



Article Principal Component Analysis of the Potential for Increased Rail Competitiveness in East-Central Europe

Szabolcs Duleba * D and Bálint Farkas

Department of Transport Technology and Economics, Budapest University of Technology and Economics, 1111 Budapest, Hungary

* Correspondence: duleba.szabolcs@mail.bme.hu

Received: 22 May 2019; Accepted: 31 July 2019; Published: 2 August 2019



Abstract: Increased rail competitiveness has been the objective of many countries around the world, including member states of the EU. Although railway market liberalization has always been accompanied by high expectations of increased efficiency and competitiveness, the overall impact of such decisions can be considered controversial. This paper aims to contribute to the scientific debate by conducting a factor analysis of some East-Central European countries from the aspect of rail freight competitiveness. Since many highly correlated factors influence competitiveness, its mathematical-statistical representation and analysis is difficult due to the high number of dimensions of the factor space. Moreover, competitiveness cannot be measured directly only as a latent variable which is a feature of Principal Component Analysis (PCA). The introduced PCA, model by way of reducing the number of dimensions, can highlight the relations among the attributes and determine the most crucial issues capable of increasing rail competitiveness in the given countries and also of clustering those national railway markets. Recommendations for structural changes in national rail freight markets of the region are also supplied. Our results show that international rail competitiveness depends rather on market efficiency than on market liberalization due to the fact that the Global Competitiveness Index and Export/Import attributes did not significantly correlate with market concentration. As for the larger domestic rail freight sectors, small freight forwarders—spawned by liberalization—are shown to play a significant role in increasing competitiveness.

Keywords: sustainable transport; market liberalization; competitiveness; rail freight market; Principal Component Analysis

1. Introduction

Economic theory states that market liberalization generally increases competition, hence motivates companies to more efficient and more productive operations that cause improved competitiveness in the market [1]. Regarding the European rail freight markets, the experience so far has not completely validated this logic of the theory [2]. On the one hand, several scientific studies confirm that the rail companies with the highest degree of independence from external influences are the most efficient and competitive players in the market [3]. Moreover, the expansion of the markets is often followed by increased technical innovation [4], which also contributes to efficiency. (In this paper, we use the term "competitiveness" in a broader sense than market efficiency. In our view, the notion of competitiveness covers both the efficiency/productivity of the operations and the demand for rail transport services. We also note that, in our paper, we only consider rail freight markets in our database and results. Although passenger rail markets undoubtedly have common characteristics with the rail freight sector such as infrastructure used and some macroeconomic attributes, there are a number of

rail freight-specific ones as well. Nevertheless, in our literature review, we apply the most relevant references from findings related to the passenger rail market taking into account the differences.) Many researchers also state [5,6] that liberalization resulted in more efficient railway undertakings and higher consumer surplus. On the other hand, there are noteworthy studies that doubt the strong correlation between productivity and market liberalization in the railway industry. One of the best examples is the work of Bougna and Crozet [7], in which the authors applied an input–output approach to measure the productivity of 17 European railway systems with the dataset of Eurostat from 1997 to 2011. In terms of methodology, they used stochastic distance functions to measure productive efficiency and change, and subsequently a limited dependent variable model to test the impact of liberalization

productivity instead of forcing market liberalization by all means. Our paper aims to contribute to this debate by focusing on East-Central European rail freight markets, having been more recently liberalized than their western counterparts within the EU but already having accumulated an experience of over a decade of free market operation. The outcome so far has not been an unqualified success story in this region. Based on recent statistics, these liberalized rail freight markets have been stagnating for the past ten years [8]. Despite the availability of new business opportunities, nowadays, the rail freight sector of the region still suffers from low attractivity and capability [9] compared to other regions in Europe. Thus, analyzing examples of this east-central cluster of EU member states may suffice to draw some conclusions contributing to the ongoing debate on the real impacts of rail freight market liberalization.

on productivity. Their conclusion was that, in Europe, policy makers should give priority to raising

For our analysis, first the proper indicators of productivity, efficiency and competitiveness need to be selected. Our objective was not only a comparison of the markets, but also their assessment from a competitiveness point of view considering the fact that pure transport volume data (export, import, transit) did not seem sufficient. The determination of the relevant attributes was partly based on the paper of Feuerstein, Busacker and Xu [10]. They examined 34 factors in five categories: technical, economic, social, political/legal and other. All factors were influencers of the open access competition in European long-distance passenger rail transport. (In our case, only freight service is examined but most of the factors are identical to that of passenger transport.) Among the 34 influencers of competition, the economic and political/legal clusters proved to be of highest importance; especially factors like market potential, access to attractive train paths and governments' attitude were among the most important ones. The authors also stated that, for different countries, different influencing attributes proved to be important. Although they applied rather different influencers from ours, the study inspired us to perform our own analysis on rail freight competition and liberalization. Since, due to their methodology, they incorporated many intangible attributes in their calculation, instead we decided to use tangible datasets or reliable indices computed based on trustworthy sources and methodology.

Thus, to ensure arrival at appropriate findings, a relevant, reliable database and a suitable analytical tool are also required. The sources of our raw data are referenced in Table 1. The Herfindahl–Hirsch index (HHI) values were calculated from publicly available governmental publications.

In terms of methodology, a multivariate factor analysis technique, the Principal Component Analysis (PCA), was selected due to its capability of determining not only complex interactions of factors but also the clusterization of the entities while highlighting dependencies from background variables. In a 15-dimensional space, the Eucledian distance of the attributes cannot be determined, so the cluster analysis of the raw data is impossible. We used an 8-year dataset to examine five East-Central European (ECE) countries including Slovakia, the Czech Republic, Hungary, Poland and Romania, representing altogether 40 entities, which all have 15 coordinates due to the selected 15 attributes, which are rather difficult to handle with the general multivariate statistical techniques. Our basic research objective was the analysis of recently transitioned countries after their railway market liberalization, so the extension of the 40 entities was impossible due to obvious reasons (the rail freight markets of these countries were opened up in the last decade, so the time frame could not be extended neither could the number of examined countries). As a substantial reduction in the number of dimensions could

significantly contribute to the efficacy of the analysis, PCA was selected as a useful methodological tool to that end. Furthermore, the research concept of analyzing the complex issue of competitiveness necessitated the assumption of "latent attributes" besides the provided 15 measured attributes. The method can be applied to various kinds of datasets not only to subjective, preference-based ones but also to "objective", measurement-based datasets [11]. Additionally, one of PCA's main areas of application is macroeconomics [12] and our dataset contained dominantly macroeconomic data.

