Classifying Countries in Terms of Government Expenditure: A Multi-criteria Approach[1]

Lich Khac Hoang[1], Moumita Chatterjee[2], Kien The Nguyen[3], Duong Anh Nguyen[4]

[1] VNU University of Economics and Business,
144 Xuan Thuy, Cau Giay, Hanoi 10000, Vietnam.
Corresponding author. E-mail: hoangkhaclich@gmail.com

[2] Aliah University,
IIA/27, New Town, Kolkata, India.
E-mail: mcmoumita8@gmail.com

[3] VNU University of Economics and Business,
144 Xuan Thuy, Cau Giay, Hanoi 10000, Vietnam.
E-mail: nguyenthekien@vnu.edu.vn

[4] Central Institute for Economic Management,
68 Phan Dinh Phung, Ba Dinh, Hanoi 10000, Vietnam.
E-mail: anhduong510@yahoo.com

This present article contributes to the literature on country classification by investigating the pattern of government expenditure across countries. Here we adopt a k-means clustering technique on indicators reflecting the actual situation of go-vernment expenditures. The variables used for the classification include GNI per capita, government effectiveness, subsidies, and other transfers, compensation of employees, goods and services expense. After implementing the unsupervised classification technique, four clusters of countries were selected. We have identified variables that vary less within a particular cluster, though the variation is higher between clusters. Our algorithm allows dropping those variables which are less

Lich Khac Hoang – PhD in Economics.
Moumita Chatterjee – PhD in Statistics.
Kien The Nguyen – PhD in Economics.
Duong Anh Nguyen – Senior Researcher.

relevant for identifying the clusters. By allowing for multi-criteria similarities in the classification, the possibility of sourcing differences in government expenditure and economic growth across countries could be enhanced. Also, our proposed method helps to study the nature and characteristics of each cluster and thereby describe the clusters in terms of the variables involved. The homogeneous clusters of countries that are being observed are likely to help future public expenditure-related research. Our study can help establish a relationship between government expenditure and economic growth in the future.

1. Introduction

Multidimensional clustering is a widely used clustering technique in various spheres of multivariate data analysis. In the present article, we will use the technique to understand and evaluate different countries all over the world, based on their expenditure structure. In the future, it will help us to establish a relationship between government expenditure and economic growth. In the beginning, let us consider the government expenditure structure of Sweden. [Bergh, Henrekson, 2011] argue that Sweden has high degrees of openness and social trust that mitigate the negative effect from its relatively large government and ultimately help it maintain above-average growth during the latter half of the twentieth century. Hence, without controlling openness and social trust, a typical regression analysis may come up with the wrong conclusion that the impact of government expenditure on economic growth in Sweden is the same as that in other high-income countries in OECD. To address this, we can either (i) add controlled variables into the regression model or (ii) implement regression analysis for countries that are similar to Sweden in terms of openness, social trust, and the other features of government expenditure. According to the second option, before implementing regression analysis, we must classify several countries using the variables that measure expenditures to find out countries similar to Sweden. Then, finding out the dependence relationship will be easier. Likewise, several latent similarities, as well as dissimilarities between the countries, can also be useful in the future for modeling.

This article contributes to the existing literature on country categorization by attempting to identify homogeneous clusters of countries in terms of government expenses. We will show a new country classification based on potential and actual government expenses. Economic researchers should answer four questions when classifying countries in terms of government expenses. *First*, how rich is the country? With both theoretical and practical reasoning, the richer

the country, the stronger its potential to finance public services. Here, [Hoang, 2019] observed that total government expenditure tends to increase along with income per capita. *Second*, how economically open is the country? Indeed, the openness may imply some deficits in revenues due to economic integration, which in turn affect the financing of budget expenditures [Baunsgaard, Keen, 2010]. *Third*, how good is its government? Bergh and Karlsson (2010) notice countries with large government expenditure require high tax rates and often have above-average institutional quality, for example, Sweden and Switzerland. *Fourth*, how does the government spend, how does its expenditure pattern look? Unlike the first three questions that investigate the pattern for government expenditure, the last one detects the actual situation of the expenses.

At first, we consider government expenditures based on GNI per capita, merchandise trade, and government effectiveness (these variables are used to answer the first three questions). Then, the actual situation of the expenses is expressed by subsidies and other transfers, for example, compensation of employees, goods and services expenses, interest payments, and other expenses (these variables are used to answer the fourth question). In this paper, we do not analyze the relationship between government expenditures and economic growth. However, we provide an important setup for regression analysis in which total government expenditures (measured in GDP or GNI) could be a key explanatory variable. Data for these indicators were taken from the World Bank's database.

This paper adopts the k-means clustering technique as the classification approach. This clustering technique is relatively simple but widely used for classification depending on different criteria. The idea of using this multidimensional technique is to construct some homogeneous groups of countries so that in the future, some unified regression type models can be implemented considering the groups (instead of considering the countries). Also, it helps us to identify the variables which are mostly varying between the groups but remain fixed within a particular group. In turn, it gives us an indication of the most important variables for the classification. The present technique is different from the popularly used classification algorithm used by the World Bank, which is primarily based on univariate classification and hence misses out on some of the important variables often. Through our work, we have tried to justify the dissimilarities between the clusters in terms of their dimensions.

