm and density forecasts: linear case

(a) Estimated model and SBBQ



(b) Fan Chart (60 step-ahead)

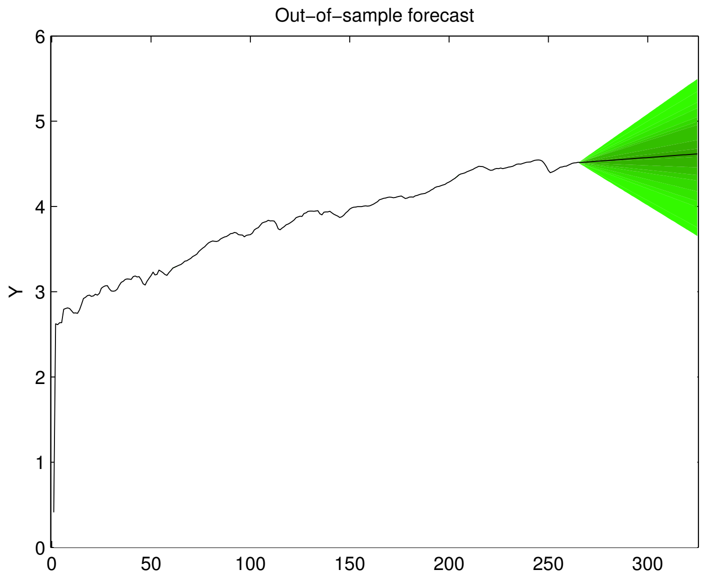
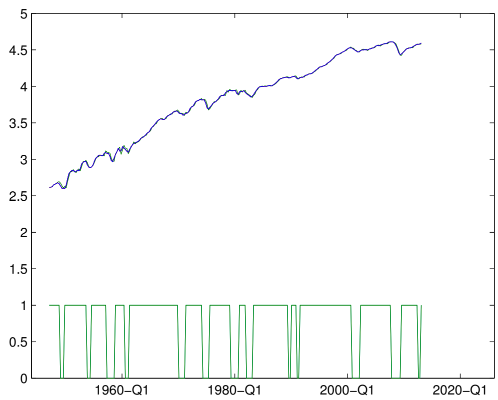
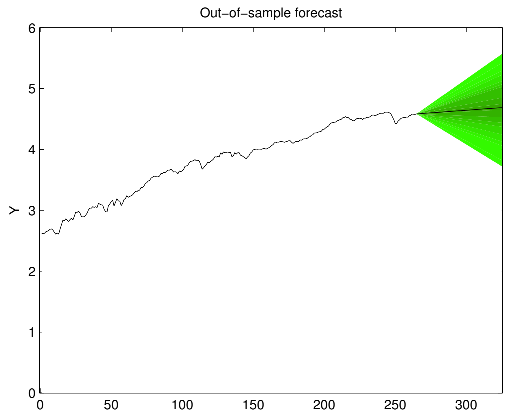


Figure 5: The SBBQ algorithm and density forecasts: nonlinear case

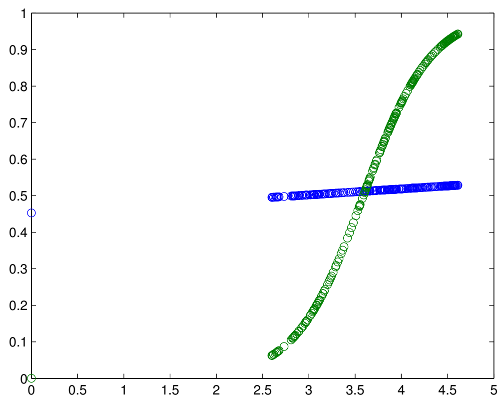
(a) Estimated model and SBBQ



(b) Fan Chart (60 step-ahead)



(c) Estimated G vs s_t



(d) Estimated G vs t

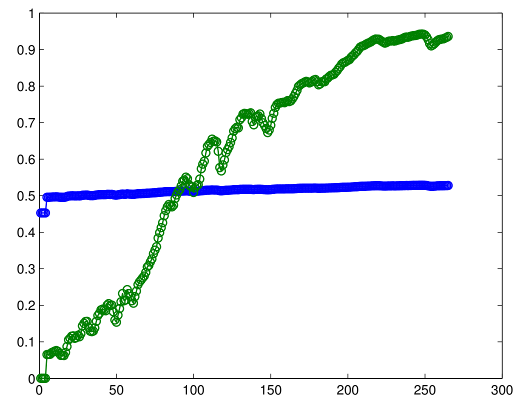
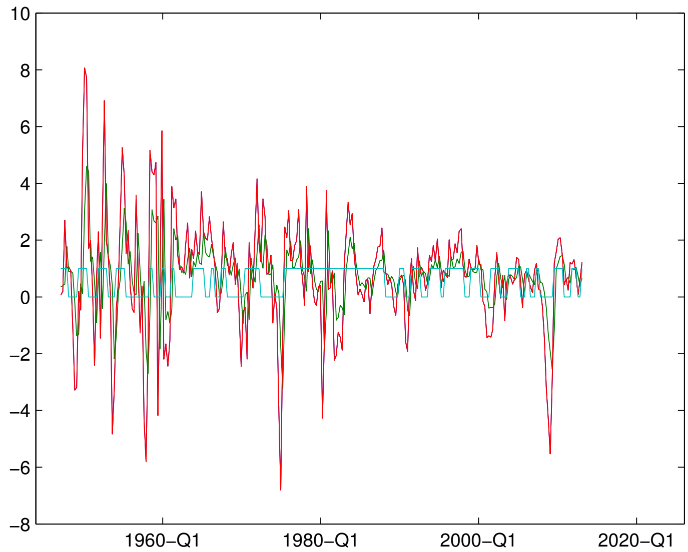


