

Chaos theory in predicting election results

M. P. Hantias^{1,*} and L. Magafas²

¹ *Dep. of Electronics, Computers, Telecommunications, and Control, University of Athens, Panepistimiopolis, Athens 15784, Greece*

² *Dep. of Electrical Engineering, Kavala Institute of Technology, St. Loukas 65404 Kavala, Greece*

Abstract

In this paper we introduce the term DemoscopoPhysics as the latest interdisciplinary research field that applies theories and methods originally developed by physicists in order to predict election results by using mainly Non Linear Dynamics (chaotic dynamics). For this purpose we have apply chaotic analysis on samples from political survey for the European Parliament in 2009, and also for the Hellenic National elections of October 4, 2009 in order to predict election results.

Keywords: Econophysics, election, results

1. Introduction

In our previous works [1,2] we have applied Physical models, especially methods from chaos theory, in order to predict and study the Euro election results of Hellas and the Hellenic National election results, defining the new scientific term called "DemoscopoPhysics", in the sense of application of physics models to social phenomena modelling. The term DemoscopoPhysics consists from two words Demoscopic and Physics. The first word is a Hellenic ancient word that means political survey. This work was inspired from the emerging field of economo-physics which mainly consists of autonomous mathematical physics models that apply to the financial markets. Now we try to use them particular aspects of the complex nonlinear dynamics of political survey in order to predict the election results for New Democracy (ND), Panhellenic Socialistic Movement (PASOK), Hellenic Communist Party (KKE), Coalition of the Radical Left (SYRIZA) and (Popular Orthodox Rally) LAOS political parties. We have approached the prediction with two different ways. Taking into account the results of opinion polls we have done regression in intension to vote. These new data are the raw data now. If we had applied statistical methods to these data we would take static results with very short horizon forecasting. For this reason we apply dynamic methods based on chaos theory in order to show the hidden potential of each political party and make predictions with a time horizon of 60 days for the Euro election results of Hellas and 30 days for Hellenic National election results.

2. Public survey time series

To construct the time series for predicting the the Euro election results of Hellas we have taken into account the assessment vote from public surveys in Hellas from 16-1-2007 to 23-04-2009 the estimation of the election behavior of the unclarified vote based on previous elections. The number of raw data is 36 for each political party, and each data is the average value of 4 polling companies with relative error 1%. In order to reconstruct of the equivalent phase space from experimental data, the timeseries that serves as experimental data should be constituted by sampled points of equal time-distances. For this purpose we interpolate with cubic spline so we take N=1000 points with a sample rate of 0.92day. We have applied the same procedure to construct the time series for Hellenic National election results and we have taken into account the assessment vote from public surveys in Hellas from 16-1-2007 to 6-09-2009. The number of raw data is 48 for each political party, and each data is the average value of 4 polling companies with average relative error 1.5%. for ND and PASOK political parties and 1% for other political parties. In order to reconstruct the phase space from experimental data, these data should be constituted by sampled points of equal time-distances. In our case this condition is not fulfilled. For this purpose we interpolate our raw data with cubic spline, as before, so we have create N=1057 points with a sample rate of 1 day. The raw data and the interpolated public survey time series of the ND and PASSOK political party are shown at Fig.1(a) and Fig 1(b) respectively, covering the period from 16-1-2007 to 23-04-2009