To select the optimal number of clusters and the appropriate variables, we used the Elbow method (and hence the scree plot), and to compare between the newly formed clusters, we used Kruskal – Wallis H test. In addition, the Silhouette score is also analyzed to assess the reasonable level of clustering. As a consequence, 119 countries were divided into four clusters based on five criteria, which are GNI per capita; government effectiveness; subsidies and other transfers; compensation of employees; and goods and services expenses. Merchandise trade, interest payments, and other expenses do not make a difference between clusters. Notably, the GNI itself represents a good classification criterion, which separates clusters. However, in the presence of the remaining four criteria, the thresholds dividing the countries into clusters based on the GNI are not the same as the World Bank classification. It is observed that the majority of low- and middle-income countries fall into the same cluster, and high-income countries are divided into three small clusters with clear differences.

Apart from the Introduction, the remaining paper is divided into four sections. Section 2 contains the literature review. Section 3 describes methodology and data collection, and section 4 discusses the results. Finally, section 5 then draws out some key conclusions from the analysis.

2. Literature Review

There is already a wide variety of indicator-based approaches to classify countries, such as development measurements. However, it covers more aspects than mere economic changes. The measurement by national income has high (if not dominant) practicality for the operational purpose of international organizations [Harris et al., 2009; Vaggi, 2017]. Bearing that in mind, the World Bank appears to offer the single most popular classification of countries based on income per capita. The World Bank has a whole history of country classifications with key characteristics for both operational and analytical purposes [Harris et al., 2009; Nielsen, 2011; Vaggi, 2017, etc.]. In particular, the World Development Report and the World Development Indicators have since the late 1980s classified countries into different categories, namely high-income, middle-income, and low-income countries. While appearing convenient and less costly to validate, the classification by income-per-capita approach suffers from several weaknesses. First, the income per capita is not the best measurement of economic capacity and well-being and does not take into account other indicators such as GDP, gross national disposable income, etc. [Vaggi, Capelli, 2014; Vaggi, 2017]. Second, the absolute threshold for different categories is accompanied by no justification [Nielsen, 2011]. Third, the country-specific characteristics underlying the income per capita – such as natural resource dependence, industrialization level, etc. – may differ across countries, even within the same category. In particular, no account of income distribution is made, although this can significantly influence the conduct of public policy in various countries [Vaggi, 2017].

UNDP has developed a classification system that covers a broader range of development-related aspects. The Human Development Index (HDI) is the single indicator for classification, but it is a composite index of three indices as a measurement of countries' progress in income, education, and life expectancy. Since the Human Development Report 2010, income is measured by gross national income per capita – which is similar to the World Bank classification. [Nielsen, 2011; Alonso et al., 2014] survey the evolution of thresholds for country categorization over time. Again, before the Human Development Report 2010, the absolute thresholds fail to be explained. The HDR in 2010 adopted relative thresholds, but the switch to relative threshold was not justified, and equal countries' country weights may offer another source of flaw [Nielsen, 2011]. Besides, even with a wider scope, the HDI still fails to comprehensively cover aspects related to people's choices, such as quality of the environment, human rights, political freedom, etc. [McGillivray, White, 1993; Alonso et al., 2014].

The International Monetary Fund (IMF) has its own classifications, which evolved over time for both operational and analytical purposes. Unlike the World Bank, however, the IMF aggregates economic and financial data, rather than just income per capita, of member countries for categorization. Reclassifications did take place, but no explanation was provided [Nielsen, 2011].

In summary, the existing approaches by international organizations to classify countries in terms of development level have some profound characteristics. First, while different clustering techniques may be used, all the approaches appear to agree on the presence of developed countries and developing countries. Second, justification for different thresholds and even reclassifications hardly exist, though the alteration of thresholds and reclassifications may take place over time for operational and analytical reasons. [Vaggi, 2017] claims that the update of thresholds does not reflect the increase in world income per capita (specifically in 1987–2015).

Third, the data required for clusters can in general be collected annually. As the donor-recipient relationship evolves, it is natural to expect that the old/existing classification need to be changed [Harris et al., 2009]. These characteristics have to some extent weakened the theoretical economic foundations for the classifications themselves, i.e. whether the countries in the same cluster are economically and structurally alike. At the same time, as information systems and data mining technology have developed, it is difficult to predict whether reclassification may be made again and, if so, how.

Various alternative frameworks for classifying countries have been proposed. [Nielsen, 2011] introduces a cluster-based approach to development taxonomy, with examples of alternative dichotomous taxonomy, and alternative trichotomous taxonomy. More generally, any number of categories (smaller than the number of countries) can be used. The optimal thresholds can also be identified and justified. [Mirzaei, Vizvari, 2011] reconstruct the classification of countries by the World Bank using a linear and/or integer-programming model known as Multi-Cluster Hierarchical Discrimination Method. The respective result sees only a marginal role of GNI per capita in the classification. [Solactive, 2017] uses a framework that covers common economic, financial, and institutional characteristics to assess the countries' level of market development. This framework by itself reflects a scorecard approach, dividing the countries into different clusters depending on whether they meet or fail to satisfy different criteria, namely rankings in terms of GNI capita, minimum HDI score, minimum of average daily volume, etc. [Vázquez, Sumner, 2012] employ cluster analysis techniques and showcase the optimum number of clusters to be five. The cluster analysis relies on various indicators of poverty, inequality, and human development. [Fantom, Serajuddin, 2016] propose various amendments to the current method of classification by the World Bank, in terms of adjusting the thresholds, redefining the thresholds. The cluster analysis technique for classification schemes is also mentioned, with the review of major critiques such as arbitrary weights, implicit trade-offs between components, and practicality in the case of indicators with poor geographic coverage and update frequency.

