Typology of multinationals in Austria: CESEE focus and foreign control as distinct features

Thomas Cernohous, Tomáš Slačík[1]

Multinational enterprises (multinationals) play an important role in every economy as they tend to be larger, more capital- and R&D-intensive, more productive and more integrated in global value chains than domestic enterprises. Focusing on multinationals active in Austria, this paper discusses essentially two research questions: Can we categorize Austrian units of multinationals in consistent groups? And can these groups be characterized by meaningful variables? To address these questions, we undertake a microdata-linking exercise to build a comprehensive dataset of multinationals in Austria and use adequate clustering techniques to identify homogeneous and distinct groups without imposing any prior knowledge regarding the number of such groups or their features. This approach enables us to characterize more than 2,500 multinationals in Austria and meaningfully identify eight types of multinationals, the main grouping factors being (1) foreign or Austrian control, (2) special purpose entity versus other form of company, (3) the share of outward investment in Central, Eastern and Southeastern Europe (CESEE) and (4) the degree of trade openness. With this basic research work, we open up a wide range of questions that may serve as the basis for future (applied) analytical work.

JEL classification: C49, F13, F14, F15, F21, F36, F41
Keywords: multinational enterprises, cluster analysis, globalization, partitioning around medoids, typology

> *"Multinational enterprises (MNEs) are a key channel of globalisation. They serve as the backbone of many global value chains by linking and organizing production across countries and are an important channel for exchanging capital, goods and services, and knowledge across countries. Foreign direct investment (FDI) is necessary for the creation of an MNE."*
> *(OECD, 2015)*

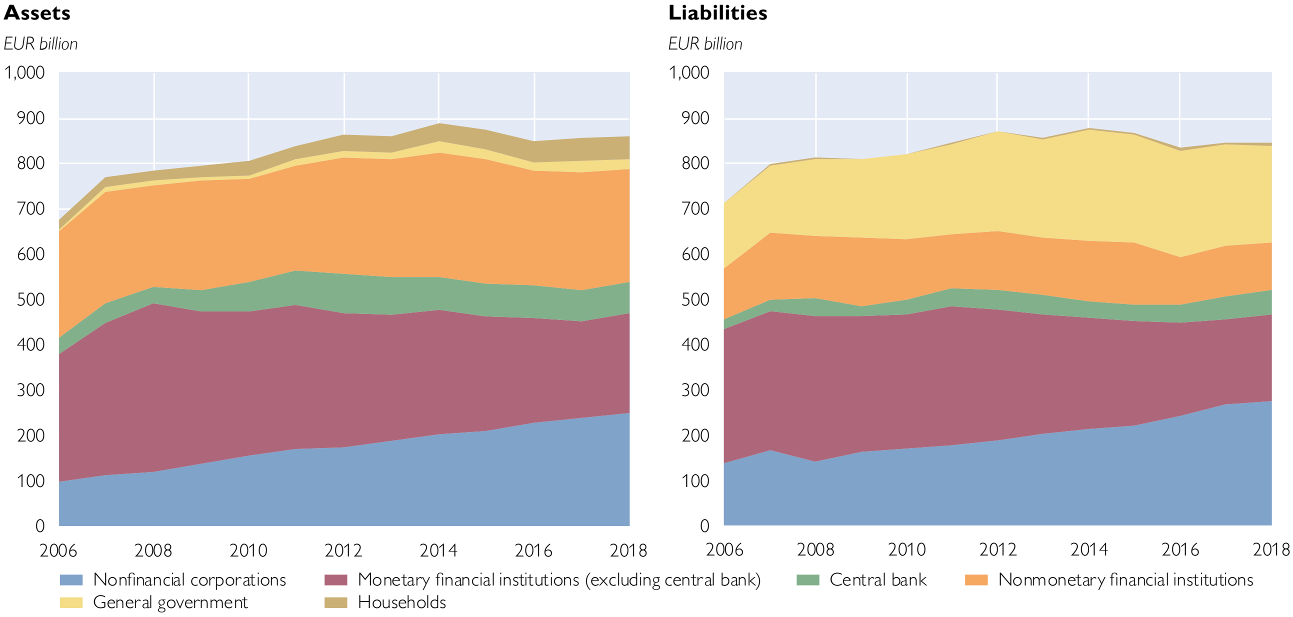
Today, multinationals account for a third of world output and two-thirds of international trade (De Backer et al., 2019). Since 2000, the global output of multinational enterprises has more than tripled (OECD, 2018). With this degree of economic power, multinationals have become a veritable political force (Kim and Milner, 2019).

