PREDICTING THE ALBANIAN ECONOMIC DEVELOPMENT USING MULTIVARIATE MARKOV CHAIN MODEL

Eralda Gjika (Dhamo) 1, Lule Basha 2, Xhensilda Allka 3, Aurora Ferrja 4

1, 2 Department of Applied Mathematics, Faculty of Natural Science, University of Tirana, Albania
3 Department of Computer Engineering, Faculty of Computer Engineering and IT, Metropolitan University of Tirana, Albania
E-mails: 1 eralda.dhamo@fshn.edu.al (corresponding author); 2 lule.hallaci@fshn.edu.al; 3 xhallka@umt.edu.al; 4 aurora.simoni@fshn.edu.al

Received 03 March 2020; accepted 05 May 2020

Abstract. In this work, the economic development and relation to social and demography indices in Albania were studied. Four time series (yearly data for the period 1995–2018) were considered: consumer price index (CPI), unemployment rate, inflation and life expectancy. In our approach, a first and fifth order multivariate Markov chain model was proposed to predict the economic situation in Albania in the proceedings years. Tests and accuracy analysis of the model were performed. The prediction probabilities fall in the interval of 0.47 to 0.52 and the accuracy of both models is 75%. Our approach is a short term probability forecast model that can be used by the policymakers to evaluate and undertake initiatives to improve the situation in the country.

Keywords: forecasting, Markov chain, economic growth, CPI, inflation, probability.

JEL Classification: C53.

1. Introduction

Economic indicators are widely used by analysts to evaluate the investment opportunities in one country and predict the movements of an economy. There are some important coincident and lagging indicators in the literature, such as Gross Domestic Product, employment indicators, interest rates, Consumer Price Index, etc. The impact of life expectancy has also been studied by many authors.

In our study, we have examined the relation of these indicators for Albania. By analyzing previous authors' work on the same topic we have compared and discussed the importance of these indicators in the economic development of the country.

In 1993, Fischer in his work analyzed the association between inflation and economic growth at a low rate. Latter, Alexander (1997) analyzing the data for OECD countries (1974–1991) found a significant negative relation between inflation and economic growth. Berber and Artan (2004) in their paper analyzed the relationship between economic growth and inflation in Turkey (period 1987–2003) and concluded that there is a negative correlation between inflation and economic growth. This negative relationship between inflation and economic growth is also supported by the empirical studies of Hodge (2006). In other studies (Hineline, 2004; Vaona & Schiavo, 2006), claimed that the link between these variables was not significant.

In 2004, Christopoulos worked on Greek economy data and suggest a possible relationship between unemployment and economic growth. Semmler and Zhang (2005) agreed that for a considerable number of countries with a similar economic development an increase in the economy is affected by the decrease in the unemployment rate. In their study, Villaverde and Maza (2009) analyzed through panel data in some areas in Spain (period 1980 to 2004 annually) and found a negative relationship between unemployment and economy. In 2018, Soylu et al. (2018) investigate the relationship between economic growth and unemployment in Eastern European Countries (period 1992–2014) and found a negative correlation between these indices.

Upreti (2015) in his work analyzed some indicators affecting economic growth in developing countries. For 76 countries he found that an increasing level of life expectancy has a positive impact on the economic growth of a country.

Demography indices such as life expectancy are foreseen to be associated with a positive impact
on the economic development of a country. Acemoglu and Johnson (2007) found that improvements in life expectancy may lead to positive behavior of economic indicators. Several studies (Lorentzen et al., 2008; Cervellati & Sunde, 2011; Todaro & Smith, 2010) found evidence of positive connection among demographic indices such as, life expectancy and income, generating improvement of the quality of the living standard.

Another interesting work was done by Cebula (2011) for the period of study 2003–2007 in OECD countries that derive that economic growth is positively correlated with freedom in monetary, investment, labor, fiscal, property rights, and freedom from corruption. Other scientific works in this field have shown that non-economic factors play a significant role in economic development (Rodrik, 2003).

Via the regression approach, Shafi and Fatima (2019) studied the impact of life expectancy and population growth rate in GDP growth of the G7 countries. They found a significant relationship between these indices.

