"New tools for qualitative analysis in economic dynamics"

SYMBOLIC DYNAMICS

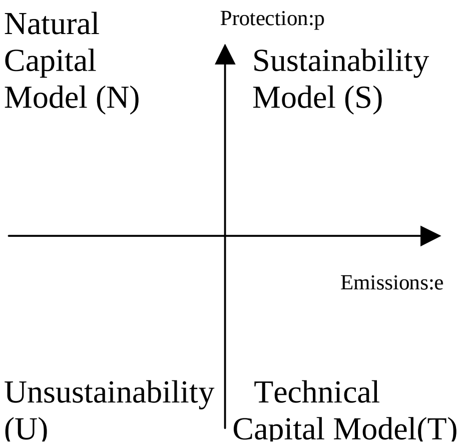
Abstract

Economic Dynamical Systems can be studied in the frame of a new approach based on the definition of regimes (symbols) and their time evolution. In the present work, the probabilistic theory of runs is suggested as a useful tool to develop qualitative analysis of these symbolic sequences. Some statistical tests are described and some economic applications are included too.

An introductory overview

The study of the dynamical systems is done using discrete time and considering iterations of the same function in order to detect particular behaviors as equilibrium points, regular fluctuations and non-regular oscillations. Going beyond, the space may be discretized too in a finite number of regions. It means that the rank of the function $R(f)$ is divided in k regions, named regimes. Each regime $R_i\ i = 1,2,\cdots k$ will be associated with a symbol and only one symbol $A_i, i = 1,2,\cdots k$ will be associated to one regime.

For instance, let us consider economic development. There is one theory that proposes to consider four types of development models: the natural capital model, the technical capital model, the "sustainable" model and the "unsustainable" one. According to these models, the space $\Re^2$ in which are defined all of them is divided in four regions (the four quadrants). The corresponding axes are: an emission index (*e*) and an environmental protection index (*p*).



Then, $\forall(e,p)\in\Re^2$ a symbol is associated, depending on the corresponding model of the piece of the partition. (A point lying on the first quadrant will be represented by an S one of the second by an N, and so on).

More formally, let's (M, f) be a dynamical system and $\sigma(x) = \{f^n(x): n \in Z\}$ be an orbit of $x \in M$. Our case corresponds to $M \equiv \Re^2$.

Consider the partition defined on M, $M = M_1 \cup M_2 \cup \cdots M_k$ with $M_i \cap M_j = \phi, i \neq j$

In the example, it corresponds to set: $M = M_S \cup M_N \cup M_U \cup M_T$. $M_S = [0;+\infty[\times [0;+\infty[$; $M_N =]-\infty;0] \times]0;+\infty[$; $M_U =]-\infty;0[\times [0;-\infty[$ and $M_T =]0;+\infty[\times]0;-\infty[$.

Then, for each value of the “real valued” orbit we will have an associated symbol, obtaining a sequence of symbols.

It can be observed that many orbits can define the same sequence of regimes, then the notion of classes of equivalence among orbits surges: two orbits will belong to the same equivalent class if they have the same symbolic representation.

The properties of the orbits in the original system will be reflected in the symbolic sequence. For instance, a periodic orbit will be “translated” in a periodic symbolic sequence.

Briefly, the idea is to analyze the dynamical behavior of these symbolic sequences, particularly:

- how much time the system remains in the same region (the same regime) before ‘to move”?
- which is the probability of changing of regime?

For instance, it can be the case that a sequence remain always in the same region (the same regime), or that it does it $\forall t \geq t_0$.

The following work is concerned about proponing a tool to proceed in this kind of approach to the dynamical systems.

RUNS

Introduction

Let us consider one row in the orchestra seats of the Teatro della Scala (Milano):

L O L L O L L L O L L L O L O L

It’s like to have a series of ***objects*** (the seats) classified in two types:

- free (*liberi*)
- occupied (*ocupati*)

We’ve used ***symbols*** to represent “the state” of the seats:

- **L**: free seat
- **O**: occupied seat

In this series we may consider the ***runs*** of each kind of symbols.

<u>Definition</u>:
The ***run*** will be a <u>sub series</u> of one or more <u>identical</u> <u>consecutive</u> symbols, preceded by another type of symbol (or by no symbol).

A :

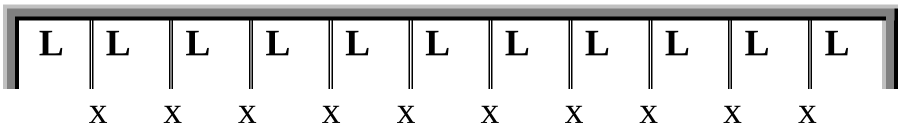
L O L L O L L L O L L L O L O L

1----2----3----------4----5----------------6----7----------------8----9----10----11

In A, we have 11 ***runs***.

