

Possibilistic clustering for crisis prediction: Systemic risk states and membership degrees

József Mezei

School of Business and Management
Lappeenranta University of Technology
Lappeenranta, Finland
Risklab Finland, Arcada University of Applied Sciences
Helsinki, Finland
Jozsef.Mezei@lut.fi

Peter Sarlin

Department of Economics
Hanken School of Economics
Helsinki, Finland
Risklab Finland, Arcada University of Applied Sciences
Helsinki, Finland
peter@risklab.fi

Abstract

Research on understanding and predicting systemic financial risk has been of increasing importance in the recent years. A common approach is to build predictive models based on macro-financial vulnerability indicators to identify systemic risk at an early stage. In this article, we outline an approach for identifying different systemic risk states through possibilistic fuzzy clustering. Instead of directly using a supervised classification method, we aim at identifying coherent groups of vulnerability with macro-financial indicators for pre-crisis data, and determine the level of risk for a new observation based on its similarity to the identified groups. The approach allows for differentiating among different possible pre-crisis states, and using this information for estimating the possibility of systemic risk. In this work, we compare different fuzzy clustering methods, as well as conduct an empirical exercise for European systemic banking crises.

1 Introduction

Recent episodes of financial turmoil have illustrated the real economic costs of systemic financial crises. A financial crisis is a state in a financial system that causes economic, social, and political costs [24] with disastrous effects on the affected economies. For this reason, developing various methods for the purpose of crisis prediction has been an important research focus in recent years. Among many other methods, machine learning algorithms and computational intelligence approaches are increasingly used in this context [5]. In this article, we describe an application of a specific class of clustering methods, namely possibilistic clustering, to crisis prediction.

Clustering is one of the most important methods ap-

plied in pattern recognition and data mining and aims at partitioning a set of data points into groups of “similar” observations. Fuzzy clustering methods rely on set-theoretical notions introduced by Zadeh [28], motivated by the imprecision present in many (if not all) real life phenomena. The main idea behind fuzzy sets (degree of belonging to sets) naturally translates to clustering algorithms: elements can belong to several (overlapping) fuzzy clusters specified by a membership value. In fuzzy clustering, the fuzzy c-means (FCM) clustering algorithm [1] is the best known and used method. Since the FCM memberships do not always explain the degrees of belonging for the data well, Krishnapuram and Keller [9] proposed a possibilistic approach to clustering to correct this weakness of FCM.

In this paper we focus on a subset of fuzzy clustering methods (objective function-based algorithms) that include the most widely applied variants [6]. Following [13], we apply fuzzy clustering approaches to indicators of financial crises, but extend our numerical study toward several directions. First, we perform a performance comparison of several fuzzy clustering methods that particularly includes possibilistic approaches. Second, rather than identifying overall financial stability states, we aim at identifying natural clusters in pre-crisis data as to provide more descriptive probabilities representation through different states and membership degrees. Third, we tackle European systemic banking crises in the past decades in contrast to their focus on excessive exchange-rate pressure in the 1990s in Asia. In a numerical study, this paper applies the chosen method as the basis of predicting whether a country at a given time is in a pre-crisis state or not based on the membership and typicality degrees of belonging to different clusters. This allows for not only differentiating among different possible pre-crisis states, but also provides a possibilistic measure of severity for each type of systemic risk.

The rest of the paper is structured as follows. Section 2 presents the preliminaries and in Section 3 (possibilistic) fuzzy clustering algorithms are discussed. The main focus is on describing how the method presented in [14] can be used as the basis of crisis prediction. The numerical analysis is presented in Section 4 to offer a novel approach utilizing possibilistic clustering as the basis of systemic risk analysis. Finally, Section 5 provides conclusions and future research directions.

2 Preliminaries

In the following we recall relevant literature and preliminary notations used in the article and related to: (i) crisis prediction, specifically focusing on systemic risk analytics, utilizing machine learning methods, (ii) the application of fuzzy sets in machine learning, and (iii) mathematical basics of different fuzzy clustering approaches.