Gjika et al. (2018) in their study have analyzed and forecasted economic indices for Albania using an econometric model (data for period 1994–2017). They conclude that the exchange rate and the number of people traveling abroad were positively correlated to CPI.

In this paper, our objectives are to understand the correlation among life expectancy, CPI, inflation rate and unemployment rate and their impact on the economic development of a country. We have used higher-order Markov chain models to identify indices associated with economic development.

### 2. Forecasting methods

Forecasting is an important issue for governments and central banks to assess a country’s economic development. Many methods are used starting from classical time series models, linear regression, neural networks, probability forecasting and also hybrid models.

Bredin and Fountas (2006) in their paper examined the relationship between inflation and its uncertainty in four European countries over 4 years using a Markov model.

Amisano and Fagan (2013) also use a Markov chain model and Bayesian techniques to model inflation in Germany, UK, USA, and Canada. Constant and Zimmermann (2014) presented a study of the labor market and employment (unemployment) by using probabilistic models such as Markov chain models. Falata and Pancikova (2015) carry out predictions for economic indices through neural network models.


### 2.1. Markov Chain models

The higher-order multivariate Markov chain model was first introduced by Ching et al. (2002, 2003, 2008) and then by Liu (2010) to model multiple time series with categorical data. The model assumes that the probability state distribution of the jth sequence at time t = i + 1 depends on the probability state distribution of other time series (including itself) at time t = i, i − 1, ..., i − q + 1 ..., where q is the order of the model.

To develop a higher-order multivariate Markov chain model we take into consideration d categorical time series, with k states and then find the inter and intra transition probability matrices with 1, 2, ..., q steps.

The higher-order (q-th order) multivariate Markov chain model is expressed in the following form:

\[
x^{(j)}_{i+1} = \sum_{l=1}^{d} \sum_{h=1}^{q} \alpha^{(h)}_{jl} Q^{(h)}_{j} x^{(l)}_{i-h+1},
\]

(1)

where \( j = 1, 2, ..., d \) and \( i = q-1, q, ..., \), initial probability distribution \( x^{(j)}_{0}, x^{(j)}_{1}, ..., x^{(j)}_{q-1} \), for \( j = 1, 2, ..., d \). Parameters, \( \alpha^{(h)}_{jl} \) satisfy the following two conditions:

1. \( \alpha^{(h)}_{jl} \geq 0; 1 \leq j, l \leq d; 1 \leq h \leq q \);

2. \( \sum_{l=1}^{d} \sum_{h=1}^{q} \alpha^{(h)}_{jl} = 1 \) for \( j = 1, 2, ..., d \).

(2)

The probability state distribution of the time series \( x_{p+1}^{(j)} \) depends on the weighted average \( \alpha^{(h)}_{jl} Q^{(h)}_{j} x_{p-h+1}^{(l)} \). Where \( Q^{(h)}_{j} \) is the probability transition matrix with \( h \) steps, whose elements are the transition probabilities with \( h \)-steps from states in
sequence \( l \) at time \( t = p + h + 1 \) to states in sequence \( j \)-th at time \( t = p + 1 \). From equation (1) we denote:

\[
X_j^{(j)} = ((x_i^{(j)})^T, (x_{i-1}^{(j)})^T, \ldots, (x_{i-q}^{(j)})^T)^T,
\]

for \( j = 1, 2, \ldots, d \) a \( dq \times 1 \) vector. So, we can write the following matrix form:

\[
X_{t+1} = \begin{bmatrix}
X_1^{(1)} \\
X_1^{(2)} \\
\vdots \\
X_1^{(d)} \\
X_2^{(1)} \\
X_2^{(2)} \\
\vdots \\
X_2^{(d)} \\
\vdots \\
X_d^{(1)} \\
X_d^{(2)} \\
\vdots \\
X_d^{(d)}
\end{bmatrix} = \mathbf{C}_t X_t.
\tag{3}
\]

Where

\[
\mathbf{C}_t = \begin{bmatrix}
C_{1t} & C_{2t} & \cdots & C_{dt} \\
C_{11} & C_{12} & \cdots & C_{1d} \\
\vdots & \vdots & \ddots & \vdots \\
C_{d1} & C_{d2} & \cdots & C_{dd}
\end{bmatrix}_{kd 	imes kd}
\]

If \( i \neq j \)

\[
\mathbf{C}_{ij} = \begin{bmatrix}
a_{ij}^{(1)} & a_{ij}^{(2)} & \cdots & a_{ij}^{(d-1)} & a_{ij}^{(d)}
\end{bmatrix}_{kd 	imes kd}
\]

The sum of elements in every column of the \( Q_{ij}^{(d)} \) matrices is equal to one.

