2016 International Conference on Information Technology Systems and Innovation (ICITSI)
Bandung – Bali, October 24 – 27, 2016
ISBN: 978-1-5090-2449-0

A Review of Missing Values Handling Methods on Time-Series Data

Irfan Pratama[1], Adhistya Erna Permanasari[2], Igi Ardiyanto[3], Rini Indrayani[4]
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
[1]Irfan.mti15@mail.ugm.ac.id, [2]adhistya@ugm.ac.id, [3]igi@ugm.ac.id, [4]Rini.mti15@mail.ugm.ac.id

***Abstract* — Missing values becomes one of the problems that frequently occur in the data observation or data recording process. The needs of data completeness of the observation data for the uses of advanced analysis becomes important to be solved. Conventional method such as mean and mode imputation, deletion, and other methods are not good enough to handle missing values as those method can caused bias to the data. Estimation or imputation to the missing data with the values produced by some procedures or algorithms can be the best possible solution to minimized the bias effect of the conventional method of the data. So that at last, the data will be completed and ready to use for another step of analysis or data mining. In this paper, we will explain and describe several previous studies about missing values handling methods or approach on time series data. This paper also discuss some plausible option of methods to estimate missing values to be used by other researchers in this field of study. The discussion's aim is to help them to figure out what method is commonly used now along with its advantages and drawbacks.**

Keywords — missing values, estimation technique, mean imputation, deletion, time series.

I. Introduction

A pack of data or dataset can be used to obtain a certain specific information which can give a new knowledge. It can be obtained as classification, trend, pattern, etc [1]. Observed data can be obtained by several methods, such as censor record (automaticly and continously) or by frequent observation (survey, medical record, etc). Each of analysis process consists of several steps, one of them is preprocessing. Preprocessing is a step or phase to identify, selection, or problem handling of the data. Missing values handling is included in the preprocessing step. Missing values is the most common problem that occured in data that is obtained by observation and censor recording.

Missing data occured because of several problems, such as technical fault or human errors (the object of observation did not give sufficient data to the observer). Some real world cases such as industrial experiment, missing values occured because of an observation devices breakdown. While in an opinion survey or an interview, missing values caused by respondents who decline to answer or complete the survey, or even because insufficient survey observation's stuff [2,3]. Both examples stated above are trully unexpected occurrence. In other word, the data are expected to be observed (censor failure cases), yet it is missing due to the setbacks. Missing values is probably can not be evaded as it is happened unexpectedly. Time series data are become one of some data that most likely to have missing values in it.

Time series is a data that observed in regular interval of times [4]. Time series form can found as a continous data or discrete because observed in a regular interval of time (example: daily, monthly, annualy). Some examples of continous and discrete time series data are sinusoidal signal which is continous time series data, while daily stock prices, temperature data are a discrete data.

Data measurements are conducted several times with different condition, and sometimes, missing data occured due to several problem that are known as the "missingness mechanism" [5,6,7]:

1) *Missing completely at random (MCAR)*: a variable is missing completely at random if the probability of missingness is the same for all units, or in other words the is no dependencies of the missingness probability related to the variable itself [7,8,9]

2) *Missing at random (MAR)*: a variable is missing at random if the probability of missingness is depending only on available information.

3) *Not missing at random (NMAR)*: the missingness probability is depending on the variable itself.

Diffferent way of data measurement will emerging different mechanism assumption, for example: censor recording failure will be treated as a missing completely at random since it has no dependencies to the data that are missing, but in a survey, a respondent who refused to report his income report will be treated as not missing at random since the surveyors already expected that such certain data are unlikely to be obtained easily. Several methods has been introduced to solve missing values according to its missing mechanism or even a general solution for any missing values mechanism from the conventional one to the more modern way to handle missing values.

Conventional methods are method that emerged on the early stage of missing values handling study and relatively a simple yet risky methods, said to be risky because it is solving a problem but creating another problem. To overcome the additional problems and flaws caused by the conventional methods, the more modern way to handle missing values called imputation is introduced. The further explanation of those missing values handling techniques wil be provided in the next section.

Table I presents the example of missing values existence on a dataset where the "?" mark represents the missing values of the dataset.