Figure 6: The SBBQ algorithm and density forecasts: nonlinear case

(a) Effect of Log Score on SBBQ: AR model



(b) Effect of Weighted Power Score on SBBQ: MR-STAR model

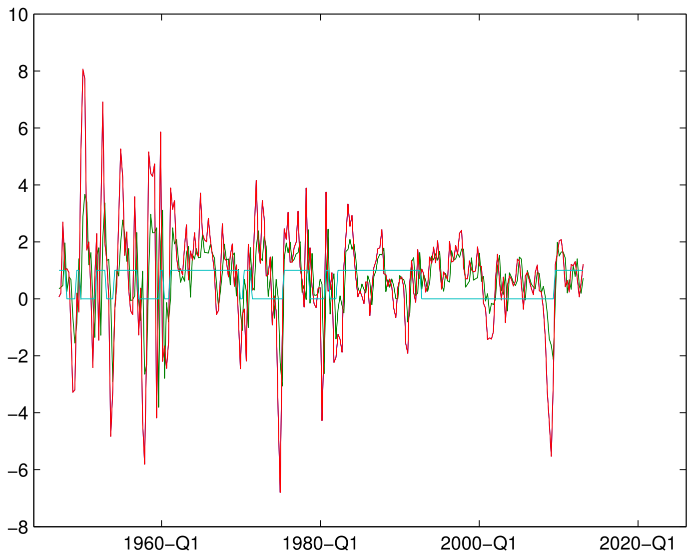
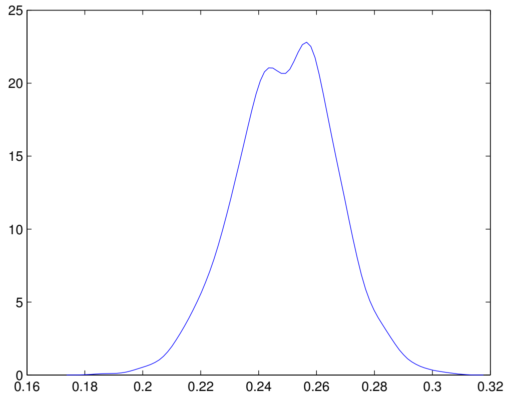
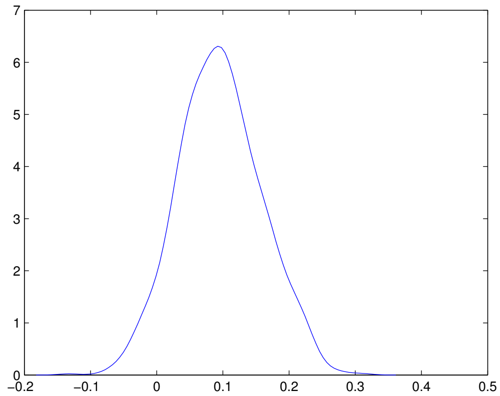


Figure 7: Bootstrapped density forecasts from estimated GSTAR model: four examples

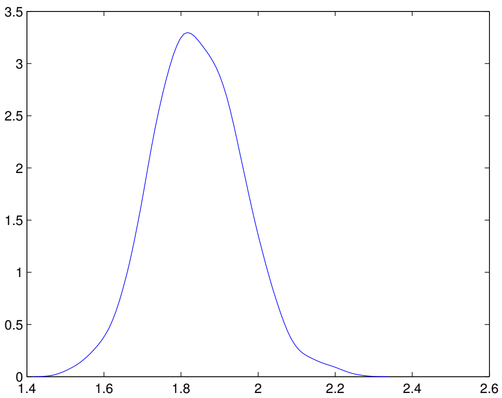
(a) U.S. IIP



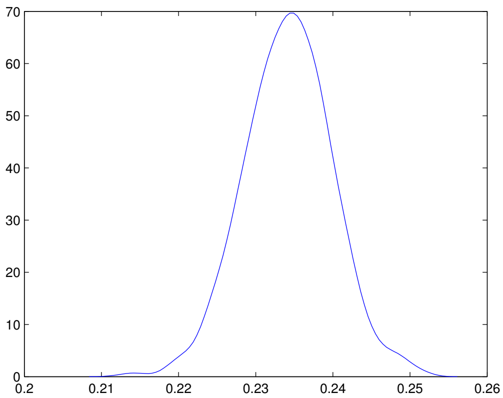
(b) U.S. UN



(c) SPR



(d) EI



NOTE: $S = 1,000$, $B = 1.000$