IS US REAL GNP CHAOTIC?

ON USING THE BDS TEST TO DECIDE WHETHER AN ARMA MODEL FOR THE US GNP GENERATES I.I.D. RESIDUALS

First version: October 1993
This version: June 1994

IS US REAL GNP CHAOTIC?[*]
ON USING THE BDS TEST TO DECIDE WHETHER AN ARMA MODEL FOR THE US GNP GENERATES I.I.D. RESIDUALS

Mauro Gallegati
Dipartimento di Metodi Quantitativi e Teoria Economica, Università "G. D'Annunzio", Pescara, Italy

Domenico Mignacca
Dipartimento di Economia, Università di Ancona, Italy
Department of Economics, University of Wisconsin-Madison, USA

Summary
In this paper we use the BDS test developed by Brock-Dechert-Scheinkman(1987) to investigate whether ARIMA models for the US real GNP generate i.i.d. residuals. The second step, after reviewing some results from Brock-Sayers(1988) and Scheinkman-LeBaron(1989), SL, we will use a different kind of specifications for the US real GNP such as a model with different volatility pre and post World War II (WWII), as in SL(1989), and a threshold autoregressive specification as in Potter(1990). The last point consists of analysing a modified threshold model that takes into account the observation made by SL(1989) and next we evaluate the forecast performance of the "best" models among those examined.

JEL classification number: C22

[*] The BDS test has been performed using the W.D. Dechert program. We wish to thank W.A. Brock, W.D. Dechert, A. Duca, S. Potter, R. Scarpa and K. West for helpful comments and suggestions.

1. Introduction

When dealing with econometric models, the aim is to estimate economic or statistical relations among variables. Obviously, the estimated relations do not perfectly "capture" the behaviour of the variables; the "trick" generally used is to treat the unexplained part of the variables as not predictable. In econometrics terminology, shocks are required to be a Martingale Difference Sequence (MDS) with respect to the information set[1]. This condition is crucial, especially if the purpose of the model is to forecast the "economy"; in fact, if the structure of the residuals is known, it is possible to reduce the forecast error until the new residuals series is a true MDS. Furthermore, if we bear in mind that the residuals are produced (partially) by a deterministic mechanism, we can explain (in part) the business cycle[2].

In this paper we apply the Brock-Dechert-Scheinkman (BDS) test[3] to the residuals produced by several specifications of ARIMA(p,d,q) models for the US real GNP series[4]; this test enables us to detect neglected nonstationarity and neglected nonlinearity as well as chaotic dynamics. Before doing so, we briefly review the literature on the topic.

2. On "white noise" and "white chaos" processes

The time series y_t is i.i.d. if y_t is strongly stationary and independent ditributed; then y_t satisfies: (a) $E(y_t) = \mu < \infty$, (b) $Var(y_t) = \sigma_y^2 < \infty$[5], (c) $Cov(y_t, y_{t-k}) = 0 \ \forall \ k \neq 0$. Conditions (a)-(c) are called "white noise conditions" and often an i.i.d. series is also called white noise.

Clearly, i.i.d. implies white noise but not vice versa[6]; in fact, it happens that non i.i.d. processes have white noise properties[7]. This is true not only for

---

[1] MDS definition (Billingsley(1986), pp.480-481):

Let $\{y_t\}_{-\infty}^{\infty}$ be a sequence of random variables on a probability space (Ω, F, P) and let $\{F_t\}_{-\infty}^{+\infty}$ be a sequence of sigma fields in F. The sequence $\{(y_t,\ F_t)\}_{-\infty}^{+\infty}$ is a martingale difference if:

(i) $F_t \subset F_{t+1}$,

(ii) y_t is measurable F_t ,

(iii) $E[y_t | F_{t-1}] = 0$ almost surely.

[2] Cf. Brock-Sayers(1988).

[3] Brock et al.(1987).

[4] 1972 prices, quarterly data, range: 1875:1-1983:4. Source: Gordon(1986).

[5] Note that mean and variance do not depend on time.

[6] Apart from the case in which the white noise process is also Gaussian (Gaussian white noise).

stochastic processes. Some deterministic chaos processes also possess white noise properties and this is why they are sometimes called "white chaos" processes. Examples of this finding are the "tent map" and the "logistic map".

Consider first the tent map whose data generating process (DGP) is the following:

$$\begin{cases} x_t = a^{-1} x_{t-1} & \text{if } 0 < x_{t-1} < a \\ x_t = (1-a)^{-1}(1-x_{t-1}) & \text{if } a \le x_{t-1} \le 1 \end{cases} ;$$

The autocorrelation functions, partial and total, for this DGP are close to those of an i.i.d. process, especially for "a" near to 0.5[8].

Similar properties apply to the "logistic map" whose DGP is given by:

$x_t=\alpha x_{t-1}(1-x_{t-1})$, with $x_0 \in (0,1)$, $x_0 \neq 0.0, .5, .75$[9] and $3.57 \le \alpha \le 4$.

We plot our deterministic chaos processes in the following figure:

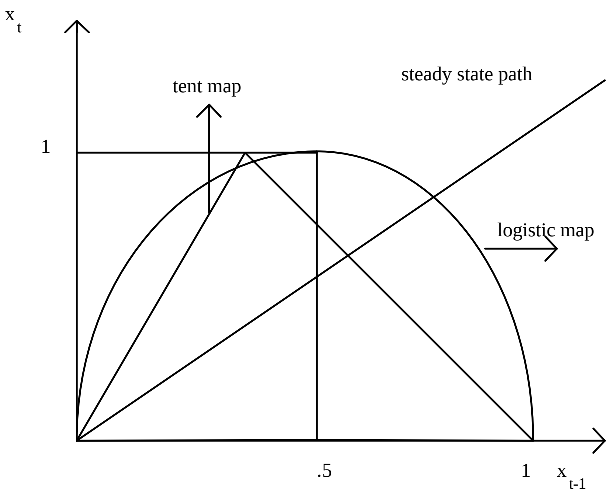

FIGURE 1

The reasons for these names are now clear. Inspection of the logistic map shows that it is symmetric around .5[10]; the tent map, instead, take its name from the fact that it has a kink in x_{t-1}=a. Furthermore, in case a=.5 the triangle becomes symmetric and lies within the logistic map function. Since this is merely a review of some specific chaos processes, we shall not examine the topic closely[11]. What we wish to stress, however, is the importance of chaos, especially in the context of business cycle theories. In fact, those who believe in i.i.d. disturbances simply state that fluctuations are determined by exogenous

---

[7] See for example Granger(1983) and Liu et al.(1992).
[8] See Sakai-Tokumaru(1980) and Liu et al.(1992).
[9] In fact: $x_0=.5 \Rightarrow x_t=0\ \forall\ t \ge 2$, $x_0=0.0 \Rightarrow x_t=0\ \forall t$ and $x_0=.75 \Rightarrow x_t=.75\ \forall t$.
[10] For α=4 as in figure 1.
[11] For initial reading see Grandmont(1986) and Baumol-Benhabib(1989).

random factors. Chaos supporters, on the other hand, disagree with a linear world and believe that the source of fluctuations is endogeneous to the economic system[12].

