Effects of the pay-out system of income taxes to municipalities in Germany

Michael C. Thrun[1], Alfred Ultsch[2]

Abstract

The methods and possibilities of data mining for knowledge discovery in economic data are demonstrated on data of the German system of allocating tax revenues to municipalities. This system is complex and not easily understandable due to the involvement of several layers of administration and legislation. The general aim of the system is that a share of income tax revenue for a municipality (system output) should be a fixed part of the total income tax yield of each municipality (system input). Tools for the scientific exploration of empirical distributions are applied to the input and output variables of the system. The main finding is that, although the critical input variables are unimodally distributed, the output variable showed a bimodal distribution. The conclusion from this finding is that the system works in two distinct states: municipalities receive either a large share or a small share of the input. Relating these states of the system to the location of the municipality a distinct east-west gradient is found. A significantly larger percentage of East-German municipalities receive in average 5% less of the taxes share. This paper focuses on the methods and tools for such type of knowledge discovery.

Keywords: *machine learning, Gaussian mixture models, density estimation, statistical analysis, income tax*

JEL Classification: C310, C380, P170

DOI: 10.14659/SEMF.2018.01.54

1 Introduction

Federal legislation determines the basic rules on how to calculate the income tax share a municipality receives (Ultsch and Behnisch, 2017). A fixed percentage of the total income tax a municipality pays to the state is restituted back to this municipality (details in Ultsch and Behnisch, 2017). The actual share of income tax is determined by the state in which the municipality is located by allocation keys which are used to specify this percentage (Ultsch and Behnisch, 2017). The overall effects of this system is investigated for all municipalities in Germany in the form of an input-output analysis by investigating distributions of extracted features and Bayesian classification of two-dimensional density.

Descriptions of distributions using a single distribution, like Lognormal or Gamma are often quite weak in describing the tails of the distribution (Dagum, 1977). They lead to separate models for the upper vs. lower parts of income distributions (c.f. Thrun and Ultsch,

[1] Corresponding author: University of Marburg, Data Bionics Research Group, Hans-Meerwein-Straße 6, D-35032 Marburg, mthrun@mathematik.uni-marburg.de.

[2] University of Marburg, Data Bionics Research Group, Hans-Meerwein-Straße 6, D-35032 Marburg, ultsch@mathematik.uni-marburg.de.

2015). These approaches can be improved and simplified using Gaussian mixture models (Thrun and Ultsch, 2015).

This approach is applied to a one- and two-dimensional analysis of the German pay-out system. This paper focuses on methods and tools. Spatial planning results were published primarily in (Ultsch and Behnisch, 2017).

2 An Approach for Distribution Analysis

A scientifically sound procedure for the identification and analysis of empirical distributions is a comparison to a known theoretic distribution. The quantile/quantile plot (QQ-plot) allows comparing an empirical distribution to a known distribution (Michael, 1983). Here, in 100 quantiles the model of a Gaussian distribution is compared to the data, and a straight line confirms a good data fit of the model. The Gaussian distribution is the canonical starting point for such a comparison. If the distribution is not Gaussian, the box-cox transformation can be applied. It aims to construct a nonlinear transformation of a variable x such that the transformed variable comes as close as possible to a Gaussian distribution (Asar et al., 2014; Box and Cox, 1964). However, the estimated powers of the box-cox transformation are seldom integers. The box-cox values may be orderly arranged by Tuckey's ladder of power relative to the nearest whole number to get understandable nonlinear transformations of a variable (Tukey, 1977).

In the case of this work, the precise form, i.e., the type, nature and parameters of the formal model of the probability density function (pdf) is the ultimate goal of this analysis. Usually, this is performed using kernel density estimators. The simplest of such a density estimation is the histogram. However, histograms are often misleading and require critical parameters such as the width of the bin (Keating and Scott, 1999). A specially designed density estimation, which has been successfully proved in many practical applications is the "Pareto Density Estimation" (PDE). PDE consists of a kernel density estimator representing the relative likelihood of a given continuous random data (Ultsch, 2005). PDE has been shown to be particularly suitable for the discovery of structures in continuous data hinting at the presence of distinct groups of data and particularly suitable for the discovery of mixtures of Gaussians (Ultsch, 2005). The parameters of the kernels are auto-adopted to the date using an information theoretic optimum on skewed distributions (Ultsch et al., 2015). PDE can be used as an effective method to estimate density in two dimensions (Ultsch, 2005) (scatter density plots) showing the relationships and dependencies between two features. In general, bivariate relations are visualized using scatterplots where two features are plotted against each other as points. In a scatter density plot the densities of these points can be estimated and visualized (c.f. Berthold et al., 2010, p. 45).