Influencer Number	Source
1, 12, (9 calculated)	The Czech Republic, Slovakia: The Railway Infrastructure Administration, state organization https://www.szdc.cz/en/o-nas/zeleznice-cr.html Románia: Compartiment Mass Media, Relatii cu Publicul CFR Marfa s.a. http://www.cfrmarfa. cfr.ro/index.php?option=com_content&view=article&id=223&Itemid=69⟨=en Poland: PKP Polskie Linie Kolejowe S.A. https://en.plk-sa.pl/for-customers-and-partners/the- rules-for-allocating-train-paths/network-statement-20192020/ Hungary: Innovációs és Technológiai Minisztérium Nemzeti Közlekedési Hatóság vasúti főosztály adatszolgáltatások: https://www.nkh.gov.hu/web/vasuti-foosztaly1/a-szervezet
2, 3, 7, 8, 13, 15	Eurostat: Goods transported by type of transport (2004–2016) http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=rail_go_typeall⟨=en
4, 5, 6	Word Economic Forum: The Global Competitiveness Report: https://www.weforum.org/reports/the-global-competitiveness-report-20172018
14	Best available data: Czech: Network Statement: https://www.szdc.cz/web/en/our-railway/how-to-operate-on-our-railway/prohlaseni-2019, Slovakia: Železnice Slovenskej republiky Network Statement, 2007–2015 https://www.zsr.sk/dopravcovia/infrastruktura/podmienky-pouzivania-zel-infrastruktury/ network-statement-2020-english-version/ Hungary: Network Statement 2007–2015 https://www2.vpe.hu/halozati-uzletszabalyzat-husz/hatalyos-husz-20072015, Polen: pkp polskie linie kolejowe s.a. network statement 2007–2015 https://www.plk-sa.pl/dla-klientow-i-kontrahentow/warunki-udostepniania-infrastruktury-i- regulaminy/regulamin-sieci/archiwum-regulaminow/ Romania: Network Statement 2007–2015, http://www.cfr.ro/index.php/ct-menu-item-126 Eurostat, https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2019:0051:FIN:HU:PDF
10	Railway enterprises—by type of enterprise [rail_ec_ent] Last update: 26 February 2018 www. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=rail_ec_ent⟨=en

Table 1. Sources of data.

In the following sections, firstly, we give an overview the existing literature references to the selected competitiveness attributes. Afterwards, the applied PCA methodology is introduced, followed by the results of the analyzed five countries. Finally, conclusions are drawn with outlined limitations and recommendations for further research.

2. Theory/Literature Overview on Competitiveness Factors of Rail Freight Markets

In order to apply a complex approach integrating technical, cost- and user-preference elements, we consider competitiveness as a combination of efficiency attributes and the demand for railway transport services. Demand is considered here as motivation for rail freight service and not as a consequence [13]. In the transportation literature, technical and cost issues are generally examined by statistical techniques [14] while, for preference analyses, multi-criteria decision-making techniques are applied [15]. In our study, influencing factors from both groups were selected and also direct competitiveness indices (Global Competitiveness Index (GCI), GCI6P, GCI2P) have been integrated to elaborate the dependencies of these influencers. Since it has been argued that single measures of competitiveness are not sufficient, several authors propose composite indices [16,17]. One of our objectives was to shed light on the connections of single factors and these composite indices—especially the Global Competitiveness Index (GCI)—in our paper.

Undoubtedly, one of the most important determining factors of rail freight and passenger transport competitiveness is the quality of the infrastructure. This assertion has been proven for passenger transport [18] and for freight [19–21]. In their exhaustive survey, Purwanto et al. [17] concluded

that increased competitiveness is among the wider benefits of transport infrastructure investments. The second pillar of the GCI is also related to the quality of rail infrastructure of the national railway.

Rail freight market competitiveness and track access fees of the country or region also have interrelations [22], hence we applied this factor among the influencers in our model.

Since one of the main objectives of this paper is to analyze the correlation between market liberalization and competitiveness, we have selected factors that reflect the assumed higher intensity of competition. Market concentration is generally measured by the Herfindahl–Hirschman index (HHI) [23] and, for rail freight markets, applicable references exist, e.g., Crozet, 2016 [24]. Although in his paper, Crozet leaves the question open as to whether high market concentration in the rail freight industry should be regulated or whether this situation leads to high(er) competitiveness. He states that dominancy of one or some major operators might be abusing and thus likely decrease competitiveness. Based on this, we applied market concentration attributes such as calculated HHI, market share of the biggest player of the total freight volume, market share of the smallest market player and number of market players.

Determining connections between regulative actions and competitiveness was also among the objectives of our survey. There have been several notable attempts in the scientific literature concerning this issue including the case of the European railways [25,26]. In our study of East-Central European countries, we decided to apply a direct regulatory factor, namely, the annual number of railway-related legislative actions in these countries.

Considering the demand side, many authors (e.g., Jarzemskis and Jarzemskiene, 2017) stated [27] that higher competitiveness in rail markets does not necessarily originate from a higher intensity of competition in the market (in Lithuania there is only one rail freight operator so the market is monopolized and still has been proven to be more competitive internationally than Poland where the competition is higher) but from the existing demand of the destination sites. Thus, we selected factors such as import, export, and transit freight volumes, domestic freight volume and total rail volumes in million tons, all measured in national levels in the examined five ECE countries.

Finally, we selected a direct competitiveness index, GCI, which is also officially measured/calculated by European countries. GCI, constructed and applied by the World Economic Forum, is accepted worldwide for indicating market and country competitiveness, including several transport- and logistics-related composites [28]. In our study, we applied GCI as an all-encompassing attribute related to competitiveness of each examined country, the second pillar of GCI related to rail infrastructure quality specifically and the sixth pillar related to market efficiency (The Global Competitiveness report 2011–2012).

