Testing and selecting local scoring rules

Emilio Zanetti Chini

UNIVERSITY OF PAVIA

Department in Economics and Management

January 23, 2015

Introduction

- Making **coherent** forecasts of **uncertain** events is one of the most critical requirements for any economic or statistical model. Uncertainty has **intrinsically probabilistic** nature and this has led an increasing number of statisticians and economists to treat it as a **subjective** fact. (De Finetti, 1937).

- Hence the need to investigate the behavior of the **whole density** of the forecast

- A consequence of the subjectivist school is that the Professional Forecaster's activity is *sensitive* to the presence of **bad incentives**

- This makes able to suspect a **strategic** behavior of the forecaster (Vovk and Shafer, 2005).

Motivations

- In statistics, one of the first tools created in order to evaluate the goodness-of-fit of density forecast p_t has been the **probability integral transform** (PIT) (Rosenblatt, 1952; Diebold *et al.*, 1998):

$$z_t = \int_{-\infty}^{x_t} p_t du_t, \tag{1}$$

which **sequence**, $\{z_t\}_t$ is **i.i.d. U(0,1)**

- **Problem**: PITs do not conveys any information about the utility underlining forecast.

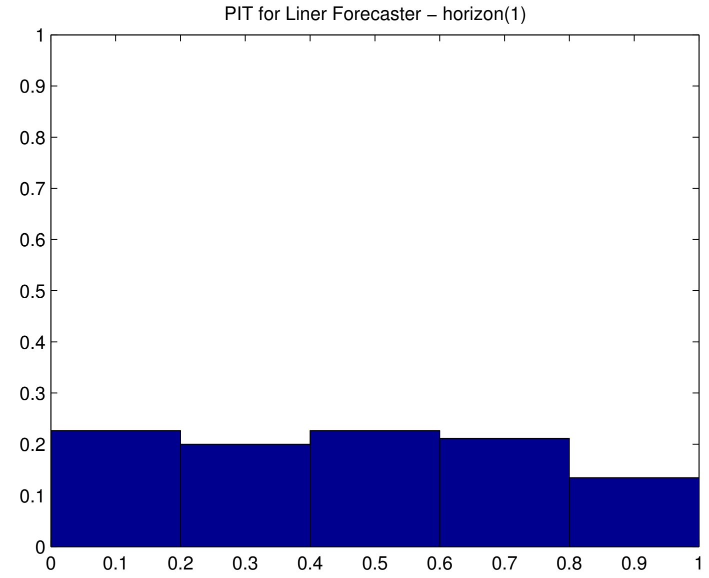
- **The Hamill (2001) Paradox**: different, misspecified forecasts produce a uniformly distributed PIT.

Example

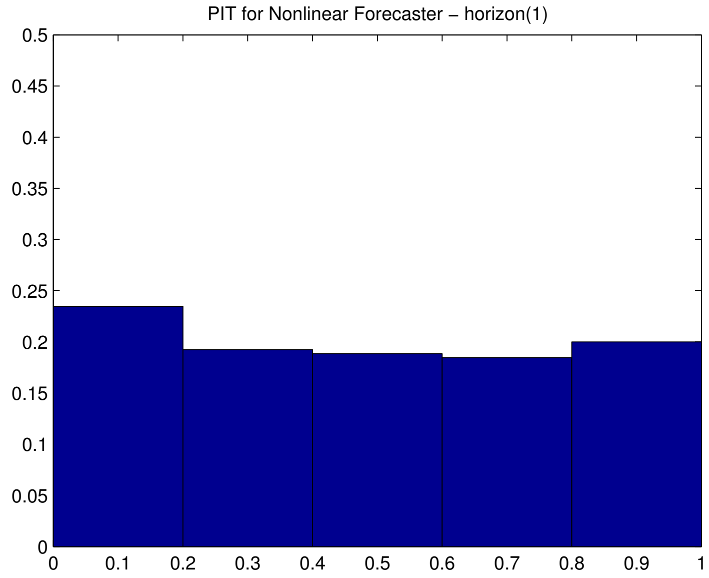
- Simulate T=265 identically distributed random numbers, compute the PITs and average them in form of histogram;
- Two Model-based Forecasts:
  - Liner AR(2): $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$, with : $\phi_0 = 0.0$; $\phi_1 = 0.9$; $\phi_2 = -0.795$;
  - STAR(2): $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + (\theta_1 = 0.4; \theta_2 = 0.25) G(\gamma_1, \gamma_2, c) + \epsilon_t$, with: $\phi_0 = 0.0$; $\phi_1 = 0.9$; $\phi_2 = -0.795$; $\theta_0 = 0.01$; $\theta_1 = 0.4$; $\theta_2 = 0.25$, $\gamma_1 = -20$; $\gamma_2 = 50$, $c = \bar{y}_t$.
- Ideal forecaster: DGP assumed i.i.d. N(0,1) and consequently estimated.
- Two Misspecified Forecasters:
  - Unfocused forecaster: $\epsilon_t \sim \frac{1}{2}[N(\mu_t, 1) + N(\mu_t + \tau_t, 1)]$;
  - Hamill's forecaster: $N(\mu_t + \delta_t, \sigma_t^2)$, $(\delta_t, \sigma_t^2) = 0.33 \cdot (0.5, 1) + 0.33 \cdot (-0.5, 1) + 0.33 \cdot (0, 167/100)$, with $(\mu)_t$, $(\tau)_t$ and $(\delta)_t$, $(\sigma^2)_t$ independent identically distributed and mutually independent.