The problem now is to find a statistical framework which will enable us to distinguish between i.i.d. processes and non-i.i.d. ones. Before introducing the BDS statistic, the concept of correlation exponent, requires specification.

3. From the correlation integral to the BDS test[13]

Consider $\{x_t\}$ a univariate time series. We define the correlation integral as:

$$C(\varepsilon) = \lim_{N\to\infty} N^{-2}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} S_{ij}\right] \text{ where } \begin{cases} S_{ij} = 1 \text{ if } \left|x_i - x_j\right| < \varepsilon \\ S_{ij} = 0 \quad \text{otherwise} \end{cases}$$

N is the number of observations and N^2 is the number of all possible pairs (x_i,x_j). Intuitively, $C(\varepsilon)$ measures the probability that the distance of any particular pair in the time series is less than ε. Suppose that for small values of ε, $C(\varepsilon)$ grows exponentially at rate v: $C(\varepsilon)\approx\varepsilon^v$. The symbol "v" is the above mentioned correlation exponent. We can now introduce a generalization of the correlation integral $C(\varepsilon)$: consider the m-vector $X_{t,m}$ to be the vector of m consecutive terms $(x_t,.......,x_{t+m-1})$. The correlation integral is defined as follows:

$$C_m(\varepsilon)=\lim_{N\to\infty} N^{-2}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} S_{ij}\right] \text{ where: } \begin{cases} S_{ij} = 1 \text{ if } \sup_k \left|X_{i,k} - X_{j,k}\right| < \varepsilon \quad \forall k = 1,\dots,m \\ S_{ij} = 0 \quad \text{otherwise} \end{cases}.$$

$S_{ij}=1$ if each term of $|X_{i,m}-X_{j,m}|$ is less than ε. For a small value of ε, $C(\varepsilon)$ grows exponentially at rate v_m: $C_m(\varepsilon)\approx\varepsilon^{v_m}$. It has been shown that for stochastic white noise processes $v_m=m$[14]. At this point the problem is how to estimate v_m. There are several methods with which to estimate:

$$v_m \approx \frac{\partial \log C_m(\varepsilon)}{\partial \log \varepsilon}.$$

[12] Cf. Brock-Sayers(1988). An excellent textbooks of statistical and mathematical tools for handling chaos theory are Malliaris-Brock(1985) and Brock-Malliaris(1989).

[13] Cf. Grassberger-Procaccia(1983).

[14] Cf. Brock-Dechert(1988), pp.256-259 and Kennedy-Gentle(1983), p.1263.

Liu et al.(1992) have proposed the "point estimator" (PE), defined as:

$$\hat{\nu}_{m,j} = \frac{\log C_m(\varepsilon_j) - \log C_m(\varepsilon_{j+1})}{\log(\varepsilon_j) - \log(\varepsilon_{j+1})} = \frac{\log C_m(\varepsilon_{j+1}) - \log C_m(\varepsilon_j)}{\log\phi} ;$$

where $\varepsilon_j = \phi^j$, with $0<\phi<1$[15].

Smith's(1992b) "binomial estimator" is defined as:

$$\hat{\nu}_m = \lim_{N\to\infty} \frac{\log\left[\sum_{j=0}^{k} C(\varepsilon_{j+1})\right] - \log\left[\sum_{j=0}^{k-1} C(\varepsilon_j)\right]}{\log\phi} .$$

Alternatively Smith(1992a) proposes:

$$\hat{\nu}_m = \frac{1}{K}\sum_{j=1}^{K}\left[\frac{\log C(\varepsilon_j) - \log C(\varepsilon_{j+1})}{\log\phi}\right].$$

Liu et al.(1992) used Monte-Carlo simulation to evaluate the properties of the point estimates of the correlation exponent[16] for both i.i.d. sampling and white chaos processes. They showed that if data were pure white chaos the PE should be equal to one for any m-value, while it should be equal to m in case of i.i.d. processes. However, statistical inference is needed to overcome the problem of not having critical values needed to test the null hypothesis.

This is an appropriate point to introduce the BDS statistic developed by Brock, Dechert and Scheinkman(1987). The BDS test statistic takes the following form:

$$W_{m,T}(\varepsilon_j) = T^{1/2}\left[C_{m,T}(\varepsilon_j) - C_{1,T}(\varepsilon_j)^m\right] / \sigma_{m,T}(\varepsilon_j) \text{ ,}^{17}$$

where $\sigma_{m,T}(\varepsilon)$ is an estimate of the standard deviation under the i.i.d. null hypothesis. Under this null hypothesis $W_{m,T} \xrightarrow{d} N(0,1)$ as $T\to\infty$[18]. The proof of this result and a suggested estimator for $\sigma_{m,T}(\varepsilon)$ are contained in Brock et al.(1987) and Brock et al.(1992). Obviously the choice of j is fundamental: in fact, if we choose a small or a too large value for j it is more likely that $C_m(\varepsilon)=C_1(\varepsilon)$ rather than $C_m(\varepsilon)=C_1(\varepsilon)^m$.

---

[15] In order to overcome problem of scale, we need to normalize the serie x_t between 0 and 1. This can be simply done by considering the new series $\frac{x_t - x_{min}}{x_{max} - x_{min}}$.

[16] Cf. Liu et al.(1992), pp.530-533.

[17] Here $\varepsilon_j=.8^j$.

[18] Except for the degenerate case of uniform data on a circle.

