Recursive Predictability Tests for Real-Time Data

ATSUSHI INOUE   BARBARA ROSSI

NC STATE   DUKE UNIVERSITY

June 2004

ABSTRACT.   We propose a sequential test for predictive ability. The test is designed for regressions in which the researcher is interested in *recursively* assessing whether some economic variables have predictive or explanatory content for another variable. It is common in the forecasting literature to assess predictive ability by using "one-shot" tests at each estimation period. We show that this practice: (i) leads to size distortions; (ii) selects overfitted models and provides spurious evidence of in-sample predictive ability; (iii) may lower the accuracy of the model selected by the test. The usefulness of the proposed test is shown in well-known empirical applications to the real-time predictive content of money for output, and the selection between linear and non-linear models.

Keywords: Sequential tests, predictive ability, model selection.

JEL Classification: C52, C53

Corresponding author: Barbara Rossi, Department of Economics, Duke University, Durham, NC27705 USA. Phone: 919 660 1801. E-mail: brossi@econ.duke.edu.

1. INTRODUCTION

Assessing whether there is predictability among macroeconomic variables has always been a central issue for applied researchers. For example, much effort has been devoted to analyzing whether money has predictive content for output. This question has been addressed by using both simple linear Granger Causality (GC) tests (e.g. Stock and Watson (1989)) as well as tests that allow for non-linear predictive relationships (e.g. Amato and Swanson (2001) and Stock and Watson (1999), among others). When parameters may be time-varying, and the objective of the researcher is to assess the presence of a relationship between two economic variables, it is tempting to use predictability tests recursively. While this procedure has the correct size at each point in time, it will not have the correct size over the whole sequence of test statistics. In particular, the overall size of the tests will approach one as the procedure is repeated more and more times. Similar problems are likely to occur when the researcher recursively tests whether inflation is under control, as many inflation-targeting Central Banks in practice do.

We propose a new recursive test for predictive ability that controls the overall size of the procedure and, hence, protects the researcher from overfitting. Our test applies to predictive regressions in which, at each point in time, the researcher tests whether a set of economic variables has predictive content for some variable of interest on the basis of an in-sample test using only observations available until that time, and the parameters are recursively re-estimated as time goes by. The outcome of the test

may be used as evidence of in-sample predictive ability as well as for out of sample forecasting purposes. Commonly used tests, whose critical values do not take into account the recursive nature of the test (referred to as "one-shot tests") will have size equal to the nominal (desired) level *at each point in time*. However, their recursive application will lead to severe size distortions. We instead derive the distribution of the test statistic under the null hypothesis by considering the recursive nature of the *whole testing procedure*. This allows us to derive the correct critical values, which can then be used to recursively test for predictive ability. The test statistics proposed in this paper are calculated as usual, but their critical values are different, and depend upon the sample size. These critical values can be easily calculated by using a table provided in the paper, so that applied researchers can directly apply the proposed test procedure. The test is similar in spirit to the fluctuation test discussed in Chu et al. (1996), but our test focuses on predictive ability. We also allow for a more general GMM framework and possibly nonlinear restrictions. The GMM framework can also be useful to select between linear and non-linear models, which is one of the empirical applications that we consider.

Our test is different from existing out-of-sample recursive tests for predictive ability (e.g. Clark and McCracken (2001, 2003d) for one step ahead predictions, and Clark and McCracken (2003c) for h-steps ahead predictions, under the maintained assumption of dynamic correct specification) or out-of-sample tests of Granger Causality (see Chao, Corradi and Swanson (2001), and Corradi and Swanson (2002) for an

out-of-sample test for Granger Causality which is consistent against generic alternatives, and which allows for dynamic misspecification under the null). In these tests, the available sample is given, i.e. it is considered *fixed*. The sample is recursively split into two subsets: one which is used to estimate the parameters, and one which is used to validate the forecasts of the model. Despite the fact that this procedure involves recursive estimation of the parameters, the test is, in essence, one-shot, because the sample size is given. Furthermore, our procedure can be applied to situations in which data available at different times vary as a result of redefinitions, a common situation for macroeconomic data (see Croushore and Stark (2001)).

Our discussion may shed some light on the fragile link between in-sample model selection and out-of-sample forecasting in real time. Stock and Watson (1989) apply in-sample Granger Causality tests and find some evidence that money has predictive content for output whereas more recent contributions find no evidence of out-of-sample predictive ability. Thus, what kind of guide do in-sample tests offer to out-of-sample predictive ability? In-sample and out-of-sample tests often provide contradictory results. These contradictory findings are often attributed to overfitting or low power of forecasting tests (Kilian and Inoue (2002)) or to the presence of parameter instability (Clark and McCracken (2003a,b,d)). This paper investigates another possible explanation, namely the fact that repeated tests for model selection might select overfitted models, thus deteriorating forecasting ability. On the other hand, the approach in this paper is valid only for comparing two nested models and,

in this sense, it cannot be viewed as a sequential alternative to the Diebold and

Mariano (1995) and West (1996) out-of-sample tests.

The paper is organized as follows. Section 2 discusses background and motivation,

Section 3 the main result of the paper: the recursive tests. Section 4 provides some

small Monte Carlo evidence on the size and power of the proposed tests, and shows

that they have both good size and power properties. Section 5 applies the recursive

tests to two important empirical applications: the relationship between money and

output, and the choice between linear and non-linear models for a few representative

macroeconomic variables. The last section concludes.

2.   BACKGROUND AND MOTIVATION

As a simple motivating example,[1] consider a researcher that has available a historical

dataset of size T. He is interested in testing a null hypothesis on a parameter at each

point in time $t > T$, that is, $t = T+1, T+2, ...$ For example, the researcher is interested

in assessing whether a scalar variable "x" has predictive content at any point in time

for another variable "y". That is, the researcher is interested in recursively testing

hypotheses on β_t in the regression: $y_{t+1} = \beta_t x_t + u_{t+1}$, where u_{t+1} satisfies the usual

linear regression assumptions. The null hypothesis is: $\beta_t = \beta_0$ at every $t \geq T+1$, and

the alternative is: $\beta_t \neq \beta_0$ for some $t \geq T + 1$. Let $\widehat{\beta}_t$ denote the recursive estimate

of β at every point in time $t = T+1, T+2, ...$, and let τ_t denote the associated t-test

statistic. To test the null hypothesis, one might simply perform a t-test using conven-

---

[1]This is just an example. The framework of this paper is much more general, as explained later.

tional (normal) large sample critical values at each t, and reject the null hypothesis if the t-test rejects at any point in time. Unfortunately, with conventional critical values, by the Law of Iterated Logarithm the probability that this test eventually rejects the null hypothesis is asymptotically one. Note that the same argument remains true if one uses any constant critical values, no matter how large. To remedy this problem, this paper derives critical values that allow to "follow" the test statistic through the whole sequence as $t = T + 1, T + 2, ...$ in such a way that the probability of rejecting the null hypothesis is under control at each t. This requires a boundary function such that the path of the test statistic crosses this boundary with the desired probability level under the null hypothesis. This is achieved by controlling the behavior of the test statistic as a function of $\pi \equiv t/T$ and exploiting results on boundary crossing probabilities (e.g. Chu et al., 1996) like $\lim_{T \to \infty} P \left\{ \sqrt{t}|\tau_t| \geq \sqrt{T}\psi\left(t/T\right), \text{ for some } t > T \right\} = \lim_{T \to \infty} P \left\{ |\frac{\tau_t}{\sqrt{\pi}}| \geq \psi\left(\pi\right), \text{ for some } \pi > 1 \right\}$, where $\psi\left(.\right)$ is the boundary function. There are various possible choices for the boundary function. For example, a fluctuation test of size α would use $\psi\left(\pi\right) = \sqrt{k_{1,\alpha}^2 + \ln\left(\pi\right)}$, where $k_{1,\alpha}$ is a constant such that $2[1 - \Phi\left(k_\alpha\right) + k_\alpha \phi\left(k_\alpha\right)] = \alpha$, α is the desired size, and Φ, ϕ are the c.d.f. and p.d.f. of a standard normal distribution. In this paper, we also propose critical values that result in more powerful tests when there is more than one restriction.