But, how many arrangements of the symbols we have can define 11 runs?

Let us consider the *free* seats. They generate 10 "holes" that we can fill with the others seats (the ones that are occupied)



so, we have C_5^{10} ways *to locate the occupied* seats *among the free* ones, and

$$P(\text{ number of runs} = 11) = \frac{C_5^{10}}{C_{11}^{16}} = 0.0577$$

$$P(\text{ number of runs} = 5) = \frac{C_5^{10}}{C_5^{16}} = 0.0577$$

The presence of a ***pattern*** in the series may be justified by a lower probability based on the existence of too many runs, runs too long, too many long runs, etc.

The general procedure will be to set as the Null Hypothesis (H_0), the randomness (absence of a pattern). In the alternative (H_1) will be state the kind of sequence's behavior we want to detect.

The Critical Region[1] based on the number of runs observed in the sequence will take into account that too many runs and too few runs will indicate that it is a behavioral pattern. If the analysis will be based on the length of the runs, too short runs or runs too longs will indicate a behavioral pattern too.

To define the Critical Region to test the lack of randomness of the sequence it is needed to obtain the exact distribution function of the tests considered. As an example, let us consider the case of tests based on the length of symbolic runs (Mood, 1940).

Tests based on the length of the runs:

Suppose *n* objects, corresponding to two symbols: *A* and *B*.
Define: $r_{A_i} = \#\{runs\ of\ A\ which\ length\ is\ i\}$ and $r_{B_i} = \#\{runs\ of\ B\ which\ length\ is\ i\}$(Capital letters ($R_{A_i}$ and R_{B_i}) will correspond to the associated random variable).
The total number of runs of type *A* (r_A) is known, the corresponding to *B* (r_B) too. (the number of symbols, n_A and n_B, of each category is obviously known too).

It can be observed that:

$$\begin{cases} r_A = \sum_{i=1}^{i=n_A} r_{A_i} & r_B = \sum_{i=1}^{i=n_B} r_{B_i} \\ n_A = \sum_{i=1}^{i=n} i.r_{A_i} & n_B = \sum_{i=1}^{i=n} i.r_{B_i} \end{cases} \text{, where } r_{A_i} \geq 0 \text{ and } r_{B_i} \geq 0 .$$

To obtain the exact distribution of the runs of type *A* which length is *i*, we proceed in the following way:

---

[1] The Critical Region is a decision rule to reject (or not) the Null Hypothesis

i) we rank the runs depending on their length (we know that there are r_A runs of type *A*: |A| |A| |A| …|A|; |AA||AA|…|AA|; … … |AAA…A|, some of $r_{A_i} = 0$

ii) this order does not necessarily be accord to reality, the r_A can be present in many ways. All the possibilities are:

$$C_{r_{A_1}}^{r_a} . C_{r_{A_2}}^{r_A - r_{A_1}} \ldots C_{r_{A_{m_A}}}^{r_A - \sum_{i=1}^{i=n_A - 1} r_{A_i}} = \frac{r_A!}{r_{A_1}!\, r_{A_2}! \ldots r_{A_{m_A}}!}$$

Remember that $r_{A_i} \geq 0$ and that $r_A = \sum_{i=1}^{i=n_A} r_{A_i}$.

In a similar way, for the symbols of type *B*, knowing that they are r_B , there are $\frac{r_B!}{r_{BA_1}!\, r_{B_2}! \ldots r_{B_{mB}}!}$ possible arrangements, subject to $r_{B_i} \geq 0$ and $r_B = \sum_{i=1}^{i=n_B} r_{B_i}$

iii) we know that $r_A = r_B$ or $r_A = r_B + 1$

iv) the total number of arrangements of the runs of type *A* and *B* of different length will be:

$$\begin{cases} \dfrac{r_A!\, r_B!}{\prod_{i=1}^{i=n_A} r_{A_i}! . \prod_{i=1}^{i=n_B} r_{B_i}!}, if\ r_A = r_B \pm 1 \\ \dfrac{2.r_A!.r_B!}{.\prod_{i=1}^{i=n_A} r_{A_i}! . \prod_{j=1}^{j=n_B} r_{B_j}!}, if\ r_A = r_B \end{cases}$$

, subject to $r_{B_i} \geq 0$, $r_B = \sum_{i=1}^{i=n_B} r_{B_i}$, $r_{A_i} \geq 0$ and $r_A = \sum_{i=1}^{i=n_A} r_{A_i}$.