RESULT: an impossible selection

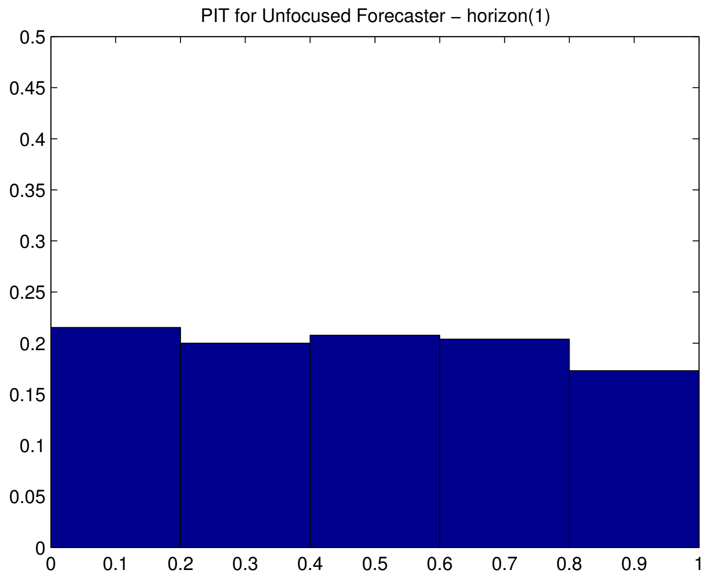
(a) PIT under linear forecast



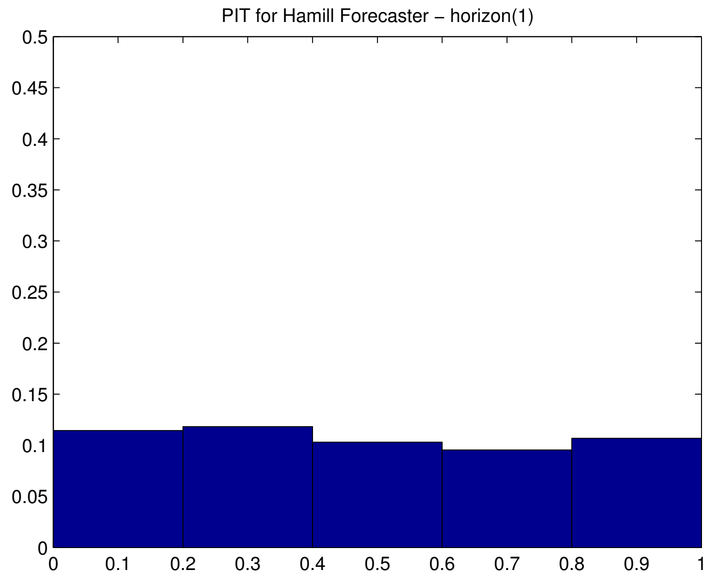
(b) PIT under GSTAR forecast



(c) PIT under unfocused forecast



(d) PIT under Hamill's forecast



Solution: Scoring Rules

- P can be evaluated by solving a decision problem originating by an experiment, consisting in the observation of a r.v. X whose probability measure belongs to a family indexed by a parameter Θ and its function $\psi = f(\Theta)$.

- A **scoring rule** (SR) is a function $S(P, Q) : \mathcal{R} \rightarrow \mathcal{R}$ assigning a numerical score to several competing (model) density forecasts of an observation that materializes, x, drawn from the quoted distribution Q, given the true (unknown) distribution P.

- 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 $P = Q$.

- **Local SR**: If Q is absolutely continuous, $S(\cdot, \cdot)$ ought to depend only on the behavior of Q, in an **infinitesimal neighborhood** of x, that is on the probability density of the true state and not upon the density of the states which could have obtained but did not. $\Rightarrow$ **Likelihood principle**!

Definitions

Definition 1. We define:

i. *Bayesian act* the action $\mathrm{a}^* := \min_{a \in \mathcal{A}} \int L(P, \mathrm{a}) p(Y_t) dY_t$

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

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

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

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

$$S(P, P) \leq S(P, Q) \;\; \forall \;\; P, Q \;\in \mathcal{P} \tag{2}$$

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

Definition 3 (Local scoring rule). A scoring rule $S(Q, x)$ is m-local if

$$S(Q, x) = s(x, q(x), q'(x), q''(x), \ldots, q^{(m)}(x)) \tag{3}$$

where: s is a q-function, s is the score function of S.