Based on the ones detailed in the literature review, concentrating merely on rail freight attributes, the following influencers were applied in the model (Table 2). For better readability, we provide the 8-year average (2008–2015) values of each country for the 15 attributes, while the entire raw data table is presented in Appendix A of the paper.

Attribute Number	Attribute Name	Contraction	[Unit]	Averages HU	Averages PL	Averages CZ	Averages SK	Averages RO
1	Biggest Market Player	BMP	percentage	77.0	50.0	80.0	80.0	47.0
2	Domestic Traffic	DT	mill. Tons	12,541	175,901	40,255	7398	46,207
3	Export	EX	mill. Tons	12,873	19,522	19,545	10,924	3207
4	GCI 6 Pillar (6) Goods Market Efficiency	GCI6P	no unit of measurement	4.3	4.3	4.6	4.5	4.1
5	GCI index	GCI	no unit of measurement	4.3	4.5	4.6	4.2	4.2
6	GCI 2. Pillar 2. 03 Quality Of Railway Infr. GCI2I		no unit of measurement	3.8	2.9	4.5	4.3	2.7
7	Import	IM	mill. Tons	13,198	32,093	20,833	16,628	5874

Table 2. Influencing attributes applied to rail freight market competitiveness.

Attribute Number	Attribute Name	Contraction	[Unit]	Averages HU	Averages PL	Averages CZ	Averages SK	Averages RO
8	International (Export + Import + Transit)	INT	mill. Tons	35,844	54,763	48,225	38,680	9521
9	Market Concentration	MC	percentage	0.6	0.4	0.7	0.7	0.3
10	Number Of Market Players	NMP	piece	24.0	55.0	27.0	18.0	22.0
11	Rail Share From The Total Vol.	RS	percentage	18.3	16.0	22.4	21.9	22.2
12	Smallest Market Player	SMP	percentage	0.0	1.3	0.5	0.0	0.0
13	Total Rail Traffic	TRT	mill. Tons	48,550	207,619	89,436	46,225	55,418
14	Track Access Fee	TAF	Euro	2.3	4.4	3.8	5.6	3.4
15	Transit Traffic	TT	mill. Tons	9772	3147	7846	11,127	439

Table 2. Cont.

3. Materials and Methods

In our research, countries are entities characterized by 15 attributes represented on different scales. The use of calculating the Euclidian distance between these entities in this high-dimensional space is non-applicable. Therefore, we have to use a method which reduces the number of dimensions of the space in a way that this reduction preserves as much information as possible.

Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert correlated variables into a set of linearly uncorrelated features which are called principal components [29].

The advantage of using PCA is twofold. On the one hand, a large attribute space can be compressed into a smaller space; on the other hand, it can be used in the visualization of the data with simple plots.

As for our case, in Table 3, it is clear that many of the attributes are highly correlated. This means that PCA can be used in order to reduce the dimensionality of the space that is the number of attributes. As a result, we will obtain a lower-dimension space defined by so-called PCA components, features or background variables.

	INT	EX	IM	DT	TT	MC	NMP	TRT	BMP	SMP	RS	TAF	GCI	GCI2P	GCI6P
INT	1	0.95	0.92	0.45	0.42	0.46	0.45	0.62	0.28	0.63	0.00	0.09	0.62	0.34	0.63
EX	-	1	0.84	0.47	0.29	0.42	0.49	0.63	0.23	0.61	0.01	-0.02	0.71	0.33	0.62
IM	-		1	0.74	0.08	0.18	0.63	0.85	-0.03	0.85	0.08	0.12	0.60	0.03	0.49
DT	-	-	-	1	-0.6	-0.43	0.80	0.98	-0.61	0.90	0.16	-0.06	0.36	-0.53	0.05
TT	-	-	-	-	1	0.76	-0.38	-0.40	0.81	-0.32	-0.19	0.13	0.02	0.75	0.40
MC	-	-	-	-	-	1	-0.44	-0.24	0.91	-0.27	0.26	0.49	0.20	0.71	0.48
NMP	-	-	-	-	-	-	1	0.81	-0.57	0.78	-0.14	-0.22	0.40	-0.33	0.24
TRT	-	-	-	-	-	-	-	1	-0.47	0.93	0.14	-0.03	0.45	-0.39	0.18
BMP	-	-	-	-	-	-	-	-	1	-0.5	0.15	0.31	0.06	0.76	0.35
SMP	-	-	-	-	-	-	-	-	-	1	0.01	-0.05	0.51	-0.26	0.21
RS	-	-	-	-	-	-	-	-	-	-	1	0.50	-0.07	-0.06	-0.02
TAF	-	-	-	-	-	-	-	-	-	-	-	1	-0.03	0.09	0.27
GCI	-	-	-	-	-	-	-	-	-	-	-	-	1	0.26	0.73
GCI2P	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0.49
GCI6P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1

Table 3. Correlation matrix (*R*) of the attributes.

The basic idea behind PCA is to compress the high-dimensional space into a lower-dimensional space such that it preserves as much variance as possible. After space compression into a convenient two-dimensional space, we can run a clustering algorithm in order to define clusters which group similar entities. This can help us analyze the connection between the different influencers (random variables) of rail freight market competitiveness and the entities. We expect that the representation in the two-dimensional space, defined by the obtained "background variables", will help us better understand market competitiveness.

In the following, we will present the main steps of the method, and we relate the steps to our practical problem.

Let us suppose we have *n* entities in an \mathbb{R}^p space. Each entity is characterized by *p* variables (X_1, \ldots, X_p) . Based on this, an **X** matrix of *nxp* dimensions can be constructed in which x_{ij} numbers denote the value of the *j*-th variable in the *i*-th entity, the rows correspond to entities, and the columns correspond to variables.

The application of PCA has two main points of interest. The first one is related to the case of the entities when we use the rows of the X matrix. The second one is related to the case of the variables (attributes) in which we deal with the columns of the X matrix.