3. ASSUMPTIONS AND THEOREMS

Assume that $\{z_{T,t}\}$ is a triangular array of random variables. Consider estimation of

the parameter θ based on moment conditions

$$E[g(z_{T,t}, \theta_0)] = 0$$

The researcher is interested in testing the null hypothesis: $H_0 : a(\theta_0) = 0$, versus

the alternative: $a(\theta_0) \neq 0$, where $g : Z \times \Theta \to \Re^q$, $Z \subset \Re^k$, $\Theta \subset \Re^p$ and $a : \Theta \to \Re^r$.

Define a sequential unconstrained GMM estimator of θ by:[2]

$$\widehat{\theta}_t = \widehat{\theta}_T(\pi) = \arg\min_{\theta \in \Theta} \widehat{Q}_T(\theta, \pi), \quad \widehat{Q}_T(\theta, \pi) \equiv \frac{1}{2} \hat{g}_T(\theta, \pi)' \hat{S}_T(\pi)^{-1} \hat{g}_T(\theta, \pi),$$

where $t = [T\pi] = T + 1, T + 2, ...$ is the time at which the monitoring takes place,

$\hat{g}_T(\theta, \pi) = (1/[T\pi]) \sum_{t=1}^{[T\pi]} g(z_{T,t}, \theta)$ is the sample analogue of $E[g(z_{T,t}, \theta)]$ and $\widehat{S}_T(\pi)$

is a sequence of consistent estimators for the long-run covariance matrix of $g(z_{T,t}, \theta_0)$.

Similarly, define a sequential constrained GMM estimator by

$$\bar{\theta}_t = \arg\min_{\theta \in \Theta} Q_T(\theta, \pi) \quad subject\ to\ a(\theta) = 0$$

Define the sequential Wald, Lagrange Multiplier and LR-like test statistics by $W_t =$

---

[2] Note that $\widehat{\theta}_T(\pi)$ is simply another way to rewrite $\widehat{\theta}_t$, as $\pi = t/T$.

$W_T(\pi)$, $LM_t = LM_T(\pi)$ and $LR_t = LR_T(\pi)$ where

$$W_T(\pi) \;\;=\;\; [T\pi]a(\hat{\theta}_T(\pi))[A(\hat{\theta}_T(\pi))(\hat{G}'_T(\pi)\hat{S}_T(\pi)^{-1}\hat{G}_T(\pi))^{-1}A(\hat{\theta}_T(\pi))']^{-1}a(\hat{\theta}_T(\pi)),$$

$$LM_T(\pi) \;\;=\;\; [T\pi]\nabla Q_T(\bar{\theta}_T(\pi),\pi)'(\bar{G}'_T(\pi)\bar{S}_T(\pi)^{-1}\bar{G}_T(\pi)))^{-1}\nabla Q_T(\bar{\theta}_T(\pi),\pi),$$

$$LR_T(\pi) \;\;=\;\; -2[T\pi](Q_T(\hat{\theta}_T(\pi),\pi) - Q_T(\bar{\theta}_T(\pi),\pi)),$$

$\hat{G}_T(\pi) \;=\; (1/[T\pi])\sum_{t=1}^{[T\pi]} \partial g(z_{T,t},\hat{\theta}_T)/\partial\theta'$, $\bar{G}_T(\pi)$ and $\bar{S}_T(\pi)$ are $\hat{G}_T(\pi)$ and $\hat{S}_T(\pi)$ with $\hat{\theta}_T(\pi)$ replaced by $\bar{\theta}_T(\pi)$, respectively, and $A(\theta) = \partial a(\theta)/\partial\theta'$.

We derive the limiting distributions of the test statistics under both the null hypothesis and a sequence of local alternatives (see (c) below) under the following:

Assumptions:

(a) z_{Tt} is strictly stationary for every T.

(b) $\Theta \subset \Re^p$ is compact and $g(.)$ is continuous in θ with probability one.

(c) For each T, there is a unique $\theta_T = \theta_0 + \delta/\sqrt{T}$ that satisfies

$$E[g(z_{Tt},\theta_T)] \;\;=\;\; 0.$$

(d) For every integer n that is greater than 1,

$$\frac{1}{[\pi T]}\sum_{t=1}^{[\pi T]} g(z_{Tt},\theta) \;\;=\;\; E[g(z_t,\theta)] + o_p(1), \tag{1}$$

$$\frac{1}{[\pi T]}\sum_{t=1}^{[\pi T]} \frac{\partial}{\partial\theta'}g(z_{Tt},\theta_0) \;\;=\;\; G + o_p(1), \tag{2}$$

$$\hat{S}_{[\pi T]} \;\;=\;\; S + o_p(1), \tag{3}$$

where $G = E[\partial g(z_{Tt},\theta_0)/\partial\theta']$ and $o_p(1)$ is uniform in $\pi \in [1,n]$ and in $\theta \in \Theta$.

(e) For every integer n that is greater than 1,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[\cdot T]} g(z_{Tt}, \theta_0) \Rightarrow S^{\frac{1}{2}} W(\cdot)$$

where the weak convergence is defined in the space of right-continuous functions with left-hand limits on $[0, n]$ equipped with the Skorohod topology.

(f) G is of rank p.

(g) $\hat{S}_{[\pi T]}$ is positive semidefinite for all $\pi \in [1, \infty)$ and T and S is positive definite.

(h) a is continuously differentiable and $A = \partial a(\theta_0)/\partial \theta'$ is of rank r.

The strict stationarity imposed in Assumption (a) is not necessary but simplifies the notation. The compactness of the parameter space in Assumption (b) can be dropped for linear regression models, but otherwise (b) ensures the existence of the estimator. Assumption (c) states that the moment conditions are correctly specified and the parameters are identified. It also specifies a sequence of local alternatives. When $c \neq 0$, z_{Tt} becomes a triangular array because the true parameter value θ_T depends on T. Assumptions (d) and (e) assume uniform convergence and weak convergence for $\pi \in [1, n]$. These assumptions are high-level assumptions; more primitive sufficient conditions can be found in the literature (e.g., Andrews, 1993). Assumptions (f), (g) and (h) are standard rank, positive definiteness and smoothness conditions. All these Assumptions are standard for extremum estimators. From these assumptions, it immediately follows that the GMM estimator converges in probability to θ_0 uniformly in $\pi \in [1, n]$ and is asymptotically normally distributed for $\pi \in [1, n]$.