New contextual development should also be considered in classifying countries. Paper [Harris et al., 2009] articulate the necessity of considering interdependence and network in future European aid policy. In terms of independence, efforts to support poorer countries may also be beneficial to the wealthy world as they may reduce the adverse impacts of global warming, illicit migration, piracy, epidemic disease, etc. The networking effect is material, as governments of wealthy countries may have to work with other new powerful agents, including influential developing countries [Harris et al., 2009; Fialho, Van Beigeijk, 2017]. [Ahmad, Nisar, 2014] argue that health and quality of environment indicators play the most important role for classification of countries, and therefore should be better reflected.

Recently, various studies have employed rigorous mathematical-statistical techniques in classification exercises. [Alshamaa et al., 2018] adopt a hierarchical clustering approach to create a two-level classification, using belief function theory. [Costa et al., 2018] propose a multiple-criteria decision aiding method for nominal classification, based on the concepts of similarity and dissimilarity. The proposed modeling method could take into account interaction effects between criteria. Other studies [Ştefan, 2012; Cuadros-Rodriguez et al., 2016; Tharwat, 2018; etc.] look into the comparison of different classification methods.

3. Methodology and Data

3.1. Elbow method and Silhouette score

The mechanism for the k-means clustering technique [Hartigan, Wong, 1979] in country classification is as follows: Clusters are formed based on the nearness of individual cells from the centroids, while the clusters themselves are as far apart as possible. Suppose there are p characteristics of a cell. Measurements on these p characteristics can be denoted by a vector of the form $x = \left(x_1, x_2, \ldots, x_p \right)$. The disparity between two cells i and j is then measured through the (unweighted) Euclidean distance between them, $d_{ij} = \sqrt{\sum_{l=1}^{p} \left(x_{il} - x_{jl} \right)^2}$. Clustering the cells can be done based on their relative distances. To start with, a supposition is that k clusters are to be formed. Next, the k farthest or most disparate cells are selected as the nodal points of these k clusters. Each of the remaining cells is then allocated to one of these k clusters depending on which of the nodal points is closest to it. Once all the cells are allocated to one or the other cluster, the new cluster centers (centroids or medoids) are re-calculated. Cells are then reallocated for the new cluster centers. This process is repeated till no further reallocation is required.

While the k-means cluster algorithm is easily applicable, the quality of the classification depends on the number of clusters (k). In particular, we can classify observations for each value of $k \left(k = 1, 2, \ldots, K \right)$, but the degree of uniformity can be distinct and rely on the number of clusters, k, which we initially defined. Therefore, selecting the optimal number of clusters is inevitably the first important task when applying this technique.

This paper adopts the Elbow method to identify the optimal number of clusters. Accordingly, we need to find the kink point on the curve representing the values of within sum of squares (WSS) or its logarithm (log (WSS)) for all cluster solutions corresponding to the values of k. The smaller the WSS, the higher the uniformity in the clusters. The number of clusters must be no larger than the average number of observations per cluster. Given that we have 119 observations, the maximum number of clusters to plot is then 10 (i.e., $k \leq 10$). The optimal number of clusters is chosen at the kink point because from this point, increasing the number of k clusters does not significantly reduce the value of the WSS. Another criterion for detecting the optimal number of clusters is the η_k^2 coefficient, which resembles the R^2 (proportion of explained and unexplained variation), or the proportional reduction of error (PRE) coefficient. η_k^2 measures the proportional reduction of the WSS for each cluster solution k compared with the total sum of squares (TSS). In contrast, PRE_k illustrates the proportional reduction of the WSS for cluster solution k compared with the previous solution with $k - 1$ clusters [Makles, 2012].

After classifying countries by the k-means technique, we analyzed the Silhouette index to examine the validity of the classification results. The silhouette analysis measures how well an observation is clustered and estimates the average distance between clusters. The silhouette plot displays a measure of how close each point in one cluster points in the neighboring clusters. The Silhouette's largest value of 1 shows a perfectly reasonable clustering. In contrast, the small-

lest value of –1 reflects a completely unreasonable arrangement, which implies that observation should be sorted into another cluster. Therefore, Silhouette analysis is sometimes used in studies to select the optimal number of clusters.

3.2. Kruskal – Wallis H test

In addition to determining the optimal number of clusters, selecting the appropriate criteria for classification is essential. Typically, classification criteria are chosen according to the analytical purpose. For example, the World Bank chooses income per capita because of its concern about the living standard of people; the UNDP chooses HDI owing to their cares about human development; the IMF selects economic and financial indicators as they consider the riskiness of public debts, etc. Similarly, this paper sorts out criteria related to actual performance and prospects of public expenditure. Based on the results of this classification, quantitative and qualitative studies related to government-spending policies can be used for reference, comparison, and analysis.

The selected criteria (variables) are required to point out dissimilarities among clusters. If it fails to do that, it means that the variable is not proper for being a basis for classification in terms of data analysis. To eliminate inappropriate variables, we use the Kruskal – Wallis H test. The Kruskal – Wallis H test is a rank-based nonparametric test to determine if there are statistically significant differences between two or more clusters of an independent variable on a continuous or ordinal dependent variable. It is considered the nonparametric alternative to the one-way ANOVA (sometimes also called the "one-way ANOVA on ranks"), and an extension of the Mann – Whitney U test to allow the comparison of more than two independent clusters. After selecting the appropriate variable by the country by Kruskal – Wallis H test, we repeated the step of determining the optimal number of clusters and analyzed the results.

3.3. Data

This paper uses several indicators measuring expenditures, such as percentages of total government expenditure, including subsidies and other transfers; compensation of employees; goods and services expense; interest payments; and other expenditures. The variables are chosen in such a way that different parts of total spending give 100% altogether. The measure for public expenditure is reflected in GNI per capita, government effectiveness, and merchandise trade. All data are taken from the World Bank's database.