* E-mail address: mhantias@gmail.com

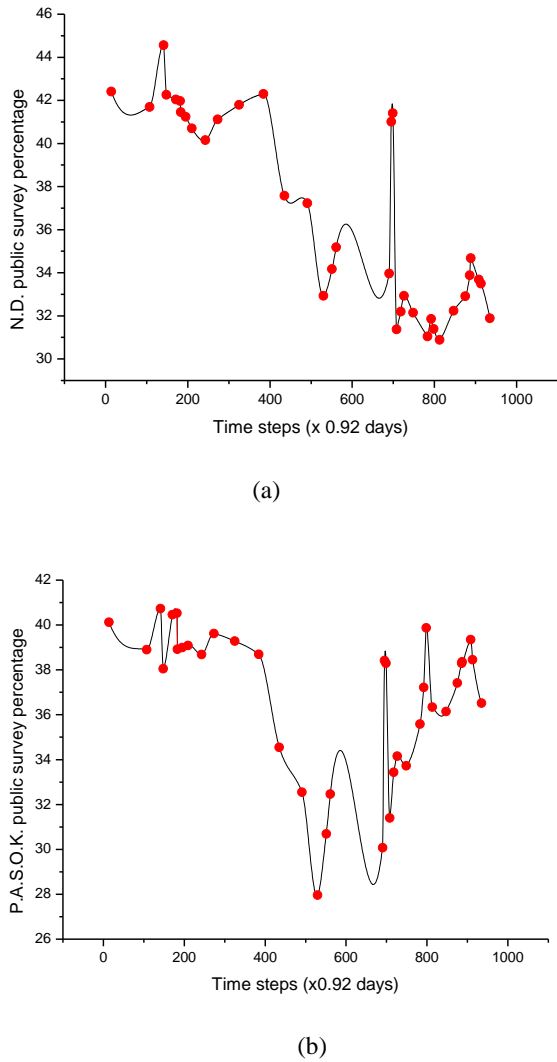


Fig. 1 (a) Interpolated time series for ND public survey for period 16-1-2007 to 23-04-2009 and interpolated time series for PASOK public survey for the same period (black line) and raw data (red dots)

3. State Space Reconstruction

For a scalar time series, in our case the gallop poll time series, the phase space can be reconstructed using the methods of delays. The basic idea in the method of delays is that the evolution of any single variable of a system is determined by the other variables with which it interacts. Information about the relevant variables is thus implicitly contained in the history of any single variable. On the basis of this an “equivalent” phase space can be reconstructed by assigning an element of the time series X_i and its successive delays as coordinates of a new vector time series \vec{X}_i . To construct a vector \vec{X}_i $i=1$ to N , in the m dimensional phase space we use the Equation 1 proposed by [1-3]

$$\vec{X}_i = \{x_i, x_{i-\tau}, x_{i-2\tau}, \dots, x_{i-(m-1)\tau}\} \quad (1)$$

\vec{X}_i represents a point to the m dimensional phase space in which the attractor is embedded each time, where τ is the time delay $\tau=i\Delta t$. Time delay is the time necessary to cancel

the correlation between two time series values. The element X_i represents a value of the examined scalar time series in time, corresponding to the i -th component of the time series. The dimension m of the reconstructed phase space is considered as the sufficient dimension for recovering the object without distorting any of its topological properties, thus it may be different from the true dimension of the space where this object lies. Use of this method reduces phase space reconstruction to the problem of proper determining suitable values of m and τ . The next step is to find time delay τ and embedding dimension m , without using any other information apart from the historical values of the indexes. This is why the methodology is labelled as a stochastic one. We can calculate the time delay by using the average mutual [3-5] information.

With the above method we found the τ to be 37, 20, 28, 40, 23 time steps for ND, PASOK, KKE, SYRIZA and LAOS, respectively. One method to determine the presence of chaos is to calculate the fractal dimension, which will be non integer for chaotic systems. Even though there exists a number of definitions for the dimension of a fractal object (Box counting dimension, Information Dimension, etc.), the correlation dimension was found to be the most efficient for practical applications. Firstly we calculate the correlation integral [7,8] for the time series for $\lim r \rightarrow 0$ and $N \rightarrow \infty$ by using the equation 3 [2].

$$C(r) = \frac{1}{N_{pairs}} \sum_{i=1, j=i+1}^N H(r - \|\vec{X}_i - \vec{X}_j\|) \quad (3)$$

In this equation, the summation counts the number of pairs (\vec{X}_i, \vec{X}_j) for which the distance, (Euclidean norm), $\|\vec{X}_i - \vec{X}_j\|$ is less than r , in an m dimensional Euclidean space. H is the Heaviside step function, with $H(u)=1$ for $u>0$, and $H(u)=0$ for $u \leq 0$, where $u = (r - \|\vec{X}_i - \vec{X}_j\|)$, N denotes the number of points and expressed in equation 4.

$$N_{pairs} = \frac{2}{(N - m + 1)^2} \quad (4)$$

where r is the radius of the sphere centered on X_i or X_j . If the time series is characterized by an attractor, then for positive values of r , the correlation function is related to the radius with a power law $C(r) \sim \alpha r^v$, where α is a constant and v is the correlation dimension or the slope of the $\log_2 C(r)$ versus $\log_2 r$ plot. Since the data set will be continuous, r cannot get to close to zero. To handle this situation, from $\log_2 C(r)$ versus $\log_2 r$ plot we select the apparently linear portion of the graph. The slope of this portion will approximate v . Practically one computes the correlation integral for increasing embedding dimension m and calculates the related $v(m)$ in the scaling region. Using the appropriate delay times for each political party i.e, 37, 20, 28, 40, 23 time steps for ND, PASOK, KKE, SYRIZA and LAOS, respectively, we reconstruct the phase space for each political party. The correlation integral $C(r)$ by definition, is the limit of correlation sum of Equation (3) for different embedding dimensions, $m=1..10$. This is shown for ND in Fig 2(a) while in Fig.2 (b), the corresponding average slopes v are given as a function of the embedding dimension m , indicating that for high values of m , v tends to

saturate at the non integer value of $\nu=1.6$. The embedding dimension m is found to be $m \geq 2[\nu]+1=3$ where $[\nu]$ is the integer part of ν [6].