Liu et al.(1992) performed several Monte Carlo experiments in order to verify the size and the power of the BDS test using generated series of length 200[19] . These experiments demonstrated that the BDS performed correctly when the null hypothesis was true[20]. But if the null was not an i.i.d. series, they found that the test had power, for certain values of j, only if the series was a deterministic chaos process (tent and logistic mapping) and for bilinear and bilinear moving average models. The power of the test is very poor if the DGP were: nonlinear sign models, nonlinear moving average models of order one and two, rational nonlinear autoregressive models and threshold autoregressive models. Further information on size and power of the BDS test, with applications to exchange rates and stock returns, can be found in Brock et al.(1992).

4. Applying the BDS test to ARIMA models for the US real GNP

In this paragraph we try to find an ARIMA model can go through the BDS test without rejecting the i.i.d. hypothesis for the residuals series. Following the suggestions of Liu et al.(1992), we use j=6,8 and m=2,3,4[21].

We start our analysis considering three different ARIMA(p,1,q) models, in which the dependent variable is the first difference of the logarithm of the US real GNP:

(a) ARI(9,1):

$\Delta y_t=\mu+a(L)\Delta y_{t-1}+\varepsilon_t$, where $\varepsilon_t \sim NID(0,\sigma^2)$[22] and $a(L)=a_1+a_2L+......+a_9L^8$,

T=435[23];

(b) IMA(1,10):

$\Delta y_t=\mu+b(L)\varepsilon_t$, where $b(L)=1+b_1L+......+b_{10}L^9$, T=435;

(c) ARIMA(3,1,3):

$\Delta y_t=\mu+a(L)\Delta y_{t-1}+b(L)\varepsilon_t$, where $a(L)=a_1+a_2L+a_3L^2$ and $b(L)=1+b_1L+b_2L^2+b_3L^3$,

[19] Cf. Liu et al.(1992) tables VI and VII, pp. S35-S36.
[20] Especially when j=6,8 and m=2,3,4.
[21] Epsilon is then 0.8^j.
[22] From now on, $\varepsilon_t \sim NID(0,\sigma^2)$
[23] T is the number of observations.

T=432.

We choose these specifications because they give (apparently) uncorrelated residuals. The following table contains the results of the BDS test on the residuals series of model (a), (b), (c).

TABLE 1: BDS test on residuals of models (a), (b) and (c)[24]

MODEL	BDS test j=6, m=2	BDS test j=6, m=3	BDS test j=6, m=4	BDS test j=8, m=2	BDS test j=8, m=3	BDS test j=8, m=4
(a)	3.20	4.32	4.59	3.56	5.87	6.88
(b)	3.16	4.34	4.58	3.26	5.59	6.63
(c)	2.75	3.94	4.27	3.73	5.89	6.64

From table 1, it is clear that we must have hidden structures in our series such as: nonlinearities, structural breaks, outliers and/or nonstationarity.

Generally, a good idea could be to perform an ARCH test[25] on residuals but, even though we are tempted to say that in our specific case the supposed nonlinearity is stochastic rather than deterministic, we have to point out that the power of the ARCH test seems to be very low when residuals are generated by a deterministic (chaotic) process (see table 2).

TABLE 2: ARCH tests on residuals and "white chaos" processes

model	ARCH(3) test	marginal significance level
(a)	106.236	≅0
(b)	106.531	≅0
(c)	107.119	≅0
tent map (x(0)=.1, a=.49)	244.69	≅0
logistic map (x(0)=.1)	287.00	≅0

Then, we try to investigate for the presence of outliers. Instead of performing our investigation using models (a), (b) and (c), we do it using the

[24] * and ** denote significance respectively at 1% and 5%. In table 1, for example, none of the statistics is significant at least at the 1%.

[25] The ARCH(q) test is performed regressing the squared residuals on its own q-lags; the ARCH test is then $T\cdot R^2$.

model that gives the best one-step ahead forecast performance; using this criterion (see table 3) the best model seems to be the ARIMA(3,1,3) model.

TABLE 3: One step ahead forecast statistics (1959:1-1983:4)[26]

model	ME* (1)	ME (max)	ME (min)	MAE** (2)	(1)/(2)	ME (%)	ME(%) (max)	ME(%) (min)	MAE (%)
(a)	-.51	-49.90	-.107	8.94	-.057	-.0262	5.6	-.029	.36
(b)	-.56	-50.69	-.023	9.13	-.061	-.0311	-3.46	-.023	.78
(c)	-.70	-48.66	-.070	8.97	-.078	-.0429	-3.33	-.013	.12

After several trials, we found evidence of two "break" periods in which we insert dummies: the first dummy is for the 1929's crisis (range: 1929:4-1933:1) while the second is for the World War II period (range: 1939:4-1945:4). Our new model is then:

(d) $\Delta y_t = \mu + a(L)\Delta y_{t-1} + \gamma_1 D_t^{1929} + \gamma_2 D_t^{WWII} + b(L)\varepsilon_t$, where a(L) and b(L)

are the same as in (c), while the meaning of the two other variables is straightforward. Unfortunately, even though we insert two dummy variables for these critical periods, we still not succeed in finding the source of "noise" (see table 4).

Table 4: BDS tests on the (d) model's residuals

model	BDS j=6, m=2	BDS j=6, m=3	BDS j=6, m=4	BDS j=8, m=2	BDS j=8, m=3	BDS j=8, m=4
(d)	2.4003*	3.9675	4.1674	3.5367	5.6509	6.1599

Then, we try to see whether or not there seems to be a structural change in our series passing from the pre-WWII period (1875:1-1938:4) to the post-WWII period (1946:1-1983:4)[27]. So the models we test are the following[28]:

(e_1) $\Delta y_t = \mu_1 + a_1(L)\Delta y_{t-1} + \gamma_1 D_t^{1929} + b_1(L)\varepsilon_{1,t}$, for t≤1938:4, T=252,

and,

[26] Note that ME and MAE are expressed in billions of 1972 dollars.
* Mean error.
** Mean absolute error.
[27] We erase the WWII period because it seems that a simple dummy variable cannot capture the complex behaviour of the rate of growth of the US real GNP during this period.
[28] Models (e_1) and (e_2) are again ARIMA(3,1,3) models.