TABLE I. DATASETS CONTAIN MISSING VALUES

	V1	V2	V3	V4	V5
Y1	0.11	?	0.42	0.25	0.12
Y2	0.12	0.23	0.25	0.12	0.53
Y3	0.13	0.25	?	0.27	0.26
Y4	0.14	0.26	?	0.23	0.15
Y5	0.14	0.62	0.62	?	0.37
Y6	0.16	?	0.36	?	0.42

Missing values can be distinguished by the two main pattern of missingness as described in Table II. The first pattern is, the data are missing monotone if we can observe a pattern among the missing values, it may be necessary to reorder or rearrange variables and/or individuals. The second pattern is the data that are missing arbritarily if the data are can not be reordered or rearranged to obtain the clear pattern of the missingness [10].

TABLE II. MISSING VALUES PATTERN [10]

Monotone

x1	x2	x3	x4
X	X	X	X
X	X	X	X
X	X	X	-
X	X	-	-
X	-	-	-

Arbritary

x1	x2	x3	x4
X	X	X	X
-	X	X	-
X	X	X	-
-	X	X	X
X	X	-	X

Table II describes missingness pattern of a dataset by its dependencies of the data structure. The existing data represented by "X" and the missing values represented by "-".

Since the missing values handling will be treated in the preprocessing step, it is important to identify what we will do to the data that contain missing values. The percentage of missingness will affect the method we use to solve it. Paper that focusing on missing values handling have their own percentage of missing values to test their methods or models. Generally, the papers uses missing values percentage of 5%-50% data [3,11]. But some certain papers uses a bit extreme percentage that ranged from 1%-80% data that contain missing values [4,12,14], and in some other paper, it starts from 10%-50% [2]. So, the significancy of the missing values percentage that are need to be treated by specific method is not clear enough. Several studies of classification, forecasting, clustering have their own claim when it comes to treating missing values by replacing it with mean or mode, or even delete the data that containing missing vlaues on their preprocessing step.

This paper will provide a review of several studies related to missing values handling methods or techniques especially on time series data. This paper consist of four sections. The first section contain explanation about related things about data and missing values. Afterwards, the second section contain some explanation about missing values handling methods from the conventional one to the modern one. The third section will contain discussion about the method that been explained in the previous section. And last section will conclude this paper including plausible option of estimation method.

II. MATERIAL AND TECHNIQUES

Missing values is not a new topic of studies in data mining or data analysis since there are so many methods, approach, and techniques are proposed by the previous researchers, from the simplest one to the complicated one and their own advantage and drawbacks. Some basic methods are proposed in [5], such as ignoring, deleting, zero or mean or mode estimation methods. These methods above have the simplicity but only effective for low percentage of missingness, but in bigger percentage of missingness, the result will affect the result of the analysis and even biased result [6]. The main disadvantage of discarding incomplete observation is the loss of efficiency and biased estimation result especially when the data missingness is systematic [6]. The quality of data mining or analysis is influenced by the quality of the data. Therefore, the data that contain missing values should be estimated to provide complete case of data to get expected result from the data [6]. Nowadays, there are many studies conducted and resulting some methods or approaches to solve missing values problem. The next section will provide an explanation about the methods and approaches to estimate missing values.

A. Conventional Method

Many methods have been invented to deal with missing values, several of them are really simple methods. Some of the methods using statistical principle as their base.

1) *Ignoring*

The first way to deal with missing values as pointed in [6] is *ignoring the missing value*, which is the simplest way to deal missing values. The missing values is completely ignored as the analysis is carry on. Though it is a simple method yet very risky if the missingness percentage of the data are big enough to disrupt the result of the analysis [5].

2) *Deletion*

As what it says, deletion method is simply delete the missing variable, or the instance of observation data to continue the analysis or data mining process [5,6].

3) *Mean/mode Imputation*

While *ignoring* and *deletion* did not give a good result of an analysis or data mining process caused by missing values (*Ignoring*) and less data to be proceed (*Deletion*), *Mean/Mode*

Imputation method comes as a solution to give a better result, it solves the missing values problem and the number of data that expected to be proceed is remain to be the same number. But the drawback of this method is the bias caused by so many values on the data are have a similar values [5,6].

B. Imputation Procedures

Imputation procedures can produce a complete sample of data, but ignores the consequences of the imputation methods side effect. Imputation or estimation is a procedures that handling missing values problem by replacing each of the missing values by some certain values (the values source is different for each methods). From the principle of statistical field, several methods have been proposed, such as *Hot and Cold Deck Imputation*, *Mean and Mode Imputation*, and *Multiple Imputation.* In this section we will give a brief explanation about these methods.