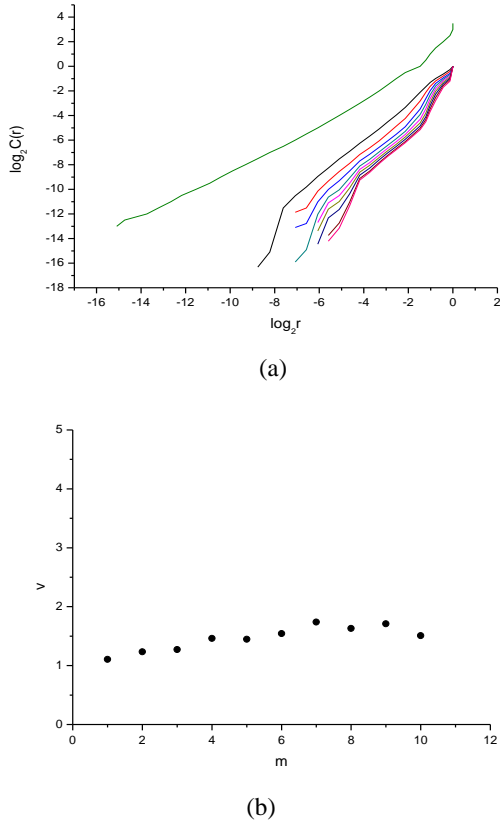


Fig.2 (a) Relation between $\log_2 C(r)$ and $\log_2 r$ for different embedding dimensions m . (b) Correlation dimension ν vs. embedding dimension m for ND

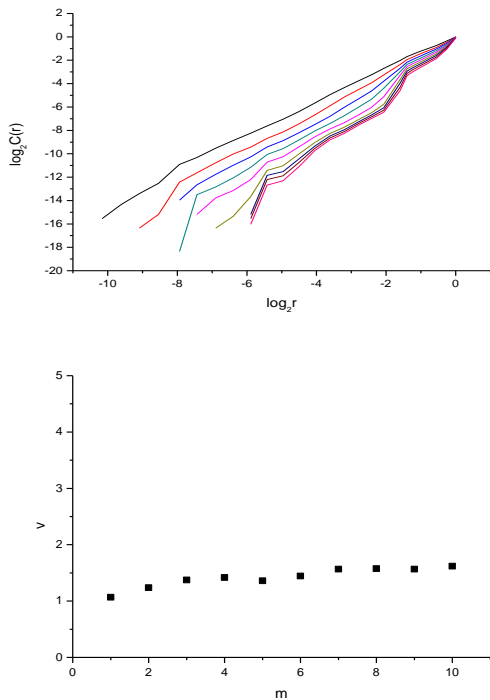


Fig.3 (a) Relation between $\log_2 C(r)$ and $\log_2 r$ for different embedding dimensions m . (b) Correlation dimension ν vs. embedding dimension m for PASOK

Table 1. The correlation dimension for Greek political parties

Political Parties	Correlation dimension ν
ND	1.6
PASOK	1.53
KKE	1.28
SYRIZA	1.29
LAOS	1.23

Applying the same procedure for PASOK political we found that for high values of m , ν tends to saturate at the non integer value of $\nu=1.53$. The embedding dimension m is found to be $m \geq 2[\nu]+1=3$ [6]. The results are shown in Fig 3. For KKE ν tends to saturate at the non integer value of $\nu=1.28$, while for SYRIZA and LAOS ν tends to saturate at the non integer value of $\nu=1.29$ and 1.23 respectively. The embedding dimension m is found to be $m=3$ for the other political parties too. [6]. We can see from Table 1 that the smaller political parties have smaller correlation dimension. We can interpret it

that the smaller political parties are more robust to keep there voters but on the other hand they cannot adapt changes as the larger parties do.