The fact that the variables are measured on different scales poses a regular problem. To overcome this drawback, we standardize the values as follows. We calculate the arithmetic means and standard deviations of each variable then the standardized values are computed by Formula 1.

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{1}$$

where $\overline{x_i}$ stands for the mean value and s_i for the variance of variable *j*.

Note that the final objective of PCA is substituting $X_1, ..., X_p$ by the principal components $C_1, ..., C_p$ and then reducing the *p* dimension by selecting the principal components with the highest explanation powers (of variance) and omitting the rest.

It is easy to see that the new variables Z_j associated with the columns will have 0 mean and all the standard deviations equal to 1.

3.1. Phase 1. PCA Applied to the Entities

Two entities are similar when they are close to each other in the *p*-dimension space.

In this subsection, we want to find the lower dimension space, where the entities (points associated with the row vectors of X) are projected in the space defined by the principal components such that as much variation as possible is preserved. First, we define the first axis for which the sum of squares of projections is maximal. We denote this by u1, then it can be proved that the two-dimensional plane with this property is determined by u1 and u2 which is orthogonal to u1. The projection of the entities in the plane are called scores. If two scores are close, then the entities have similar properties.

Now we consider the projections of each entity *i* to the component *k*. We denote this by F_{ik} . The vector with coordinates F_{ik} , i = 1, ..., n is denoted by $F_{.k}$, k = 1, ..., p. We obtain, for each component, an *n* dimensional vector since we projected *n* points.

3.2. Phase 1. Variables Represented in the Space Defined by the Principal Components

In this subsection, we express the correlation between $x_{.k}$, k = 1, ..., p, which are the column vectors of the **X** matrix corresponding to the variables (attributes) and $F_{.k}$, k = 1, ..., p. If we project only into the two-dimensional space defined by the two principal components, then we calculate the correlations between $x_{.k}$, k = 1, ..., p and $F_{.1}$, respectively $x_{.k}$, k = 1, ..., p and $F_{.2}$. In this way, variable $x_{.k}$ has the coordinates $r(F_{.1}, x_{.k})$ respectively $r(F_{.2}, x_{.k})$. This means that the variables are projected inside a circle with a unit radius of 1 known as the correlation circle.

Since one of the main objectives of PCA is to determine which variables are correlated to each other, we need the correlations matrix between the variables. The covariance of the standardized *j*-th variable and the standardized *l*-th variable is:

$$s_{ij} = \frac{1}{n} \sum_{i=1}^{n} z_{ij} z_{il} j, \ l = 1, \dots p$$
⁽²⁾

The strength of the linear relation between two variables (attributes) of a particular phenomenon is expressed by the correlation coefficient:

$$r_{jl} = \frac{s_{jl}}{s_j s_l} \tag{3}$$

where $0 \leq |r_{il}| \leq 1$.

The closer the absolute value of r_{jl} lies to the unit, the stronger the linear relationship between the two variables is. This is the basis behind formulating the model through linear functions. From a geometrical point of view, the correlation between the centered variables represents the cosine of the angle θ_{kl} between the vectors associated with the two variables.

$$\cos\left(\theta_{kl}\right) = r(x_l, x_k) \tag{4}$$

The problem is that these variables are in a high-dimensional space. So, it is impossible to interpret the angles and therefore the closeness between the variables. The idea is to project these variables onto a two-dimensional space. Based on their projection and with the help of the angles we obtain between the projections, we can obtain information on which of the variables are "close" to each other.

We emphasize here that only those variables which have an associated vector width and length close to 1 are deemed well represented. If not, then those variables are not well represented and the angles which appear are not meaningful. We used this in our data analysis.

The low-dimensional space can also be created in a way that it retains as much variance as possible. Let us consider the correlation matrix:

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \dots & r_{1l} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2l} & \dots & r_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ r_{j1} & r_{j2} & \dots & r_{jl} & \dots & r_{jp} \\ \vdots & \vdots & & \vdots & \dots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pj} & \dots & 1 \end{pmatrix}$$
(5)

Accordingly, the correlation matrix and for Formulas (2) and (3) the correlation matrix can also be written as follows:

$$\mathbf{R} = \frac{1}{n} \mathbf{Z}^T \mathbf{Z}$$
(6)

The **R** matrix for the standardized variables will be used to determine the principal components.

3.3. Phase 2. Determination of the Principal Components

The idea is to find the component v1 such that

$$v_1 = \operatorname{argmax}_v \sum_{k=1}^p r(v, z_k)^2 \tag{7}$$

And then find the direction v_2 orthogonal to it and so on. These directions are orthogonal to each other and have the property of maximizing the variance of the projections.

The principal components can be found by singular value decomposition or by diagonalizing the correlation matrix to extract the eigenvectors and the eigenvalues. We denote by u_s the eigenvectors corresponding to the eigenvalues λ_s . Let us note that λ_s represents the variance of the entities projected to the u_s component. The eigenvalues will appear in decreasing order.

3.4. Phase 2. The Representation of the Entities and Variables

The representation of the entities with respect to the attributes are expressed by the scores with respect to the loadings in the following two formulae:

$$F_{is} = \frac{1}{\lambda_s} \sum_{k=1}^p z_{ik} G_{ks}, \text{ coordinate s for entity } i$$
(8)

$$G_{ks} = \frac{1}{\lambda_s} \sum_{k=1}^{p} z_{ik} F_{is} \text{ coordinate } s \text{ of the variable } k$$
(9)

These formulae imply the very important conclusion that entities are on the same side as their corresponding variables with high values.

Some important concluding assertions:

- The entities which have close projections have similar attributes, the entities which are far from each other are different in their subsets of attributes.
- Attributes (variables) that are close to each other correlate in a similar way with the principal components, i.e., their angle with the principal components is similar (see Figure 1).
- The attributes are projected in the subspace of the two principal components such that their length is maximum 1. Variables that have lengths close to 1 are well represented and can be interpreted in this subspace (Figure 1).
- The entities that are on the same side as each other have corresponding variables with high values. If the entities are in the opposite direction to each other, then they have small values for these variables.



Figure 1. Attributes and entities in the space generated by the selected principal components.