1. *Hot and Cold Deck Imputation*

In hot deck the imputation is done through replacing the missing values X by matching the values from diffferent observed data Y with similar variables X and take the values. The cold deck is in contrast from hot deck method, the imputation is done through replacing the values by the external source, like the values from the previous observation dataset with the same domain of observation. These types of method is simple, but it has a bad effect when the dataset have large amount of dataset, moreover when the assumption of missing data are MAR [14].

2. *Mean Imputation*

One of the easiest way to impute or estimate missing values to get a complete sample is replacing each of the missing values with the mean of the observed data for that variable, or known as unconditional mean imputation. Aside of mean imputation, there is median and mode that can used for replacing missing values [14].

Kevin strike et al. [15] did an evaluation of three missing data handling techniques: listwise deletion, mean imputation, hot deck imptutation for the certain purpose. Listwise deletion (one of deletion method) can handle missing values for below 15%, however its accuracy is different to the increasing number of missingness percentage. Euclidean distance and z-score standarization are employed by hot deck method and proves to provide a consistent and accurate compared to the other methods. But the three methods are less effective for non-ignorable missing values compared to MCAR and MAR [14].

3. *Multiple Imputation*

Other statistical method of missing values imputation is multiple imputation which proposed by rubin [5]. In this method missing values imputed n-times to represent the uncertainty of possibile values that are to be imputed. The n-times values then analyzed to get a single combined estimates [14].

The result that produced by statistical method is somewhat questionable because just effective in small amount percentage of the missing values. For example for mean imputation, the data that is imputed by this method will suffer from high bias because the newly imputed data are the same with the mean of the observed data.

To overcome the bias problem caused by the statistical method result of missing values estimation mentioned above, several methods have been proposed in the previous studies that concerning in missing values estimation. The method is adapted from the other field of research to be able to solve problem in this research field. The latest technique than the statistical one is soft computing techniques that also including method which is need learning phase to be able to estimate missing values.

1) *Autoregressive*

Sridevi et al. [11] proposed a method that focuses on autoregressive-model-based (ARLSimpute) to estimate missing values. This method is claimed to be able to handle missing values where a particular time point contains many missing values or where the entire point time is missing. The mechanism of the method are, find the K-similar data from the dataset, then calculate the AR Coefficients for the K-similar data and last, estimating the missing values. If the result coverged then the result will evaluated using Normalized Root Mean Square Error (NRMSE), if not coverged, the step rollback to find K-similar data. This method then compared with KNNimpute method by measuring the NRMSE result. The NRMSE result of the ARLSimpute and KNN Impute shown that ARLSimpute performed better than KNN Impute generally. However, ARLSimpute performing better in lower order because the error rate is increasing as the order increased.

2) *Genetic Algorithm Optimization Based*

Genetic algorithm (GA) have been widely applied as a global search in various domains, data mining problems and optimization problems are among the domain. The main concerns of choosing GA as a solution for the problem are: (1) GAs explore large search spaces while exploit optimal solution; (2) the GAs paradigm is scalable and can be effectively parallelized; (3) GA is relatively easy to implement and adapt for different domains [16].

Lobato et al. [16] introduced a multi-objective genetic algorithm method based on NSGA-II called Multi-objective genetic algorithm imputation (MOGAImp). In many problems, GA operators is very simple and have limited impact on processing time. However, the fitness calculation may demand high computation time. To solve the problem, paralell approach also proposed to reduce processing time without interfering with other algorithm properties. Therefore, the fitness processing will not take long time to done since each individual is assigned to one thread and done at the same time.

Tang et al. [12] proposed GA as an optimization method for the fuzzy c-means. FCM as a clustering method needs to be optimized to get the best result from the clustering process. GA was choosen as the optimization method because of its advantages above. GA used on the study to make sure the optimum solution (minimum errors) is obtained, the the missing values is estimated.