The following Theorem extends weak convergence over $\pi \in [1, n]$ to weak convergence over $\pi \in [1, \infty)$:

Theorem 1: Under Assumptions (a)-(h) we have:

$$\sqrt{[T\cdot]}(\hat{\theta}_T(\cdot) - \theta_0) \Rightarrow (G'S^{-1}G)^{-1}G'S^{-1}\left(\left(\frac{S}{\cdot}\right)^{\frac{1}{2}}B_q(\cdot) + G\delta\right), \qquad (4)$$

$$\sqrt{[T\cdot]}(\bar{\theta}_T(\cdot) - \theta_0) \Rightarrow (G'S^{-1}G)^{-\frac{1}{2}}M(G'S^{-1}G)^{-\frac{1}{2}}G'S^{-1}\left(\left(\frac{S}{\cdot}\right)^{\frac{1}{2}}B_q(\cdot) + G\delta\right), \quad (5)$$

$$W_T(\cdot) \Rightarrow \Psi, \quad LM_T(\cdot) \Rightarrow \Psi, \quad LR_T(\cdot) \Rightarrow \Psi \qquad (6)$$

where $\Psi \equiv (\frac{1}{\sqrt{\cdot}}B_q(\cdot) + S^{-\frac{1}{2}}G\delta)'C'(CC')^{-1}C(\frac{1}{\sqrt{\cdot}}B_q(\cdot) + S^{-\frac{1}{2}}G\delta)$, $B_q(\cdot)$ *is the q-dimensional*

standard Brownian Motion, $C \equiv A(G'S^{-1}G)^{-1}G'S^{-1/2}$, $M \equiv I_p - (G'S^{-1}G)^{-1/2}A'$

$(A(G'S^{-1}G)^{-1}A')^{-1}A(G'S^{-1}G)^{-1/2}$.

Remarks: 1) For fixed π, both GMM estimators are asymptotically normally distributed, and the three statistics are asymptotically distributed as $\chi^2(r)$ because $C'(CC')^{-1}C$ is idempotent with rank r. 2) Theorem 1 also shows that, when π is fixed and $\delta \neq 0$, the test statistics are asymptotically distributed as non-central $\chi^2(r)$ distributions with non-centrality parameter equal to $\bar{\delta} \equiv (G\delta)'S^{\frac{1}{2}}C'(CC')^{-1}CS^{\frac{1}{2}}G\delta$. Thus, the tests have nontrivial power provided $\bar{\delta} > 0$.

The next Theorem provides the probability of accepting H_0 when it is true.

Theorem 2: Under Assumptions (a)-(h) with $\delta \equiv 0$, as $T \rightarrow \infty$, we have:

$$P(W_t \leq c^2 + r\ln\left(\frac{t}{T}\right), \forall\, t \geq T) \rightarrow \xi$$

$$P(LM_t \leq c^2 + r\ln\left(\frac{t}{T}\right), \forall\, t \geq T) \rightarrow \xi$$

$$P(LR_t \leq c^2 + r\ln\left(\frac{t}{T}\right), \forall\, t \geq T) \rightarrow \xi,$$

where $\xi \equiv P(B_r(\pi)'B_r(\pi)/\pi \leq c^2 + r\ln(\pi), \forall\pi \geq 1) = \int \Pi_{i=1}^r \{1 - 2[1 - \Phi(a_i) +$

$a_i\phi(a_i)]\}da$, and $B_r(\cdot) = (CC')^{-1/2}CB_q(\cdot)$ is the r-dimensional standard Brownian motion and the integrations are taken over the set $\{a = (a_1, ..., a_r) : 0 \leq a_i \leq c, \forall i = 1, 2, ..., r, \sum_{j=1}^r a_j^2 = c^2\}$.

To make the tests operational, one needs c. For example, to test at the 5% significance level, one needs to find c such that ξ is equal to 0.95. Because it is difficult to evaluate the integral numerically for mildly large r, we estimate c by Monte Carlo simulation. First, we generate r-dimensional random vectors $\{a^{(j)}\}_{j=1}^J$ that are uniformly distributed over the set $\{a = (a_1, a_2, ..., a_r) : 0 \leq a_i \leq c, \forall i = 1, 2, ..., r, \sum_{j=1}^r a_j^2 = c^2\}$. Second, we estimate the integral by averaging

$$\frac{1}{J} \sum_{j=1}^J \Pi_{i=1}^r [1 - 2(1 - \Phi(a_i^{(j)}) + a_i^{(j)}\phi(a_i^{(j)}))] \tag{7}$$

Lastly, we choose c so that the Monte Carlo integral is close to $1 - \alpha$, where α is the desired significance level. Table 1 readily provides values for c to be used in the construction of the critical values when $r \leq 14$. Since these depend on both r and α, we denote them by $c_{r,\alpha}$. Critical values for the Recursive Wald, LM and LR tests of size α at time t with r restrictions and first monitoring time $T + 1$ are obtained from Table 1 as $c_{r,\alpha}^2 + r \ln(\frac{t}{T})$. Thus, the critical values depend upon both the actual sample size, t, and the historical sample size, T.

It is very easy to implement the test in practice. The researcher calculates the Wald, LM or LR test as usual (and, thus, with commonly used packages). The

critical values of size α for the test statistic calculated at time t with r restrictions and first monitoring time $T + 1$ are obtained as $c_{r,\alpha}^2 + r \ln(\frac{t}{T})$, where the values of $c_{r,\alpha}$ are obtained from Table 1. Note that this formula and Table 1 allow the applied researcher to readily calculate the critical value to be used in a particular economic application without having to calculate (7), so that no simulations nor complicated computations are required. An example of how to implement the test in practice is presented in the next section.

Note that the test statistic proposed in this paper is different from the fluctuation test obtained by applying Chu et al. (1996) to our problem. First of all, it is more powerful. We investigate its local power and compare with that of the fluctuation test in the next section. Finally, we consider a more general framework (GMM and possibly nonlinear restrictions).

<div align="center">INSERT TABLE 1</div>

<div align="center">4. A SMALL MONTE CARLO EXPERIMENT</div>

We are interested in the situation in which the researcher is making real-time tests for Granger Causality. At each point in time $t = T, .., T^{\max}$, the researcher chooses between two models. For simplicity, we assume that the models are nested and linear:

$M_0 : y_{t+1} = \epsilon_{t+1}$, vs. $M_p : y_{t+1} = \beta_t' x_t^{(p)} + \epsilon_{t+1}$, $p = 1, 2$ and $T = 250$, $T^{\max} = 500$. Note that the superscript identifying the model ("p") is equal to the number of estimated parameters and also to the number of restrictions. We assume that the DGP is $y_{t+1} = \beta' x_t + \epsilon_{t+1}$, where $\beta = 0$, ϵ_t is an i.i.d. scalar standard normal, $x_t^{(p)}$

is a p-dimensional normal with zero mean and covariance $\Sigma_X^{(2)} = \begin{pmatrix} 1 & \rho_X \\ \rho_X & 1 \end{pmatrix}$ for

$p = 2$ and $\Sigma_X^{(1)} = 1$ for $p = 1$. The choice between the two models at time t is based

on an in-sample test using observations available until time t. As models are nested,

the model selection device is an F-test for $\beta = 0$, where β is a rolling estimate.

We will consider the following tests, and show that, among them, only the Re-

cursive Wald test has both good size and power properties:

- the *"Fluctuation test"*:

$$F_t = \sqrt{t}\widehat{\Sigma}_t^{-1/2}\widehat{\beta}_t \tag{8}$$

where $\widehat{\Sigma}_t \equiv \widehat{\sigma}^2 \left(\frac{1}{t}\sum_{s=1}^t x_s x_s\right)^{-1}$ and $\widehat{\sigma}^2$ is the estimated variance of the residuals of

the regression up to time t. For $t = [T\pi]$, from the functional central limit theorem, F_t

converges (weakly) to $B_p\left(\pi\right)/\sqrt{\pi}$, where $B_p\left(\pi\right)$ is a standard p-dimensional Brownian

motion. This test statistic is similar in spirit to that proposed by Chu et al. (1996).

These authors developed a test statistic for recursively testing for structural breaks;

our test is a recursive test for predictive ability. Boundary crossing probabilities for

(8) can be derived by using the methods discussed in Chu et al. (1996, p. 1055).