(e_2) $\Delta y_t = \mu_2 + a_2(L)\Delta y_{t-1} + b_2(L)\varepsilon_{2,t}$, for t≥1946:1 and T=149.

Table 5: BDS tests on models (e_1) and (e_2)

model	BDS j=6, m=2	BDS j=6, m=3	BDS j=6, m=4	BDS j=8, m=2	BDS j=8, m=3	BDS j=8, m=4
(e_1)	0.754**	1.684**	1.977*	0.967**	2.265*	2.708
(e_2)	0.890**	0.543**	0.731**	0.811**	0.632**	0.936**

From table 5, we can draw an important conclusion: we did not solve our original problem but at least we have an idea of why we have non-i.i.d. residuals. First of all, it seems that marginalizing the analysis with respect to the WWII, "things get better"; secondly, the sensation is in favour of a strong structural change in the GNP's data generation process. For the moment, we think that this can be due to a different DGP taking place after WWII. Anyway, our opinion is not completely supported by our results for two main reasons: firstly, we have contraddictory results for model (e_1) and, secondly, inference in model (e_2) is based on a sample size of 149 that is a small sample size, expecially in light of our testing procedure based on Monte Carlo simulations results by Liu et al.(1992).

To try to overcome this problem, we decide to do 1000 bootstrap experiments[29]. In this way, we try to recover the empirical distribution of the BDS test, under the null hypothesis of models (e_1) and (e_2), and decide whether or not the calculated value belongs to the acceptance region. Summarizing, we use the following procedure:

(1) we estimate the model and recover the estimated parameters ($\hat{\beta}$) and the residuals series ($\hat{\varepsilon}_t$);

(2) we resample, with repetition, the series $\hat{\varepsilon}_t$ obtaining $\tilde{\varepsilon}_{i,t}$ (i=1,....,1000);

(3) we rebuild the series y_t in the following way: $\tilde{y}_{i,t} = \hat{\beta}' x_t + \tilde{\varepsilon}_{i,t}$, given the initial conditions and i=1,.....1000;

(4) we estimate $\tilde{y}_{i,t} = \tilde{\beta}' x_t + \bar{\varepsilon}_{i,t}$, and get a thousand of BDS tests from the estimated residuals $\bar{\varepsilon}_{i,t}$;

(5) the last point consists of evaluating the BDS test result using $\hat{\varepsilon}_t$ in light of the empirical BDS distribution obtained from step (4).

We decide not to use model (e_1) and (e_2) for the bootstrap experiment because it is too time consuming to compute ARIMA models rather than a simple AR model. So we reproduce the same analysis of model (e_1) and (e_2) in the following ARI(9,1) model:

** At least 5% marginal significance level.
* Marginal significance level between 1% and 5%.
[29] See also Scheinkman-LeBaron(1989).

$(f_1)\ \Delta y_t = \mu_1 + a_1(L)\Delta y_{t-1} + \gamma_1 D_t^{1929} + \varepsilon_{1,t}$, $t \leq 1938{:}4$;

$(f_2)\ \Delta y_t = \mu_2 + a_2(L)\Delta y_{t-1} + \varepsilon_{2,t}$, $t \geq 1946{:}1$.

Table 6: BDS tests on (f_1) and (f_2) models

model	BDS j=6, m=2	BDS j=6, m=3	BDS j=6, m=4	BDS j=8, m=2	BDS j=8, m=3	BDS j=8, m=4
(f_1)	1.4042**	2.3125*	2.4323*	1.6893**	3.4272	3.7192
(f_2)	1.0371**	0.8644**	1.4245**	0.6500**	0.5412**	1.0164**

Table 7: Empirical marginal significance levels for models (f_1) and (f_2)[+]

model	m.s.l. j=6, m=2	m.s.l. j=6, m=3	m.s.l. j=6, m=4	m.s.l. j=8, m=2	m.s.l. j=8, m=3	m.s.l. j=8, m=4
(f_1)	0.728 (4.475)	0.378 (4.182)	0.298 (3.972)	0.740 (4.655)	0.207 (4.566)	0.130 (4.512)
(f_2)	0.903 (6.100)	0.912 (5.829)	0.779 (5.673)	0.946 (6.500)	0.950 (6.364)	0.840 (6.374)

[+] 5% critical values in parenthesis.

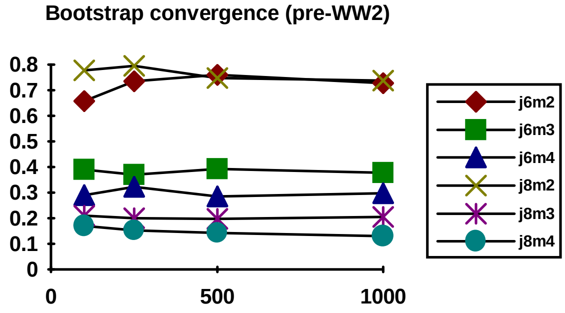

Figure 2

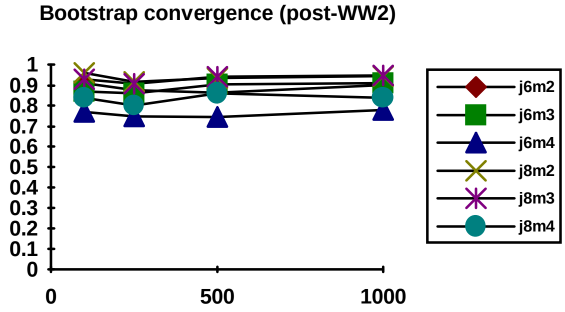

Figure 3

Table 7 gives us encouraging results, it seems quite clear that by splitting the series in two parts we can get i.i.d. residuals. From this, we could say that treating the US real GNP series as generated by a unique linear DGP can be misleading. In figure 2 and 3, we perform an experiment to give an idea[30]of how fast the BDS test converges towards its 1000 experiments p-value; in order to do this, we consider bootstrap samples of 100, 250, 500, 1000. The first impression we receive is that, considering the scale factor, increasing the number of experiments we would not obtain significantly different results, expecially for the post-WW2 period. Probably, for a conclusive opinion we would need to get p-value for 2000 experiments and maybe more, but we could not do it because of computational constraints.