2.1 Machine learning in financial risk analysis

When utilizing machine learning, the main goal is to build algorithms that can learn from the available data and utilize the results of the learning process, typically for predictions. Various machine learning methods can be classified in different ways, with the two most widely used categories being the following:

- Supervised learning: in a general supervised learning algorithm, there is an attribute of particular interest (e.g. indicator of crisis period, credit score of a customer), and the aim is to identify patterns in the other attributes that influence the target attribute. The target variable can be continuous (typically a general prediction problem and usually approached by a type of regression model) or categorical (classification problem). The most widely used methods belonging to this group include decision trees [29], neural networks [19] and support vector machines [26].
- Unsupervised learning: methods belonging to this group are applicable to problems with datasets with out without target variables, and aim at identifying underlying structures in the data. While in case of the problem of crisis prediction, a class variable is usually present, unsupervised learning methods can still offer an important tool, with the most frequently used methods belonging to this group include self-organizing maps [23] and c -means clustering [12]

In a recent and literature review, a complete picture of the application of machine learning in financial crisis pre-

diction was given by Lin et al. [11]. Additionally to the above mentioned two groups, they found statistics-based learning methods, for example logistic regression [24] or discriminant analysis [18], and other methods, such as genetic algorithms [19], as widely used in the crisis prediction literature.

Unsupervised learning methods, in particular clustering, have been used less frequently compared to other methodologies, especially when compared to supervised learning methods, and the contributions mainly originate from the domain of individual's credit evaluation. Additionally, it is important to observe that most of the contributions are restricted to a handful of clustering methods, mainly utilizing c -means clustering. Kuo et al. [8] present an evaluation of clustering algorithms for financial risk analysis. They choose six clustering algorithms for the study: c -means, expectation-maximization (EM), COB-WEB, repeated-bisection approach, graph-partitioning algorithm, and density-based method.

2.2 The role of fuzzy sets in machine learning

In the last decades, the tools of fuzzy set theory has been increasingly applied to develop new machine learning approaches and address various practical problems with these tools. As it was pointed out by Hüllermeier [7], the main reason behind this lies in the paradigm change of model development: from knowledge-based to data-driven fuzzy modelling. While the traditional applications of fuzzy sets relied on the manual design of fuzzy systems by utilizing the knowledge of human experts (for example in approximate reasoning in control applications), this is in contrast with the automatic development of such systems by refining fuzzy models based on available data.

Hüllermeier [7] specified the three most important potential advantages of fuzzy sets in machine learning, present also in clustering models, as:

- creating models defined in terms of fuzzy concepts, e.g. using fuzzy similarity measures to define the distance among observations or using complex aggregation operators in model building;
- the use of non-inductive inference, e.g. applying knowledge learnt in a different context;
- representation of uncertainty (partial truth), e.g. representing linguistic expressions or imprecision (non-random uncertainty) present in the data.

Fuzzy clustering, as one of the most popular applications of fuzzy sets in machine learning, has been addressed in several articles focusing on both model devel-

opment and practical applications. From the modelling perspective, there exist several families of models [20], for example hierarchical cluster analysis or fuzzy c -shell algorithm; still the most widely used variant, the objective function-based clustering dominates the literature.

2.3 Clustering

In clustering, the general input data consists of n observations: $\mathbf{x}_i = [x_{i1}, x_{i2} \dots x_{im}]$ for $i = 1, \dots, n$, where every observation includes measurement by m variables. The following representation can be used:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

While the process of utilizing clustering does not require the availability of a target or class variable, in many applications, as it is the case with this study, there is a class variable x_c , assigning one of a predefined classes to each observations. The obtained clustering structure can afterwards also aid in estimating class information for new observations; this will be discussed later in the paper.

The output of a clustering algorithm is a partition of the datapoints. Observations that are more similar to each other in terms of a predefined distance measure, are assigned to the same partition subset or cluster. The traditional approach assumes that each observations is assigned to a single cluster. Formally, if we denote the number of clusters with c , a crisp partition can be described by the matrix $\mathbf{U} = [\mu_{ij}]_{c \times n}$, where

$$\mu_{ij} \in \{0, 1\}, 1 \leq i \leq c, 1 \leq j \leq n$$

$$\begin{aligned} \sum_{i=1}^c \mu_{ij} &= 1, 1 \leq j \leq n \\ 0 < \sum_{j=1}^n \mu_{ij} &< n, 1 \leq i \leq c \end{aligned}$$

For example, in the traditional c -means clustering, the partition is determined by assigning an observation to the cluster with the cluster center minimizing the distance from the observation.

As it was reasoned the first time in [21] using the famous ‘‘butterfly’’ example, in many cases it is reasonable to release the assumption of an observation belonging to a single cluster to assign observations on the boundary of clusters to several clusters to some degree. This observation lead to the development of various fuzzy clustering approaches [4].