An Example of Strictly Proper SR (Predd et al., 2009)

Assume **two events**: E (recession) and F (downturn), respectively, where $E \subseteq F$. You attribute probabilities $.6$ and $.9$, respectively. It turns out that F comes to pass but not E.

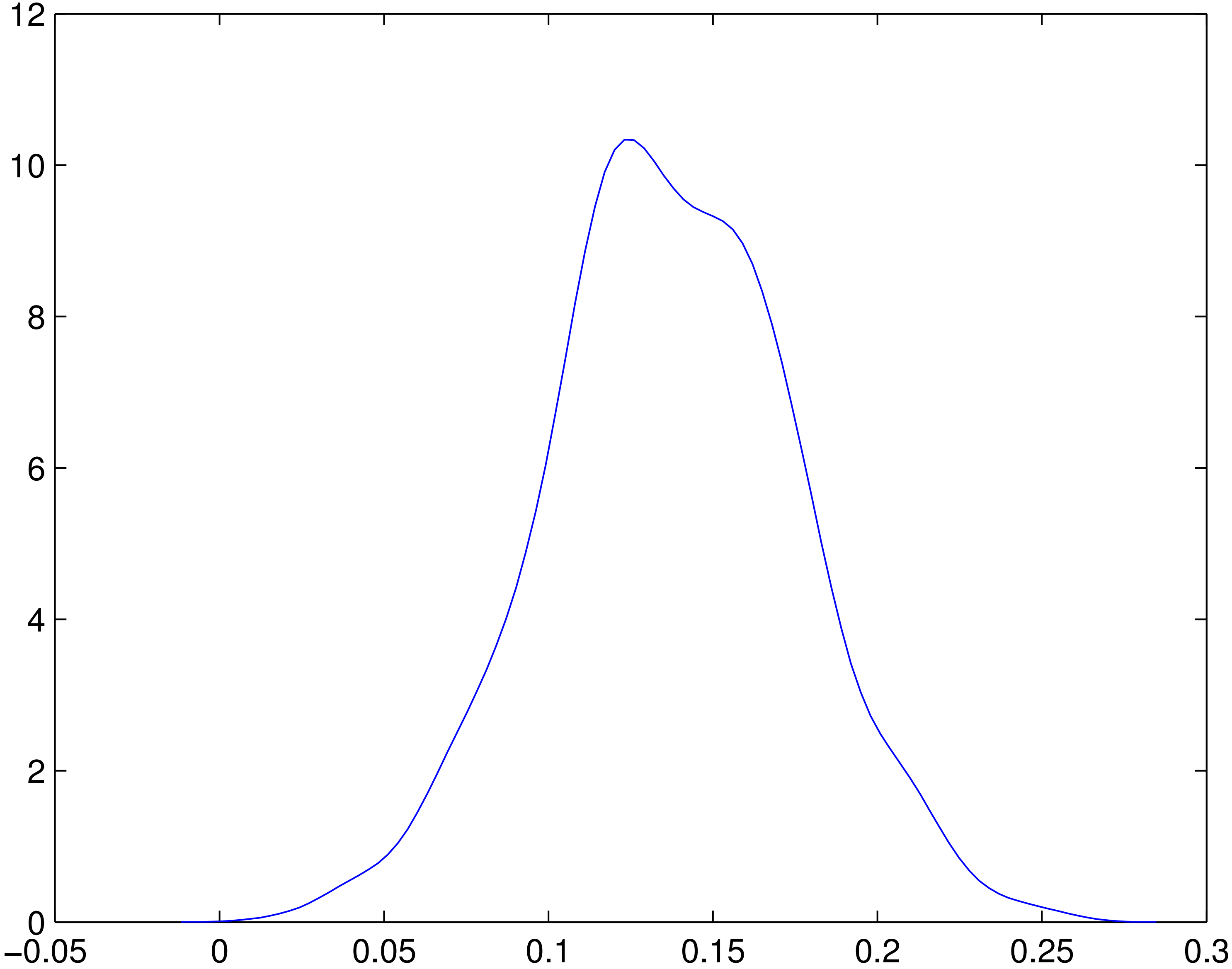
Expected loss for E (prior to discovering the facts):
expected score for **truly** quoting recession: $(1 - .6)^2$;
" " for **wrongly** quoting recession: $(0 - .6)^2$;
overall expectation: $.6(1 - .6)^2 + .4(0 - .6)^2 =$ **.24**.

Now suppose that you attempted to improve (lower) this expectation by **insincerely announcing .65** as the chance of E, even though your real estimate is $.6$.
Then your expected loss would be $.6(1-.65)^2+.4(0-.65)^2 =$ **.2425**, **worse than before**.

In general: Suppose your probability for an event E is p, that your announced probability is x, and that your loss is assessed according to the rule: $(1 - x)^2$ if E comes out true; $(0 - x)^2$ otherwise. Then your expected loss is uniquely minimized by choosing $x = p$.

Counter-example (improper SR): substitute the absolute deviation to quadratic loss!