The reasons for choosing these variables for public expenditure are as follows. First, the rich and poor countries may have distinct concerns. For example, the rich pay more attention to the issues that may be less direct to generating budget revenues such as social security, environment improvement, maintenance and development of culture, etc. while the poor concentrate relatively more on growth-related issues such as technology transfer, manufacturing, and infrastructure development, etc. Second, the openness of the economy, on the one hand, shows the integration and contribution of imports and exports to GDP. On the other hand, deeper economic integration (reflecting higher openness) also requires the country to phase out tariff barriers via bilateral and multilateral agreements. Therefore, economic openness affects the budget revenues, and in turn the capacity to finance public spending in concerned countries. Third, the effective-

ness of public policy (including budget management) is a key factor affecting public trust in the government. As documented by [Bergh, Henrekson, 2011], public spending can be expanded even when its current scale is already high, for which Sweden sets an example.

<div align="right">**Table 1.**</div>

List of variables for classification

Label	Indicator	Theoretical reasons
TRFT	Subsidies and other transfers (% of expense)	These variables show the structure of government expense. The rich countries tend to spend a high fraction of expense on subsidies and other transfers, while the poor countries have a high fraction of expense on compensation of employees and good/services
COMP	Compensation of employees (% of expense)	
GSRV	Goods and services expense (% of expense)	
INTP	Interest payments (% of expense)	
OTHR	Other expense (% of expense)	
TRADE	Merchandise trade (% of GDP)	Merchandise trade as a share of GDP is the sum of merchandise exports and imports divided by the value of GDP, all in current US$. This variable may imply some losses/ increases of revenues due to economic integration which in turn affect the financing of budget expenditures
GE_EST	Government Effectiveness: Estimate (ranked from –2.5, less effective, to 2.5, more effective)	Government Effectiveness captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies
GNI	GNI per capita, Atlas method (current US$)	GNI on this own helps classify countries into different clusters

Source: Authors synthesized.

In the present analysis, after eliminating the missing observations, we use the cross-section for the 2015 dataset (with 119 countries) for classification because it has the largest number of countries. In this dataset, Burundi is the poorest country, with an average income of US$ 260 per person, while Norway is the wealthiest country with US$ 93,080 per capita. The country with the lowest government efficiency index is the Marshall Islands, while Switzerland has the highest government efficiency index. Belgium looks like the most economically open country, as this indicator accounts for 169.4% of GDP. Timor – Leste is categorized as the least open one as its economic openness indicator accounts for 19.27% of GDP. In terms of public expenditure structure, only the subsidies and other transfers variable roughly follow a normal distribution (as shown in Fig. 1), represented by 95% of observations located between 37% and 47% of total

expense. The remaining variables have positively skewed distribution (as shown in Fig. 1), which means that most countries witnessed low ratios of compensation of employees, expenditures on goods and services, interest payments, and other expenses.
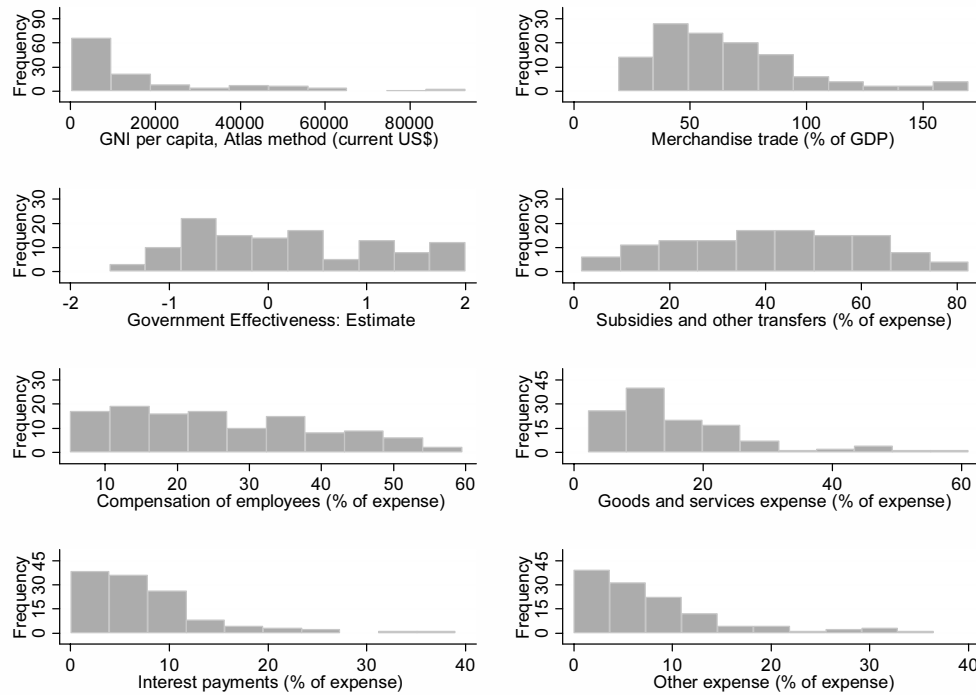
Fig. 1. Frequency distributions of data

Source: Authors' calculations based on the World Bank's database.