The classical, usually termed as probabilistic, fuzzy partition can be described with the same constraints with

extending the possible values for the cluster memberships, μ_{ij} , to the $[0, 1]$ interval as follows:

$$\mu_{ij} \in [0, 1], 1 \leq i \leq c, 1 \leq j \leq n$$

$$\sum_{i=1}^c \mu_{ij} = 1, 1 \leq j \leq n$$

$$0 < \sum_{j=1}^n \mu_{ij} < n, 1 \leq i \leq c$$

Furthermore, a possibilistic fuzzy partition releases the normalization constraint on the membership values and it is replaced by

$$\exists \mu_{ij} > 0, 1 \leq j \leq n,$$

which ensures that every observation belongs to at least one cluster to some degree.

While probabilistic fuzzy clustering can improve clustering based on crisp partitions from various perspectives, it has an inherent limitation regarding the assumption requiring the sum of membership values to be equal to 1 for every observation. This property forces each point to belong to at least one cluster to a large degree, which can lead to misleading results when dealing with outliers. To tackle this problem, possibilistic partitions as the basis of clustering can be defined as in [9], by modifying the definition of a possibilistic fuzzy partition into

$$\mu_{ij} \in [0, 1], 1 \leq i \leq c, 1 \leq j \leq n$$

$$\mu_{ij} > 0, 1 \leq j \leq n$$

$$0 < \sum_{j=1}^n \mu_{ij} < n, 1 \leq i \leq c$$

This type of partition, while ensures that each datapoint belongs to at least one cluster, does not constrain an observation to necessarily belong to a high degree to clusters overall. While in case of probabilistic fuzzy clustering μ_{ij} is termed as membership value, in possibilistic clustering this assignment value is termed as typicality value, and denoted by t_{ij} , a notation and terminology we will also employ in the paper. An important example of the usefulness of typicality values arises when we deal with outliers. While traditional c -means clustering would force an outlier to belong to at least one cluster to a high degree, possibilistic clustering allows for low typicality values associated to all the clusters.

3 Possibilistic clustering for crisis prediction

In the following, the mathematical preliminaries necessary to formulate various clustering methods are presented. Our main goal is to apply (fuzzy/possibilistic)

clustering as the basis of a classification task with binary outcomes. The presented method summarized in the second part of this section utilizes and adapts the general approach described in [14, 15] to the problem of crisis prediction.

3.1 Objective function-based possibilistic fuzzy clustering for classification

Based on the above discussed types of partitions, there are numerous clustering methods proposed in the literature. The most widely applied methods belong to one of two groups of approaches [20]: (i) objective function-based methods aiming to find the optimal partition using optimization techniques, and (ii) methods generalizing the objective function-based approach by specifying update equations for the cluster centers and membership degrees and perform iterative updates. In this work, we focus on the more frequently utilized group of methods, the objective function-based approach. In these methods, a set of variables, usually the membership values, are optimized holding the other group (e.g., the cluster centers) fixed and vice versa following an iterative updating scheme. A general objective function, incorporating many of the proposals from the literature was formulated in [14, 15] as

$$J(\mathbf{X}, \mathbf{U}, \mathbf{B}) = \sum_{i=1}^c \sum_{j=1}^n \left(a\mu_{ij}^m + bt_{ij}^\lambda \right) d^2(x_i, c_j) + \sum_{i=1}^c \eta_i \sum_{j=1}^n (1 - t_{ij})^\lambda + \sum_{i=1}^c \gamma_i \sum_{j=1, j \neq i}^c \frac{1}{\zeta d^2(c_i, c_j)} \quad (1)$$

where μ_{ij} is the membership, t_{ij} is the typicality, and the other parameters modify the distance measure in the model and the shape of the resulting clusters. The special cases of this function include the ones used in: (i) the fuzzy c -means [1] ($a = 1, b = 0, \eta_i = 0, \gamma_i = 0$); (ii) possibilistic fuzzy clustering in [9] ($a = 0, b = 1, \gamma_i = 0$); (iii) mixed c -means clustering model [16] ($a = 1, b = 1, \eta_i = 0, \gamma_i = 0$); (iv) possibilistic c -means clustering model [17] ($\gamma_i = 0$); (v) the extended possibilistic clustering model [25] ($a = 0$).

In this work, we only focus on fuzzy c -means clustering and the original possibilistic clustering methodology. In the first case, the algorithm uses the update formulas

$$\mu_{i,j} = \frac{1}{\sum_{k=1}^c \left(\frac{d(x_i, c_j)}{d(x_i, c_k)} \right)^{\frac{2}{m-1}}} \quad (2)$$

$$c_j = \frac{\sum_{k=1}^c \mu_{k,j}^m x_k}{\sum_{k=1}^c \mu_{k,j}^m} \quad (3)$$

while in the possibilistic case, the typicality values and the cluster centers are calculated in every iteration as

$$t_{i,j} = \frac{1}{\sum_{k=1}^c \left(\frac{d(x_i, c_j)}{\eta_k} \right)^{\frac{2}{m-1}}} \quad (4)$$

$$c_j = \frac{\sum_{k=1}^c t_{k,j}^m x_k}{\sum_{k=1}^c t_{k,j}^m} \quad (5)$$

The fuzzy exponent value $m > 1$ specifies the extent of overlapping between clusters, with the two extreme being the classical c -means (when m approaches 1) and completely overlapping clusters when $m \rightarrow \infty$.