Most multinational enterprises are classified in the nonfinancial corporate sector, whose fast-growing role is evident from international investment statistics (see chart 1 with data for Austria).

[1] *Oesterreichische Nationalbank, External Statistics, Financial Accounts and Monetary and Financial Statistics Division, thomas.cernohous@oenb.at; Foreign Research Division, tomas.slacik@oenb.at. Opinions expressed by the authors do not necessarily reflect the official viewpoint of the OeNB or the Eurosystem.* *The authors would like to thank Julia Wörz, Kujtim Avdiu and Martin Feldkircher (all OeNB) and an anonymous referee for helpful comments and valuable suggestions.*

Chart 1

External assets and liabilities in Austria by domestic sector



Source: OeNB.

This paper sheds some light on the relevance and role of multinational enterprises active in Austria. After all, a better understanding will enable better informed and more targeted decision-making and thus more efficient and effective policymaking on issues like regional development, employment or taxes. Our approach is to first identify those Austrian companies that satisfy the OECD definition of multinational enterprises with a view to compiling a dataset covering all Austrian companies that are on the giving or receiving end of foreign direct investment (FDI). For every company with a direct investment relationship, i.e. with a cross-border equity participation of at least 10%, we then identify further Austrian companies that are part of the same multinational group and assign them to the initial enterprise. Hence, we build a dataset of so-called "truncated enterprise groups"[2] of multinationals in Austria, for which we use the term "multinationals" in this paper for the sake of readability. To populate the database further, we link a set of variables, both numerical and categorical, from different statistical areas and sources to the individual enterprises and calculate aggregated values for each multinational.

In a second step, we partition this dataset into homogeneous and distinct groups. Rather than imposing prior knowledge about the number of such groups or their features, we want to "let the data speak" and suggest meaningful subsets of the dataset. We do this by breaking the dataset into clusters using the "partitioning

[2] *As defined in Regulation (EC) No. 177/2008 of the European Parliament and of the Council of 20 February 2008 establishing a common framework for business registers for statistical purposes, Article 2(e).*

around medoids" algorithm, which is a rather robust and well-designed method for clustering mixed datasets (see annex).[3]

The resulting dataset contains more than 2,500 multinationals consisting of 3.6 enterprises on average. The clustering algorithm allows us to meaningfully identify eight types of multinationals, with the key grouping factors being foreign or Austrian control; special-purpose entity (SPE) versus other company forms; the share of outward investment in Central, Eastern or Southeastern Europe (CESEE); and the degree of trade openness.

The remainder of the paper is structured as follows. In section 1, we describe how we set up the dataset and which data sources we used. In section 2, we explain the clustering algorithm and our choice of key input parameters. Following analysis of the results in section 3, we interpret the results and suggest further avenues of research in section 4. Section 5 concludes.

1 Data sources

To build a viable dataset, we used 2017 data on different aspects of globalization to (1) identify any multinational enterprises operating in Austria, (2) link these data with appropriate microdata and (3) calculate input variables *t* for the clustering algorithm. Generally speaking, we combined foreign direct investment (FDI) statistics[4] with foreign affiliates statistics (FATS), as FDI statistics typically cover only financial items and employment data for *direct* investment relationships whereas FATS statistics also cover *indirect* controlled companies. In addition, we used the Austrian business register data as well as international trade, balance of payments and international investment statistics data.

1.1 Identifying multinationals in Austria

Identifying the population of multinationals in Austria[5] meant that we were looking for

- any domestic companies with at least one direct investment relationship to another economy; or
- any groups of domestic companies controlled by a domestic group head, of which at least one unit maintains an FDI relationship with another economy.