Based on these assertions, we can analyze our rail freight-related dataset and formulate some findings.

4. Results

Following the general procedure of the PCA, first, the standardized raw data matrix must be constructed. In our case, the 15 influencing factors, as Table 2 demonstrates, are the variables. So, the value of *p* is 15. The entities are the countries, whose freight market data we gathered from 2008 to 2015 (we were able to obtain a reliable complete dataset only for this period to include all countries

for all 16 factors). So, the number of entities is five times eight (the number of years). Thus n = 40 in our model. All five countries had liberalized their rail freight market by the initial date of the examined period.

Afterwards, we constructed the **R** correlation matrix of the 15 variables, presented in Table 3. Elements under the main diagonal represent the density plots of the pair of random variables. Elements on the diagonal represent the histograms of the univariate variables.

For the scope of this study, the elements above the main diagonal are the most essential; they are the pairwise correlations of the variables.

Table 3 justifies the selection of PCA, since it demonstrates not only the high degree of correlation of the analyzed attributes, but also the complexity of these interrelations. This complexity can be interpreted more appropriately by the reduction in dimensions applying PCA. The data represented in this table are official data of the five Central-European countries through eight years. Strong correlation (over 0.8) could be detected—except for the rather trivial ones—between some pairs of attributes. Interestingly, the percentage share of the smallest market player out of the total rail freight volume of the country (SMP) is in strong correlation with domestic traffic (DT, intensity: 0.9); import (IM, intensity: 0.85) and total rail traffic (TRT, intensity: 0.93). Moreover, the number of market players (NMP) is strongly correlated with domestic traffic (DT, intensity: 0.8) and total rail traffic (TRT, intensity: 0.8). Furthermore, between the percentage of the biggest market player (BMP) and transit traffic (TT, intensity: 0.81), a strong correlation could also be detected. This could suggest that the composition of the rail freight market in terms of the size of its competitors has an impact (or at least interrelation since we do not know the direction of the effect) on domestic, import and transit traffic (Figure 2). It is also exhibited in Table 3, that the directly measured market efficiency (by GCI 6th pillar, GCI6P) is correlated with international and export freights (which was quite expectable), but also interrelated with market concentration (MC) with the third strongest intensity, 0.48. The direct measure of rail infrastructure quality (GCI 2nd pillar, GCI2P) is bounded to total traffic, which is quite obvious. However, there is also significant correlation between GCI2P and market concentration (0.7) and the biggest market player (0.76). Based on the above, PCA is worth performing in order to reveal more complex interrelations of the variables and their connections to the entities.

Applying Formulas (4) and (5) described in the Methodology section, the steps of principal components determination can be conducted. In our case, the 16 influencers are substituted by 16 principal components. After reducing the dimensions, only those principal components remain in the model, which explains the dominance of the total variance (this could be reached by determining the eigenvalues because it can be proven that the standard deviation of the principal components equals the square root of the eigenvalues). Having calculated the unstandardized first and then the standardized principal components, the following was deduced.

Figure 3 shows that the first two principal components explain 67% of the total variance. Thus, selecting these two can be sufficient for the analysis. Therefore, the space generated by PC1 and PC2 (which is a rotated and projected space now compared to the original 15-dimensional space of the raw variables) is applied to describe the interrelations of the variables and entities. The first three principal components would have explained 79.5% of the total variance but the contribution of the third one would only have been 10.8% and the interpretation of the results would have been much more difficult in a three-dimensional space so we decided to consider the first two principal components. Figure 1 exhibits the situation of the 15 attributes in this two-dimensional space.

The PCA technique constructed two axes, PC1 and PC2, and the proximity of the attributes demonstrates the weight of each variable in the linear combination for the construction of PC1 and PC2. Consequently, in the case of PC2, the attributes 'Biggest market player' and 'Market concentration' play the most significant roles (these have the highest PC2 coordinates). Based on this, the "shadow attribute" is most likely the degree of liberalization (not legislative but the real situation). For PC1, the most significant attributes are 'Total rail traffic', 'Smallest market player', 'Import' and 'Domestic traffic'.

Evidently, PC1 "shadow attribute" is the quantity of rail transport, measured in tons. It is interesting that the 'Smallest market player' has significant weight in constructing the PC1 axis.



Figure 2. Positions of data by countries in the principal component space.



Figure 3. Order of the calculated principal components and their explanation of variance.

We note that some conclusions can be drawn by the directions of the vectors in this space. Opposite direction means negative correlation of an attribute on another or between an attribute and an axis. Also, the position of the entity points means positive or negative correlation with the others in the

constructed PCA space. In our case, the Romanian points (marked with light blue color) are remote from the other ones in the negative area of the coordinate space, which means that their data correlate negatively with the other countries' data.

Another remarkable issue in PCA is that in the created space, the position of the attributes shows the degree of their correlation. If two variables are close to each other, that means they are strongly interrelated. Based on this, the strong interrelation of transit traffic and the quality of railway infrastructure is visible and also the connection between the smallest market player and the total rail traffic. The proximity of 'International traffic' and 'Export' indicates that the majority of railway transport in Eastern-Central European countries is motivated by export transportation.

In Figure 1, we consider only the variables for which the associated vectors have lengths close to one (in the current scaling, 0.1). Only those vectors are well represented, so 'Rail share' and 'Track access fee' can be omitted from the PCA analysis, their projection is not significant in the constructed vector space of PC1 and PC2 axis. One can observe that in Figure 3, three main attribute groups can be determined.

The first consists of biggest market player (BMP), transit traffic (TT), quality index of infrastructure (GCI2P), market concentration (MC) and track access fee (TAF) can be also connected here. This group demonstrates best the intensity of freight competition. If the market is very concentrated and the biggest company is very dominating in the C-E region, then high infrastructure quality and low fees could be expected with big volumes of transit traffic.