In applying clustering approaches, a crucial issue is to determine the appropriate number of clusters. For this purpose, there are several cluster validity measures proposed to assess cluster configurations, with many of them specific to fuzzy clustering as traditional measures evaluating hard clusterings are not applicable. As it was pointed out by Wang and Zhang [27], there is not a single validity measure that shows good performance in every situation, it is always recommended to consider several measures as the basis of evaluating the optimal number and structure of clusters. Typical measures defined for fuzzy clustering approaches include the partition coefficient [1] and the separation index [6]. In this article, as it will be described later in more details, we use a measure of classification performance to select the optimal number of clusters.

3.2 Fuzzy clustering for crisis prediction

In the following, we describe how the obtained clusters can be utilized as the basis of crisis prediction with the main focus on describing the case of possibilistic clustering and typicality values. Our basic setting is that we have a set of datapoints with several components representing specific financial indicator measures characterizing an entity at a given time point, and a binary class value specifying whether the entity experienced a pre-crisis period at that time point. The task is to predict the pre-crisis class based on the information from the indicator values.

To utilize the output of fuzzy clustering as a basis of solving a classification problem, the following steps need to be performed: (i) specifying a hard clustering; and (ii) assigning a class to each cluster. In case of traditional fuzzy c -means clustering, the associated hard clustering is obtained by assigning the observation to the cluster with the highest corresponding membership value. While this first step is trivial to perform, the second step of assigning a class to each cluster that results in optimal classification accuracy requires determining the optimal threshold value on the required percentage of observations in a crisis state assigned to a cluster. This threshold can be found

by sorting the clusters based on the frequency of observations in crisis state, and evaluate the accuracy when the top k clusters are assigned the crisis class, for $k = 1, \dots, c$, where c is the number of clusters.

Possibilistic clustering can be applied in a conceptually different way as motivated in [14]. Namely, the clustering is applied only to the data points classified as being in a pre-crisis state, termed as positive cases from now on. When the clusters are created, the new observations are assigned to some (or none) of the clusters in terms of the typicality values. As initially every cluster contains only crisis cases, one would expect observations being in a tranquil state to have large distance from all the specified clusters, meaning that all the associated typicality values are low. Formally, we calculate the minimum of the typicality values for each point, and if it is smaller than a specific threshold value, then the observations is assigned to the tranquil class. The threshold can be identified with the Usefulness measure (i.e., a preference weighted average of type I and II errors [22]).

As a summary, the following steps capture the most important components of the approach. These steps are based on the specifications of the general possibilistic clustering approaches described in [14, 15] applied to the context of crisis prediction. Our main contribution lies in the way we combine membership and typicality degrees compared to the original paper.

- Initialize the clustering process with the membership values generated by a run of fuzzy c -means clustering. This step is necessary, because as it was mentioned previously, with random initialization of typicality values possibilistic clustering tends to result in identical cluster centers. Additionally, the clustering parameters (i.e., maximum number of iterations, range for the possible number of clusters to be tested) are defined.
- Update the typicality values based on the formulas specified above, and calculate the center of the clusters.
- When a specified stopping criterion is reached (in terms of the number of iterations or change in the typicality values), stop the updating and record the final cluster centres.
- Predict the class for new observations by comparing the minimum of typicality values among all the clusters to a specified threshold.

A pseudo code capturing the main steps of the algorithm can be specified as follows.

1: **procedure** POSSIBILISTIC CLUSTERING FOR CRISIS PREDICTION(X, m, c_{max}, n)

```

2:   Initialize  $t_{ij}$  values through a run of traditional  $c$ -
   means clustering using crisis observations from the
   dataset
3:   for  $c \in [2, c_{max}]$  do
4:     while iteration number is less than  $n$  do
5:       Update  $t_{ij}$ 
6:       Recalculate the cluster centers
7:     end while
8:     Assign the non-crisis cases to the clusters using
   the average of membership and typicality
9:     Determine a threshold value for belonging to
   clusters based on usefulness measure
10:    end for
11:    return  $c$ , the optimal number of clusters and the
   assignment of the observations to one of the two
   classes
12: end procedure

```

4 Predicting European banking crises with possibilistic clustering

In the following we present the numerical experiment: the data used, the classification accuracy measures employed, and the results.