Let α denote the desired size of the test. Then, the test rejects when $max_p|F_t| >$

$\sqrt{(k_{p,\alpha})^2 + r\ln\left(t/(T)\right)}$, where $k_{p,\alpha}$ solves $1 - [1 - 2\left(1 - \Phi\left(k_{p,\alpha}\right) - k_{p,\alpha}\phi\left(k_{p,\alpha}\right)\right)]^p = \alpha$.

- the *"Recursive Wald test"* proposed in this paper:

$$W_t = t\widehat{\beta}_t'\widehat{\Sigma}_t^{-1}\widehat{\beta}_t \tag{9}$$

By noting that $W_t = F_t'F_t$ and by applying the continuous mapping theorem to (9): $W_t \Rightarrow R_p(\pi)^2/\pi$, where $R_p(\pi)$ is the Bessel process, that is the square root of a sum of p-dimensional squared Brownian Motions. Boundary crossing probabilities and critical values can be obtained as described in the previous section.

   - the "*One-shot Granger Causality (GC) test*", where, for each $t = [\pi T]$, as $T \to \infty$, $W_t = t\widehat{\beta}_t'\widehat{\Sigma}_t^{-1}\widehat{\beta}_t$ has a χ_p^2 asymptotic distribution.[3]

To clarify how to implement these tests, let's consider the following concrete example. Let $p = 1$, $\alpha = 0.05$, and we monitor at $t = 251,..500$ starting from the historical sample $T = 250$. To implement the Recursive Wald test at time t, calculate the usual Wald test at time t and then reject if, at any t, the estimated Wald test statistic is bigger than $c_{1,\alpha}^2 + \ln(t/250)$. To implement the Fluctuation test at time t, calculate F_t and $k_{1,\alpha}$ such that $2(1 - \Phi(k_{1,\alpha}) - k_{1,\alpha}\phi(k_{1,\alpha})) = \alpha$, and reject if, at any t, $|F_t| > \sqrt{k_{1,\alpha}^2 + \ln(t/250)}$. Finally, to implement the one-shot GC test, reject if, at any t, the estimated Wald test is bigger than the $1 - \alpha$ quantile of a χ_1^2.

We will consider the following Monte Carlo experiments.

Local power comparison. We perform a simple local power comparison of the tests that have the correct size asymptotically (F_t and W_t) and a size-adjusted (by simulation) One-shot test. Let $p = 1$ or 2, $\rho_X = 0$ or 0.7, $T^{\max} = 500$, $T = 250$ and the number of Monte Carlo replications be 1000. Figure 1 shows the resulting power functions (see also Table 2 for details). It is clear that the recursive Wald test has

---

[3] That is, for any π, W_t has a $B_p(1)'B_p(1)$ asymptotic distribution, which is a χ_p^2.

higher power than the Fluctuation test when $p > 1$. Note that the size-adjusted One-shot test has higher power than W_t, but it would have the wrong size if its critical values, which are computed for a given $T^{\max}$, are used recursively with a bigger total sample size.

<div align="center">INSERT FIGURE 1 AND 2, AND TABLE 2</div>

Recursive model selection. To investigate size, we perform the following Monte Carlo experiment. At times $t = T + 1, ..., T^{\max}$ (where $T = 250$) we use both the one-shot and the recursive tests to choose between Models 1 and 2. Figure 2 plots the fraction of times in which each test rejected at least once. Note that, for this design, the one-shot Wald test has a 0.10 probability of (incorrectly) rejecting the true model at least once after 30 periods, whereas the proposed recursive procedures successfully control size.

<div align="center">5.   EMPIRICAL APPLICATIONS</div>

The real-time predictive content of money for output.

Amato and Swanson (2001) test in-sample marginal significance of money for output in autoregressive models by using Wald and SIC tests based on recursive estimation. They find evidence that money is significant in revised data, (although the evidence is somewhat weaker if one uses un-revised data). Here below we consider a similar exercise. First, we recursively test whether money has predictive content for output at each point in time between 1978:1 and 2002:4. Second, based upon

the results of the recursive test, we select the forecasting model. Forecasts use only data available at the time in which the forecast is made and are one-step ahead.[4] We consider the realistic situation in which data are released in real-time. We use quarterly data for money (M1 and M2) and industrial production (IPT) for the U.S. provided by the Federal Reserve Bank of Philadelphia, as well as CPI and 3-month Treasury Bills as a measure of interest rates. The data are discussed in Croushore and Stark (2001). The first estimation period is from 1959:01 to 1978:01, and forecasts are made until 2002:4. All variables (except interest rates) are in logarithms, and first differenced. See Appendix 2 for details.

Results for regressions with different lag structure and different regressors are reported in Table 3. The first four columns report the number of lags included of: interest rate (i), money supply M1 or M2 (m), industrial production (ip) and prices (p). If a variable is not included as a regressor, the column reports "0". Rows denoted by "AIC" ("SIC") report the results if the researcher recursively selects the lag length by AIC (SIC). Every regression includes a constant. MSE_F and MSE_W denote, respectively, the ratio of RMSE (i.e. Root Mean Squared Error) of forecasts based on models selected by the Fluctuations and Wald tests *relative to* the RMSE of the one-shot GC test. R_F, R_W and R_{GC} respectively denote the fraction of times

---

[4]That is, real time observations that would have been available to someone at that point in time, the vintage date. We assume a floating-date information set, in which the exact day of the information set is given by the 15th when the day of the release of the industrial production report falls on the 15th or before, or by the date of the industrial production report, when that date exceeds the 15th.

the Fluctuations, Wald and one-shot GC tests rejected the null hypothesis.

The results show that, for M1, the recursive tests reject the hypothesis that money does not Granger Cause output less often than the one-shot test. For M2, instead, the results of the recursive and one-shot tests are very similar. Furthermore, forecasts based on the recursive tests are, on average, better than those based on the one-shot GC test for M1, but not for M2. In fact, for M1, over 27 estimated regressions, the forecasts based on the recursive Wald test are better than the one-shot test approximately seventy percent of the times. For M2, the former are always better than or equal to the latter when based on the Fluctuation test, and when based on the Wald test they are worse only once. As it is clear from the table, this is generally due to the fact that the recursive tests reject less often the null hypothesis of no Granger Causality, thus relying on the restricted (smaller) model more often than the one-shot test. This result may shed some light on the fragile link between in-sample model selection and out-of-sample forecasting in real time: one-shot tests, when used recursively, tend to select overfitted models, which deteriorate forecasts.

<div align="center">INSERT TABLE 3, AND FIGURE 3</div>

By comparing the R_F, R_W and R_{GC} columns of Table 3, our results suggest that M1 does have some predictive content for output, but M2 does not. Also, Figure 3 shows which dates the tests do (do not) reject Granger Causality for a particular model specification in Table 3 (one in which the Recursive Wald and the one-shot tests disagree over some portion of the sample). The picture plots the asymptotic

critical values of the Recursive Wald test and the one-shot tests. Note that the critical values of one-shot tests are constant, whereas those of the recursive test are generally increasing with the sample size (the only times in which they are not, it is because some data on any of the variables used in the regression were not available at that time, and this reduced the sample size, e.g. in 1981). According to our method, there is no evidence of Granger Causality of money for output especially for most of the 1980's and the late 1990's.[5]

INSERT TABLE 4

Linear versus non-linear models, and forecasting

Many nonlinear models have a linear model nested in the nonlinear structure. Before using a nonlinear model, it is advisable to test whether the linear model is an adequate description of the economic relationships of interest. If that turns out to be the case, the researcher can avoid estimating unnecessarily complicated models. Furthermore, a non-linear model might appear to fit the data better than a linear model even if the true economic relationship is linear, but its forecasts would be worse out-of-sample. Indeed, Stock and Watson (1999) compare forecasts of linear and non-linear models and conclude that linear models appear to forecast better than non-linear models, on average. Marcellino (2002) finds similar results. Swanson and White (1997a,b) also found that multivariate adaptive linear VAR models often

---

[5]In results not reported, we also find less evidence of Granger causality of other macroeconomic variables (not only money) for output in the early 1980s.

outperform (in terms of forecasting) adaptive nonlinear models. As the true economic relationship may change over time from linear to non-linear (and vice-versa), Swanson and White (1997b, p. 540) allow their econometric model to be recursively chosen over time between linear and non-linear, and analyze whether this possibility can improve forecasting ability out-of-sample. Following their approach, we call a model *adaptive* if a new specification is chosen before each new rolling forecast is constructed.