The next question we try to answer is: are we able to find a "unique" model for the entire sample period able to go through the BDS test?

5. Modelling the whole period with a unique framework

In this paragraph, we revise some models proposed to describe the US real GNP. We focus our attention on two specifications: Scheinkman-LeBaron(1989) and Potter(1990)[31]. To simplify the exposition, the SL model takes into account the change in volatility of the series using GLS technique. In fact, even though we do not report it here, by looking at the residual plot of any linear model we considered, it is possible to see a remarkable reduction in

[30] For a similar application, see Brock et al.(1992), p. 1751 and p.1754.
[31] Henceforth respectively SL and P.

volatility of the series after the WWII period. We can express the SL model in the following way:

$$\text{(g) } \Delta y_t^* = \mu_1^* + a(L)\Delta y_{t-1}^*, \text{ where:} \begin{cases} \Delta y_t^* = \dfrac{1}{\hat{\sigma}_1}\Delta y_t, \text{ and } \mu_1^* = \dfrac{1}{\hat{\sigma}_1}\mu_1 \text{ for } t \geq 1946{:}1 \\ \Delta y_t^* = \dfrac{1}{\hat{\sigma}_2}\Delta y_t, \text{ and } \mu_1^* = \dfrac{1}{\hat{\sigma}_2}\mu_1 \text{ for } t \leq 1938{:}4 \end{cases};$$

where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are respectively the estimated standard errors for the post and pre-WWII.

The P model instead, is a threshold autoregressive model with the threshold fixed at zero. These kind of models represent the existence of two regimes: the expansionary regime for $y_{t-i} \geq 0$ and, the contractionary regime for $y_{t-i} \leq 0$. The P model, after considering i=2 as suggested in Potter(1990), is:

$$(h_1)\ \Delta y_t = \mu_1 + a_{11}\Delta y_{t-1} + a_{12}\Delta y_{t-2} + a_{13}\Delta y_{t-5} + \varepsilon_{1,t} \text{ for } y_{t-2} \leq 0, T = 115;$$

$$(h_2)\ \Delta y_t = \mu_2 + a_{21}\Delta y_{t-1} + a_{22}\Delta y_{t-2} + a_{23}\Delta y_{t-5} + \varepsilon_{2,t} \text{ for } y_{t-2} \geq 0, T = 313;$$

Table 8: BDS tests on SL and P models

model	BDS j=6, m=2	BDS j=6, m=3	BDS j=6, m=4	BDS j=8, m=2	BDS j=8, m=3	BDS j=8, m=4
(g)	1.3512**	2.2890*	2.3159*	1.6864**	2.8133	3.0210
(h_1)	1.4335**	1.1613**	1.0226**	1.6313**	1.4013**	1.4108**
(h_2)	4.3950	5.0465	4.8968	4.2800	5.7326	5.9310

Again, from table 8 we get mixed results expecially if we consider the P model. We try to find out whether or not we can improve the specification of the threshold autoregressive model. Erasing the WWII period from (h_1) and (h_2)[32]does not help in improving the BDS performance (see table 9).

Table 9: BDS tests on (h_{1b}) and (h_{2b})

[32] We call these new models respectively (h_{1b}) and (h_{2b}).

model	BDS j=6, m=2	BDS j=6, m=3	BDS j=6, m=4	BDS j=8, m=2	BDS j=8, m=3	BDS j=8, m=4
(h_{1b})	1.0448**	0.7410**	0.6076**	1.0282**	0.5971**	0.4849**
(h_{2b})	5.0149	5.6947	5.5580	4.0709	5.3934	5.5687

We think that a good idea could be to try to nest the observation made by SL into the P model[33], furthermore we introduce a dummy variable for the 1933:4 that, we think, can capture the end of depression effect[34]. Furthermore, we also considered y_{t-1} instead of y_{t-2}[35].

(h_{1c}) same as (h_{1b}); [36]

$$(h_{2c})\ \Delta y^*_t = \mu^*_1 + a_1\Delta y^*_{t-1} + a_2\Delta y^*_{t-2} + a_3\Delta y^*_{t-5} + \gamma\ D^{1933}_t + \varepsilon_t \text{ for } y_{t-2} > 0;$$ [37]

and:

$$(h_{1d})\ \Delta y_t = \mu_1 + a_{11}\Delta y_{t-1} + a_{12}\Delta y_{t-2} + a_{13}\Delta y_{t-5} + \varepsilon_{1,t} \text{ for } y_{t-1} \le 0;$$

$$(h_{2d})\ \Delta y^*_t = \mu^*_2 + a_{21}\Delta y^*_{t-1} + a_{22}\Delta y^*_{t-2} + a_{23}\Delta y^*_{t-5} + \gamma\ D^{1933}_t + \varepsilon_t \text{ for } y_{t-1} > 0;$$

Table 10: BDS tests on models (h_{2c}), (h_{1d}) and (h_{2d})

model	BDS j=6, m=2	BDS j=6, m=3	BDS j=6, m=4	BDS j=8, m=2	BDS j=8, m=3	BDS j=8, m=4
(h_{2c})	0.8212**	2.1383**	2.4160*	0.097**	1.6357**	2.2868*
(h_{1d})	0.6349**	1.3099**	1.3656**	1.5002**	2.3973*	2.5342*
(h_{2d})	-0.944**	-0.007**	0.1565**	-0.432**	0.4953**	1.1491**

From table 10, we can see how much we improved the performance of the BDS test just using y_{t-1} as the threshold variable instead of y_{t-2}. Model (h_{2d}) gives a strong answer in favour of i.i.d. residuals while (h_{1d}) gives mixed results. We have to be careful in evaluating model (h_{1d}) because it has just 115 observations and we did not take into account the variance change problem.

Before evaluating the forecast performance of our "best" models, we

---

[33] We consider a change in the volatility only for the expansionary regime for a problem of degrees of freedom in the contractionary regime.

[34] Models (h_{1c}) and (h_{2c}). Note that $(h_{1c})= (h_{1b})$. See note 30.

[35] Models (h_{1d}) and (h_{2d}).

[36] A superscript * has the same meaning it has in models (g).