4.1 Data

The dataset used in this paper covers as many European economies as possible on a quarterly frequency and spans from 1976Q1 to 2014Q3. The sample is an unbalanced panel with 15 European Union countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden, and the United Kingdom. In total, the sample includes 15 crisis events, which cover systemic banking crises. The dataset consists of two parts: crisis events and vulnerability indicators. The crisis events cover country-level distress in the financial sector with systemic implications and rely on the IMF's crisis event initiative by [10]. The second part of the dataset consists of country-level macro-financial vulnerability indicators. In most cases, we have relied on the most commonly used transformations, such as ratios to GDP or income, growth rates, and absolute and relative deviations from a trend. We include the following measures in the analysis (in parenthesis the mean and standard deviation of the variables):

- House prices to income (mean: 86.86, standard deviation: 24.42)
- Current account to GDP (mean: -0.13 , standard deviation: 5.79)

- Government debt to GDP (mean: 60.96, standard deviation: 33.29)
- Debt to service ratio (mean: 22.28, standard deviation: 16.31)
- Loans to income (mean: 114.71, standard deviation: 57.86)
- Credit to GDP (mean: 127.20, standard deviation: 61.70)
- GDP growth (mean: 2.53, standard deviation: 3.26)
- Bond yield (mean: 2.93, standard deviation: 3.34)
- Credit growth (mean: 4.83, standard deviation: 5.99)
- Inflation (mean: 5.78, standard deviation: 11.08)
- House price growth (mean: 1.88, standard deviation: 8.53)
- Stock price growth (mean: 6.37, standard deviation: 28.94)
- Credit to GDP gap (mean: 3.78, standard deviation: 13.63)
- House price gap (mean: 1.30, standard deviation: 13.72)

The above measures cover asset prices (e.g., house and stock prices), leverage (e.g., mortgages, private loans and household loans), business cycle indicators (GDP and inflation), measures from the EU Macroeconomic Imbalance Procedure (e.g., current account deficits and government debt), and the banking sector (e.g., loans to deposits). In most cases, we have relied on the most commonly used transformations, such as ratios to GDP or income, growth rates, and absolute and relative deviations from a trend. A more detailed description of the data can be found in [5]

4.2 Experiment design

The numerical experiment utilizes the procedure described in the previous section. As the objective of the models in this paper is to support systemic risk analysis, we have a set-up that strictly follows the use of data in a real-time manner. The ultimate objective is two-fold: to identify systemic risk states (clusters) and their possibilistic likelihood (membership/typicality degree). To achieve this, we make use of the following analysis procedure according to the general approach described above:

- Identify optimal number of clusters and apply possibilistic clustering to pre-crisis data (i.e., solely observations representing systemic risk or vulnerability states).
- Compute typicality values for all data to the clusters, in which data distant to all cluster centroids exhibits low typicality values to all clusters.
- Identify optimal thresholds on the possibilistic likelihood with the Usefulness measure (i.e., a preference weighted average of type I and II errors).

We apply the above in multiple ways. The first differentiation is in-sample vs out-of-sample analysis. In this work, we start with in-sample analysis, and also target out-of-sample analysis by splitting the data into train and test sets at a specific year. While in-sample analysis can illustrate the usefulness of the method, relying only on this in an evaluation process can result in various drawbacks, mainly in the problem of overfitting. For this reason, it is important to employ also out-of-sample analysis, where the model built on a specific part of the dataset (training set) is evaluated on a different part (test set) not used in building the clusters. This also reflects the main intended use of the proposed method, namely using it to predict pre-crisis state for a country, as a prediction in real-time can only rely on already observed data.

4.3 Classification performance measures

The traditional way of evaluating binary classification problems is through the confusion matrix which is specified through the following four values: (i) true positive (TP), i.e. positive cases classified correctly as positive; (ii) true negative (TN), i.e. negative cases classified correctly as negative; (iii) false positive (FP), i.e. negative cases classified incorrectly as positive; (iv) false negatives (FN), i.e. positive cases classified incorrectly negative. Based on these notations, one can define type I error as $T_1 = FN/(FN + TP)$ (the share of misclassified positive cases to the total number of positive cases), and type II errors $T_2 = FP/(TN + FP)$ (the share of misclassified negative cases to the total number of negative cases). By specifying the preference between making type I and type II errors as μ and using the notations P_1 and P_2 for the probabilities of positive and negative cases, respectively, the loss function can be defined as $L(\mu) = \mu T_1 P_1 + (1 - \mu) T_2 P_2$. The absolute usefulness of a classification model can be defined by comparing the loss function to using the model of assigning every observation to the most frequent class: $U_a(\mu) = \min(\mu P_1, (1 - \mu) P_2) - L(\mu)$. Relative usefulness compares absolute usefulness with a per-

fect model (model with loss function value 0)

$$U_{rel}(\mu) = \frac{\min(\mu P_1, (1 - \mu) P_2) - L(\mu)}{\min(\mu P_1, (1 - \mu) P_2)}.$$