We revisit this empirical evidence by using our testing procedure to recursively select between nested linear and non-linear models for a series of macroeconomic variables described in Appendix 2. We focus on Artificial Neural Network (ANN) versus linear models, both because ANN are commonly used (see Swanson and White (1997b)) and because Stock and Watson (1999) found that ANN outperform other non-linear model specifications. Let $\varsigma_t = [y_{t-1}, y_{t-2}, ..., y_{t-p}]$; the single layer neural network model (with 1 hidden unit) is:

$$y_t = \beta_0' \varsigma_t + \gamma_1 g\left(\beta_1' \varsigma_t\right) + \epsilon_t, \quad g\left(z\right) = (1 + e^z)^{-1} \tag{10}$$

We perform a general test for non-linearities based on LM principles due to Teräsvirta et al. (1993). It is a test for $H_0 : \varphi_{ij} = 0$ in the auxiliary regression:

$$\widehat{u}_t = \beta_0' \varsigma_t + \sum_{i=1}^{p} \sum_{j=1}^{p-1} \varphi_{ij} \varsigma_t^{(i)} \varsigma_t^{(j)} \tag{11}$$

where the dependent variable, $\widehat{u}_t$, is the residual of a linear regression of y_t onto ς_t,

and $\varsigma_t^{(i)}$ is the i-th column of ς_t (constant excluded). Teräsvirta et al. (1993) show that this test performs better than tests explicitly designed for neural networks. If the test does reject the hypothesis that the model is linear, we estimate (11) or (10).[6]

Table 4 reports the results. The table focuses on four basic macroeconomic variables (unemployment, real and nominal GDP, and consumption) and a variety of univariate autoregressive models, characterized by different lag lengths p. R_F, R_W and R_{GC} respectively denote the frequency by which the Fluctuations, Recursive Wald and one-shot tests rejected the null hypothesis over the sample. MSE_F^{SW} denotes the ratio (relative to one-shot tests) of the RMSEs of forecasts obtained as follows. At each point in time, we use our method and the test statistic (11) to select between a linear and a non-linear model; if the test does reject linearity, forecasts are made by using the Stock and Watson method. MSE_F and MSE_W instead use model (10) to forecast; the former chooses the model by using the Fluctuation test whereas the latter uses the recursive Wald test. "Lags" denotes the number of lags used in the autoregression. The results show that the recursive linearity tests based on our critical value reject the linear model less often than one-shot tests. The latter end up selecting an over-parametrized non-linear model too often than desired. The table also shows that forecasts based on our recursive test for linearity are better

---

[6]The ANN model is estimated by a modified version of an algorithm developed by Stock and Watson (1999), see M. Watson's webpage: http://www. econ.princeton.edu/ ~mwatson. The only difference is that we use the faster Gauss-Newton algoritm provided by Matlab (fminunc.m) to estimate the parameters. It is possible to use other tests for non-linearities such as Hong and Lee (2003). These tests are more powerful to reject non-linearities and would reinforce our results.

than those based on usual one-shot critical values.

6. CONCLUSIONS

This paper proposes tests for recursive predictive ability. The test statistics are the same as those commonly used in the literature, but the critical values are different. These critical values can be easily calculated by using Table 1, so that the applied researcher can directly implement the proposed test. We show in Monte Carlo simulations that our test has good power and can significantly improve forecasts based on recursive model selection procedures. By using our test, we find weaker empirical evidence of predictive ability of money for output, and of non-linear relationships between some representative macroeconomic variables.

While the method proposed in this paper is easy to implement and performs well in our Monte Carlo simulations, alternative approaches to inference are possible. In particular, one might implement LIL bounds, as proposed by Altissimo and Corradi (2003). As shown by these authors, LIL bound procedures are straightforward to calculate and implement, but require some computational adjustments in small samples. While these procedures have both zero size and power one for fixed alternatives, they will have no power against the local alternatives considered in this paper. They are therefore not considered here, but nevertheless remain an interesting option for future research.

7.   REFERENCES

Altissimo, F. and V. Corradi (2003), "Strong Rules for Detecting the Number of Breaks in a Time Series", *Journal of Econometrics* 117, 207-244.

Amato, J. and N. Swanson (2001), "The real-time predictive content of money for output", *Journal of Monetary Economics* 48, 3-24.

Andrews, D.W.K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica,* 61, 821–856.

Billingsley, P. (1968), *Convergence of Probability Measures,* New York: John Wiley & Sons.

Billingsley, P. (1995), *Probability and Measure*, Third Edition, New York: John Wiley & Sons.

Chao, J., V. Corradi and N. Swanson (2001), "An out-of-sample test for Granger Causality", *Macroeconomic Dynamics* 5(4)*, 598-620.*

Chu, C. J. , M. Stinchcombe and H. White (1996), "Monitoring structural change", *Econometrica* 64(5), 1045-1065.

Clark, T. and M. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models", *Journal of Econometrics*, 105, 85-110.

Clark, T. and M. McCracken (2003a), "The Predictive Content of the Output Gap for Inflation: Resolving In-Sample and Out-of-Sample Evidence", *mimeo.*

Clark, T. and M. McCracken (2003b), "Forecast Accuracy and the Choice of Observation Window", *mimeo*, Kansas City Fed.

Clark, T. and M. McCracken (2003c), "Evaluating Long Horizon Forecasts", *mimeo.*

Clark, T. and M. McCracken (2003d), "The Power of Tests of Predictive Ability in the Presence of Structural Breaks", *Journal of Econometrics,* forthcoming.

Corradi, V. and N. Swanson (2002), "A Consistent Test for Out of Sample Nonlinear Predictive Ability", *Journal of Econometrics* 110, 353-381.

Croushore, D. and T. Stark (2001), "A real-time data set for macroeconomists", *Journal of Econometrics* 105(1), November, 111-130.

Diebold, F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics* 13, 253-263.

Hong, Y. and Y. Lee (2003), "Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroskedasticity of Unknown Form", *mimeo.*

Inoue, A. and L. Kilian (2002), "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?", *European Central Bank W.P. No. 195.*

Marcellino, M. (2002), "Instability and non-linearity in the EMU", *mimeo.*

Stock, J.H. and M.W. Watson (1989), "Interpreting the Evidence on Money-Income Causality", *Journal of Econometrics* (40), 161-182.

Stock, J.H. and M.W. Watson (1999), "A comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series", in: R. Engle and H. White (eds.), *Cointegration, Causality, and Forecasting - Festschrift in honour of Clive W. J. Granger*, Oxford University Press.

Swanson, N., and H. White (1997a), "Forecasting economic time series using flexible versus fixed specification and linear versus non-linear econometric models", *International Journal of Forecasting* 13, 439-461.