Aydilek et al. [6] introduced a hybrid method that combine FCM, support vector regression (SVR) and GA. Similar to what Tang et al.[12] proposed, but the estimation mechanism is done through machine learning as SVR is used. The purpose of SVR and GA usage is to minimize the error after the estimation of FCM is done. The mechanism of the method is as follows : (1) SVR algorithm is trained using the complete dataset rows; (2) the incomplete dataset rows estimated using FCM, and compare the FCM output and the SVR output vector; (3) GA works on the minimizing the difference between FCM and SVR output and obtained optimized *c* and *m* parameter; (4) estimate the missing values using optimized parameters using FCM.

Azadeh et al. [17] conducted a study to test some estimation methods to be applied in randomized complete block design table. One of the method is GA, GA run over 30 times of iteration to make a complete table and claimed to be better than other method that are compared in the study. The correlation between the estimated data and the actual data emerged GA and ANN as the best method and been encouraged to uses GA approach for setimating missing values in RCBD tables.

3) Support Vector Machine

Wu et al. [3] proposed a method to handling missing values namely least square support vector machine, the basic idea is "Local Time Index" (LTI) which ignores the missing values and adopts temporal information in addition to put in training patterns. Any missing values methods that basicaly manipulating the existing time-series data may cause another problem. The LTI idea is ignores the missing values yet it want to keep the continuity of the series itself. The solution is providing temporal information to replace the missing values, the temporal information of a pattern somehow can be obtained through some forecasting model. That hipothesis is the basis of this method. At the comparison section, this method is compared to several conventional method such as, hot-decking, mean, and Sridevi et al. [11] proposed method's AR. This method is shows a better result than those methods.

4) Interpolation

Shao et al. [18] conducted a study to handle missing values by introducing an interpolation method based on NURBS geometry modeling. The study is discussed about interpolation method using window interpolation adjustment based on time series features in assumption that the most time series of the system have certain cyclical characteristic, which reflects on seasonal volatility. The window interpolation adjustment is based on such periodic correlation sequence. This method then evaluated using two evaluation index NRMSE (normalized root mean square error) and COEF (Pearson's product-moment correlation coefficient) and the result outperformed Sbspline as comparison. The Snurbs interpolation combined with window adjustment is superior than traditional interpolation method. The Snurbs introduction on the interpolation give a flexibility and provide unified mathematica expression for time series, and the window adjustment improves the initial interpolation result effectively.

Sree dhevi [14] proposed an interpolation method namely inverse distance weigted(IDW). By this method, the missing values can be predicted using the measured data, the influence of the distance is affected by the distance of missing values and known values sample time, the weight increases as the sample time distance of the data that is missing is closer to the known value and vice versa. IDW then evaluated by compare it with mean imputation and see the RMSE score result, the performance shown that IDW outperformed mean imputation by big margin.

5) Maximum Likehood

Banbura et al. [19] conducted a study of missing values imputation on arbitrary pattern of missingness data. The proposed method is maximum likehood (EM) based method as a general method of missing values imputation method. The steps on the proposed method are fill in the missing data in the like of expectation step, and re-optimize the expectation step in the maximization step. On certain regularity condition the EM algorithm converges toward the local optimum. The advantage of the maximum likehood is the method allow us to impose the restrictions on the parameter in a relatively straightforward manner. This methodology is well performed on the large fraction of missing data and small sample.

6) Fuzzy-rough set

Amiri et al. [1] intoduced fuzzy-rough methods to impute missing values on time series data. The main reason to use this method is that it give an exellent framework to deal with uncertainty that is often occuring in imputation problem. It simply calculate the fuzzy similarities of instances and can work with noise and can deal with missing data. The method of fuzzy-rough used as lower and upper approximators, and the prediction work is given to KNN algorithm. The algorithm is work as follows: for every instances of the dataset "*a*" that containing at least at least one missing values for each feature "*b*", the algorithm find its KNN and puts them in set "*X*". The approximation done by using *a*'s nearest neighbor. This method outperformed other variant in the same basis,ordered weighted average-based fuzzy-rough sets with nearest neighbor (OWANNI) and vaguely quantified nearest neighbor imputation (VQNNI) in missing values estimation domain.

7) Similarity Measure

Sitaram et al. [4] introducted a method of time series missing values estimation based on Mahalanobis distance that is known as a method for handwritten character recognition. It is also used widely in data mining domain such as classification and clustering techniques. The method developed in two phases. In the first phase, the candidate with missing values and the query time series is assumed to be exact matched. The result is formed as best match and exact match for each algorithm. The mahalanobis shown the same results as exact match with a good number missing values that implying the reasonable way to extend the algorithm for missing values. In the second phase, extend the algorithm to also acomodate warped time series data and proposed and algorithm that handles warped data using the same way as dynamic time warping (DTW) did. The result from this methods is a complete dataset without performing any imputation in the candidate to handle missing values. It is just simply replace the missing series by the one that best matched.