[37] D^{1933}_t is a dummy variable for 1933:4.

wonder if we can improve the threshold specifications (h_{1d})-(h_{2d}). What we intend to do is to introduce a close threshold related model; we call it TOM model[38]and it has the following representation:

$$\begin{cases} (i_1)\ \Delta y_t = \mu_1 + a_{11}\Delta y_{t-1} + a_{12}\Delta y_{t-2} + a_{13}\Delta y_{t-5} + \gamma\ D_t^{1933} + \varepsilon_{1,t}, \text{ if } y_{t-1} > 0 \\ (i_2\ )\ \Delta y_t = \mu_2 + a_{21}\Delta y_{t-1} + a_{22}\Delta y_{t-2} + a_{23}\Delta y_{t-5} + \varepsilon_{2,t}, \text{ if } y_{t-1} \text{ and } y_{t-2} \le 0\ ; \\ (i_3)\ \Delta y_t = 0, \text{ otherwise} \end{cases}$$

The reason why we called the above model TOM is straightforward if we consider that the contractionary regime[39]applies only when we have two consecutive negative quarters. Considering that not all the TOM equations use more than 100 observations, we are going to apply the BDS test in case we have at least 100 observations (see table 11). Furthermore, we will consider also the change in variance for (i_1) and we refer to this model as the (i_{1b}) model.

Table 11: BDS tests on TOM models

model	BDS j=6, m=2	BDS j=6, m=3	BDS j=6, m=4	BDS j=8, m=2	BDS j=8, m=3	BDS j=8, m=4
(i_1)	0.0642**	1.6303**	2.1184*	1.6616**	3.4495	4.4147
(i_2)	na	na	na	na	na	na
(i_3)	na	na	na	na	na	na
(i_{1b})	-1.002**	-0.014**	0.1036**	-0.311**	0.7517**	1.2293**

Considering the results of table 11, we may conclude that, according to the BDS statistics, the “best” representation of the US real GNP is given by models (h_c), (h_d) and (i_b).

6. Forecast analysis

The last step of this paper consists in evaluating the forecast performance, based on the root mean squared prevision error[40] at different forecast horizons, for models: (f_2), (g), (h_b), (h_c), (h_d), (i) and (i_b)[41].

[38] Threshold Optimistic “M” model.

[39] Regime (i_2).

[40] RMSPE= $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\Delta\overline{y}_{t+h+i} - \Delta y_{t+h+i})^2}$, where $\Delta\overline{y}_{t+h+i}$ is the forecast h-periods ahead and n is the number of forecasts.

[41] Out-of-sample forecast. Period considered 1958:4-1983:3. Forecast horizon steps: 1, 4 and 12.

Table 12: Forecast performance of different models for US real GNP

model	(1) RMSPE step=1	(2) RMSPE steps=4	(3) RMSPE steps=12	(4) ME[*]	(5) MAE[**]	(6) (4)/(5)
(f_2)	0.01042	0.01074	0.01051	-0.00081	0.00811	-0.09963
(g)	0.01006	0.01045	0.01036	-0.00025	0.00770	-0.03294
(h_b)	0.00983	0.01017	0.01040	0.000428	0.00757	0.05648
(h_c)	0.00952	0.01110	0.01061	0.00001	0.00752	0.01325
(h_d)	0.00970	0.01038	0.01031	-0.00039	0.00731	-0.0531
(i)	0.00971	0.01185	0.01168	0.000315	0.00739	0.04263
(i_b)	0.00969	0.01180	0.01167	0.000083	0.00738	0.00112

Focusing the attention on columns (1) and (5) of table 12[42], the forecast analysis tells us that, expecially if we consider the short-term, both threshold and TOM models give encouraging results.

If we consider the RMSPE of model (h_c) for the one step ahead forecast horizon, we discover that it is the 7.4% lower than that of (f_2), the 4.1% lower than that of (g), and the 2% lower than that of (h_b), while it is comparable to those of the remaining models.

Considering the MAE statistics, model (h_d) is the 10% lower than that of (f_2), the 5% lower than that of (g), the 3.5% lower than that of (h_b), the 2.5% lower than that of (h_c) and comparable to those of the TOM's family.

Reminding that we are dealing with rates of growth, it seems that in the long-run, the best performance is obtained using models (g), (h_b) and (h_d).

In what follows, we use the framework developed in West-Cho(1994)[43], and analysed in more generality in West(1994), to test various hypotheses of equality constraints on MSPE's across different models. Before this, we had better review this technique.

Suppose that with $\hat{\sigma}_{s,h}$[44] we indicate the RMSPE generated by model s for a forecast horizon equal to h[45]; we denote with $\hat{\sigma}_h$ the vector containing all the RMPSE's for the forecast horizon h. Following West-Cho(1994), we consider a function q_t, which under suitable regularity conditions is a zero mean, weak stationary random vector with:

$$[6.1] \quad n^{-1/2}\sum_{t=1}^{n} q_t \xrightarrow{d} N(0,\ S), \text{ where } S \equiv \sum_{i=-\infty}^{+\infty}\Gamma_i \text{ and } \Gamma_i \equiv E(q_t q'_{t-i}).$$

---

[*] Mean error (forecast horizon=1).

[**] Mean absolute error (forecast horizon=1).

[42] The sixth column is a necessary statistic when we want to verify the existence of systematic forecast errors. In fact, as $-1 \leq (ME/MAE) \leq +1$: (ME/MAE) equals -1 when the model systematically underevaluate the variable while the contrary applies when (ME/MAE) equals +1.

[43] See also Newey-West(1987,1993) for further details.

[44] $s=f_2$, g, h_b, h_c, h_d, i, i_b. h=1, 4, 12.

[45] RMSPE is defined in footnote 39.

Consider now that $q_t=q_t(\theta)$, where θ is a vector of parameters of interest. We estimate θ with $\hat{\theta}$ by setting:

$$[6.2]\quad n^{-1}\sum_{t=1}^{n} q_t(\hat{\theta}) = 0.$$

Given [6.1]-[6.2] and considering, for our purpose, that:

$$[6.3]\quad \dim(q_t)=\dim(\theta),$$

and that:

$$[6.4]\quad \partial q_t(\theta)/\partial\theta,$$

is a matrix of known constants, we obtain[46]:

$$[6.5]\quad n^{1/2}(\hat{\theta}-\theta)\xrightarrow{d} N(0,V),\ \text{where } V=[(\partial q_t/\partial\theta)^{-1} S\,(\partial q_t/\partial\theta)^{-1'}].$$

Newey-West(1987), suggests the following estimator for S:

$$[6.6]\quad \hat{S} = \hat{\Gamma}_0 + \sum_{i=1}^{k}[1-i/(k+1)]\,(\hat{\Gamma}_i + \hat{\Gamma}_i')\ ^{47}, \text{ with } \hat{\Gamma}_i = n^{-1}\sum_{t=i+1}^{n} g_t(\hat{\theta})g_{t-i}(\hat{\theta})'.$$

Now, we have the tools we need to set up a test of equality constraints of the MPSE's across our "r" models.