Swanson, N., and H. White (1997b), "A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks", *Review of Economics and Statistics*, 540-550.

Swanson, N. (1998), "Money and output viewed through a rolling window", *Journal of Monetary Economics* 41, 455-473.

Teräsvirta, T., C.-F. Lin and C.W.J. Granger (1993), "Power of the neural network linearity test", *Journal of Time Series Analysis,* 14, 209-220.

van der Vaart A., and J.A. Wellner (1996), *Weak Convergence and Empirical Processes with Applications to Statistics,* Springer-Verlag: New York.

West, K. (1996), "Asymptotic Inference about Predictive Ability", *Econometrica* 64(5)*, 1067-1084.*

8. Appendix 1: Proofs

Proof of Theorem 1: Let $D[0,n]$ and $D[0,\infty)$ denote the spaces of right-continuous functions with left-hand limits on $[0,n]$ and $[0,\infty)$, respectively. Equip $D[0,n]$ with the Skorohod topology. For $z_1, z_2 \in D[0,\infty)$, define $D[0,\infty)$ with $\rho(z_1, z_2) = \sum_{n=1}^{\infty} 2^{-n} \min(\rho_n(z_1, z_2), 1)$, where $\rho_n(z_1, z_2)$ is the Skorohod topology for the restrictions of z_1 and z_2 to $D[0,n]$.

First, we prove that the GMM estimator is consistent uniformly in $\pi \in [1, \infty)$:

$$\hat{\theta}_T(\pi) = \theta_0 + o_p(1) \tag{12}$$

where $o_p(1)$ is uniform in $\pi \in [1, \infty)$. By Assumption (d) it follows that

$$\sup_{\pi \in [1,n]} \sup_{\theta \in \Theta} |\hat{Q}_{[\pi T]}(\theta) - Q(\theta)| = o_p(1), \tag{13}$$

where $Q(\theta) = \lim_{T \to \infty} E[g(z_{Tt}, \theta)']S^{-1}E[g(z_{Tt}, \theta)]$. It follows from Assumptions (c), (d), (g) that, for every neighborhood $\Theta_0 \subset \Theta$ of θ,

$$\inf_{\pi \in [1,n]} \inf_{\theta \in \Theta/\Theta_0} |Q(\theta) - Q(\theta_0)| > 0 \tag{14}$$

A slight modification of Lemma A.1 of Andrews (1993, p.846), (13) and (14) imply

$$\hat{\theta}_T(\pi) = \theta_0 + o_p(1), \tag{15}$$

where $o_p(1)$ is uniform in $\pi \in [1, n]$. By Theorem 25.2 of Billingsley (1995, p. 330) it follows from (15) that

$$\hat{\theta}_T(\cdot) \Rightarrow \theta_0 \tag{16}$$

where the weak convergence is uniform in $\pi \in [1, n]$. Note that $D[1, \infty) = \cup_{n=2}^{\infty} D[1, n]$ and that elements in $D[1, n]$ are uniformly bounded for $n = 2, 3, ...$ (see e.g. Billingsley, 1968, p.110). Thus, by Theorem 1.6.1 of van der Vaart and Wellner (1995, p. 43), it follows that (16) holds with weak convergence uniform in $\pi \in [1, \infty)$. Thus it follows that

$$\sup_{\pi \in [1,\infty)} \left(\hat{\theta}_T(\pi) - \theta_0 \right) \Rightarrow 0 \tag{17}$$

By applying Theorem 25.3 of Billingsley (1995, p. 331) to (17), the uniform convergence (12) follows. The proof for the uniform consistency of the constrained GMM estimator is analogous and is thus omitted.

Next we will prove that the unconstrained GMM estimator is asymptotically normal. Suppose that n is a fixed positive integer greater than 1. Then it follows

from analogous arguments used in the proof of Theorem 1 of Andrews (1993) that

$$
\begin{aligned}
\sqrt{T}(\hat{\theta}_T(\cdot) - \theta_0) &= [\hat{G}_T(\hat{\theta}(\cdot))'\hat{S}_T^{-1}(\cdot)\hat{G}_T(\hat{\theta}(\cdot))]^{-1}\hat{G}_T(\hat{\theta}(\cdot))'\hat{S}_T^{-1}(\cdot)\sqrt{T}\cdot\hat{g}_T(\cdot) + o_p(1) \\
&= (G'S^{-1}G)^{-1}GS^{-1}\sqrt{T\pi}\hat{g}_T(\cdot) + o_p(1), \qquad\qquad (18) \\
\sqrt{T}\cdot(\bar{\theta}_T(\cdot) - \theta_0) &= [\hat{G}_T(\bar{\theta}(\cdot))'\bar{S}_T(\cdot)^{-1}\hat{G}_T(\bar{\theta}(\cdot))]^{-1/2}\hat{M}_T[\hat{G}_T(\bar{\theta}(\cdot))'\bar{S}_T(\cdot)^{-1}\hat{G}_T(\bar{\theta}(\cdot))]^{-1/2} \\
&\quad \times \hat{G}_T(\hat{\theta}(\cdot))'\hat{S}_T^{-1}(\cdot)\sqrt{T}\cdot\hat{g}_T(\theta_0,\cdot) + o_p(1) \\
&= (G'S^{-1}G)^{-1/2}M(G'S^{-1}G)^{-1/2}G'S^{-1}\sqrt{T}\cdot\hat{g}_T(\theta_0,\cdot) + o_p(1), \quad (19)
\end{aligned}
$$

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\cdot]} g(z_{Tt}, \theta_T) \;\Rightarrow\; S^{\frac{1}{2}}W_r(\cdot) + G\delta \qquad\qquad (20)
$$

on $D[1, n]$ where $o_p(1)$ is uniform. Next, it follows from Theorem 1.6.1 of van der Vaart and Wellner (1996, p.43) that (18), (19) and (20) hold on $D[1, \infty)$, which proves (4) and (5). (6) follows by applying the continuous mapping theorem.□

Proof of Theorem 2: Under the null hypothesis, i.e, $\delta \equiv 0$, Theorem 1 implies

$$
W_T(\pi) \Rightarrow B_r(\pi)'B_r(\pi)/\pi
$$

and a similar expression holds for $LM_T(\pi)$ and $LR_T(\pi)$. Thus,

$$
W_T(\pi) - r\ln(\pi) \Rightarrow B_r(\pi)'B_r(\pi)/\pi - r\ln(\pi), \qquad\qquad (21)
$$

and similar expressions hold for $LM_T(\pi)$ and $LR_T(\pi)$. Since $B_r(\pi)'B_r(\pi)/\pi - r\ln(\pi)$ is the sum of r stochastically independent processes, that is, the sum of (squares of independent Brownian motions minus $\ln(\pi)$), we have

$$
P(B_r(\pi)'B_r(\pi)/\pi - r\ln(\pi) \le c^2, \; \forall\pi \ge 1) = P(\sum_{i=1}^{r}(B_r^{(i)2}(\pi)/\pi - \ln(\pi)) \le c^2, \; \forall\pi \ge 1)
$$

$$
= \int \Pi_{i=1}^{r} P(B_r^{(i)2}(\pi)/\pi - \ln(\pi) \le a_i^2, \; \forall\pi \ge 1)da, \qquad\qquad (22)
$$

where the integration is taken over the set $\{a = (a_1, a_2, ..., a_r) : \; 0 \le a_i \le c, \forall i = 1, 2, ..., r, \; \sum_{j=1}^{r} a_j^2 = c^2\}$. By equation (7) of Chu et al. (1996, p.1052), we have $P(B_r^{(i)2}(\pi)/\pi - \ln(\pi) \le a_i^2, \; \forall\pi \ge 1) = 1 - 2(1 - \Phi(a_i) + a_i\phi(a_i))$ for $i = 1, 2, ..., r$, and Theorem 2 immediately follows from this and (21).□