PROBLEM 1: Multimodality of nonlinear PDF



NOTE: Example from previously simulated GSTAR(2)

PROBLEM 2: A plethora of SRs

Score	S(P,x)	H(p,x)	Measure	$d(P,Q)$	Brègman type	Reference
QS	$2p(x) - \lVert p \rVert_2^2$	$\lVert p(x) \rVert_2^2$	L_2	$\lVert p - q \rVert_2^2$	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	$\int \{\lvert Q(A_t) - 1_{A_t}(x))\}^2 d\mu(t)$	$\int P(A_t)\{1 - P(A_t)\} d\mu(t)$	μ	$\int \{P(A_t) Q(A_t)\}^2 d\mu(t)$	No	Epstein (1969)
PseudoSph	$\frac{p(x)^{\alpha-1}}{\lVert p \rVert_\alpha^{1-\alpha}}$	$\lVert p \rVert_\alpha$	L_α	$\lVert p \rVert_\alpha$	No	Good (1971)
IntS	$(u-l) + \frac{2}{\alpha}(l-x)I_{(x<l)} +$ $+ \frac{2}{\alpha}(x-u)I_{(x>u)}$	$\int S_\alpha^{int} dp(x)$	$\mathcal{P}$	$\lVert p \rVert_\alpha$	No	Winkler (1972)
CRPS	$\frac{1}{2} E_F \lVert X - X' \rVert - E_F \lVert X - x \rVert$	$\frac{1}{2} E_F \lVert X - X' \rVert$	$\mathcal{P}_1$	$\int_{-\infty}^{+\infty} (F(x) - G(x))$	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	$-\lvert \phi_P(\mathbf{y}) - e^{i<\mathbf{x},\mathbf{y}>} \rvert^2$	$-\lvert \phi_P(\mathbf{y}) \rvert$	$\mathcal{P}$	$\int_u \lVert \alpha - \beta \rVert^2$	No	Eaton *et al.* (1996)
DispersionS	$K(Q_V) + tr\{\boldsymbol{V}_P - \boldsymbol{V}_Q \boldsymbol{\Gamma}_Q\} -$ $-(\mathbf{x} - \boldsymbol{\mu}_P)' \boldsymbol{\Gamma}_P^{-1} (\mathbf{x} - \boldsymbol{\mu}_P)$	$-\log\det \boldsymbol{\Gamma}_P - mK$	$\mathcal{P}$	$tr(\boldsymbol{\Gamma}_P^{-1} \boldsymbol{\Gamma}_Q) - \log\det(\boldsymbol{\Gamma}_P - 1 \boldsymbol{\Gamma}_Q) +$ $+ (\boldsymbol{\mu}_P - \boldsymbol{\mu}_Q)' \boldsymbol{\Gamma}_P^{-1} (\boldsymbol{\mu}_P - \boldsymbol{\mu}_Q) - K$	Yes	Dawid and Sebastiani (1999)
Hyvaärinen	$((\ln q)'(x)^2 + 2(\ln q)''(y)$	$E_P p(x) \bigtriangledown \ln p(x)$	L	$\frac{1}{2} \int p(x) \lvert \bigtriangledown \ln p(x) - \bigtriangledown \ln q(x) \rvert dx$	Yes	Hyvärinen (2005)
ES	$\frac{1}{2} E_F \lVert \mathbf{X} - \mathbf{X}' \rVert^\beta - E_F \lVert \mathbf{X} - \mathbf{x} \rVert^\beta$	$\frac{1}{2} E_F \lVert \mathbf{X} - \mathbf{X}' \rVert$	$\mathcal{P}_\beta$	$\int_{-\infty}^{+\infty} (F(\mathbf{x}) - G(\mathbf{x}))$	No	Gneiting and Raftery (2007)
GES	$\frac{1}{2} E_F \lVert \mathbf{X} - \mathbf{X}' \rVert_\alpha^\beta - E_F \lVert \mathbf{X} - \mathbf{x} \rVert_\alpha^\beta$	$\frac{1}{2} E_F \lVert \mathbf{X} - \mathbf{X}' \rVert_\alpha^\beta$	$\mathcal{P}$	$\int_{-\infty}^{+\infty} (F(\mathbf{x}) - G(\mathbf{x}))$	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/Yes	Jose *et al.* (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/Yes	Jose *et al.* (2008)
QuantS	$2(I_{[x \leq F^{-1}(\alpha)]} - \alpha)(F^{-1}(\alpha) - y)$	$\int S(\alpha; x) dp(x)$	$\mathcal{P}$	$\lVert p - q \rVert_2^2$	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	$\int_t p_t(x) \ln(\frac{q(x)}{p(x)}) dx$	Yes	Diks *et al.* (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	-	Yes	Diks *et al.* (2011)
TW-CRPS	$\frac{1}{2} w(z) E_F \lVert X - X' \rVert - E_F \lVert X - x \rVert$	$\frac{1}{2} E_F \lVert X - X' \rVert$	$\mathcal{P}_1$	$\int_{-\infty}^{+\infty} (F(x) - G(x))$	No	Gneiting and Ranjan (2011)
QW-CRPS	$2(I_{[x \leq F^{-1}(\alpha)]} - \alpha)(F^{-1}(\alpha) - y) w(\alpha) d\alpha$	$\frac{1}{2} E_F \lVert X - X' \rVert$	$\mathcal{P}_1$	$\int_{-\infty}^{+\infty} (F(x) - G(x))$	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