Consider $\hat{\theta}_h = [\hat{\sigma}^2_{f_2,h},\dots,\hat{\sigma}^2_{i_b,h}]'$ and $g_t(\hat{\theta}) = [\hat{u}^2_{f_2,h,t} - \hat{\sigma}^2_{f_2,h},\dots,\hat{u}^2_{i_b,h,t} - \hat{\sigma}^2_{i_b,h}]'$ [48]. Equality constraints, across our r models can be imposed by considering the null hypothesis: $R\theta = 0$, where R is a (r-1) x r full rank matrix of the following form:

$$R = \begin{bmatrix} -1 & \\ \cdot & \\ \cdot & I_{r-1} \\ \cdot & \\ -1 & \end{bmatrix}.$$

Under H_0:

[46] This result can be easily proved using a Taylor's expansion together with hypotheses [6.3] and [6.4].

[47] The value of k is given, case by case, using the procedure suggested in West-Cho(1994), p.10.

[48] $\hat{u}^2_{s,h,t} = (\Delta\bar{y}_{s,\,t+h} - \Delta y_{s,\,t+h})^2$. Note that by construction we have $\sum_{t=1}^{n} g_t(\hat{\theta}) = 0$.

[6.7] $n\,[\hat{\theta}' R' (R \hat{S} R')^{-1} R \hat{\theta}] \xrightarrow{d} \chi^2_{(r-1)}$.

A difference between our procedure and West and Cho procedure is the following: when we want to impose restrictions on a coefficients' subset, we consider the $\hat{S}$ estimate for all r models contemporaneously; West and Cho, instead, propose to estimate a different S for each hypothesis[49]. In general, we test hypothesis like $R\theta=0$, where rank(R)=q<r and q is the number of restrictions under investigation. For h=1, 4, 12 , we investigate the following hypotheses (cf. tables 14-15-16):

$H_{a.h}$: all MSPE are equal;

$H_{b.h}$: Threshold's MSPE are equal ($\hat{\sigma}^2_{h_b,h}, \hat{\sigma}^2_{h_c,h}, \hat{\sigma}^2_{h_d,h}$);

$H_{c.h}$: TOM's MSPE are equal ($\hat{\sigma}^2_{i,h}, \hat{\sigma}^2_{i_b,h}$);

$H_{d.h}$: $H_{b.h} \cup H_{c.h}$;

$H_{e.h}$: linear's MSPE ($\hat{\sigma}^2_{g,h}$) is equal to the lowest MSPE;

$H_{f.h}$: SL's MSPE is equal to the lowest MSPE.

Before the forecast analysis, we present Monte Carlo experiment results to evaluate the size of the West and Cho procedure. The design of our experiment (EXP 1) is the following: we consider T=407, n=100[50] and 10000 experiments using time series drawn from two independent standard normal distribution[51]. As reported in table 13, the size of the test is near to the upper bound of the 95% region; even though we consider this result acceptable, we noticed that from a theoretical point of view we must have a k equal to 0 but in our case k is equal to zero only in the 41% of the cases. We decided to perform another Monte Carlo exeperiment (EXP 2) but now we set k equal to zero; in this case the size of the test is inside the 95% region (5.58%). Obviously, a "good guess" of k is desirable and a misspecified value of k does not seem to affect the test to much.

TABLE 13: Size of the West-Cho(1994) test procedure

Experiments	# of exp.	RMSPE	nature of k	size (%)	% of k=0

[49] The two strategies are equivalent for a given value of k. West-Cho(1994) hypothesis can be expressed in the following way: $R_1\theta_1 = 0$, where R_1 is a full rank matrix of order $(r_1\text{-}1) \times r_1$ and r_1 ($\le r$) is the number of coefficients under investigation. It is easy to verify that, we can always find a W matrix of order $(r_1\text{-}1) \times (r\text{-}1)$ such that: $WR\theta = R_1\,\theta$ and so $WR\,\mathrm{Var}(\theta)\,R'W' = R_1\,\mathrm{Var}(\theta_1)\,R_1'$.
[50] T is the number of observation while n is the number of out-of-sample forecasts.
[51] For each experiment, we compare two series which have the same RMSPE and so we expect a size of the test of 5% (which 95% region is approximately 0.037-0.063).

EXP 1	10000	1.0	variable	6.46	41
EXP 2	10000	1.0	fixed to 0	5.58	100

TABLE 14: Tests on forecast performance (h=1)

H_0	χ^2-value	DF	p-value (%)	k
$H_{a.1}$	15.6	6	1.6	8
$H_{b.1}$	1.9	2	38.0	8
$H_{c.1}$	0.2	1	66.3	8
$H_{d.1}$	6.4	4	17.2	8
$H_{e.1}$	6.4	1	1.1	8
$H_{f.1}$	2.1	1	14.9	8

TABLE 15: Tests on forecast performance (h=4)

H_0	χ^2-value	DF	p-value (%)	k
$H_{a.1}$	113.4	6	0.0	6
$H_{b.1}$	3.5	2	17.4	6
$H_{c.1}$	0.03	1	85.4	6
$H_{d.1}$	73.6	4	0.0	6
$H_{e.1}$	1.4	1	23.3	6
$H_{f.1}$	0.6	1	43.1	6

TABLE 16: Tests on forecast performance (h=12)

H_0	χ^2-value	DF	p-value (%)	k
$H_{a.1}$	21.8	6	0.1	6
$H_{b.1}$	1.0	2	61.6	6
$H_{c.1}$	0.007	1	93.4	6
$H_{d.1}$	11.0	4	2.7	6
$H_{e.1}$	3.4	1	6.4	6
$H_{f.1}$	0.8	1	38.0	6

From tables 14-16 we can see that, using a 5% significance level, for none of the horizons considered we can accept the hypothesis $H_{a.h}$ and this is evident expecially for h=4,12. Furthermore: (i) within the class of threshold and TOM models we cannot see a significant difference in performance (see $H_{b.h}$, $H_{c.h}$), this is an evidence against the utility of considering a change in the volatility of the US real GNP in the TOM models type; (ii) hypothesis $H_{d.h}$ is not rejected only for h=1 and this is due to the high RMSPE produced by the TOM models for h=4,12; (iii) the best model is always a threshold model[52] and comparing it with the linear model it rejects the equality constraint only for h=1

[52] h_c for h=1, h_b for h=4 and h_d for h=12.

and (iv) comparing the best model with the SL model (g) it never seems to be statistically superior to the former.