For \( q = 1 \) model (1) is the first-order multivariate Markov chain model which was first presented by Ching et al. (2002). Their model takes into consideration \( d \) categoric time series and the forecasting model is build considering only one step inner and intra transition probability matrices. Using the same notations as above, the form of the model is:

\[
x_j^{(t+1)} = \sum_{i=1}^d a_{ij} Q_{ij}^{(i)} x_i^{(t)},
\tag{4}
\]

for \( j = 1, 2, \ldots, d \) and \( t = 0, 1, \ldots \) where \( Q_{ij}^{(i)} \) is the one-step probability transition matrix from \( i \)th sequence to the \( j \)th sequence, while \( x_i^{(t)} \) is the probability state distribution at time \( t \) of the \( i \)th time series. The parameters \( a_{ij} \) satisfy the conditions (2). In matrix form model (4) is:

\[
X_{t+1} = \begin{bmatrix}
X_1^{(1)} \\
X_1^{(2)} \\
\vdots \\
X_1^{(d)} \\
X_2^{(1)} \\
X_2^{(2)} \\
\vdots \\
X_2^{(d)} \\
\vdots \\
X_d^{(1)} \\
X_d^{(2)} \\
\vdots \\
X_d^{(d)}
\end{bmatrix} = \begin{bmatrix}
a_{11} Q_{11}^{(1)} & \cdots & a_{1d} Q_{1d}^{(1)} \\
a_{21} Q_{21}^{(2)} & \cdots & a_{2d} Q_{2d}^{(2)} \\
\vdots & \cdots & \vdots \\
a_{d1} Q_{d1}^{(d)} & \cdots & a_{dd} Q_{dd}^{(d)}
\end{bmatrix}_{kd 	imes kd} X_t.
\]

For \( d = 1 \) model (1) is the higher-order univariate Markov chain model which was developed by Ching et al. (2003). In their model, they considered only one time series with categorical data and 1, 2, ..., \( q \) step probability transition matrices. So, the state probability distribution at time \( t + 1 \) depends on the state probability distribution at time \( t, t - 1, \ldots, t - q + 1 \), where \( q \) is the order of the model. Using the same notations as above the form of the Markov chain model is:

\[
x_{t+1} = \sum_{i=1}^N a_{i} Q_{i} x_{t-i+1}, \quad t = q-1, q, \ldots.
\tag{5}
\]

Where \( Q_{i} \) is the \( i \)th step probability transition matrix, while \( x_{t+i} \) is the probability state distribution of the time series at time \( t + i \). Parameters \( a_{i}, \ i = 1, 2, \ldots, q \) satisfy conditions (2).

### 2.2. Prediction accuracy

When working on a forecasting model it is important to understand and test the accuracy and effectiveness of the model. For evaluating the performance of the Markov chain models presented above, the prediction accuracy is measured as a percentage of matching values and is calculated from the formula:

\[
Acc = \frac{1}{N-n} \left( \frac{\sum_{i=n+1}^N \varepsilon_i}{N} \right) \cdot 100\%.
\tag{6}
\]

where \( N \) is the size of our data and \( n \) is the size of data considered when constructing the model. While \( \varepsilon_i \) is defined as below:

\[
\varepsilon_i = \begin{cases} 
1 & \text{if } x_p = x_i \\
0 & \text{if } x_p \neq x_i 
\end{cases}
\]
where $x_p$ is the predicted value and $x_i$ is the real value.