PROBLEM 3: Some of the most important SRs are nested! (Jose et al., 2008)

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

The Literature

- The current literature (Christoffersen and Diebold, 1996, 1997; Patton and Timmermann, 2007a) **do not** take the the nesting problem as testable hypothesis, and tends to **circumvent** it by:

  i. improving the flexiblity of the assumed Loss function Patton (2011);

  ii. combining different forecasters' outcomes in the spirit of Bates and Granger (1969)
     - *via* Kullback-Liebler Information Criterion (Bao *et al.*, 2007; Mitchell and Hall, 2005; Hall and Mitchell, 2007; Amisano and Giacomini, 2007) or
     - *via* bayesian methods (Kascha and Ravazzolo, 2010; Ravazzolo and Vahey, 2013).

- Two valuable exceptions:

  i. Patton and Timmermann (2007b) the loss function is assumed unknown with consequences for optimality of FED predictions of U.S. output gap;

  ii. Boero *et al.* (2011): the loss function is selected via Giacomini and White (2006) test for equal predictive ability.

- **Locality remains unexplored!**

The Contribution

- **The ECONOMIC research question**:
  How to understand the utility (loss) which drives **effectively** the process in order to assess the **rationality** of the forecast ?

- **The STATISTICAL research question**:
  Do data allow for the **null** of **local**, *that is logarithmic*, **structure**? And, additionally, *is really* the **most likely** in the set of known scoring rules?

- The literature potentially able to answer to this question **exists but is purely theoretical**: Parry *et al.* (2012); Dawid *et al.* (2012); Ehm and Gneiting (2012)

- **We convey this information in a classical LM test to evaluate the underlying model**

The Key Condition

- Parry *et al.* (2012) provide three sufficient conditions for s to be proper.

Theorem 1. *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 conditions:*

i. $\mathcal{L}s = 0$,

ii. $s = (I - \mathcal{L})g$, *g being a 0-homogeneous q-function,*

iii. $s = \Lambda\phi$, *with ϕ being a 1-homogeneous q-function and being a operator*

where :

$\mathcal{L} := \sum_{k\geq 0}(-1)^k \mathcal{D}^k q_0 \frac{\partial}{\partial q_k}, \quad \mathcal{D} := \frac{\partial}{\partial x} + \sum_{j>>0} q_{j+1}\frac{\partial}{q_j}, \quad \Lambda = \sum_{k\geq 0}(-1)^k D^k \frac{\delta}{\delta q_k},$

$\mathcal{L} = I + \sum_{k\geq 0}(-1)^{k+1} D^k \frac{\delta}{\delta q_k} q_0 \odot$, *and I being the Identity operator*

A local scoring rule fulfilling condition (i) is called *key score*.

- **The Key condition is a testable hypothesis!**

Assumptions

- **A1** - **A3**: There exists a **unique** a^* that maximizes the expected utility (minimizes the Expected Loss).

  Scoring rules representation:

- **A4**: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.

- **A5**: $S(P,Q)$ is affine, real-valued for all $P, Q \in \mathcal{P}$ and minimized in Q at $Q = P$.

- **A6**: $D(P,Q) - D(P,Q_0)$ is affine in P , and $D(P,Q) \geq 0$, with equality achieved at $Q = P$

REMARK: **A6** links Local SR to *Brègman* divergences!

The model

$$y_t = \boldsymbol{\phi}'\mathbf{z}_t + \boldsymbol{\theta}'\mathbf{z}_t G(\boldsymbol{\gamma}, h(\mathbf{w}_t, \mathbf{c}_k)) + \epsilon_t \tag{4}$$

$$G(\boldsymbol{\gamma}, h(\mathbf{w}_t, \mathbf{c}_k)) = \left(1 + \exp\left\{ - \prod_{k=1}^{K} h\big(\mathbf{w}_t, \mathbf{c}_k\big)\right\}\right)^{-1}, \tag{5}$$

where:

$$\mathbf{1}_{\{\mathbf{w}_t - \mathbf{c}_k > 0\}} h(\mathbf{w}_t, \mathbf{c}_k) = \begin{cases} \gamma_1^{-1} \exp(\gamma_1 |\mathbf{w}_t - \mathbf{c}_k| - 1) & \text{if } \gamma_1 > 0, \\ 0 & \text{if } \quad \gamma_1 = 0, \\ -\gamma_1^{-1} \log(1 - \gamma_1 |\mathbf{w}_t - \mathbf{c}_k|) & \text{if } \gamma_1 < 0, \end{cases} \tag{6}$$

$$\mathbf{1}_{\{\mathbf{w}_t - \mathbf{c}_k \leq 0\}} h(\mathbf{w}_t, \mathbf{c}_k) = \begin{cases} -\gamma_2^{-1} \exp(\gamma_2 |\mathbf{w}_t - \mathbf{c}| - 1) & \text{if } \gamma_2 > 0, \\ 0 & \text{if } \gamma_2 = 0, \\ \gamma_2^{-1} \log(1 - \gamma_2 |\mathbf{w}_t - \mathbf{c}|) & \text{if } \gamma_2 < 0, \end{cases} \tag{7}$$