4. Time series prediction

The next step is to predict the evolution of the percentages of votes, for each political party, by computing weighted average of evolution of close neighbors of the predicted state in the reconstructed phase space [1,2]. The reconstructed m -dimensional signal projected into the state space can exhibit a range of trajectories, some of which have structures or patterns that can be used for system prediction and modeling. To predict the Euro election results of Hellas we used the values of τ and m from our previous analysis. We had better results using as embedding dimension the value of $2*m = 6$ [1,2] for all predictions. At table 2 we present our out of sample estimation. At the same table we present the results and the deviation from Euro election results of Hellas.

Table 2. Political survey estimation for Euro election results of Hellas

Political parties	Political survey estimation %, 23/4/	Election Results % 7/6/2009	Deviation from election results
ND	32.33	32.29	0.04
PASOK	37.19	36.65	0.54
KKE	8.52	8.35	0.17
SYRIZA	7.4	4.7	2.7
LAOS	6.8	7.15	0.35

At this point we mark that until 23/4/2009 we had not data for Ecological Party. This political party, generally speaking, is in cognation with Coalition of the Radical Left (SYRIZA) so its presence can affect SYRIZA's percentage. The % percentage for Ecological Party was 3.49 and it is included to SYRIZA's estimation.

We used the same procedure for the Hellenic National election results Actual and predicted time series for $k=30$ time steps ahead are presented at Figs 4.(a),(b) for ND, PASOK respectively.

In order to capture the polarization of voters we have decreased the degrees of freedom to 4. We use as embedding dimension $m = 4$, keeping the values of delay time the same

from our previous analysis and the number of near neighborhoods nn , as a rule of thumb [2] equal to $3 \cdot m = 12$, for all political parties. Actual and predicted time series for $k=30$ time steps ahead are presented at Figs 5.(a),(b) for ND, PASOK respectively.

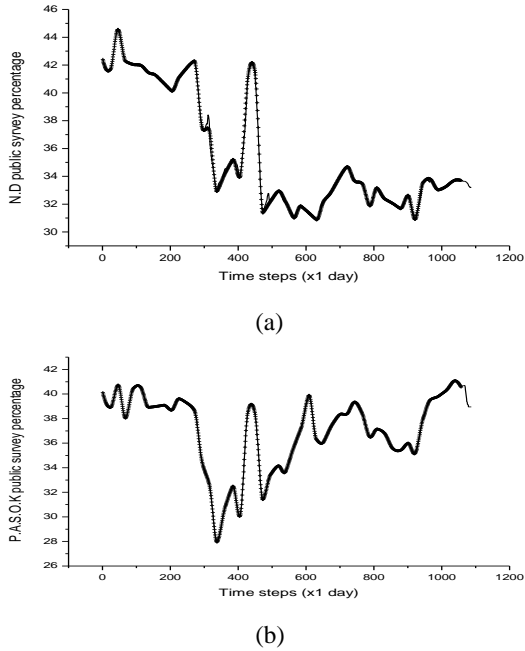


Fig 4. Actual (crosses) and predicted (solid line) time series for $n=30$ time steps ahead for ND(a),PASOK (b) political parties. The embedding dimension is $m=6$.

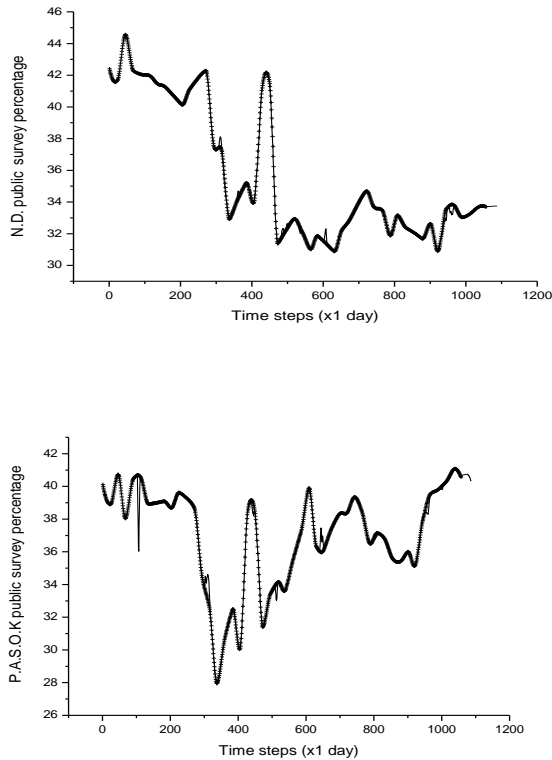


Fig 2. Actual (crosses) and predicted (solid line) time series for $n=30$ time steps ahead for ND(a),PASOK (b) political parties. The embedding dimension is $m=4$.