We compiled this information from the OeNB's annual survey on inward and/or outward FDI[6] and from the Austrian business register. The OeNB's annual FDI survey, which is the main building block of FDI statistics in Austria, identifies the "entry points"[7] of foreign investors in Austria and serves as a starting point for

[3] *The term medoid refers to an object within a cluster for which average dissimilarity between it and all the other the members of the cluster is minimal. It corresponds to the most centrally located point in the cluster. These objects (one per cluster) can be considered as a representative example of the members of that cluster. www.datanovia.com/en/lessons/k-medoids-in-r-algorithm-and-practical-examples/*

[4] *Compiled in Austria since 1968. Direct investment relationships have deepening economic effects on involved economies.*

[5] *Technically speaking "truncated enterprise groups."*

[6] *For details, see the handbook on the balance of payments and the international investment position according to BPM6 rules published by the OeNB at www.oenb.at/dam/jcr:b46b2770-83c9-4281-9f20-bcb73d86c8e8/ZABIL-Handbuch_V1.0.pdf (so far available in German only, English version scheduled for August 2020).*

[7] *For example: A German investor has a 100% subsidiary in Austria. This entity has 100% stakes in two further domestic companies on its own, which are hence indirectly controlled by the German group head. The Austrian company in the middle of this participation chain is defined as "entry point" in Austria, since the two indirectly controlled entities can only be identified if this "entry point" is known.*

identifying connected group members in Austria. The survey pools answers from approximately 3,000 respondents, either inward FDI respondents or domestic companies[8] that hold outward FDI. The defining element for selection of companies into our database of multinationals in Austria was a controlling participation chain (50%+ of voting rights on each step in the chain) back to the Austrian group head. In a second step, we relied on the Austrian business register to identify domestic relationships between companies. Based on end-2017 data, we thus identified 2,555 multinational groups operating in Austria with a total of 9,096 companies, which yields an average group size of 3.6 enterprises.

1.2 Using microdata-linking to enrich the database

In a next round, we were able to enrich the multinationals database with microdata from other statistics, using the business register number and an identifier issued and managed by the OeNB as the connecting link. Specifically, we added selected variables from the following sources:

- Annual FDI survey
- Business register
- Structural business statistics
- Foreign trade statistics
- Services according to the balance of payments
- External statistics compilation system

The OeNB's annual FDI survey served to provide structural information and enabled us to source data on direct investment itself, especially regional breakdown details and data on intracompany loans. From the business register we extracted variables on economic activity and age. Structural business statistics provided figures on employment and turnover. Foreign trade statistics provided us with microdata on exports and imports of goods (global values, no regional breakdown). Services according to the balance of payments were available on a more granular level, allowing us to form the following service groups: technological services, financial and insurance services, and other services. Finally, the OeNB's external statistics compilation system offers the opportunity to calculate assets and liabilities for "other investment" and to some extent "portfolio investment" and "financial derivates" at the company level.[9]

2 Data clustering with partitioning-around-medoids (PAM) algorithm

2.1 Selection and weighting of input variables

For reasons detailed in the annex, we picked the "partitioning around medoids" (PAM) algorithm from the cluster analysis toolbox to divide our dataset into meaningful clusters. Intuitively,[10] the PAM algorithm proceeds in the following iterations:

1. The starting point are a set of *k* random observations in the dataset. These observations, called medoids, represent centers of *k* clusters which, at this point, consist of single observations.

[8] *Some Austrian-controlled multinationals may also include individuals (as the respective group's head).*

[9] *For portfolio investment assets, the use of microdata was limited: only banks' own holdings were available.*

[10] *For a more detailed technical description of the PAM method which would go beyond the scope of this paper, see e.g. Kaufman and Rousseeuw (1987).*

2. All observations are (re-)assigned to their closest medoid.
3. In each of the clusters thus built, the algorithm finds the observation that would yield the lowest average distance to its cluster members. If this is a different observation than the one in step 1, this observation becomes the new medoid.
4. If at least one out of *k* medoids has changed, the algorithm goes back to step 2; otherwise the process ends.

The variables to be entered into the algorithm need to be selected with caution to avoid the presence of noisy noninformative and/or redundant, correlated variables, which may produce multicollinearity (Fraiman et al., 2009). Unlike in regression analyses where multicollinearity spoils the beta coefficients, in clustering multicollinearity implies that some variables get a higher weight than others. As Sambandam (2003) puts it: "If two variables are perfectly correlated, they effectively represent the same concept. But that concept is now represented twice in the data and hence gets twice the weight of all the other variables. The final solution is likely to be skewed in the direction of that concept."