Then, when H_0 (the null hypothesis) holds, independence holds too, then all of these arrangements is equiprobable, and there are $C_{n_A}^{n}$ of them, then the <u>joint distribution</u> of $\left(R_{A_i}, R_{B_j}\right)$ will be:

$$p_{R_{A_i}, R_{B_j}}\left(r_{A_i}, r_{B_j}\right) = \begin{cases} \dfrac{r_A!.r_B!}{C_{n_A}^{n} \prod_{i=1}^{i=n_A} r_{A_i}! \prod_{j=1}^{j=n_B} r_{B_j}!}; r_A = r_B \pm 1 \\ \dfrac{2.r_A!.r_B!}{C_{n_A}^{n} . \prod_{i=1}^{i=n_A} r_{A_i}! . \prod_{j=1}^{j=n_B} r_{B_j}!}, if\ r_A = r_B \end{cases}$$

, $r_{B_i} \geq 0$, $r_B = \sum_{i=1}^{i=n_B} r_{B_i}$, $r_{A_i} \geq 0$ and $r_A = \sum_{i=1}^{i=n_A} r_{A_i}$.

Summing up in all the possible r_{B_i} contained in r_B runs of type *B*, we get:

$$P\left(R_{A_i} = r_{A_i}, R_B = r_B\right) = \begin{cases} \dfrac{C_{r_B - 1}^{n_B - 1}}{C_{n_A}^{n}} \dfrac{r_A!}{\prod_{i=1}^{i=n_A} r_{A_i}!}; r_A = r_B \pm 1 \\ 2.\dfrac{C_{r_B - 1}^{n_B - 1}}{C_{n_A}^{n}} \dfrac{r_A!}{C_{n_A}^{n} . \prod_{i=1}^{i=n_A} r_{A_i}!}, if\ r_A = r_B \end{cases}$$

Now, we need to sum over all the possible values of R_B ($r_A = r_B + 1$; r_A ; $r_A = r_B - 1$) to get the marginal distribution of $R_{A_j} \mid R_A = r_A$.

Because, $C_{r_A}^{n_b - 1} + C_{r_A - 2}^{n_b - 1} + 2.C_{r_A - 1}^{n_b - 1} = C_{r_A}^{n_b - 1} + C_{r_A - 1}^{n_b - 1} + C_{r_A - 2}^{n_b - 1} + C_{r_A - 1}^{n_b - 1} = C_{r_A - 1}^{n_b} + C_{r_A}^{n_b} = C_{r_A}^{n_b + 1}$, finally the result is:

$$P\left(R_{A_j} = r_{A_j} \mid R_A = r_A\right) = \left\{ \dfrac{C_{r_A}^{n_B + 1} . r_A!}{C_{n_A}^{n} . \prod_{j=1}^{j=n_A} r_{A_j}!} \right.$$

, the exact probability distribution of runs of type *A* which length is *j*, knowing that the total number of runs of this symbol is r_A

Depending on the Alternative Hypothesis, will be the kind (one left tail, on right tail or two tail) of critical region used. The probability that the empirical number of runs of one type and of one length lies in this region will be computed using the probability distribution function obtained before.

In which follows other examples of symbolizations are depicted.

Runs "up and down":

In this case, we study the "dynamics" of the series, comparing the increase (or decrease) of each symbol with respect to its previous one. This means that the analysis is on changes on the growth evolution of the considered system.

Remark: in the time series of the observed values the **order of appearance matters**.

$$\forall X_i \begin{cases} \uparrow, \text{if } X_{i+1} > X_i \\ \downarrow, \text{if } X_{i+1} < X_i \end{cases}$$

The ***runs*** are defined in an indirectly way, using $d_i = sign(X_{i+1} - X_i)$

The presence of ***r*** runs in the series which length is a given ***t*** or more, will indicate that *there is a pattern* in the evolution of the symbols: the change between consecutive symbols is not random ("chaotic")

Example: Value Added/Employment

1980 - 1996

(in 1.000:000.000 £ of 1990)

Taking the data corresponding to MEZ:

261957.70	261122.00	264318.60	273431.80	281115.70
(-)	**(+)**	**(+)**	**(+)**	**(+)**
287599.90	294311.40	303521.90	313387.70	320985.40
(+)	**(+)**	**(+)**	**(+)**	**(+)**
327126.00	334101.60	334541.50	330521.50	333425.60
(+)	**(+)**	**(-)**	**(+)**	**(+)**
337214.60	336804.70			
(-)				

so, the corresponding *run* "up and down" will be:

(-)(+)(+)(+)(+)(+)(+)(+)(+)(+)(+)(+)**(-)**(+)(+)**(-)**

Then we test:

$$\left. \begin{array}{r} H_0 \text{) randomness} \\ H_1 \text{) non randomness} \end{array} \right\}$$

Randomness, means that the independent identically distributed (a "chaotic" behavior).