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

How to Do: the Locality Test (1)

Assume y_t follows model (4) with $h(\mathbf{w}, \mathbf{c})$, $\mathbf{w} = a'\mathbf{z}_t \odot s_t$, $a = [a_1, \ldots, a_p]$, with a stressing the fact that the delay is unknown.

Linearize the model via Taylor expansion:

$$y_t = \phi'\mathbf{z_t} + \theta'\mathbf{z_t}T_3G(\cdot) + \epsilon'_t, \tag{8}$$

and regress the resulting auxiliary model:

$$\hat{u}_t = \hat{\mathbf{z}}'_{\mathbf{1t}}\tilde{\boldsymbol{\beta}}_{\mathbf{1}} + \sum_{j=1}^{p} \beta_{2j} s y_{t-j} y_{t-d} + \sum_{j=1}^{p} \beta_{3j} s y_{t-j} y_{t-d}^2 + \sum_{j=1}^{p} \beta_{4j} s y_{t-j} y_{t-d}^3 + v_t\,, \tag{9}$$

Hence:

$$H_0 : \beta_{2j} = \beta_{3j} = \beta_{4j} = 0,\ j = 1, \ldots, p, \tag{10}$$

$$S^{LM} = T(SSR_0 - SSR)/SSR0 \overset{as}{\sim} \chi^2_{p+n} \tag{11}$$

This is a simple adaptation of the **Luukkonen et al. (1988) test**.

How to Do: the Locality Test (2)

- The F-version of S^{LM} may be preferable when testing H_0 in order to preserve power in low samples.

- In practice the form of G is unknown.

- Teräsvirta (1994) proposes a battery of F-tests on the auxiliary model :

$$H_{01} : \beta_4 = 0 \text{ vs } H_{11} : \beta_4 \neq 0 \quad (12)$$

$$H_{02} : \beta_3 = 0 | \beta_4 = 0 \text{ vs } H_{12} : \beta_3 \neq 0 | \beta_4 = 0 \quad (13)$$

$$H_{03} : \beta_2 = 0 | \beta_3 = 0 \text{ and } \beta_4 = 0 \text{ vs } H_{22} : \beta_2 \neq 0 | \beta_3 = 0 \text{ and } \beta_4 = 0. \quad (14)$$

Monte Carlo Experiment: Simulation Design

- Two different processes:

$$
\begin{aligned}
y_{1,t}^{(s)} &= 0.4y_{1,t-1}^{(s)} - 0.25y_{1,t-2}^{(s)} + (0.02 - 0.9y_{1,t-1}^{(s)} + 0.795y_{1,t-2})^{(s)}G^{(s)}(\mathbf{\Xi}) + \epsilon_{1,t}^{(s)} \\
y_{2,t}^{(s)} &= 0.00 + 0.2y_{2,t-1}^{(s)} - 0.3y_{2,t-2}^{(s)} + (0.02 - 0.2y_{2,t-1}^{(s)} + 0.12y_{2,t-2}^{(s)})G^{(s)}(\mathbf{\Xi}) + \epsilon_{2,t}^{(s)},
\end{aligned}
$$

where:
$G^{(s)}(\mathbf{\Xi}) = \big(1 + \exp\big\{-\big[h(\boldsymbol{\gamma})^{(s)}I_{(\boldsymbol{\gamma}\leq 0)}(y_{t-1}^{(s)} - \bar{y}_t^{(s)}) + h(\eta_t)^{(s)}\big]\big\}\big)^{-1}$
with $\eta_t = (y_{t-1}^{(s)} - \bar{y}_t^{(s)})$, $\bar{y}_t = \frac{1}{T}\sum_{t=1}^{T} y_t$ and y_{t-1} being the transition variable,
$s = \{1, \ldots, S\}$, $S = 1,000$.

- $y_{1,t}^{(s)}$ ("DGP 1") is an additive nonlinear model with accentuated nonlinear behavior, due to the high autoregressive parameters driving $G(\mathbf{\Xi})$ (Ex: a macroeconomic indicator affected by an unexpected shocks)

- $y_{2,t}^{(s)}$ ("DGP 2") describes an unclear scenario.

- $T = \{100, 300, 1000\}$ and $\alpha = \{0.01, 0.05, 0.10\}$ and a set of couples $(\gamma_1,\ \gamma_2)$

Monte Carlo Experiment: Results

SIMULATION 1:

- Well-behaving for what concern the empirical size;
- Empirical power is poor if an almost linear model is used, and in general for DGP 2
- **Model Dependance:** power is sensitive to the signs of the slope parameters;
- The **(generalized) logistic** transition function is the only really performant.