### 3. Case study

In this section, we have considered four time series: CPI, life expectancy, unemployment rate, and inflation. The time series were collected yearly from 1995 to 2018 and are obtained from the World Bank website. When we work on the model, we have considered as a training set the data from 1995 to 2015, and as a testing dataset the data from 2016 to 2018. The missing data for the unemployment rate and life expectancy are obtained from the links indicated at the end of the references.

Regarding World Bank indicator definitions, inflation is measured as the annual percentage of consumer price. The two indicators are expressed in two different units, one shows the annual growth (Inflation) and the other the real value (CPI). We have studied them to understand which one has the highest impact on the other variables considered in the study.

When we work on a forecasting model, the aim is not only to predict future values but also to capture the patterns of the time series considered. So, it is important to analyze the time series over time and their correlation.

![Figure 1. Time series of Life Expectancy, CPI change, unemployment and inflation rate in Albania (yearly from 1995 to 2018) (source: authors)](image)

As may be observed from Figure 1, the CPI and life expectancy have a linear increasing trend; unemployment and inflation, on the other hand, have not a stationary pattern.

In this study we have constructed multivariate Markov chain models, a first-order and a fifth-order, considering the economic indices presented above. A comparative analysis between these models and other classical forecasting methods is done. Based on the accurate evaluation of the models, the best one is considered for prediction.

A Markov chain forecasting model may work for categorical time-series data. In the first step, we must discretize our data in a reasonable and justified way so they meet the model conditions. Observing the summary output of our time series and their behavior we have used the first quartile and the mean value to determine the breakpoints of the intervals. The discretization is done by classifying each value depending on the interval it falls. So, values lower than the first quartile are indexed as “1” considering them as Low values; values falling in between the first quartile and the mean are indexed as “2” considered as Medium values and values greater than the mean are indexed as “3”, considered as High values. Table 1 shows the intervals for each time series.

#### Table 1. Intervals considered for the discretization of the time series (source: authors)

<table>
<thead>
<tr>
<th>Indicator</th>
<th>LOW (1)</th>
<th>MEDIUM (2)</th>
<th>HIGH (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>&lt;74.14</td>
<td>[74.14, 85.29]</td>
<td>&gt;85.29</td>
</tr>
<tr>
<td>Unempl</td>
<td>&lt;13.67</td>
<td>[13.67; 15.08]</td>
<td>&gt;15.08</td>
</tr>
<tr>
<td>Inflation</td>
<td>&lt;1.9254</td>
<td>[1.9254; 5.5348]</td>
<td>&gt;5.5348</td>
</tr>
<tr>
<td>LifeExp</td>
<td>&lt;73.95</td>
<td>[73.95; 75.22]</td>
<td>&gt;75.22</td>
</tr>
</tbody>
</table>

Using R software packages we fitted to our data a first-order multivariate Markov chain model. The final model is expressed in (7).

$$
\begin{align*}
X_1^{(t+1)} &= X_3^{(t)} \cdot Q_{13} \\
X_2^{(t+1)} &= 0.0188 \cdot X_1^{(t)} \cdot Q_{21} + 0.9624 \cdot X_3^{(t)} \cdot Q_{23} + 0.0188 \cdot X_4^{(t)} \cdot Q_{24} \\
X_3^{(t+1)} &= 0.5 \cdot X_1^{(t)} \cdot Q_{31} + 0.5 \cdot X_4^{(t)} \cdot Q_{34} \\
X_4^{(t+1)} &= X_3^{(t)} \cdot Q_{43}
\end{align*}
$$

where $X_1$ – represents the life expectancy time series; $X_2$ – represents the inflation ratio time series; $X_3$ – represents the unemployment ratio time series; $X_4$ – represents the CPI time series; and, $Q_{ij}$ represents one step probability matrix from time series $j$ to time series $i$. $X_i^{(t)}$ is the probability distribution of $i$th time series at time $t$.