If the distribution of data is more complicated than a single distribution, the underlying process which produces the data may operate in different states, which have their factors to influence the data. Then mixture models can be applied as the standard statistical tool (Fraley and Raftery, 2002; Ultsch et al., 2015). Gaussian mixture models (GMM) are a superposition of a weighted sum of single modes. Each mode (component) consists of a single Gaussian parameterized with the mean and standard deviation (Bishop, 2006). The GMM is a weighted sum of *M* component Gaussian densities as given by the equation

$$p(\vec{r}) = \sum_{i=0}^{M} w_i\, N(\vec{r}|\vec{m}_i, \vec{s}_i) = \sum_{i=1}^{M} w_i \cdot \frac{1}{\sqrt{2\pi s_i}} \cdot e^{-\frac{(\vec{r}-\vec{m}_i)^2}{2*\vec{s}_i^2}} \tag{1}$$

where $N(\vec{r}|\vec{m}_i, \vec{s}_i)$ denotes the two-dimensional Gaussian probability densities (component, mode) with means $\vec{m}_i$ and standard deviations, $\vec{s}_i$. The w_i are the mixture weights indicating the relative contribution of each component Gaussian to the overall distribution, which add up to a value of 1. *M* denotes the number of components in the mixture. Usually, the parameters of the GMM, including the number of components M, are optimized using the expectation maximization (EM) algorithm (Dempster et al., 1977). A GMM represents the presence of subclasses within a complete data set. Precise limits for these classes can be calculated using the theorem of Bayes (Duda et al., 2001). A GMM can be visually verified by a QQ-plot and statistically by Xi-Quadrat test (cf. Thrun and Ultsch, 2015).

Datasets

The dataset of (Ultsch and Behnisch, 2017) has been used. In 2010 there were 11,669 municipalities, and 228 so-called “unincorporated areas” generally forested areas, lakes and larger rivers, of 16 states resulting in thirteen territorial states: Schleswig-Holstein and Hamburg (1,117), Lower Saxony and Bremen (1,026), North Rhine-Westphalia (396), Hessen (426), Rhineland-Palatinate (2,306), Baden-Wuerttemberg (1,101), Bavaria (2,056), Saarland (52), Brandenburg and Berlin (420), Mecklenburg-Vorpommern (814), Saxony (485), Saxony-Anhalt (300) and Thuringia (942). The number of taxpayers per municipality has been given in this data set.

Data on the income tax per taxpayer collected by each municipality from its population and transferred to the state (Municipality Income Tax Yield, MTY) has not directly been available. To estimate MTY, Tax2007 and Tax2010 of the Regional Database Germany (see Ultsch and Behnisch, 2017) has been used. The sum of the income tax revenues paid by the state to a municipality has been obtained from Regional Database Germany (see Ultsch and Behnisch, 2017). The income tax share (**ITS**) of a municipality per taxpayer in 2010 has been calculated as TaxShare divided by the number of taxpayers in the year 2010.

3 Results

MTY can be estimated through a scatter density plot between Tax2007 and Tax2010. It is presented in Fig. 1 that the data on the wage and income statistics after-tax return for the year 2007 (Tax2007) as well as for the year 2010 (Tax2010) allows a reasonable estimation of MTY for 2010. The true MTY value is located somewhere between Tax2007 and Tax2010. The Pearson Correlation between the two variables exceeded 92% in a range of up to 10,000 EUR. Therefore, Tax2010 divided by the number of taxpayers in a municipality is taken as a proper estimation for MTY in 2010. However, MTY could not be obtained for n=247 municipalities due to restrictions on data protection and problems of data availability. These municipalities are disregarded in the following distribution analysis. In Fig. 2 and 3 the distributions of ITS and MTY are analyzed. MTY is unimodally distributed, but ITS is multimodal distributed. The first maximum of ITS lies at 300 EUR and the second at 640 EUR.

