Islamic World Science Citation Center

Islamic Azad University
Marvdasht Branch

Journal of Regional Planning

ISSN (Print): 2251-6735
ISSN (Online): 2423-7051

**Vol. 10/ No. 37/ Spring 2020**

Research Paper

# Clustering Based on Forecasting Density: Case Study of Unemployment Rate in Iran's Provinces

***Ramin Khochiani***[1]***:*** *Assistant Professor of Economics, Ayatollah Borujerdi University, Borujerd, Iran*
***Seyed Mohammad Hosseini:*** *Assistant Professor of Mathematics, Ayatollah Borujerdi University, Borujerd, Iran*

*Received: 2018/11/11* *PP: 13- 16* *Accepted:2019/4/26*

## Abstract

*It is important for regional planners and policymakers to be aware of the unemployment rate of the provinces in specified time horizons. In this paper, clustering of time series based on their forecasting density to a specified horizon is investigated. In this algorithm, we use the bootstrap process to approximate the distribution of predictions. The differences between each pair of bootstrap densities generate a dissimilarity matrix that is used for clustering. For this purpose, seasonal unemployment data was used in the spring of 2005 to fall of 2017, and according to the forecasting density algorithm, we will cluster the unemployment rate of Iran's provinces for two horizons of 4 steps (one year) and 10 steps (two and a half years). The best situation will be in the 4 steps or 10 steps (two and a half years), the provinces of Semnan and Zanjan, and the worst situation in the provinces of Lorestan and Kermanshah. Also, in the two horizons studied, except for some provinces, the rest were fixed in their main clusters. The spatial distribution of unemployment in Iran, based on forecasting density clustering, shows that western and southwestern provinces will have the highest unemployment rates. Therefore, the need for regional planning and serious attention to the employment of these provinces is recommended. At the same time, the provinces that are in an unfavorable situation have high unemployment neighbors, and the provinces with low unemployment rate have predominantly neighborhoods with a low unemployment rate. In other words, there is a positive spatial correlation between the neighboring provinces and the unemployment rate.*

***Key words:*** *Forecast Density Clustering, Time Series, Sieve Bootstrap, Unemployment Rate.*

## Extended Abstract

### Introduction:

*Managing raw data and extracting useful information plays an important role in decision making. Clustering as one of the descriptive data mining methods is followed by organizing the data into a number of clusters in such a way that objects in the same cluster are more similar to each other than to those in other clusters. Alonso et al. (2006) proposed a concept of dissimilarity measure based on the forecast densities, for each one of the observed series in the sample for a given future horizon. They combined a smoothed sieve bootstrap procedure with nonparametric kernel density estimation ideas to approximate the distribution of the predictions. Villar et al. (2010) developed this method and also covered nonparametric nonlinear autoregressive models.*

[1] *Corresponding Author's Email: khochiany@abru.ac.ir, Tel: +989133728231*

*In this paper, both feature-based and model-based approaches are used to cluster time series data. The main purpose of this study is to cluster time series data based on complete prediction densities for each series in the set, rather than focusing on point predictions. Here, time series clustering is performed based on full forecast densities. Time series fall into a cluster in a forecast density distribution in the future time horizon with other similar time series in the same future time horizon. Clustering and dissimilarity based on the forecast densities can be easily interpreted. Unemployment is a good measure for the state of the balance between the key pillars of the country's economy. Therefore, it is important for the authorities to address it. In this study, in order to show the efficiency of the mentioned clustering method, the problem of unemployment rate clustering in the provinces of Iran was considered and the provinces were clustered in terms of unemployment rate in the next four and ten steps horizon.*

**Methodology:**

*In this paper, clustering is performed based on full forecast densities instead of focusing on point predictions. Suppose* $X_T$ *and* $Y_T$ *are two stationary processes with the following autoregressive representation:*

$S_t = \varphi(S_{t-1}) + \varepsilon_t$ *Where* $\varepsilon_t$ *is a white noise and* $\varphi(.)$ *is a smooth function which is not constrained by any predetermined parametric model. At a specified future time T+h, let*

$$d_{PRED,h}(X_T, Y_T) = \int \left|\hat{f}_{X_{T+h}}(u) - \hat{f}_{Y_{T+h}}(u)\right| du$$

*where* $\hat{f}_{X_{T+h}}$ *and* $\hat{f}_{Y_{T+h}}$ *denote the density function of the forecasts* $X_{T+h}$ *and* $Y_{T+h}$*, respectively.* $d_{PRED,h}(X_T, Y_T)$ *denotes the distance between* $X_T$ *and* $Y_T$ *at the specified future time T+h. The correct forecast densities are replaced by kernel (nonparametric) estimates based on bootstrap predictions.*

**Results and discussions:**

*According to the proposed algorithm, the unemployment rate of the Iran's provinces will be clustered for two time horizons of 4 seasons and 10 seasons. Seasonal data of provincial unemployment rate during spring 2005 to autumn 2017 were extracted from the statistics center. Some of the results are:*

*Clustering for 4 future seasons:*

- *cluster 1) Zanjan, Semnan, Kerman, Mazandaran, Khorasan Razavi, North Khorasan, South Khorasan, Yazd, East Azarbaijan, Golestan, Markazi, Hormozgan, Qazvin, Qom, Sistan and Baluchestan, Bushehr and Tehran.*
- *cluster 2) Chaharmahal & Bakhtiari, Kohkiluyeh & Boyerahmad, Kermanshah, Lorestan, Kurdistan, West Azarbaijan, Hamedan, Ilam, Gilan, Ardabil, Fars, Isfahan, Khuzestan.*

*Forecasting at 10 future seasons, the provinces of West Azerbaijan, Hamedan, Kurdistan, Ardebil, Khuzestan and Isfahan were moved from main cluster 2 to main cluster 1. Since the situation in the provinces of cluster 1 is better than cluster 2, it can be said that these provinces will have better situation in the next two years compared to the next one. Kermanshah and Lorestan have the worst situation in both 4 and 10 time horizons.*

*In the current situation, the provinces of East Azerbaijan and Yazd have the lowest average unemployment rate among the provinces of the country. However, in the above time horizons, although they are located in cluster 1 (provinces with good status), the prediction density of Zanjan and Semnan proves better. According to the results of the clustering, west and southwestern provinces will also have the highest unemployment rate.*

**Conclusion:**

*A new density-based clustering method is represented for forecasting. The time series of unemployment rates of the provinces were clustered in two time horizons of 4 seasons and 10 seasons. The results show that Semnan and Zanjan provinces will have the best situation in 4 or 10 steps and* *the worst situation will be in Lorestan and Kermanshah provinces. Also in the two time horizons studied, with the exception of a few provinces, the rest were fixed in their original clusters.*

*The spatial distribution of unemployment in Iran based on the forecast density-based clustering shows that the western and southwestern provinces will still have the highest unemployment rate. There is also a positive spatial correlation between the neighboring provinces and the unemployment rate. The results of the research of Razvani et al (2013) indicate high unemployment rate in the south and west of the country. As can be seen from the results of the present study, one can expect the high unemployment rate to remain in the same areas with the exception of southeastern provinces such as Sistan and Baluchestan, Kerman and Hormozgan as well as Bushehr.*

*The results of this study recommend attention to provincial-based economic policies to reduce employment inequalities. It is also suggested that other macroeconomic variables be clustered with this new approach to provide a clearer horizon in economic policy making. In addition, the literature results show that the unemployment has a huge impact on immigration, crime and suicide. Therefore, addressing the problem of high unemployment rates in the next 4 or 10 time horizons in the provinces involved can minimize the impact of such phenomena.*