III. Discussion

Estimation techniques is required to handle dataset that have non-ignorable missing data and large amount of missing values. Estimation techniques that have been used by previous researchers shown good performance to the method that are compared to. From the literature review, we can summarize the papers based on method, and dataset as described in Table III.

TABLE III. List of time series missing values papers

Estimation Techniques			
Paper	*year*	*Method*	*dataset*
Sridevi et al.[11]	2011	Auroregressive-model-based missing values estimaton method	Stock UK statistic Sales Weather
Aydilek et al.[6]	2013	Optimized fuzzy c-means with support vector regression and genetic algorithm	Glass Haberman Iris Musk1 Wine Yeast
Azadeh et al.[17]	2013	Genetic algorithm	-
Sree dhevi [14]	2014	Inverse distance weighted	Istanbul stock exchange
Banbura et al.[19]	2014	Maximum likehood	-
Shao et al.[18]	2014	Combination of Snurbs with window interpolation adjusment	-
Wu et al.[3]	2015	Least squares support vector machine	Sine function Sinc function Poland electricity load Sunspot Jenkins-box gas furnace
Tang et al.[12]	2015	Fuzzy c-means and genetic algorithm method integration	Traffic volume
Lobato et al.[16]	2015	Multi-objective genetic algorithm	Public benchmarking dataset
Sitaram et al.[4]	2015	Mahalanobis distance	-
Amiri et al.[1]	2016	Fuzzy-rough set	-

In most paper, the accuracy measurement is done by using NRMSE or RMSE method because they use comparison as their way to measure the method's performance from the other method since NRMSE and RMSE score represent the accuracy of the method. In terms of simplicity, Sitaram et al. [4] provides the most simple way to complete the dataset by just replacing the missing values with a part of another values in the closest dataset without any imputation or estimation process. It means less mathematic equation. However, the method is fully based on threshold on matching process, in case of the threshold can not met, the missing values can not be solved. In several time series data there are anomaly conditions or the deviation between periods is high. So, the condition when threshold is can not fulfilled is very possible. The possible solution is decreasing the threshold, but it may reduce the quality of the method. Mathematical approach is apparently better that the previous approach, because the parameter is can be vary depend on the method itself. Sridevi et al. [11] presented probably the most basic and simple estimation method that are based on autoregressive model. The method is claimed to be able to solve missing values at several time points or columns of the dataset. To prove the method is doing well, it is compared to KNN based estimation method by the score of NRMSE. The method is better of KNN based imputation method in terms of handling such type of missing values. From the papers above, the most used method by previous researchers in interpolation and genetic algorithm combination method [6,12,14,16,17,18]. The main reason to choose those methods is their easy to use mechanism and the flexibility of manipulating the step. The other reason is mainly because these method is can be easily implemented on various domain.

From the methods explained, genetic algorithm and the combination that include GA [6,12,16,17] and interpolation methods [14,18] are mostly used to solve missing values problems. But the most frequently used method can not guarantee that those are the best method for general missing values problems, since some problem in some certain dataset have its own characteristic.

In terms of flexibility, genetic algorithm [6,12,16,17], fuzzy c-means [12,16], and autoregressive [11] are considered as the best among the menthods explained above. Beside those methods can be manipulated according to the needs of the user, those method can combined with another complementary method.

IV. Conclusion

Missing values can be treated in various way depends on the problem caused by the missing values. The method of missing values can affect the result of the analysis or data mining process. Estimation technique is probably the best option of missing values handling, since to accomplish certain work the complete dataset is required and some dataset have dependend variable which is imposssible to delete the missing values as it can disrupt the data itself. Although the recent methods performance is good, the better methods to cut out the time complexity or demanding computational resource that required by certain method.

This paper figures missing values handling method from the conventional one to the modern one. Methods such as deletion, mean imputation, and hot decking are considered as conventional method which can be used for more general dataset, while estimation techniques is modern method that specifically picked to handle missing values in time series data.