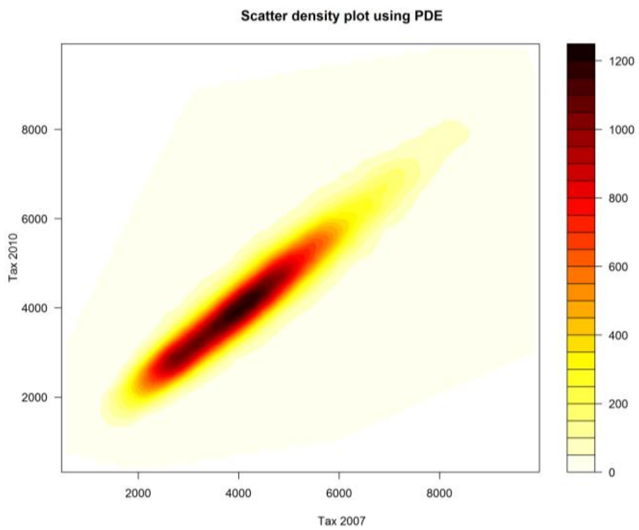

Fig. 1. Scatter density plot of Tax 2010 versus Tax 2007 shows only one mode with a high correlation verified by the Pearson correlation value of 92%.

In the next step, the scatter density plot between the input MTY and the output ITS is presented in Fig. 4. In Contrast to Fig. 1 two modes are visualized by using the PDE. Therefore, a two-dimensional GMM is calculated, and from this GMM a Bayesian classification is derived. The result of the classification is presented in a simple scatter plot in Fig. 5. It is apparently visible that in the scatter plot the two modes are not recognizable if the PDE is disregarded.

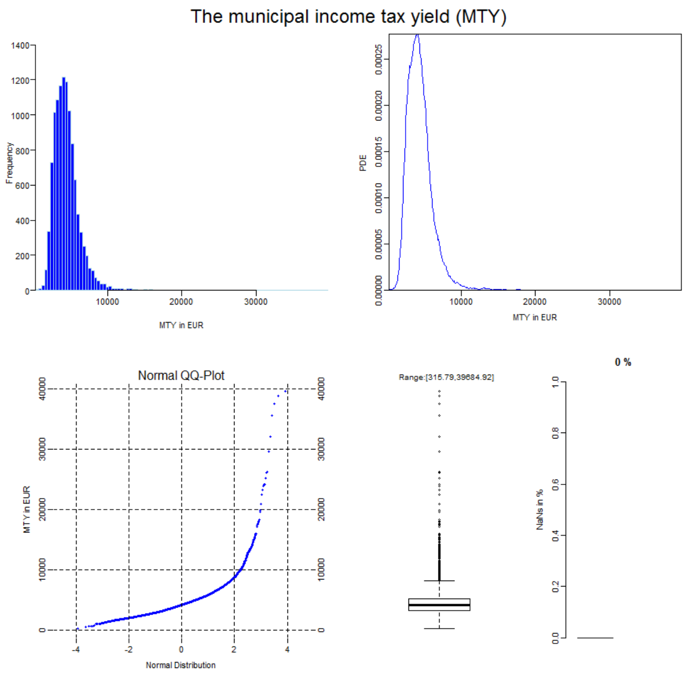

Fig. 2. The distribution of MTY is unimodal shown in the histogram and PDE plot. However, it is not Gaussian because the QQ-plot does not have a straight line. The box plot shows some outliers, and the feature has no missing values.