9. Appendix 2: Data description

The data used in the real-time predictive content of money for output exercise are real-time data for money (M1 and M2) and industrial production (IPT) for the U.S. from the database provided by the Federal Reserve Bank of Philadelphia (Federal Reserve Bank of Philadelphia, "Notes on the Philadelphia Fed's Real Time Data Set for Macroeconomists", at: http://www.phil.frb.org/econ/forecast/reaindex.html). In addition, we also use 3-month Treasury Bills as a measure of interest rates (like all financial data, they are never revised). The data are discussed in Croushore and Stark (2001). Monthly data are available from 1959:01 to 2002:12 and at quarterly vintages starting from 1965:11 to 2002:11 for M1 and from 1973:2 to 2002:11 for M2. All data are seasonally adjusted. We merged quarterly and monthly data following the suggestions of the Notes on the Philadelphia Fed's Real Time Data Set for Macroeconomists, sec. IV, "The relationship between monthly industrial production vintages and core quarterly vintages". All our results are thus based on quarterly data. The first estimation period is from 1959:01 (or the earliest available date after 1959:01 in which all variables are reported) to 1978:01. Data for CPI (not seasonally adjusted) is available from Norman Swanson's webpage (CPI_NSA.xls).

The data used in the linear versus non-linear empirical work have been chosen among the macroeconomic variables considered by Swanson and White (1997b) and Stock and Watson (1999). All the following data are from the St. Louis Fed database. Mnemonics are provided.

Unemployment (U): Civilian U.S. Unemployment Rate, mnemonic "UNRATE", is seasonally adjusted, available at monthly frequency, in percentage units, from 1948:01 to 2003:07. Concerning these data, the BLS announced several revisions to the Household Survey on February 2003, with the release of the January 2003 data. The changes affect data back to 2000, and are mainly due to a new seasonal adjustment procedure and new seasonal factors back to January 1998. This series is ultimately taken from the U.S. Department of Labor, Bureau of Labor Statistics.

Real GDP (RGDP) and Nominal GDP (GDP): Nominal U.S. GDP, mnemonic "GDP", is seasonally adjusted, at annual rate, available at quarterly frequency, in billions of dollars units, from 1947:1 to 2003:4. Real U.S. GDP, (RGDP), mnemonic "GDPC1" is seasonally adjusted, available at quarterly frequency, in billions of chained 1996 Dollars, from 1947:1 to 2003:4. Both series are ultimately taken from the U.S. Department of Commerce, Bureau of Economic Analysis.

Consumption (C): Real Personal Consumption expenditures, mnemonic "PCEC96", is a seasonally adjusted annual rate, available at monthly frequency, in billions of chained 1996 Dollars, from 1967:1 to 2003:6. This series is ultimately taken from the U.S. Department of Commerce, Bureau of Economic Analysis.

All variables except unemployment are first differences of logs; unemployment is in levels.

Table 1: The Critical Value $\mathbf{c}_{r,\alpha}$ for the Sequential Tests

r	$\alpha = 0.05$	$\alpha = 0.10$
1	2.7955	2.5003
2	3.1070	3.0548
3	3.4253	3.3949
4	4.0745	4.0479
5	4.5980	4.5746
6	5.0584	5.0372
7	5.4701	5.4507
8	5.8655	5.8475
9	6.1605	6.1440
10	6.4066	6.3912
11	7.3913	7.3752
12	7.3494	7.3348
13	7.4186	7.4057
14	8.9039	8.8884

Note to Table 1. The table reports the values of c to be used to obtain critical values for the Recursive tests described in Theorem 2. Critical values for the Recursive Wald, LM and LR tests of size α at time t with r restrictions and first monitoring time T are obtained from this table as: $c_{r,\alpha}^2 + r \ln(\frac{t}{T})$. The number of simulations used to obtain these critical values are 1000000 for $r = 1,..,5$ and 10000 for $r > 5$.

Table 2. Local power comparison of correctly sized tests.

$p =$		1			2			2	
				$\rho_X =$	0		$\rho_X =$	0.7	
δ	F_t	W_t	LM_t	F_t	W_t	LM_t	F_t	W_t	LM_t
0	.014	.014	.014	.024	.034	.034	.028	.034	.034
.025	.032	.032	.028	.050	.062	.058	.046	.086	.08
.05	0.08	0.08	.076	.102	.148	.138	.212	.296	.288
.075	.188	.188	.178	.292	.400	.388	.518	.664	.646
.10	.348	.348	.342	.458	.638	.634	.812	.916	.910
.125	.566	.566	.552	.702	.862	.854	.978	.988	.988
.15	.76	.76	.756	.886	.944	.944	.998	1	1
.175	.906	.906	.904	.968	.99	.988	1	1	1
.20	.954	.954	.954	1	1	1	1	1	1

Note to Table 2. The table reports the power function (as a function of the local alternative δ) of the Fluctuations test (F_t), the Recursive Wald (W_t) and LM (LM_t) tests, with 1,000 Monte Carlo simulations and p regressors. The DGP is: $x_t \sim N(0, \Sigma_X^{(p)})$, where $\Sigma_X^{(1)} = 1$, $\Sigma_X^{(2)} = \begin{pmatrix} 1 & \rho_X \\ \rho_X & 1 \end{pmatrix}$.

<u>Table 3. Recursive tests</u>

				M1					M2				
m	i	ip	p	MSE_F	MSE_W	R_F	R_W	R_{GC}	MSE_F	MSE_W	R_F	R_W	R_{GC}
3	0	0	0	0.984	0.996	0.45	0.74	0.89	1	1	1	1	1
3	1	0	0	1.011	1.017	0.39	0.5	0.73	1	1	0.97	1	1
3	3	0	0	0.991	0.992	0.38	0.46	0.68	1	1	1	1	1
3	0	0	1	1.013	0.998	0.58	0.86	1	1	1	1	1	1
3	1	0	1	1.02	0.989	0.43	0.68	1	1	1	1	1	1
3	3	0	1	1.04	0.992	0.47	0.69	1	1	1	1	1	1
3	0	0	3	0.979	0.995	0.59	0.86	1	1	1	1	1	1
3	1	0	3	1.001	0.986	0.38	0.70	1	1	1	1	1	1
3	3	0	3	1.011	0.991	0.40	0.68	1	1	1	1	1	1
3	0	1	0	1.029	1.018	0.36	0.54	0.75	1	1	0.97	1	1
3	1	1	0	1.022	1.006	0.34	0.47	0.67	0.999	1	0.94	1	1
3	3	1	0	1.017	1.009	0.24	0.37	0.51	0.999	1	0.97	1	1
3	0	1	1	1.058	0.992	0.43	0.82	1	1	1	1	1	1
3	1	1	1	1.057	0.982	0.31	0.65	0.98	1	1	1	1	1
3	3	1	1	1.071	0.986	0.21	0.58	0.95	1	1	1	1	1
3	0	1	3	1.018	0.98	0.33	0.79	1	1	1	1	1	1
3	1	1	3	1.072	0.986	0.26	0.56	1	1	1	1	1	1
3	3	1	3	1.11	1.005	0.15	0.43	0.89	1	1	0.99	0.99	0.99
3	0	3	0	1.028	1.015	0.32	0.52	0.71	0.999	1	0.99	1	1
3	1	3	0	1.024	1.016	0.33	0.46	0.67	0.999	1	0.95	1	1
3	3	3	0	1.023	1.006	0.16	0.36	0.46	1	1	1	1	1
3	0	3	1	0.986	0.996	0.47	0.83	1	1	1	1	1	1
3	1	3	1	1.028	0.986	0.36	0.67	0.97	1	1	1	1	1
3	3	3	1	1.04	1.001	0.18	0.61	0.9	1	1	0.99	0.99	0.99
3	0	3	3	0.959	0.985	0.42	0.71	1	1	1	1	1	1
3	1	3	3	0.961	0.987	0.31	0.60	0.99	1	1	0.99	0.99	0.99
3	3	3	3	0.996	0.771	0.24	0.47	0.85	1	1.018	0.99	1	0.99
S	I	C		1	1	1	1	1	1	1	0.74	0.74	0.74
A	I	C		1	1	1	1	1	1	1	0.74	0.74	0.74