The optimal threshold for the classification problem is chosen as the value which results in maximal relative usefulness.

A basic measure of prediction performance is accuracy ($TP + TN / (TP + TN + FP + FN)$): the proportion of correctly classified cases. A popular performance measure is the area under the curve (*AUC*), which is based on the Receiver Operating Characteristic (*ROC*). The *ROC* curve [2] depicts the true positive and false positive rates based on the threshold chosen in case of a probabilistic classifier output to determine the output class, and *AUC* measures the area under the *ROC* curve. The maximum value of *AUC* is 1, and the closer the value is to 1, the higher the probability that the classifier assigns the right class to the data point. It is important to note that the value of *AUC* can be misleading in the case of a dataset with imbalanced classes, but still this is one of the most widely used evaluation measures.

4.4 Results

In the following, the results of the numerical experiment are presented. In [14], some preliminary results were presented using various fuzzy clustering methodologies, but with inconclusive results regarding the suitability of the approaches for different datasets. In the considered dataset, we have 238 positive (crisis) and 1129 negative observations, resulting in an unbalanced dataset with 17.4 % of positive cases. The small amount of datapoints puts a limitation on the capability of the possibilistic clustering methodology to result in very high accuracy and *AUC* values, especially when we consider out-of-sample analysis. The problem of working with imbalanced datasets in general is discussed for example in [3]. Considering this, instead of purely focusing only on overall classification accuracy, in the analysis we decided to pay specific attention to checking the effect of utilizing typicality values together with memberships in the clustering model, analyse the impact of the sample being unbalanced, and offering some basic guidelines on how to utilize the output of clustering in a more descriptive type of analysis while always paying special attention to reducing the number of false negatives, as a misclassification of a pre-crisis state is much more costly than a false alarm.

As the first exercise, we performed in-sample analysis. In this step, we utilized all the 238 positive observations to perform fuzzy clustering, then we identified the optimal number of clusters based on the usefulness measure. The number of clusters tested in the experiment ranged from

2 to 25. We have tried several combinations of model parameters, and the presented results are obtained with setting $a = 1, b = 1, m = 2$, and $\eta = 2$. Here we note that without initializing the algorithm with values from an initial run of the traditional fuzzy *c*-means, we always obtain identical cluster centers, which implies that the classification part of the approach results in either each observation classified as positive or as negative, depending on the parameter of the usefulness measure. The results of the analysis are presented in Table 1. In all the tables presenting the results, the percentage values are specified in a way that $TP + FN = TN + FP = 100\%$. As one can observe, while classification accuracy is not particularly high, the usefulness value shows that the approach offers improvements compared to the baseline model. Additionally, the number of false negatives is less than 2%, emphasizing the important result that the likelihood of missing a crisis is very low. Lastly, we can observe that by combining the information from the typicality and membership degrees as the minimum of the two values can slightly improve the performance of the algorithm.

In order to check the effect of the dataset being unbalanced, we performed an in-sample analysis focusing on the years after 2000, where the class frequencies are 237 positive and 563 negative. As we can see from Table (2), the results are better, especially considering the 50% improvement in usefulness. This results combined with the observations of higher accuracy for more balanced datasets in [14] indicate that the initial class distribution can significantly affect the quality of the output of the algorithm.

In the out-of-sample analysis, we divided the dataset into training and test set. The training set consists of the years preceding 2007, resulting in approximately 70-30% division of the positive cases into the training and test set, respectively. The results can be seen in Table (3). As one can observe, while the accuracy regarding only the positive cases decreased, which is a natural tendency in out-of-sample analysis, the usefulness and overall accuracy have improved.