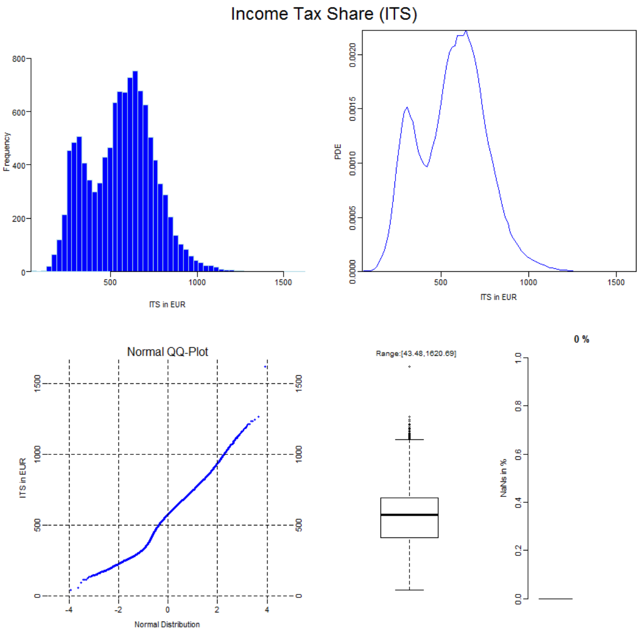

Fig. 3. The Distribution of ITS is multimodal indicated by an s-curve in the QQ plot additionally to the two modes in the histogram and PDE plot. The boxplot shows one large outlier.

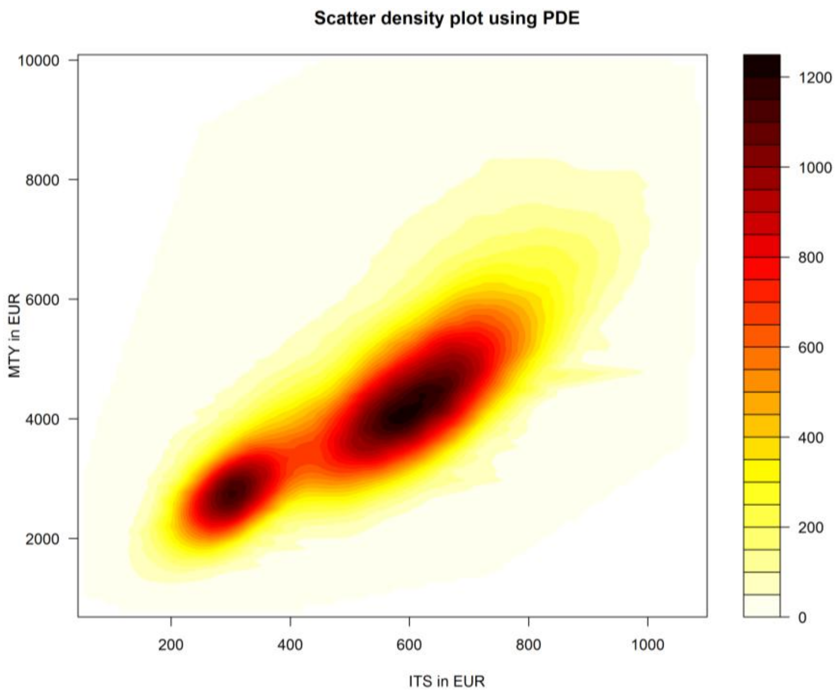

Fig. 4. Scatter density plot using Pareto density estimation of MTY versus ITS shows two distinctive modes.