Note to Table 3. We report empirical evidence on the predictive ability of lagged money for output from regressions with various lag structure and regressors. The first four columns report the number of lags included of: interest rate (i), money supply (m), industrial production (ip) and prices (p). Every regression includes a constant. "AIC" ("SIC") denote recursive lag length selection by AIC (SIC). MSE_F and MSE_W denote the ratio of RMSE of forecasts based on models selected by the Fluctuations and Wald tests relative to the RMSE of the one-shot GC test. R_F, R_W and R_{GC} respectively denote the fraction of times the Fluctuations, Recursive Wald and one-shot GC tests rejected H_0. One-step ahead forecasts begin at 1978:1.

Table 4. Linear versus non-linear models

Variable	Lags	MSE_F	MSE_W	MSE_F^{SW}	R_F	R_W	R_{GC}
Unemployment	2	0.9981	0.9981	- -	0.9968	0.9968	1
	3	1	1	- -	1	1	1
	4	1	1	0.0149	1	1	1
	5	1	1	0.0148	1	1	1
Consumption	2	1	1	1	0	0	0
	3	1	1	1	0	0	0
	4	0.8524	1	0.8100	0.3529	1	1
	5	0.9073	1	0.8108	0.5936	1	1
GDP	2	0.7589	0.7589	0.7589	0	0	0.0971
	3	0.9074	0.9074	0.9074	0	0	0.0291
	4	1	1	0.4708	1	1	1
	5	1	1	0.4758	1	1	1
Real GDP	2	1	1	1	0	0	0
	3	1	1	1	0	0	0
	4	0.7836	0.7836	0.7836	0	0	0.8350
	5	0.9936	0.9738	0.7131	0.7961	0.9612	1

Note to Table 4. MSE_F^{SW} denotes the ratio of the RMSE (relative to one-shot Granger Causality tests) of forecasts obtained as follows. At each point in time, we use our method and the test statistic (11) to select between a linear and a non-linear model; if the test does reject linearity, forecasts are made by using the Stock and Watson method. MSE_F and MSE_W instead use model (10) to forecast; the former chooses the model by using the Fluctuation test whereas the latter uses the recursive Wald test. R_F, R_W and R_{GC} denote, respectively, the fraction of times the Fluctuations, Wald and one-shot Granger Causality tests rejected the null hypothesis. "Variable" denotes the macroeconomic variable used in the univariate model and "Lags" denotes the number of lags used in the autoregression. "- -" means "not available".

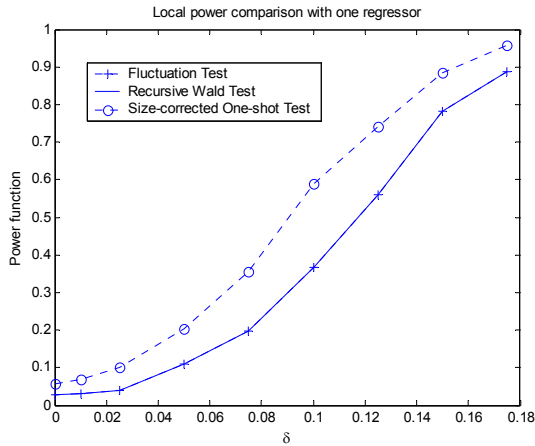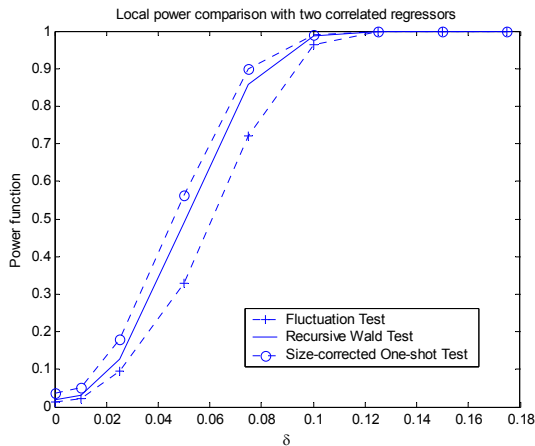**Figure 1** : **Local power comparison of correctly sized tests**.

Local power comparison with one regressor

Figure 2. **Rejection probabilities of recursive and one − shot tests**.

Rejection probabilities of recursive tests (i.e. the test rejects at least once)

Figure 1(**continued**)

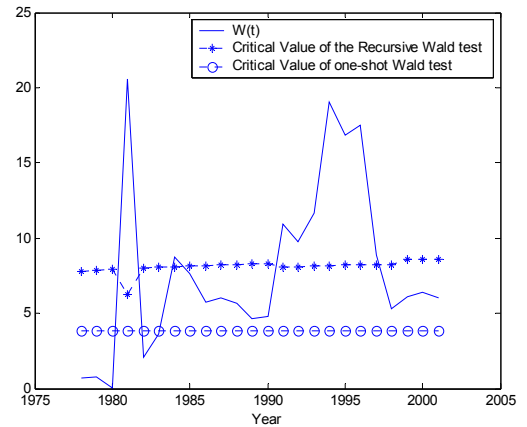Local power comparison with two correlated regressors
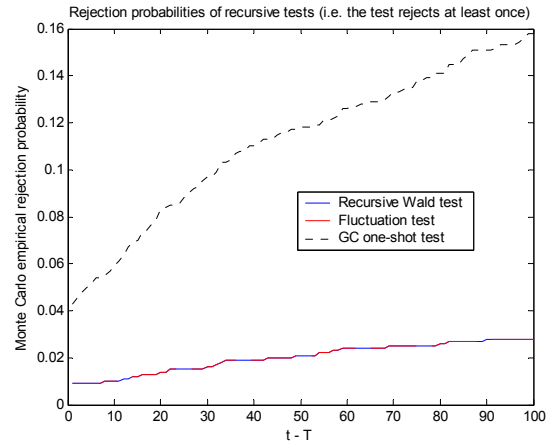
Figure 3. **Empirical evidence on predictive content of M1 for output in real time**.

Notes to figures. Figure 1 compares the local power of the Fluctuation, the Recursive Wald and a size-adjusted (by simulation) One-shot test with one (upper panel) or two correlated (lower panel, $\rho_X = 0.7$) regressor(s). Figure 2 shows the rejection probabilities of the GC one-shot, Recursive Wald and Fluctuation tests. The latter two perfectly overlap. Figure 3. We recursively test whether money GC output during the last decade. The figure shows when the recursive Wald ($W(t)$) test rejects the null hypothesis that some macroeconomic variables do not jointly GC output (it rejects when $W(t)$ is bigger than its critical value). For comparison, we also show the critical values of the one shot test. The regression includes a constant, 1 lag of money, and no lags of interest, output and prices.