Market efficiency (GCI6P), international traffic (INT), export (EX), import (IM), rail share (RS) and the global competitiveness index (GCI) constitute the second group of variables. This group can be called international competitiveness indicators, since GCI is a direct measure of it and the others all reflect the global capability of the national rail system. Based on the conducted PCA, however, market liberalization and thus market concentration have no serious impact on the international competitiveness, because there is no significant correlation between this second group and the 'Market Concentration' and 'Number of Market players' attributes. However, increasing market efficiency instead of further liberalization might have a positive impact on the international competitiveness of the countries, which supports the findings of Bougna and Crozet (2016) cited in the Introduction. The third group consists of domestic traffic (DT), smallest market player (SMP), number of market players (NMP) and total rail traffic (TRT). Hence, we can conclude that for national rail freight, the existence of small freight forwarders is essential and has a positive impact on the total rail freight performance. In conclusion, market liberalization implementation should concentrate on motivating the small forwarders in order to increase domestic traffic, but it does not contribute to international competitiveness.

Figure 1 also includes the positions of the examined five Eastern-Central European countries in the two-dimensional principal component space. The results strengthen the findings of Feuerstein et al. [14], who stated that different competitiveness influencers have different impact and importance in EU countries. Based on the attribute effects, the clusterization of the examined national rail freight markets can be completed—for better visualization, we constructed Figure 2.

National characteristics can be easily detected in this figure. Since the rail freight market of Poland is very much related to the other examined countries, in their case domestic factors are dominant. For Slovakia and Hungary, transit traffic is very important. So, their position is situated near the transit factors. The Czech Republic is also situated next to transit-like factors but also close to international competitiveness variables. Romania seems to be in the opposite direction to the variables, meaning that it is characterized by low values for these variables, which can be explained by the relative closeness of its rail freight market (lower proportion of international freight volumes) and different features compared to the other Eastern-Central European countries.

Regarding the recommended national transport policy implications, Figure 1 is also rather telling. For Slovakia and Hungary, measures related to infrastructure developments and market concentration (focusing on the biggest participants of the competition) can be the most effective in terms of raising competitiveness. As for the Czech Republic, market efficiency issues have to be prioritized. In Poland, the situation of small freight forwarders is the most crucial issue and many more small rail freight companies would be necessary. For Romania, further expansion of the railway to international traffic can be recommended.

The analysis revealed the interrelation of market liberalization and international competitiveness in a way that the roles of cause and effect are not trivial. Based on the results of the correlations of influencers and positions of countries, we can state that market liberalization issues do not necessarily cause increased international competitiveness in the rail market. Merely opening the market and motivating new players to enter in the market does not automatically cause higher competitiveness; see the different/opposite cases of Poland and Hungary. Because of the complex relations between railway attributes and the relative remote positions of liberalization and competitiveness influencers demonstrated by PCA, and also considering the significant distance between the examined countries, we emphasize the need for country-specific analyses before taking market liberalization measures.

5. Conclusions

Improving railway markets is essential both from economic and sustainability perspectives, which have been recognized by many states all over the world. However, the best ways of raising rail competitiveness are not trivial and need thorough analysis because a general solution may not exist for the different cases of national markets.

This paper aimed to contribute to the ongoing debate about the relation of rail freight market liberalization and competitiveness by applying the complex multivariate method of PCA. The contradiction between the high expectations of railway market liberalization and the stagnation of the Central-Eastern European rail freight sector has been palpable in recent years. The findings of our analysis can partly explain this discrepancy.

Our results support the findings of Bougna and Crozet in terms of international competitiveness; dealing with market efficiency questions appears to be more important than the drive for liberalization. However, for domestic rail freight markets—especially in countries with big volumes of rail freight and a large national market—focusing on small forwarder companies can be even more advisable.

Rail infrastructure developments contribute most directly to transit traffic and, in these cases, the biggest rail forwarders play the most significant role.

National characteristics of railway markets could be also detected, which supports the robustness of our study results.

Furthermore, the statements of Zunder et al. (2013) [30], that liberalization alone cannot explain the different rail freight performance of the countries in the EU, have also been verified. They found that organizational and managerial bottlenecks strongly determine this performance and, thus, market competitiveness. Our results support this idea as market concentration and market efficiency issues have not been positioned closely in our PCA vector space. Thus, liberalization measures cannot be the only means to improve rail freight performance.

Involving more data from databases of other countries (from other regions beyond the territory of the European Union or analyzing the data of old EU members but regarding a different period due to their earlier liberalization) might be a promising subject of further research and also selecting and integrating other attributes might improve on our findings in the future.

Author Contributions: Conceptualization, S.D. and B.F.; methodology, S.D.; Validation, B.F.; formal analysis, B.F.; investigation, S.D.; resources, B.F.; data curation, B.F.; writing—original draft preparation, S.D.; writing—review and editing, S.D.; visualization, B.F.

Funding: This research received no external funding.

Acknowledgments: We wish to acknowledge the help and support provided by the GYSEV Ltd. and we would like to offer our special thanks to Szilárd Kövesdi, the Chief Executive Officer of GYSEV Ltd.

Conflicts of Interest: The authors declare no conflict of interests. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Table A1. The raw data table and graphical representation of the examined attributes.