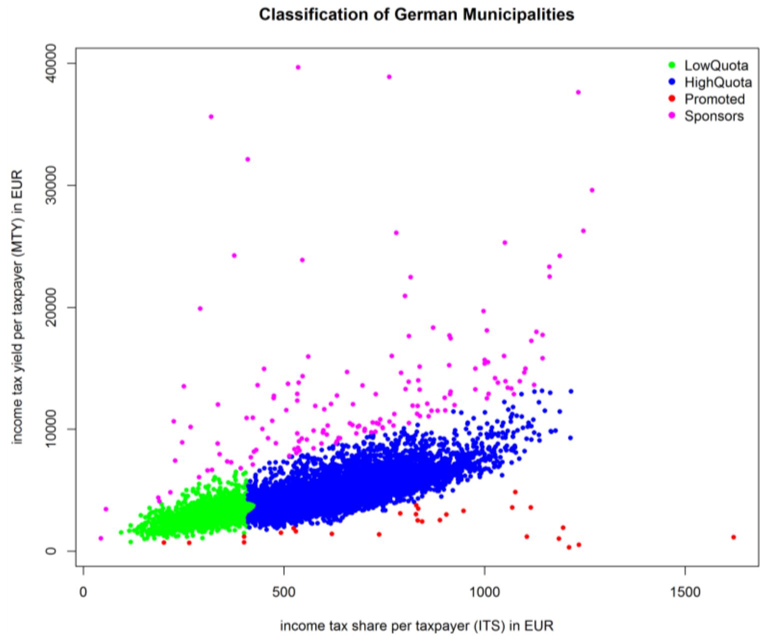

Fig. 5. A scatterplot of MTS versus ITS colored by the labels of a Bayesian classification showing two main groups of low quota and high quota municipalities. Additionally, outliers are manually classified into two separated groups called sponsors and promoted.

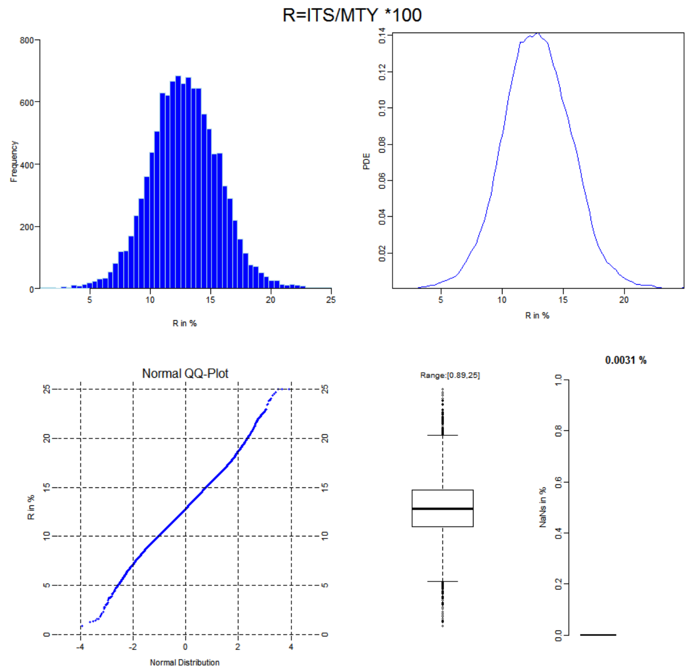

Fig. 6. The ratio R= ITS/MTY*100 is Gaussian distributed. To better visualize the relevant range, values above 25 % as manually set as missing values (0.0031% of data).

Fig. 6 presents the ratio R = ITS/MTY*100 of the payments the municipalities received from the respective state governments (ITS) to the income taxes yield the municipality transferred from its taxpayers to the state (MTY). The units are percentages [%]. The amount of the collected tax yields described by the ratio has a mean of 13% ± 3%. Accordingly, a German municipality expects to receive 13% funding from the state on average (Fig. 6). Hence, the ratio R allows to describe the four classes as deviations from the expectation (see Fig. 5): high-quota, low-quota, sponsors and promoted.

4 Discussion

Ideally, the system input and output should be proportional, i.e., the income tax share per taxpayer (ITS) should be proportional to the municipal income tax yield (MTY). The fact that this ratio is Gaussian distributed implies that there should be only unintentional deviations from this general percentage. Therefore, the payment of income tax revenues should result in a direct proportionality between tax payments to and funding from the state government. A municipality should expect a certain fixed percentage of the taxes it delivers (Fig. 6). If the deviations from the base percentage ITS are as unintentional, the distribution of ITS should be unimodal. Instead, distribution analysis reveales that the pdf of ITS consists of two modes. Thus, the payout system of income taxes discriminates between low quota and high quota classes of municipalities.