This overview may give researcher *state of the art* about missing values estimation which is performed by other researcher in various methods. This overview may also provide a view of each method for what its advantage and disadvantage to take into consideration of future research in this field of study.

References

[1] M. Amiri and R. Jensen, "Missing data imputation using fuzzy-rough methods," *Neurocomputing*, pp. 1–13, 2016.

[2] C. Yozgatligil, S. Aslan, C. Iyigun, and I. Batmaz, “Comparison of missing value imputation methods in time series: The case of Turkish meteorological data,” *Theor. Appl. Climatol.*, vol. 112, no. 1–2, pp. 143–167, 2013.

[3] S. Wu, C. Chang, and S. Lee, “Time Series Forecasting with Missing Values,” *2015 1st Int. Conf. Ind. Networks Intell. Syst.*, pp. 151–156, 2015.

[4] D. Sitaram, A. Dalwani, A. Narang, M. Das, and P. Auradkar, “A Measure of Similarity of Time Series Containing Missing Data Using the Mahalanobis Distance,” *Proc. - 2015 2nd IEEE Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2015*, pp. 622–627, 2015.

[5] D. Little, RJA and Rublin, “Statistical Analysis with Missing Data,” *Wiley, New York.*, p. 381, 1987.

[6] I. B. Aydilek and A. Arslan, “A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm,” *Inf. Sci. (Ny).*, vol. 233, pp. 25–35, 2013.

[7] Y. S. Afrianti, “Imputation Algorithm Based on Copula for Missing,” pp. 252–257, 2014.

[8] A. Briggs, T. Clark, J. Wolstenholme, and P. Clarke, “Missing... presumed at random: cost-analysis of incomplete data.,” *Health Econ.*, vol. 12, no. 5, pp. 377–92, May 2003.

[9] J. Sensch, “Book Reviews,” *Br. J. Math. Stat. Psychol.*, vol. 55, no. 1, pp. 193–196, 2002.

[10] Y. Yuan, “Multiple imputation for missing data: Concepts and new development,” *SAS Inst. Inc, Rockville, MD*, pp. 1–13, 2010.

[11] S. Sridevi, S. Rajaram, C. Parthiban, S. SibiArasan, and C. Swadhikar, “Imputation for the analysis of missing values and prediction of time series data,” *2011 Int. Conf. Recent Trends Inf. Technol.*, pp. 1158–1163, 2011.

[12] J. Tang, G. Zhang, Y. Wang, H. Wang, and F. Liu, “A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation,” *Transp. Res. Part C Emerg. Technol.*, vol. 51, pp. 29–40, 2015.

[13] X. Zhu, S. Zhang, Z. Jin, Z. Zhang, and Z. Xu, “Missing value estimation for mixed-attribute data sets,” *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 1, pp. 110–121, 2011.

[14] A. T. Sree Dhevi, “Imputing missing values using Inverse Distance Weighted Interpolation for time series data,” *6th Int. Conf. Adv. Comput. ICoAC 2014*, pp. 255–259, 2015.

[15] K. Strike, K. El Emam, and N. Madhavji, “Software Cost Estimation with Incomplete Data,” *IEEE Trans. Softw. Eng.*, vol. 27, no. 10, pp. 890–908, 2001.

[16] F. Lobato, C. Sales, I. Araujo, V. Tadaiesky, L. Dias, L. Ramos, and A. Santana, “Multi-objective genetic algorithm for missing data imputation,” *Pattern Recognit. Lett.*, vol. 68, pp. 126–131, 2015.

[17] A. Azadeh, S. M. Asadzadeh, R. Jafari-Marandi, S. Nazari-Shirkouhi, G. Baharian Khoshkhou, S. Talebi, and A. Naghavi, “Optimum estimation of missing values in randomized complete block design by genetic algorithm,” *Knowledge-Based Syst.*, vol. 37, pp. 37–47, 2013.

[18] C. Shao, “An interpolation method combining Snurbs with window interpolation adjustment,” *2014 4th IEEE Int. Conf. Inf. Sci. Technol.*, pp. 176–179, 2014.

[19] M. Bańbura and M. Modugno, “MAXIMUM LIKELIHOOD ESTIMATION OF FACTOR MODELS ON DATASETS WITH ARBITRARY PATTERN OF MISSING DATA,” *J. Appl. Econom.*, vol. 29, no. 1, pp. 133–160, 2014.