From equation (7) we may observe that time series are dependent one step backward on time from each other. Life expectancy, CPI and unemployment are dependent on each other. The inflation rate is correlated at the same time with life expectancy, unemployment rate, and CPI. The
predicting the Albanian economic development using multivariate Markov chain model

forecasting values using the model (7) are presented in Table 2. The state with the highest probability (in bold) will be the state considered at time \( t + 1 \). Table 2 shows for three years (2016, 2017 and 2018) the interval where the value falls and the probabilities obtained from the Markov model. The last column of the table shows the prediction probabilities of falling in each of the three states obtained from model (7) for 2019.

So, the predictions for 2019 of life expectancy are 51.9% chance of falling in the High interval; for the inflation rate, there is 49.86% chance of falling in Medium interval; for the unemployment rate, there is 47% chance of falling in the High interval and for CPI is 51.9% of falling in High interval. For CPI and life expectancy we expect these values considering the upward trend of these time series.

The second approach was building a multivariate model. With the same notations as in the first-order multivariate model, we have constructed the fifth-order multivariate Markov chain model:

\[
X_i^{(t+1)} = X_1^{(t)} \cdot Q_{11} + X_2^{(t)} \cdot Q_{21} + \ldots + X_4^{(t)} \cdot Q_{44} + \text{error},
\]

where \( Q_{ij}^{(h)} \) is the \( h \)-step probability transition matrix from state \( j \) to the state \( i \).

Table 2. State probability for each of states in the years 2016, 2017, 2018 and 2019 using the model (7) (source: authors)

<table>
<thead>
<tr>
<th>Time series</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Markov</td>
<td>Real</td>
<td>Markov</td>
<td>Real</td>
</tr>
<tr>
<td>X_1</td>
<td>0.1802</td>
<td>0.2137</td>
<td>0.2149</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>0.2973</td>
<td>0.2641</td>
<td>0.2659</td>
<td>0.266</td>
</tr>
<tr>
<td></td>
<td>0.5225</td>
<td>0.5222</td>
<td>0.5192</td>
<td>0.519</td>
</tr>
<tr>
<td>X_2</td>
<td>0.3052</td>
<td>0.2844</td>
<td>0.289</td>
<td>0.2893</td>
</tr>
<tr>
<td></td>
<td>0.53</td>
<td>0.5059</td>
<td>0.499</td>
<td>0.4986</td>
</tr>
<tr>
<td></td>
<td>0.1648</td>
<td>0.2097</td>
<td>0.212</td>
<td>0.2121</td>
</tr>
<tr>
<td>X_3</td>
<td>0.3077</td>
<td>0.3128</td>
<td>0.313</td>
<td>0.313</td>
</tr>
<tr>
<td></td>
<td>0.2298</td>
<td>0.2176</td>
<td>0.217</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>0.4625</td>
<td>0.4696</td>
<td>0.47</td>
<td>0.47</td>
</tr>
<tr>
<td>X_4</td>
<td>0.1802</td>
<td>0.2137</td>
<td>0.215</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>0.2973</td>
<td>0.2641</td>
<td>0.266</td>
<td>0.266</td>
</tr>
<tr>
<td></td>
<td>0.5225</td>
<td>0.5222</td>
<td>0.519</td>
<td>0.519</td>
</tr>
</tbody>
</table>

Table 3. State probability for each of states in the years 2016, 2017, 2018 and 2019 using the model (8) (source: authors)