4. Result

4.1. Optimal number of clusters

The Elbow analysis for all the variables is shown in Fig. 2a. At $k = 4$, there is a kink in both WSS and log (WSS). η^2 points to a reduction of the WSS by 47% and PRE to a reduction of about 19% compared with the $k = 3$ solution. It indicates the reason why we have used $k = 4$ as the optimum number of clusters. Here, we can see a substantial gain while considering $k = 4$ as compared to $k = 3$. So, after the countries are classified, we next move forward to investigating which variables are instrumental in creating these clusters. To achieve that, we investigated how similar or dissimilar are the variables (that are used for classification), coming from different clusters using a Kruskal –Wallis H test. Here we have chosen to adopt this test as for most of the variables normality is not assured. For these four clusters, the Kruskal – Wallis H test shows a statistically significant difference in GNI, GE_EST, TRFT, COMP, and GSRV between in four clusters

(p-value < 0.05). In contrast, TRADE, INTP, and OTHR exhibit statistically insignificant differences (p-value equals 0.3106, 0.0723, and 0.6732 respectively). To cross-verify, the task of determining the optimal number of clusters is then repeated with the reduced number of the aforementioned five variables.

The results show that these five variables are appropriate and the optimal number of clusters suggested by the Elbow method becomes more obvious at position $k = 4$ (Fig. 2b). From Fig. (4b) also, we can see that at $k = 4$, there is a sharp elbow shape, indicating a sharp reduction of within (cluster) sum of squares. It also gives us the justification of dividing the countries into 4 clusters. The classification results show that there are five countries in Cluster 1, 17 are in Cluster 2, 26 are in Cluster 3, and 71 are in Cluster 4 (see Table 2 in Appendix).
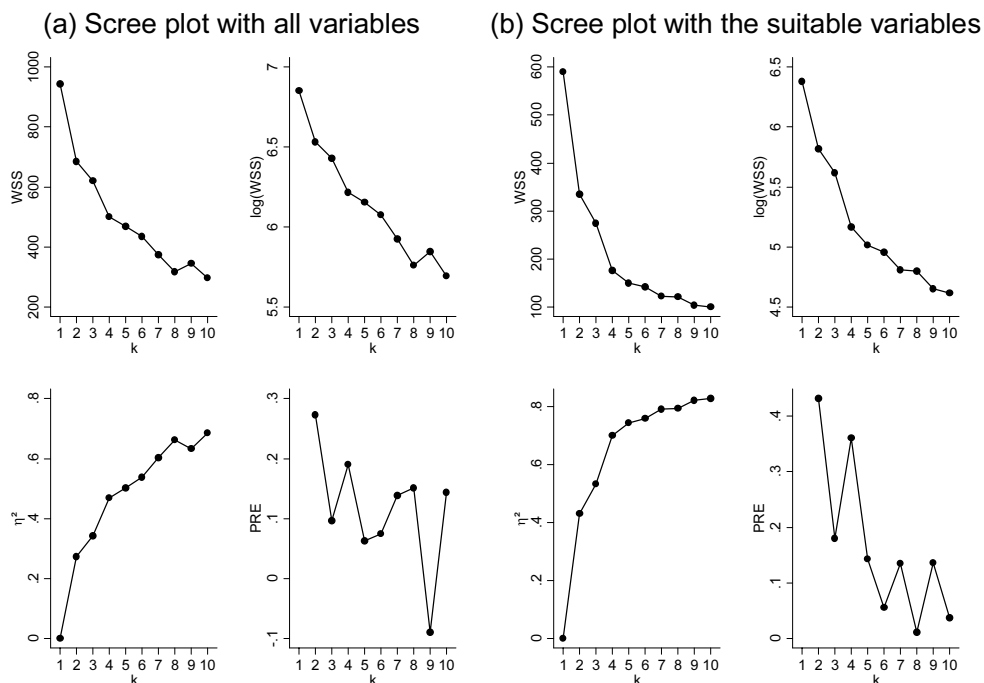

Fig. 2. WSS, log(WSS), η^2, and PRE for all k cluster solutions

Source: Authors' calculations based on the World Bank's database.

In addition, Silhouette analysis shows that the average Silhouette index reached a high level, approximately 0.81. Silhouette index of clusters (arranged in order from 1 to 4) are respectively 0.37, 0.82, 0.67 and 0.91. This means that cluster 4 experienced the highest homogeneity between the countries, followed by cluster 2, cluster 3, and cluster 1 respectively. This indicates that the countries which are there in cluster 4 are the most similar within themselves compared to the countries in the other clusters. Countries in cluster 2 are more similar (within themselves) than the countries in clusters 1 and 3. Finally, the countries in cluster 3 are more similar than the countries in cluster 1. Only 3 out of 119 cases lead to the Silhouette index being