To exemplify the additional use of clustering, we looked at the output of the out-of-sample analysis for two countries, Portugal and Germany, for the years 2007-2014 (31 quarters). The following discussion presents an analysis that could be performed for any of the countries present in the dataset, we chose these two cases because they reflect two specific general behaviours as described below. Both countries have been in a pre-crisis state from 2007 Q1 until 2008 Q3, and the cluster structure built based on data before 2007 correctly identified all but one (Portugal in 2007 Q1) of this observations as being in a pre-crisis state. However, when we look at the remaining datapoints, we obtain very distinct results for

	c	TN	FP	TP	FN	AUC	Acc	U_{rel}
t_{ij}	15	269 (24%)	860 (76%)	233 (98%)	5 (2%)	0.63	0.37	0.20
$\min(t_{ij}, \mu_{ij})$	13	283 (25%)	846 (75%)	233 (98%)	5 (2%)	0.65	0.38	0.21

Table 1: Results of the in-sample analysis

	c	TN	FP	TP	FN	AUC	Acc	U_{rel}
t_{ij}	15	185 (33%)	378 (67%)	230 (97%)	7 (3%)	0.67	0.52	0.29
$\min(t_{ij}, \mu_{ij})$	8	224 (40%)	339 (60%)	221 (93%)	16 (7%)	0.68	0.56	0.32

Table 2: Results of the in-sample analysis based on data after 2000

the two countries. In case of Portugal, only 5 out of the remaining 24 tranquil quarters are misclassified as being in a pre-crisis state (2008 Q4 and 2009Q2-2010Q1, all of them just after the true pre-crisis periods), while in case of Germany we find 17 false positive cases. One possible explanation for this can be found when we look at the original structure of the clusters created based on the training set to which the test set observations of these two countries are assigned. We found that pre-crisis states of Portugal were assigned to clusters which additionally contained several observations from other countries, in this sense representing a set of datapoints being in crisis independently of the country specific features. On the other hand, many pre-crisis observations from Germany belong to clusters that almost exclusively contain observations from Germany, indicating a set of observations in pre-crisis state together with the unique features of the German economy. Considering this, the misclassification in case of Germany can be mainly attributed to the tranquil states being similar to the pre-crisis states in terms of the specific attributes of Germany. At the same time in case of Portugal, the structure of clusters containing pre-crisis states of this country are not so much affected by the unique features of Portugal, but by general features of a country being in a pre-crisis state. This reasoning shows an example of discovering an underlying structure in the clusters that would not be possible with other methodologies.

5 Conclusions

In the literature, there are numerous proposals to predicting systemic risk. In this article, we outlined a possible approach using different vulnerability measures as a basis of clustering. We propose to use possibilistic fuzzy clustering as it has various benefits compared to traditional clustering approaches that better allow for describing the (dis)similarity to systemic risk states. We outlined the approach for systemic risk analysis, and offer some prelim-

inary results in a numerical experiment focusing on crisis prediction.

This study extends previous work in [13] along several directions. First, we tackle European systemic banking crises in the past decades, in contrast to their focus on excessive exchange-rate pressure in the 1990s in Asia. Additionally, rather than identifying overall financial stability states, we aim at identifying natural clusters in pre-crisis data in order to provide more descriptive representations through different states and membership degrees. The proposed approach can provide, beyond a crisis probability provided by standard early-warning models, more descriptive output that allows coupling the characterization of and membership to different systemic risk states with concrete policy actions. As the most important future research direction, this angle of the study will be extended, with a special focus of the visualization of the resulting clusters. This could offer a structured way of performing analysis similar to the one presented in the end of Section 4. Another main future goal would be to obtain a descriptive profile of various possible crisis scenarios in terms of distinct clusters and use it to obtain a more qualitative type of understanding (an example of this is the discussion on the case of Germany and Portugal). While the classification performance is sufficiently good in terms of positive cases, it is nearly not as good for tranquil states, consequently, in its form the method would not be used (and this was not our original intention either) as an early warning model per se, but as a support tool to provide a more detailed understanding on crisis states, and we can see that for those cases the model works very well. Additionally, in the future the approach can be utilized in systemic risk analytics with various extensions of the classical possibilistic clustering algorithm, incorporating various, higher-level families, of fuzzy sets.

	c	TN	FP	TP	FN	AUC	Acc	U_{rel}
t_{ij}	13	247 (67%)	124 (33%)	55 (62%)	34 (38%)	0.66	0.65	0.26
$\min(t_{ij}, \mu_{ij})$	14	251 (68%)	120 (32%)	54 (61%)	35 (39%)	0.67	0.66	0.27