Next, the input of the system (MTY) was connected to the output of the system (ITS) in order to understand this effect of inequity. A clear separation into two distinct distributions can be observed in Fig. 4. For a vast majority of German municipalities that pay income tax per taxpayer of approx. 2,500 to 4,500 EUR to the state, the refund can be either low (~ 10%) or high (~ 20%). Two 2-dimensional Gaussian mixtures could efficiently model the two modus operandi of the pay-out system of income taxes to municipalities. Using the deviation from the expected ratio, the groups of the classification are distinguished as low quota vs. high quota classes. This classification is expanded to include a class of “promoted”-municipalities receiving a substantially larger share of income taxes (above 30%) and a class of “sponsors”-municipalities receiving a considerably smaller share of income taxes (less than 8%).

Conclusions

This work shows that distribution analysis of features itself allows generating insights besides being the prerequisite for cluster analysis (Thrun, 2018, pp. 16-17). Further, the correct estimation of density is crucial for estimating the pdf and improves a scatterplot significantly. The methods are exemplarily applied to Germany’s complex system of allocating tax revenues. The geographical distribution of the low-quota vs. high-quota municipalities reveals an evident east-west disparity (cf. Ultsch and Behnisch, 2017, p. 29, Fig. 9). The percentage of income tax revenues which a municipality received per taxpayer depends on the location of the municipality. If the municipality is located in western Germany, the municipality could expect about 15-30%. On the contrary, if the municipality is located towards the east, its share is more likely to be only 10% or less. All visualizations are available in the CRAN package “DataVisualizations” in R. The German Research Foundation partly funded the research under grant agreement (BE4234/3-1, UL159/10-1). The DFG had no role in the study design, data selection, and analysis, the decision to publish or preparation of the manuscript.

References

Asar, Ö., Ilk, O. & Dag, O. (2014). Estimating Box-Cox power transformation parameter via goodness of fit tests. *Communications in Statistics-Simulation and Computation, 46*(1), 91-105.

Berthold, M. R., Borgelt, C., Höppner, F. & Klawonn, F. (2010). *Guide to intelligent data analysis: how to intelligently make sense of real data*: Springer Science & Business Media.

Bishop, C. (2006). *Pattern recognition and machine learning*. New York: Springer.

Box, G. E. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211-252.

Dagum, C. (1977). New model of personal income-distribution-specification and estimation. *Economie appliquée, 30*(3), 413-437.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B, 39*(1), 1-38.

Duda, R. O., Hart, P. E. & Stork, D. G. (2001). *Pattern Classification. 2nd. Edition.* New York.

Elsner, H. (1979). *Das Gemeindefinanzsystem: Geschichte, Ideen, Grundlagen*. Köln: Kohlhammer Verlag.

Fraley, C. & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of American Statistical Association, 97*(611-631).

Keating, J. P. & Scott, D. W. (1999). A primer on density estimation for the great homerun race of 1998. *STATS, 25*, 16-22.

Michael, J. R. (1983). The stabilized probability plot. *Biometrika, 70*(1), 11-17.

Thrun, M. C. (2018). *Projection Based Clustering through Self-Organization and Swarm Intelligence* (A. Ultsch & E. Hüllermeier Eds.). DOI: 10.1007/978-3-658-20540-9, Heidelberg: Springer.

Thrun, M. C. & Ultsch, A. (2015). *Models of Income Distributions for Knowledge Discovery*. Paper presented at the European Conference on Data Analysis, DOI: 10.13140/RG.2.1.4463.0244, Colchester.

Tukey, J. W. (1977). *Exploratory data analysis* (1st ed.). ISBN: 978-0201076165, Pearson Education (US), United States.

Ultsch, A. (2005). Pareto density estimation: A density estimation for knowledge discovery. In: Baier, D. & Werrnecke, K. D. (Eds.), *Innovations in classification, data science, and information systems* (Vol. 27, pp. 91-100). Berlin, Germany: Springer.

Ultsch, A. & Behnisch, M. (2017). Effects of the payout system of income taxes to municipalities in Germany. *Applied Geography, 81*, 21-31.

Ultsch, A., Thrun, M. C., Hansen-Goos, O. & Lötsch, J. (2015). Identification of Molecular Fingerprints in Human Heat Pain Thresholds by Use of an Interactive Mixture Model R Toolbox (AdaptGauss). *International journal of molecular sciences, 16*(10), 25897-25911.