less than 0, namely Australia, Denmark (cluster 1), and Kazakhstan (cluster 3). Another 3 cases produce the index being positive but close to 0, namely Italy (cluster 2), Russian Federation, and Turkey (cluster 3). These might have been considered as a potential outlier among the data points. Despite some cases experiencing low levels (reflected by low Silhouette), the majority of others have a score higher than 0.5 (105 cases accounting for 88% of the overall). Thus, our classification arguably succeeds in identifying homogenous clusters considering the types of expenditures like GNI per capita, government effectiveness, subsidies, and other transfers, compensation of employees, goods and services expense. Here, we checked for multicollinearity between the variables with condition numbers and it came out to be < 30 for all the variables, indicating no sign of multicollinearity.
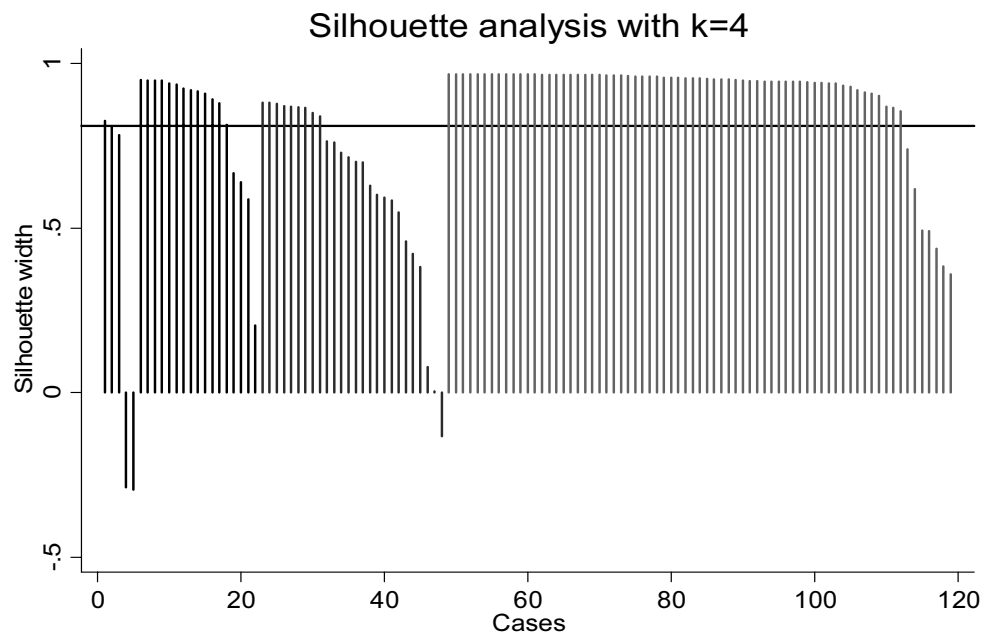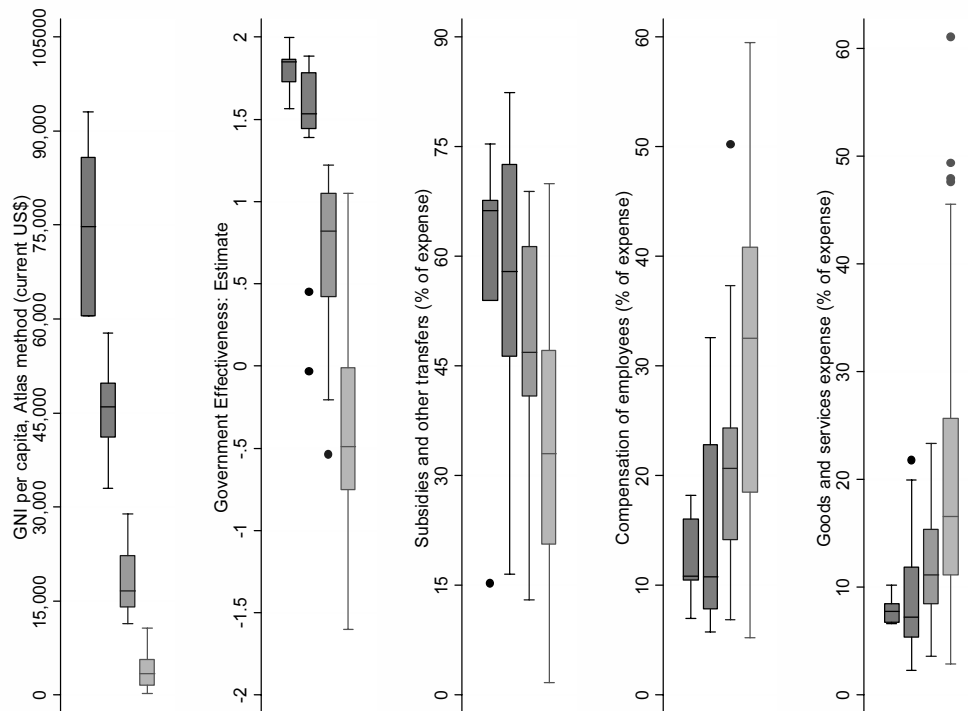

Fig. 3. Silhouette plot for the four clusters,
where different colours represents different country clusters

Source: Authors' calculations based on the World Bank's database.

4.2. Within Cluster (group) analysis

To determine the characteristics of national clusters, we use boxplots to show the distributions as well as quantiles in each cluster. This will lead us to know how similar or dissimilar the variables are between the four clusters. The plots give us an idea regarding why separate clusters are needed. Figure 4 shows that the GNI per capita clearly distinguishes the clusters. The figure for Cluster 1 includes extremely rich countries (goes over US$ 60,000); Cluster 2 experiences from $ 30,000 to $ 60,000; Cluster 3 contained countries from US$ 11,000 to US$ 30,000;

and below US$ 11,000 countries are gathered in Cluster 4. These thresholds differ from those determined by the World Bank (US$ 1,025; US$ 4,035; US$ 12,475 following 4 blocks of low, lower-middle, upper-middle, and high-income countries). Under our proposed classification, most of the countries with low, lower-middle, upper-middle-income levels (according to the World Bank classification, also called developing countries) are in Cluster 4. High-income countries in World Bank's classification are meanwhile distributed into Clusters 1, 2, and 3 in this categorization.

Fig. 4. Box plot for the five variables over the four clusters, where different colours represent different country clusters

Source: Authors' calculations based on the World Bank's database.

The key difference in our results and the one from World Bank is that ours is derived from the multi-criteria classification method. The additional criteria, along with average income, have changed the threshold values delineated between clusters. The threshold value created by GNI per capita is more straightforward than that in other variables although they still detect differences among countries. For example, two countries with the same level of Government effectiveness (or subsidies and other transfers, or compensation of employees, or goods and services expense) can be categorized into two different clusters. Unlike a classification based on a single variable, this classification identifies clusters of countries with similarities in many other criteria. In other

words, the countries that belong to the same cluster in this paper have uniformity in many different aspects related to the performance and future capacity of public expenditure.