Table 3: Results of the out-of-sample analysis

References

- [1] James C Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers, 1981.
- [2] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.
- [3] Nitesh V Chawla. Data mining for imbalanced datasets: An overview. In *Data mining and knowledge discovery handbook*, pages 853–867. Springer, 2005.
- [4] Christian Döring, Marie-Jeanne Lesot, and Rudolf Kruse. Data analysis with fuzzy clustering methods. *Computational Statistics & Data Analysis*, 51(1):192–214, 2006.
- [5] Markus Holopainen and Peter Sarlin. Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Bank of Finland Research Discussion Paper*, (6), 2015.
- [6] Frank Höppner, Frank Klawonn, Rudolf Kruse, and Thomas Runkler. *Fuzzy cluster analysis: methods for classification, data analysis and image recognition*. John Wiley & Sons, 1999.
- [7] Eyke Hüllermeier. Does machine learning need fuzzy logic? *Fuzzy Sets and Systems*, 281:292–299, 2015.
- [8] Gang Kou, Yi Peng, and Guoxun Wang. Evaluation of clustering algorithms for financial risk analysis using mcdm methods. *Information Sciences*, 275:1–12, 2014.
- [9] Raghuram Krishnapuram and James M Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110, 1993.
- [10] L. Laeven and F. Valencia. Banking crisis database: An update. *IMF Economic Review*, 2013.
- [11] Wei-Yang Lin, Ya-Han Hu, and Chih-Fong Tsai. Machine learning in financial crisis prediction: a survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42(4):421–436, 2012.
- [12] Shu-Ting Luo, Bor-Wen Cheng, and Chun-Hung Hsieh. Prediction model building with clustering-launched classification and support vector machines in credit scoring. *Expert Systems with Applications*, 36(4):7562–7566, 2009.
- [13] Dorina Marghescu, Peter Sarlin, and Shuhua Liu. Early-warning analysis for currency crises in emerging markets: A revisit with fuzzy clustering. *Intelligent Systems in Accounting, Finance and Management*, 17(3-4):143–165, 2010.
- [14] József Mezei and Peter Sarlin. On a generalized objective function for possibilistic fuzzy clustering. In *Proceedings of the Information Processing and Management of Uncertainty in Knowledge-based systems (IPMU2016)*, 2016.
- [15] József Mezei and Peter Sarlin. On interval-valued possibilistic clustering with a generalized objective function. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2016)*, 2016.
- [16] Nikhil R Pal, Kuhu Pal, and James C Bezdek. A mixed c-means clustering model. In *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, volume 1, pages 11–21. IEEE, 1997.
- [17] Nikhil R Pal, Kuhu Pal, James M Keller, and James C Bezdek. A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, 13(4):517–530, 2005.
- [18] Savas Papadopoulos, Pantelis Stavroulias, and Thomas Sager. Systemic early warning systems for eu15 based on the 2008 crisis. Technical report, Bank of Greece, 2016.
- [19] Pediredla Ravisankar, V Ravi, and Indranil Bose. Failure prediction of dotcom companies using neural network–genetic programming hybrids. *Information Sciences*, 180(8):1257–1267, 2010.
- [20] Thomas A Runkler and James C Bezdek. Alternating cluster estimation: a new tool for clustering

- and function approximation. *IEEE Transactions on Fuzzy Systems*, 7(4):377–393, 1999.
- [21] Enrique H Ruspini. A new approach to clustering. *Information and control*, 15(1):22–32, 1969.
- [22] P. Sarlin. On policymakers’ loss functions and the evaluation of early warning systems. *Economics Letters*, 119(1):1–7, 2013.
- [23] Peter Sarlin and Dorina Marghescu. Visual predictions of currency crises using self-organizing maps. *Intelligent Systems in Accounting, Finance and Management*, 18(1):15–38, 2011.
- [24] Cuneyt Sevim, Asil Oztekin, Ozkan Bali, Serkan Gumus, and Erkam Guresen. Developing an early warning system to predict currency crises. *European Journal of Operational Research*, 237(3):1095–1104, 2014.
- [25] Heiko Timm, Christian Borgelt, Christian Döring, and Rudolf Kruse. An extension to possibilistic fuzzy cluster analysis. *Fuzzy Sets and Systems*, 147(1):3–16, 2004.
- [26] Chih-Fong Tsai. Financial decision support using neural networks and support vector machines. *Expert Systems*, 25(4):380–393, 2008.
- [27] Weina Wang and Yunjie Zhang. On fuzzy cluster validity indices. *Fuzzy Sets and Systems*, 158(19):2095–2117, 2007.
- [28] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.
- [29] Huimin Zhao, Atish P Sinha, and Wei Ge. Effects of feature construction on classification performance: An empirical study in bank failure prediction. *Expert Systems with Applications*, 36(2):2633–2644, 2009.