SIMULATION 2:

- **Score Invariance:** the power of the test is not affected by different scoring rules (Lindley, 1982)
- In general, more extreme than the logarithmic case

Table 1: Empirical Size of LM test for Locality for different slopes parameters

DGP 1											
T	γ_1	γ_2	F_1			F_2			F_3		
			$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
100			0.0001	0.0026	0.0162	0.0001	0.0036	0.0156	0.0000	0.0001	0.0058
300			0.0017	0.0071	0.0188	0.0009	0.0045	0.0136	0.0009	0.0060	0.0073
1000			0.0015	0.0121	0.0326	0.0004	0.0092	0.0285	0.0001	0.0017	0.0082

DGP 2											
T	γ_1	γ_2	F_1			F_2			F_3		
			$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
100			0.0024	0.0164	0.0185	0.0039	0.0039	0.0123	0.0024	0.0027	0.0082
200			0.0001	0.0081	0.0272	0.0001	0.0017	0.0221	0.0005	0.0019	0.0032
300			0.0036	0.0460	0.1018	0.0032	0.0441	0.0811	0.0001	0.0040	0.0069

Empirical Power of LM test for Locality for different slopes parameters with DGP 1

T	γ_1	γ_2	F_1			F_2			F_3		
			$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
300	5	2	0.0530	0.1534	0.2503	0.0150	0.0927	0.1340	0.0009	0.0031	0.0061
	-5	2	0.0178	0.1033	0.1773	0.0054	0.0193	0.0665	0.0009	0.0029	0.0054
	5	-2	0.0257	0.1221	0.1998	0.0071	0.0314	0.1064	0.0009	0.0029	0.0115
	50	20	0.1252	0.2900	0.3935	0.0747	0.2005	0.2803	0.0009	0.0014	0.0034
	-50	20	0.0262	0.1312	0.2043	0.0074	0.0466	0.1093	0.0009	0.0014	0.0057
	50	-20	0.0348	0.1433	0.2218	0.0085	0.0581	0.1299	0.0009	0.0014	0.0072
	50	0	0.0353	0.1339	0.2059	0.0065	0.0482	0.1012	0.0009	0.0014	0.0071
	-50	0	0.0190	0.1033	0.1476	0.0009	0.0209	0.0497	0.0009	0.0051	0.0078
	0	50	0.0467	0.1606	0.2284	0.0077	0.0424	0.1040	0.0012	0.0012	0.0063
	200	100	0.1493	0.3309	0.4407	0.0986	0.2437	0.3328	0.0009	0.0010	0.0022
	-200	100	0.0360	0.1500	0.2150	0.0074	0.0548	0.1216	0.0009	0.0014	0.0057
	200	-100	0.0413	0.1513	0.2269	0.0085	0.0597	0.1356	0.0009	0.0014	0.0062
	200	0	0.0380	0.1416	0.2281	0.0078	0.0513	0.1152	0.0009	0.0014	0.0079
	-200	0	0.0190	0.1033	0.1464	0.0009	0.0209	0.0500	0.0009	0.0048	0.0078
	0	200	0.0475	0.1614	0.2349	0.0080	0.0513	0.1197	0.0012	0.0027	0.0063
1000	5	2	0.0475	0.1614	0.2349	0.0080	0.0513	0.1197	0.0012	0.0027	0.0063
	-5	2	0.0491	0.1537	0.2253	0.0224	0.0713	0.1195	0.0001	0.0047	0.0089
	5	-2	0.1004	0.2120	0.3370	0.0380	0.1245	0.1845	0.0007	0.0053	0.0140
	50	20	0.5337	0.7927	0.8732	0.4438	0.6777	0.7814	0.0015	0.0076	0.0135
	-50	20	0.1150	0.2338	0.3527	0.0545	0.1566	0.2257	0.0000	0.0040	0.0113
	50	-20	0.1334	0.2772	0.4222	0.0748	0.1721	0.2688	0.0006	0.0040	0.0090
	50	0	0.1344	0.3275	0.4455	0.0479	0.1467	0.2243	0.0006	0.0031	0.0084
	-50	0	0.0446	0.1193	0.2482	0.0140	0.0375	0.0606	0.0001	0.0029	0.0089
	0	50	0.1636	0.3471	0.4766	0.0633	0.1604	0.2469	0.0024	0.0050	0.0254
	200	100	0.6660	0.8601	0.9307	0.5699	0.7839	0.8639	0.0045	0.0072	0.0195
	-200	100	0.1290	0.2733	0.4298	0.0696	0.1810	0.2699	0.0006	0.0040	0.0105
	200	-100	0.1344	0.2871	0.4381	0.0760	0.1883	0.2863	0.0006	0.0040	0.0109
	200	0	0.1412	0.3459	0.4640	0.0605	0.1636	0.2435	0.0006	0.0044	0.0100
	-200	0	0.0462	0.1223	0.2390	0.0140	0.0377	0.0598	0.0001	0.0029	0.0089
	0	200	0.1692	0.3616	0.5052	0.0727	0.1825	0.2619	0.0021	0.0050	0.0242

Empirical Power of LM test for Locality for different Scoring functions and $(\gamma_1, \gamma_2) = (200, 100)$ with DGP 1.