Our classification yields two important results. First, by characterizing the four clusters, high-income clusters often have an effective government, devoting a large proportion of public spending to subsidies and other transfers, but only set aside a small portion for compensation of employees, and goods and services expenses. As a note, this paper sorts descending wealth from Cluster 1 to Cluster 4. Second, by allowing for multi-criteria similarities in the classification, we could enhance the possibility of sourcing differences in government expenditure and economic growth across countries. This in turn implies more material opportunities to learn and share experiences between countries, rather than just setting up the donor-recipient relationship solely based on GDP per capita.

Investigating the cases where the Silhouette index is less than 0 gives further insights. *First*, Australia with a Silhouette index of –0.29 is categorized in Cluster 1. This country has the lowest government efficiency index in the cluster, reaching 1.56, while the remaining countries in the cluster are higher than the score of 1.72. In addition, Australia experiences the highest rate of goods and services expense in the cluster, accounting for 10.16% of total spending, while the remaining countries are below 8.48%. *Second*, Denmark has a Silhouette index of –0.29, similar to that of Australia. The difference lies in the rate of spending on subsidies and other transfers. While countries in Cluster 1 spend over 54% of their total expenditure on the above variable, Denmark just pays 15% out. From the first look, this low number may require recalculation by the World Bank. However, this paper proposes the explanation for Denmark, having the highest income level, which is a key index to classify and determines an income threshold of US$ 60,000 to identify Cluster 1. *Third*, Kazakhstan with a Silhouette index of –0.1 is incorporated in Cluster 3 despite the lowest income of US$ 11,420. In addition, the country has its government efficiency index ranked third from the bottom at –0.07 and the lowest proportion of expenditure for compensation of employees at 6.9% compared to more than 10% of the rest of the countries) but the share of goods and services expense in total expenses is the highest (23%). In short, countries with low Silhouette index are often ones experiencing the lowest value in one of the component indexes, so they are located at the boundary delimiting between clusters.

5. Conclusions

Depending on their own operational and analytical requirements, each international organization (such as WB, IMF, or UNDP) proposes different criteria for identifying homogeneous clusters of countries. Economic researchers should have a good understanding of the purpose of the classification to wisely use these results. Efforts should then be made to identify the number of criteria and involve techniques, to balance out the trade-off between higher costs associated with more classification criteria and (benefits from) smaller errors when comparing countries in the resulted categories.

This paper proposes a practical approach in applying the k-means clustering technique to classify countries based on actual performance and potential of public expenditure in 2015. The analysis aims to identify homogeneous clusters of countries for future public expenditure-related research. For example, quantitative research on the effects of expanding public spending or austerity on economic growth across countries; or study experience in public expenditure management to get practical lessons for a certain case. The actual situation of public expenditure is

reflected in its structure, including subsidies and other transfers, compensation of employees, goods and services expenses, interest payments, and other expenses. The measures of public expenditure are indicated in GNI per capita, government effectiveness, and merchandise trade. The results show that there are four clusters of countries with homogeneity, classified under the set of five most appropriate criteria. Specifically, Cluster 1 includes 05 countries; Cluster 2 includes 17 countries; Cluster 3 includes 26 countries, and Cluster 4 includes 71 countries. Even under a multi-criteria classification framework, this paper confirms that GNI per capita remains the single best criterion for classification because of its clear thresholds. Based on this paper, further specific studies can apply similar techniques and approaches to identify homogenous country clusters according to a different purpose-driven set of specific criteria.

However, this paper is subject to a couple of limitations. *First,* while considering the countries to classify, the paper can only classify 119 out of 217 countries in the world. Several countries are not listed because of missing data. To a certain extent, this issue may be reduced by repeating the classification over several consecutive years while using additional techniques (such as interpolation). *Second,* despite looking into the space to finance public expenditure, the analysis excludes the possibility for the government to borrow from the capital market. In particular, the "implications of economic openness" concept is only assessed for budget revenues collected from trade, while ignoring the possibility of international borrowings. Future studies should look into this possibility, at best to validate the clustering of countries in terms of government expenditure, i.e. to confirm that governments of homogeneous countries under a multi-criteria classification framework should also be treated relatively equally in the international capital market. Third, the variables that we are considering might often be related to others. For example: since the variable "Government effectiveness" is an estimate gained from the country's citizens, it could be connected to other variables such as a share of transfers (which include pensions, social assistance benefits, and so on). So, one should check for multicollinearity between the variables and should have a provision of eliminating the problem of multicollinearity, if it exists.

Appendix

Table 2.

Country classification by the five variables

Cluster 1		Cluster 2		Cluster 3	
No.	Country name	No.	Country name	No.	Country name
1	Australia	6	Austria	23	Bahamas, The
2	Denmark	7	Belgium	24	Bahrain
3	Luxembourg	8	Canada	25	Barbados
4	Norway	9	Finland	26	Chile
5	Switzerland	10	France	27	Croatia
		11	Germany	28	Cyprus

Continues

Cluster 1		Cluster 2		Cluster 3	
No.	Country name	No.	Country name	No.	Country name
		12	Iceland	29	Czech Republic
		13	Ireland	30	Estonia
		14	Israel	31	Greece
		15	Italy	32	Hungary
		16	Japan	33	Kazakhstan
		17	Kuwait	34	Korea, Rep.
		18	Netherlands	35	Latvia
		19	New Zealand	36	Lithuania
		20	Sweden	37	Malta
		21	United Kingdom	38	Palau
		22	United States	39	Poland
				40	Portugal
				41	Russian Federation
				42	Seychelles
				43	Slovak Republic
				44	Slovenia
				45	Spain
				46	Trinidad and Tobago
				47	Turkey
				48	Uruguay