<table>
<thead>
<tr>
<th>Time series</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Markov</td>
<td>Real</td>
<td>Markov</td>
<td>Real</td>
</tr>
<tr>
<td>X_1</td>
<td>0.1802</td>
<td>0.2115</td>
<td>0.1846</td>
<td>0.1879</td>
</tr>
<tr>
<td></td>
<td>0.2973</td>
<td>0.2438</td>
<td>0.271</td>
<td>0.2714</td>
</tr>
<tr>
<td></td>
<td>0.5225</td>
<td>0.5447</td>
<td>0.5444</td>
<td>0.5407</td>
</tr>
<tr>
<td>X_2</td>
<td>0.3052</td>
<td>0.2491</td>
<td>0.2698</td>
<td>0.273</td>
</tr>
<tr>
<td></td>
<td>0.53</td>
<td>0.5432</td>
<td>0.5462</td>
<td>0.5411</td>
</tr>
<tr>
<td></td>
<td>0.1648</td>
<td>0.2077</td>
<td>0.184</td>
<td>0.1859</td>
</tr>
<tr>
<td>X_3</td>
<td>0.2948</td>
<td>0.1928</td>
<td>0.2057</td>
<td>0.2002</td>
</tr>
<tr>
<td></td>
<td>0.3047</td>
<td>0.3712</td>
<td>0.352</td>
<td>0.3628</td>
</tr>
<tr>
<td></td>
<td>0.4005</td>
<td>0.436</td>
<td>0.4423</td>
<td>0.437</td>
</tr>
<tr>
<td>X_4</td>
<td>0.1802</td>
<td>0.2115</td>
<td>0.1846</td>
<td>0.1879</td>
</tr>
<tr>
<td></td>
<td>0.2973</td>
<td>0.2438</td>
<td>0.271</td>
<td>0.2714</td>
</tr>
<tr>
<td></td>
<td>0.5225</td>
<td>0.5447</td>
<td>0.5443</td>
<td>0.5407</td>
</tr>
</tbody>
</table>
From the representation of the model (8), we note that the probability state distribution of unemployment at time $t + 1$ depends on the probability state distribution of the other time series at time $t - 4$. We also observe that life expectancy is explained from the unemployment rate at one step backward. The inflation rate is explained from all time series at one step backward. The unemployment rate as well is explained from all time series at five steps backward. CPI is explained only by the unemployment rate at one step backward.

The forecasting values using the model (8) are presented in Table 3. The state with the highest probability (in bold) will be the state considered at time $t + 1$.

Predictions for 2019 obtained from model (8) show that for the time series of life expectancy there is $54\%$ chance being at state “3”, for the inflation rate a chance of $54\%$ being at state “2”, for unemployment rate a chance of $43.7\%$ being at state “3” and for CPI a chance of $54\%$ being at state “3”.

Using equation (6) the prediction accuracy for both Markov chain models are: $100\%$ prediction accuracy for life expectancy; $66.7\%$ prediction accuracy for inflation rate; $33.3\%$ prediction accuracy for unemployment rate and $100\%$ prediction accuracy for CPI. Calculating the prediction accuracy of the first order and fifth-order multivariate Markov Chain model in both cases we obtain a $75\%$ prediction accuracy. This is a considerable value to believe that the model will predict accurately as a short term forecast.

**4. Conclusions**

In this work, we have analyzed the correlation of some important economic indices such as inflation rate, Consumer Price Index, life expectancy and unemployment rate and their impact on the economic development of a country. These time series were selected among other economic and demographic indices that have been analyzed in previous research from other authors.

We observed a significant correlation between life expectancy, inflation rate, and CPI. The unemployment rate was not significantly correlated with these time series. The time series of CPI and life expectancy will exhibit an upward trend in upcoming years. Our approach was to build a probabilistic model that would predict the interval where the forecast will be at time $t + 1$. We build a first-order and a fifth-order multivariate Markov chain model and used them to predict probabilities for the next observations. Both models strengthen the suspicion of the correlation among these time series.

What was also observed and confirmed during the forecasting procedure was that the time series of life expectancy, inflation rate and CPI due to their trend and periodic nature are forecasted with more confidence compared to the unemployment rate. Regarding the use of the inflation rate and CPI indices in both models, we came into the conclusion that CPI is the most appropriate economic indicator to use compared to inflation.

The first and fifth order Markov chain models were fitted to examine the relationship between selected time series. The prediction probabilities for each time series fall in the interval of 0.47 to 0.52. The prediction accuracy of both Markov chain models is $75\%$. The values of CPI and life expectancy in the coming years are expected to be in the High interval. And, for the unemployment rate and inflation are expected to fall in the medium interval.

The absence of information before 2001 made it challenging to add other economic, social and demographic indices to our model. The high order Markov chain models may be considered in countries with supplementary data. In another view, this study brings into attention the use of Markov chains as auxiliary models combined with other forecasting models such as neural or Bayesian networks.

**Disclosure statement**

The authors declare that there is no conflict of interest.

**References**


E. Gjika (Dhamo), L. Basha, X. Allka, A. Ferrja