The alternative hypothesis will capture the effect we want to check:

- a tendency to cluster:
  - if the prevailing sign is +, there is an upward trend
  - if the prevailing sign is -, there is a downward trend
- a tendency to mix will express cyclical variations

Then, looking for the number of runs, and defining adequately the Rejection Region, the behavior of a symbolic series may be tested.

Remark: Another approach to time series in the context of *runs* is to compare the data available to some standard for the period in study (the median, for instance), and thus to define *runs above and below the median*.

The Wald & Wolfowitz Runs Test:

Consider two sets of numerical observations:

$$X_1, X_2, \cdots X_m$$

$$Y_1, Y_2, \cdots Y_n$$

We may think that they correspond to the same phenomena, so we joint them into an ordered set, i.e.:

$$X_3, X_m, Y_1, X_2, \cdots Y_i, X_1, \cdots Y_n, \cdots X_4$$

If they really correspond to the same distribution, we expect that the **X** and the **Y** will be well mixed in the pooled set.

So, the number of *runs* will be a good indicator of the degree of mixing.

Too few *runs* will indicate that the observations **X** and **Y** correspond to different phenomena.

Example: Value Added/Employment

1980 - 1996

(in 1.000:000.000 £ of 1990)

Taking the data corresponding to MEZ and CE:

no sort:		sorted:	
MEZ	**CE**	**MEZ**	**CE**
261957.70	215713.00	261122.00	215713.00
261122.00	217688.40	261957.70	217688.40
264318.60	221086.90	264318.60	221086.90
273431.80	224531.30	273431.80	224531.30
281115.70	229076.20	281115.70	229076.20
287599.90	236350.00	287599.90	236350.00
294311.40	244031.50	294311.40	244031.50
303521.90	250564.90	303521.90	250564.90
313387.70	256306.30	313387.70	256306.30
320985.40	262802.70	320985.40	262802.70
327126.00	269865.20	327126.00	269865.20
334101.60	274713.60	330521.50	274713.60
334541.50	277902.30	333425.60	274971.00
330521.50	274971.00	334101.60	277902.30
333425.60	278247.60	334541.50	278247.60
337214.60	284750.10	336804.70	284750.10
336804.70	287937.80	337214.60	287937.80

0 0 0 0 0 0 0 0 0 **1 1** 0 **1** 0 **1** 0 0 0 0 **1** 0 **1** 0 **1 1 1 1 1 1 1 1 1 1 1** 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 **1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1**

Computing the $P(R=14)$, we can state if the MEZ and CE data correspond to the same evolution of the AV/E in 1980-1996.

Some final comments

Why this tool is suggested?

It is suggested because of the volume of really data. In general, when working with economic variables, their value is monthly, quarterly and, the most of time, yearly, making too difficult to work with asymptotic distributions like the normal one.

But…

The theory exposed above **cannot be directly used** in The Symbolic Time Series Analysis.

In fact, all that have been said is referred to *two* symbols, and we need at least work with *three*.

Why at least* three *symbols?

To set one of them for the *increasing* state, another for the *decreasing* state and the last one for the *steady* state in analyzing speed of transformations.

Some questions arise:

- Would it be possible by using the multinomial distribution to extend the results exposed above?
- If so: two symbolic series starting from the same "point" correspond to the same phenomena?
- … and in the case of sub-sequences?

Celina A. Gutiérrez
Siena, 15th December 2000

Bibliography:

- Bimonte, Salvatore & Punzo, Lionello (2000) –
  Turismo e Svillupo Sostenibile Locale Quale Modello per I Microsistemi Insulari (*Facoltà di Economia "R.M. Goodwin", mimeo.*)

- Brida, Gabriel (2000) –
  PhD. Thesis. (*Facoltà di Economia "R.M. Goodwin", mimeo.*)

- Feller, William (1978) –
  "Introducción a la Teoría de Probabilidades y sus Aplicaciones". Ediciones Limusa, 2da. edición

- Gibbons, J. D. (1985) –
  Nonparametric Statistical Inference, Mc.Graw-Hill, NY

- Mood, A. M. (1940) –
  The distribution Theory of Runs
  (*Ann. Math. Statistics* 11, pp. 367-392)

- Swed, F. & Eisenhart, C. (1943) Tables for Testing Randomness of Grouping in a Sequence of Alternatives
  (*Ann. Math. Statistics*, 14, pp. 66-87)

- Wald, A. & Wolfowitz, J. (1940)
  On a Test Whether Two Samples are from The Same Population
  (*Ann. Math. Statistics* 11, pp. 147-162)