Cluster 4					
No.	Country name	No.	Country name	No.	Country name
49	Afghanistan	75	Georgia	101	Paraguay
50	Albania	76	Guatemala	102	Peru
51	Angola	77	Honduras	103	Philippines
52	Armenia	78	India	104	Romania
53	Azerbaijan	79	Indonesia	105	Samoa
54	Bangladesh	80	Jordan	106	Sao Tome and Principe
55	Belarus	81	Kiribati	107	Serbia
56	Benin	82	Lao PDR	108	Sierra Leone
57	Bhutan	83	Lebanon	109	Solomon Islands
58	Bosnia and Herzegovina	84	Lesotho	110	South Africa
59	Brazil	85	Malawi	111	Sri Lanka
60	Bulgaria	86	Malaysia	112	Suriname
61	Burkina Faso	87	Mali	113	Thailand
62	Burundi	88	Marshall Islands	114	Timor-Leste
63	Cabo Verde	89	Mauritius	115	Togo
64	Cambodia	90	Mexico	116	Tunisia
65	Colombia	91	Micronesia, Fed. Sts.	117	Ukraine
66	Congo, Rep.	92	Moldova	118	Vanuatu
67	Costa Rica	93	Mongolia	119	Zambia
68	Cote d'Ivoire	94	Morocco		
69	Dominican Republic	95	Mozambique		
70	Egypt, Arab Rep.	96	Myanmar		
71	El Salvador	97	Namibia		
72	Eswatini	98	Nepal		
73	Ethiopia	99	Nicaragua		
74	Fiji	100	Pakistan		

* *
*

References

Ahmad Z., Nisar L. (2014) Classifications of Countries Based on Their Standard of Living. *Pakistan Journal of Commerce and Social Sciences (PJCSS)*, 8, 1, pp. 74–98.

Alonso J.A., Cortez A.L., Klasen S. (2015) LDC and Other Country Clusterings: How Useful Are Current Approaches to Classify Countries in a More Heterogeneous Developing World? *Global Governance and Rules for the Post-2015 Era: Addressing Emerging Issues in the Global Environment* (eds. J.A. Alonso, J.A. Ocampo), 424.

Alshamaa D., Chehade F.M., Honeine P. (2018) A Hierarchical Classification Method Using Belief Functions. *Signal Processing*, 148, pp. 68–77.

Baunsgaard T., Keen M. (2010) Tax Revenue and (or?) Trade Liberalization. *Journal of Public Economics*, 94, 9–10, pp. 563–577.

Bergh A., Henrekson M. (2011) Government Size and Growth: A Survey and Interpretation of the Evidence. *Journal of Economic Surveys*, 25, 5, pp. 872–897.

Bergh A., Karlsson M. (2010) Government Size and Growth: Accounting for Economic Freedom and Globalization. *Public Choice*, 142, 1–2, pp. 195–213.

Capelli C., Vaggi G. (2014) *Why Gross National Disposable Income Should Substitute Gross National Income.* Pavia, Italy: Department of Economics and Management, University of Pavia.

Costa A.S., Figueira J.R., Borbinha J. (2018) A Multiple Criteria Nominal Classification Method Based on the Concepts of Similarity and Dissimilarity. *European Journal of Operational Research*, 271, 1, pp. 193–209.

Cuadros-Rodríguez L., Pérez-Castaño E., Ruiz-Samblás C. (2016) Quality Performance Metrics in Multivariate Classification Methods for Qualitative Analysis. *TrAC Trends in Analytical Chemistry*, 80, pp. 612–624.

Fantom N., Serajuddin U. (2016) *The World Bank's Classification of Countries by Income*. The World Bank.

Fialho D., Van Bergeijk P.A. (2017) The Proliferation of Developing Country Classifications. *The Journal of Development Studies*, 53, 1, pp. 99–115.

Harris D., Moore M., Schmitz H. (2009) *Country Classifications for a Changing World*. IDS Working Papers, 326, pp. 01–48.

Hartigan J.A., Wong M.A. (1979) Algorithm AS 136: A k-means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28, 1, pp. 100–108.

Makles A. (2012) Stata Tip 110: How to Get the Optimal k-means Cluster Solution. *The Stata Journal*, 12, 2, pp. 347–351.

McGillivray M., White H. (1993) Measuring Development? The UNDP's Human Development Index. *Journal of International Development*, 5, 2, pp. 183–192.

Mirzaei N., Vizvari B. (2011) Reconstruction of World Bank's Classification of Countries. *African Journal of Business Management*, 5, 32, pp. 12577–12585.

Nielsen L. (2011) *Classifications of Countries Based on Their Level of Development: How it is Done and How it Could be Done* [Electronic resource]. IMF Working Paper. Mode of access: http://www.relooney. fatcow. com/0_NS4053_1504. pdf

Solactive (2017) *Solactive Country Classification Framework*. Available at: https://www.solactive.com

Ştefan R.M. (2012) A Comparison of Data Classification Methods. *Procedia Economics and Finance*, 3, pp. 420–425.

Tharwat A. (2018) Classification Assessment Methods. *Applied Computing and Informatics*.

UNDP (2011) *Human Development Report 2010*.

Vaggi G. (2017) The Rich and the Poor: A Note on Countries' Classification. *PSL Quarterly Review*, 70, 280.

Vázquez S.T., Sumner A. (2012) *Beyond Low and Middle Income Countries: What If There Were Five Clusters of Developing Countries?* IDS Working Papers, 2012, 404, pp. 1–40.

World Bank. *World Development Reports*. Various years.