	2008	HU	PL	CZ	SK	RO	2009	HU	PL	CZ	SK	RO
INT		41,474	60,443	50,925	40,841	11,665		29,916	43,670	39,856	32,027	5130
EX		12,615	22,369	21,228	11,070	3415		10,257	14,874	18,049	9182	1931
IM		15,345	34,328	21,875	17,538	7773		11,040	26,732	15,807	14,785	2859
DT		10,069	188,417	44,148	7069	45,465		12,362	157,150	36,859	5557	45,967
TT		13,514	3746	7822	12,233	477		8619	2064	6000	8060	340
TRT		51,543	248,860	95,073	47,910	57,130		42,278	200,820	76,715	37,584	51,097
MC		0.658	0.51	0.8123	0.8543	0.2897		0.6249	0.43	0.7915	0.8104	0.2694
BMP		90.1	51.47	89.02	90.5	59		85.6	45.31	88.77	78.14	48.6
SMP		0.012	1.56	0.01	0.0224	0.014		0.018	1.59	0.01	0.0226	0.018
NMP		2	43	23	12	69		2	47	24	15	68
KS		14.98	18.5	23.3	23.4	19		13.94	15.4	22.2	19.6	19.4
		2	4.5	4.1	9	3.9		2.7	4.4	4.2	10.3	3.3
CCUP		4.2	4.5	4.7	4.5	4.1		4.5	4.5	4.0	4.2	4.2
GCI2P CCI4P		4	4.3	4.3	3.9	4.2		4.4	5.6	4.0	4.2	3.4 4.1
00101	2010	HU	4.5 PL	4.0 CZ	SK	RO	2011	HU	PI	4.0 CZ	SK	RO
INT	2010	34,396	54 257	45.822	37.918	6964	2011	36.661	58 252	46 893	36 701	8721
EX		11.859	17.709	19,748	11.166	2739		14.068	16.844	19,401	10.407	3354
IM		12,768	32.816	18,790	17.142	3842		13,728	38,183	19,391	15.825	4903
DT		11.398	162.510	37.078	6409	52.003		10,763	190.354	40.203	7009	47.637
TT		9769	3732	7284	9610	383		8865	3225	8101	10,469	464
TRT		45,794	216,767	82,900	44,327	58,967		47,424	248,606	87,096	43,710	56,358
MC		0.657	0.43	0.7586	0.7891	0.2421		0.5768	0.41	0.717	0.743	0.2691
BMP		80.53	54.31	86.84	37.2	45.3		75.249	56.37	84.36	77.84	45.23
SMP		0.021	1.35	0.1	0.002	0.01		0.025	1.43	0.36	0.0229	0.0111
NMP		16	48	24	16	-		19	57	28	14	-
RS		16.35	15.4	21	22	23.5		17.66	16.9	20.7	20.9	28
TAF		2.7	4.4	4.3	10.3	3.3		2.18	4.1	4.2	4.1	3.2
GCI		4.4	4.5	4.5	4.2	4.1		4.3	4.5	4.5	4.1	4.1
GCI2P		3.6	2.5	4.5	4.5	2.4		3.5	2.4	4.6	4.5	2.2
GCI6P	0010	4.3	4.4	4,6	4.4	4	0010	4.3	4.4	4.5	4.4 CK	3.9
INT	2012	HU 25.229	PL 52.070	4E 014	5K	RO 8120	2013	HU 26.622	PL 53.332	CZ	5K 40.210	KO 0081
		33,328 13 796	52,970	45,914	36,243	8120 2784		36,623	52,223	46,687	40,219	3259
IM		12 233	32 127	18,698	15 135	4938		12 750	29 296	20 318	16 693	5405
DT		11 556	177 908	37 054	6356	41 268		12,750	18 0412	37 270	8182	41 549
TT		9299	3120	8117	11.256	398		9607	2995	7557	11 798	417
TRT		46.884	230.878	82.968	42,599	49.388		49.084	232.635	83.957	48,401	50.630
MC		0.523	0.38	0.6387	0.6947	0.2782		0.4703	0.36	0.5917	0.539	0.251
BMP		71.57	50.42	79.38	73.48	43.59		67.581	48.64	76.33	70.65	42.4
SMP		0.08	1.11	0.3	0.0251	0.01		0.03	1.2	0.92	0.0203	0.02
NMP		47	58	30	14	22		52	69	32	17	67
RS		18.77	15	21.8	19.8	24.2		19.24	14.6	20.3	21.4	21.9
TAF		2.2	4.2	3.8	3.3	3.2		2.2	3.1	3.8	2.8	3.2
GCI		4.3	4.5	4.5	4.1	4.1		4.3	4.5	4.5	4.1	4.3
GCI2P		3.5	2.4	4.6	4.5	2.2		3.8	2.9	4.5	4.4	2.9
GCI6P		4.3	4.4	4.5	4.4	3.9		4.4	4.5	4.5	4.4	4.2
-	2014	HU	PL	CZ	SK	RO	2015	HU	PL	CZ	SK	RO
		35,573	54,258	50,900 18 508	40,563	9190		35,923	52,611	18 828	39,302	11,875
EA IM		13,940	20,074	10,390	16,299	5095		12,017	20,000	10,030	11,743	4324
DT		12,731	173 563	23,372 40.656	10,956	43 431		14,903	29,303	42,467	8055	43 500
		8902	2660	40,000 8930	12 306	193		10.003	2566	42,009 8886	11 549	43,300 827
TRT		50,593	227 821	91,556	50,997	52,621		50,332	224 320	97,280	47,357	55,375
MC		0.554	0.34	0.5076	0.507	0.2681		0.5093	0.33	0.4585	0.5483	0.2577
BMP		73.4	47.94	69.9	70.63	41.4		70.29	47.48	66.52	68.35	39.78
SMP		0.08	1.21	0,94	0.0237	0.03		0.014	1.36	0.94	0.0256	0.02
NMP		52	69	32	16	67		52	72	43	16	68
RS		18	14.4	-	-	-		19.72	14	-	-	-
TAF		2.3	3.3	1.8	1.8	3.12		2.2	2.9	3.6	2.5	3.14
GCI		4.2	4.5	4.7	4.2	4.3		4.2	4.6	4.7	4.3	4.3
GCI2P		3.8	3.1	4.5	4.6	2.8		3.5	3.3	4.6	4.6	2.4
GCI6P		4.3	4.5	4.6	4.4	4.3		4.4	4.6	4.7	4.5	4.2











Figure A3. Transit Traffic (mill Tons).

14 of 19





15 of 19











Figure A9. Track Access Free (€).







Figure A11. GCI Index.



Figure A12. GCI 2. Pillar 2. 03 Quality Of Railway Infr.