At table 3 we present our out of sample estimation about political survey estimation for two embedding dimensions. The first is $m=6$ and the second is $m=4$. We estimate the mean error as 1.5% for ND and PASOK and 1% for the other political parties. At the same table we present the results and the deviation from Hellenic National election results

Table 3. Political survey estimation for Hellenic National election

Political parties	Election Results %	Embedding dimension $m=6$ political survey estimation %	Deviation from election results	Embedding dimension $m=4$ political survey estimation % for 4/10/2009	Deviation from election results
ND	33.48	33.2 (Range 31.7 - 34.70)	0.28	34 (Range 32.5-35.5)	0.52
PASOK	43.92	39 (Range 37.5 - 40.5)	4.92	41.1 (Range 39.6 - 42.6)	2.82
KKE	7.54	8.6 (Range 7.6 - 9.6)	1.06	8.6 (Range 7.6 - 8.6)	1.06
SYRIZA	4.60	4 (Range 3-5)	0.6	4.1 (Range 3.1 - 5.1)	0.5
LAOS	5.63	7.8 (Range 6.8 - 7.8)	2.17	7.4 (Range 5.4 - 7.4)	1.77

5. Conclusions

In this paper, we use a chaotic analysis to predict Greek political parties election results. After estimating the dependence of correlation dimension on embedding dimension, we point out that the system is a deterministic chaotic. A separate attractor for each political party, embedded in 3-dimension space, is derived from the analysis. However the election system is obviously a complex multi-variable system with strong inter-relationship between variables. In this sense, we have model each separate time series and never the whole election system, whose attractor is obviously much more complex. From absolute values of correlation dimension we see that the smaller political parties have smaller correlation dimension. We can interpret it that the smaller parties are more robust to keep there voters but on the other hand they cannot adapt changes as the larger parties do. From reconstruction of the systems' strange attractors, we achieve a 60 time steps out of sample prediction, for Euro election results of Hellas and a 30 time steps out of sample prediction for the Hellenic National election results. As the time horizon increases the prediction becomes weak. This depends on strange attractor's structure and the number of raw data. As this number increases the influence of cubic spline is reduced and the results will be more precise. As seen before the in sample prediction works well so we believe that the out of sample prediction gives satisfactory results. Of course if we could include data for Ecological Party our prediction will be more accurate. The prediction with different degrees of freedom shown that, when the value of embedding dimension is 4 instead of 6 there is an increase of the percentage of larger political parties as N.D. and P.A.S.O.K. are. This can be interpreted as a polarization of voters which corresponds to a lesser degrees of freedom. The only exception is the percentage of SYRIZA political party. Speculate that this behavior is due to that SYRIZA is composed of components that tend to come together.

Using tools and principles from Physics, we try modeling an open humanitarian system as the National and Euro election system is. In any case this work opens new

directions in order to study social behaviour using Physical models, since human society obey to Physical laws. This is the first attempt to predict election results with Physical models. However our predictions 60 days and 30 days, before the elections , were the most sucesfull among predictions and exit polls from Greek gallop companies. To establish a new term as DemoscopoPhysics we need more data for testing and more tools from Physics to apply as entropy and criticality and phase transition are. The future researches may concentrate on the alternative models (i.e. parametric and nonparametric ones) for prediction. In

addition, to reflect the time-scaling effects and wavelet theory which can be combined with the chaos theory.

References

1. M.Haniyas and L.Magafas "Application of Physics Model in prediction of the Hellas Euro election results" Journal of Engineering Science and Technology Review Volume 2, Number 1, Pages 104-111, (2009).
2. M.Haniyas and L.Magafas "Application of Physics Model in prediction of the Hellas election results" Journal of Engineering Science and Technology Review Volume 2, Number 1, Pages 112-117, (2009).
3. Abarbanel H.D.I. (1996), Analysis of observed chaotic data, Springer, New York.
4. Fraser A.M., Swinney H.L. (1986), Independent coordinates for strange attractors from mutual information, Phys. Rev. A, 33, 1134.
5. Kugiumtzis D., Lillekjendlie B., Christophersen N.: Chaotic time series, Part I, Modeling Identification and Control 15, 205, (1994).
6. Kantz H. and T.Schreiber (1997), Nonlinear Time Series Analysis, Cambridge University Press, Cambridge