7. Conclusions

In this paper we have applied the BDS statistic, on different ARIMA models for the US real GNP, in order to test whether residuals can be considered independent and identically distributed. Using ARIMA frameworks to fit the whole period, we highly rejected this hypothesis, while, helping our analysis with a bootstrap experiment, we accepted the i.i.d. hypothesis splitting the series in two (pre and post World War II). This finding can either be due to a change in the volatility of the series across the two periods, or to structural breaks and/or hidden nonlinearities.

Using the results obtained by other authors, we studied the Scheinkman-LeBaron(1989) and Potter(1990) specifications for the US real GNP. We found mixed results in using their specifications but, things improved when we introduced the variance change problem found by SL in the threshold model of Potter. In addition to these, we performed the BDS test modifying the threshold of Potter and trying a new model that we called TOM. A limitation in our work is that, expecially for the TOM specification, we took into account variance change only for the expantionary regime since we faced a problem of degrees of freedom.

Considering the results of our out-of-sample forecast analysis, we found that SL and nonlinear specifications are superior to the linear specification only for a forecast horizon equal to one. Furthermore, there seems to be no statistical significant benefit in choosing either SL or Potter type specification models.

REFERENCES

Baumol, W., J. Benhabib, (1989), "Chaos: Significance, Mechanism, and Economic Applications", Journal of Economic Perspectives, 3.

Billingsley, P., (1986), Probability and Measure, Wiley, New York.

Brock, W.A., W.D. Dechert, J.A. Sheinkman, (1987), "A Test of Independence Based on the Correlation Dimension", SSRI no. 8702, Department of Economics, University of Wisconsin, Madison.

Brock, W.A., W.D. Dechert, (1988), "Theorems on Distinguishing Deterministic Random Systems", in W.A. Barnett, E.R. Berndt, and H. White (eds.), Proceedings of the Third International Symposium in Economic Theory and Econometrics.

Brock, W.A., D.A. Hsieh, B. LeBaron, (1992), Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence, The MIT Press.

Brock, W.A., J. Lakonishok, B. LeBaron, (1992), " Simple Technical Trading Rules and the Stochastic Properties of Stock Returns", Journal of Finance, vol. XLVII, no. 5, pp. 1731-1764.

Brock, W.A., A.G. Malliaris, (1989), Differential Equations, Stability and Chaos in Dynamic Economics, for the series Advanced Textbooks in Economics, C.J. Bliss and M.D. Intriligator (eds), North Holland.

Brock, W.A., C.L. Sayers, (1988), "Is the Business Cycle Characterized by Deterministic Chaos?", Journal of Monetary Economics, 22, pp.71-90.

Gordon, R., (1986), "Appendix Historical Data", in R. Gordon (ed.), The American Business Cycle, Chicago University Press.

Grandmont, J.M., (1986), (ed.), Nonlinear Economic Dynamics, Academic Press, New York.

Granger, C.W.J., (1983), "Forecasting White Noise", in A. Zellner (ed.), Applied Time Series Analysis of Economic Data, US Department of Commerce, Bureaus of the Census, Washington D.C..

Grassberger, P., I. Procaccia, (1983), "Measuring the Strangeness of Strange Attractors", Physica, 9D, pp.189-208.

Kennedy, W., J. Gentle, (1983), Statistical Compunding, Marcel Dekker, New York.

Liu, T., C.W.J. Granger, W.P. Heller, (1992), "Using the Correlation Exponent to Decide Whether an Economic Series is Chaotic", Journal of Applied Econometrics, 7, pp.S25-S39.

Newey, W.K., K.D. West, (1987), "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", Econometrica, 55, pp.703-708.

Newey, W.K., K.D. West, (1993), "Automatic Lag Selection in Covariance Matrix Estimation", NBER Technical working paper no. 144.

Malliaris, A.G., W.A. Brock, (1985), Stochastic Methods in Economics and Finance, for the series Advanced Textbooks in Economics, C.J. Bliss and M.D. Intriligator (Eds), North Holland.

Potter, S., (1990), Nonlinear Time Series and Economic Fluctuations, unpublished Ph.D. thesis, Department of Economics, University of Wisconsin-Madison.

Sakai, H., H. Tokumaru, (1980), "Autocorrelations of a Certain Chaos", IEEE Transactions on Acoustic, Speech, and Signal Processing, ASSP-28(5), pp.588-590.

Scheinkman, J.A., B. LeBaron, (1989), "Nonlinear Dynamics and GNP Data", in Barnett W.A., J. Geweke and K. Shell (eds), Economic Complexity: Chaos, Sunspots, bubbles, and Nonlinearity, International Symposia in Economic Theory and Econometrics, Cambridge University Press.

Smith, R.L., (1992a), "Optimal Estimation of Fractal Dimension", in M. Casdagli and S. Eubank (eds), Nonlinear Modeling and Forecasting, SFI Studies in the Science of Complexity, Proceeding, vol.12, Addison-Wesley, Reading, MA.

Smith, R.L., (1992b), "Estimating Dimension in Noisy Chaotic Time Series", Journal of Royal Statistical Society B, 54, pp.329-352.

West, K.D., D. Cho, (1994), " The Predictive Ability of Several Models of Exchange Rate Volatility", SSRI no.9317R, University of Wisconsin-Madison, forthcoming, Journal of Econometrics.

West, K.D., (1994), "Asymptotic Inference About Predictive Ability", SSRI no. 9417, University of Wisconsin-Madison.