17 of 19





References

- Xin, X.; Li, M. Correlation Analysis of China's Urban Rail Transit Industry. *Period. Polytech. Transp. Eng.* 2017, 45, 148–156. [CrossRef]
- 2. Abramović, B.; Šipuš, D. *A Case Study of the Railway Engineering Programme; Sustainable Rail Transport;* University High Education in Croatia; Springer: Berlin/Heidelberg, Germany, 2019; pp. 101–109.
- 3. Cantos, P.; Pastor, J.M.; Serrano, L. Vertical and horizontal separation in the European railway sector and its effects on productivity. *J. Transp. Econ. Policy* **2010**, *44*, 139–160.
- 4. Bergantino, A.; Capozza, C.; Capurso, M. The impact of open access on intra- and inter-modal rail competition. A national level analysis in Italy. *Transp. Policy* **2015**, *39*, 77–86. [CrossRef]
- 5. Holvad, T. Review of railway policy reforms in Europe. Built Environ. 2009, 35, 24–42. [CrossRef]
- 6. Nash, C.A. Passenger railway reform in the last 20 years-European experience. *Res. Transp. Econ.* **2008**, *22*, 61–70. [CrossRef]
- 7. Bougna, E.; Crozet, Y. Towards a liberalized rail transport: Analyzing and modelling the impact of competition on productive efficiency. *Res. Transp. Econ.* **2016**, *59*, 358–367. [CrossRef]
- 8. Dolinayova, A.; L'och, M.; Camaj, J. Liberalization of the railway freight market in the context of a sustainable transport system. *Transp. Res. Procedia* **2016**, *14*, 916–925. [CrossRef]
- 9. SCI/Verkehr Berlin. *European Rail Freight Market—Developments, Volumes, Players;* Extract from the study; SCI Verkehr Gmbh: Berlin, Germany, 2016.
- 10. Feuerstein, L.; Busacker, T.; Xu, J. Factors influencing open access competition in the European long-distance passenger rail transport—A Delphi study. *Res. Transp. Econ.* **2018**, *69*, 1–10. [CrossRef]
- 11. Jolliffe, I.T.; Cadima, J. Principal Component Analysis: A review and recent developments. *Phylosophical Trans. R. Soc.* **2016**, 374, 1–16. [CrossRef] [PubMed]
- 12. Fifield, S.G.M.; Power, D.M.; Sinclair, C.D. Macroeconomic factors and share returns: An analysis using emerging market data. *Int. J. Financ. Econ.* **2002**, *7*, 51–62. [CrossRef]
- Hansut, L.; David, A.; Gasparik, J. The Critical Path Method as the Method for Evaluation and Identification Of The Optimal Container Trade Route Between Asia And Slovakia. *Bus. Logist. Mod. Manag.* 2017, 17, 29–42.
- 14. Wiegmans, B.; Janic, M. Analysis, modeling and assessing performances of supply chains served by long-distance freight transport corridors. *Int. J. Sustain. Transp.* **2019**, *13*, 278–293. [CrossRef]
- 15. Sakalys, R.; Sivilevicius, H.; Miliauskaite, L.; Sakalys, A. Investigation and evaluation of main indicators impacting synchromodality using ARTIW and AHP methods. *Transport* **2019**, *34*, 300–311. [CrossRef]
- 16. Simlar, L.; Wilson, P.W. Estimating and bootstrapping Malmquist indices. *Eur. J. Oper. Res.* **1999**, *115*, 459–471. [CrossRef]
- 17. Purwanto, A.J.; Heyndrickx Kiel, J.; Betancor, O.; Socorro, M.P.; Hernandez, A.; Eudenio-Martin, J.L.; Pawlowska, B.; Borkowski, P.; Fiedler, R. Impact of transport infrastructure on international competitiveness of Europe. *Transp. Res. Procedia* **2017**, *25*, 2877–2888. [CrossRef]
- 18. D'ovidio, F.D.; Leogrande, D.; Mancarella, R.; Schinzano, A.; Viola, D. A multivariate analysis on the quality of transport services. *Procedia Econ. Financ.* **2014**, *17*, 238–247. [CrossRef]

18 of 19

- 19. Vickerman, R. Measuring changes in regional competitiveness: The effects of international infrastructure investments. *Ann. Reg. Sci.* **1989**, *23*, 275–286. [CrossRef]
- 20. Odolinski, K.; Wheat, P. Dynamics in rail infrastructure provision: Maintenance and renewal costs in Sweden. *Econ. Transp.* **2017**, *14*, 1–10. [CrossRef]
- 21. Aditjandra, T.P. Europe's freight transport policy: Analysis, Synthesis and Evaluation. *Adv. Transp. Policy Plan.* **2018**, *1*, 197–243.
- 22. Marschnig, S. Innovative track access charges. Transp. Res. Procedia 2016, 14, 1884–1893. [CrossRef]
- 23. Matsumoto, A.; Merlone, U.; Szidarovszky, F. Some notes on applying the Herfindahl-Hirschman Index. *Appl. Econ. Lett.* **2012**, *19*, 181–184. [CrossRef]
- 24. Crozet, Y. Rail freight development in Europe: How to deal with a doubly imperfect competition. In Proceedings of the World Conference on Transport Research WCTR 2016, Shanghai, China, 10–15 July 2016.
- 25. Gathon, H.J.; Pestieau, P. Decomposing efficiency into its managerial and its regulatory components: The case of European railways. *Eur. J. Oper. Res.* **1995**, *80*, 500–507. [CrossRef]
- 26. Oum, T.H.; Yu, C. Economic efficiency of railways and implications for public policy: A comparative study of the OECD countries' railways. *J. Transp. Econ. Policy* **1994**, *28*, 121–138.
- 27. Jarzemskis, A.; Jarzemskiene, I. Comparison of rail freight transportation markets in Lithuania and Poland. *Procedia Eng.* **2017**, *187*, 492–497. [CrossRef]
- 28. Cemberci, M.; Civelek, M.E.; Canbolat, N. The moderator effect of Global Competitiveness Index on dimensions of Logistics Performance Index. *Procedia Soc. Behav. Sci.* **2015**, *195*, 1514–1524. [CrossRef]
- 29. Szelényi, L. *Multivariate Methods of Econometrics;* Educational Material; Szent István University Faculty of Economics and Social Sciences: Gödöllő, Hungary, 2009.
- 30. Zunder, T.H.; Islam, D.Z.; Mortimer, P.N.; Aditjandra, P.T. How far has open access enabled the growth of cross border pan European rail freight? A case study. *Res. Transp. Bus. Manag.* **2013**, *6*, 71–80. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).