	F_1			F_2			F_3		
$T = 100$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
$T = 300$									
QSR	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	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
WPowerS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	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
IntS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	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
ES	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	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
PSpctr	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	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
QuantS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	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
CsLS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	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
LCS	0.0582	0.0024	0.1440	0.0000	0.0300	0.0602	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
WPowerS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	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
WPSph	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	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
TsallisS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	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
GES	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	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
CRPS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	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
CLS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	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
HS	0.8539	0.9876	1.0000	0.8350	0.9808	0.9995	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

Application 2: Assessing univariate forecasts

- DATA
  - U.S. industrial production in growth rates from 1948:01 to 2013:01
  - U.S. unemployment rate from 1948:01 to 2013:03
  - Long-term interest spread between Italy and Germany 1991:03 to 2013:05
  - Norway's Inflation rate from 1985:01 to 2013:12.

  - RESULT: Null hypothesis of logarithmic reward is strongly rejected in all the cases.

Table 2: 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

Unexpected result: further analysis

	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

RESULTS

- 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**
  - **Logarithmic** scores.

- The preference for linear specifications seems generally confirmed;
- **Important exceptions**:
  i. 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;
  ii. In three cases (Energy, Generalized Energy and Hyvaarinen and PseudoSpherical Scores with $\alpha = 1$) the superiority of AR is not so evident.

- This last case does not coincide with its limiting case, the LogS.

Conclusions

- Locality of SRs is a desirable property for inferential purposes.

- An LM test is able to detect the coherency of the give forecast with respect to a given model .

- Nonlinear-based model's density forecasts seem to be **more reliable** than linear-based ones;

- The hypothesis of logarithmic structure of the forecaster reward, is generally a **too strong assumption**.

- **Open question**: relation between locality of the SR and the stability of the forecast in the sense of Giacomini and Rossi (2010).
- **Next developments**:
  (a) relaxing the implicit assumption that SR is **observed** via unobserved component analysis.
  (b) **Local Estimation** of SRs.

References

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

BAO, Y., LEE, T. and SALTOGLU, B. (2007). Comparing Density Forecast Models. *Journal of Forecasting*, **26** (3), 203–225.

BATES, J. and GRANGER, C. (1969). The combination of forecasts. *Operations Research Quarterly*, **20**, 451–468.

BOERO, G., SMITH, J. and WALLIS, K. (2011). Scoring rules and survey density forecasts. *International Journal of Forecasting*, **27**, 379–393.

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

CERVERA, J. and MUNOZ, J. (1996). Proper Scoring Rules for Fractiles. In J. Bernardo, J. Berger, A. Dawid and A. Smith (eds.), *Bayesian Statistics 5*, Oxford, UK: Oxford University Press, pp. 513–519.

CHRISTOFFERSEN, P. F. and DIEBOLD, F. X. (1996). Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics*, **11** (5), 561–572.

— and — (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, **13** (6), 808–817.

DAWID, A., LAURITZEN, S. and PARRY, M. (2012). Proper Local Scoring Rules on Discrete Sample Space. *The Annals of Statistics*, **40** (1), 593–608.

Dawid, P. and Sebastiani, P. (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.

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

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

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

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

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

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

Giacomini, R. and Rossi, B. (2010). FORECAST COMPARISONS IN UNSTABLE ENVIRONMENTS. *Journal of Applied Econometrics*, **25**, 595–620.

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

— and RANJAN, R. (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 V. Godambe and A. Sprott (eds.), *Foundations of Statistical Inference*, Toronto: Holt, Rinehart and Winston, pp. 337–339.

HALL, S. and MITCHELL, J. (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.

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

JOSE, V., NAU, R. and WINKLER, R. (2008). Scoring Rules, Generalized Entropy, and Utility Maximization. *Operation Research*, **56**, 1146–1157.

KASCHA, C. and RAVAZZOLO, F. (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., SAIKKONEN, P. and TERÄSVIRTA, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, **75** (3), 491–499.

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

Mitchell, J. and Hall, S. (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.

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

Patton, A. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, **160** (1), 246–256.

— and Timmermann, A. (2007a). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, **140** (2), 884–918.

— and — (2007b). Testing Forecast Optimality Under Unknown Loss. *Journal of American Statistical Association*, **102**, 1172–1184.

Predd, J., Seiringer, R., Lieb, E., Osherson, D., Poor, H. and Kulkarni, S. (2009). Probabilistic coherence and proper scoring rule. *IEEE Transactions on Information Theory*, **55**, 4786–4792.

Ravazzolo, F. and Vahey, S. (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.

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

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

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

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