Hence, it is crucial to strike a balance between including all major variables that are of interest in clustering the data, and not choosing too many[11] and/or highly correlated variables. Table A1 in the annex displays and describes the 20 variables that we carefully selected as input for the PAM algorithm. While the focus lay on numerical characteristics, three binary attributes were assessed by expert judgment to be crucial for grouping multinationals, namely "SPE" (special purpose entity), "FOREIGN_CONTROL" (multinationals controlled by nonresidents) and "BANK" (one of the units is classified as a bank). In other words, we created a mixed-type dataset consisting of numerical and categorical data. In addition, while not being part of the clustering procedure, other nominal scaled attributes were important for the ensuing analysis and interpretation of results. In particular, the variables economic activity[12] and controlling region[13] were of high explanatory value.

In weighting the variables, we basically followed the concept of equal weights, assigning each variable a weight of 100% divided by the number of variables. Exceptions were made only for the numerous balance of payments/international investment position variables[14] because of their high correlation to each other (see chart 7 in the annex), and the fact that they cover similar aspects (external funding). To avoid the excessive influence issues described above, these attributes were assigned only one-quarter of the weight other variables have.

[11] *The variable space can be reduced by dimension reduction techniques such as the principal component analysis (PCA) (Fraiman et al., 2009). However, after some experiments we decided not to go that way for two reasons. First, principal components (i.e. linear combinations of variables) that result from the PCA are difficult to interpret, which is impractical if our aim is to identify and describe types of multinationals. Second, and more importantly, dimensions which explain the maximum variation in the data and are thus retained by the PCA need not necessarily be the same dimensions that are decisive for clustering the data.*

[12] *Predominant economic activity of multinationals in Austria.*

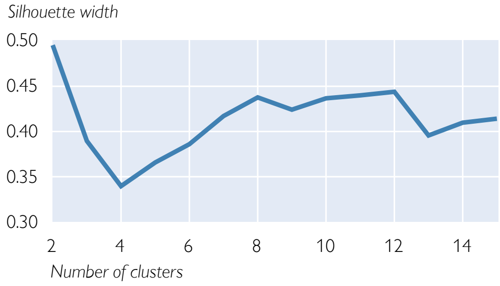
[13] *World region of a multinational's ultimate controlling unit.*

[14] *ODI, IDI, OI_A, OI_P, FININS_EXP, FININS_IMP, TECH_EXP, TECH_IMP, OtherS_EXP, Others_IMP.*

2.2 Choice of key input parameters

Chart 2

Silhouette plot



Source: Authors' calculations.

For the PAM function, some key parameters need to be defined as inputs, in particular a distance metric and the number of clusters. For the distance matrix, feeding the algorithm with the data matrix and some gauge of dissimilarity would have sufficed if we had worked with numerical data alone. As we used mixed data, we had to provide a dissimilarity matrix directly. In line with common practice, we used the so-called Gower distance matrix (see annex for details).

Regarding the number of clusters *k* to be defined, a large number of methods and indices has been proposed in the literature for identifying the optimal number of clusters (Mirkin, 2011). Yet, most of these indices and evaluation methods are not applicable to mixed data. For this reason, we picked one of the few indicators available also for mixed data, the popular silhouette plot, which indicates the so-called silhouette width for a given number of clusters. The silhouette width is a normalized ratio between the average dissimilarity within clusters relative to the nearest neighboring cluster. It is normalized to a range between –1 and 1, with values closer to 1 suggesting good clustering. The purpose of the silhouette plot is to find the relative maximum silhouette width for a reasonable range of possible numbers of *k*.[15] In our case, the silhouette plot shown in chart 2 suggested either eight or – with an even slightly better value – twelve as the optimal number of clusters.[16] While we had a close look at both suggestions, we considered eight clusters to be the more reasonable choice, since interpreting and comparing four more clusters bears the risk of losing focus.

3 Results

3.1 Clustering multinationals in Austria

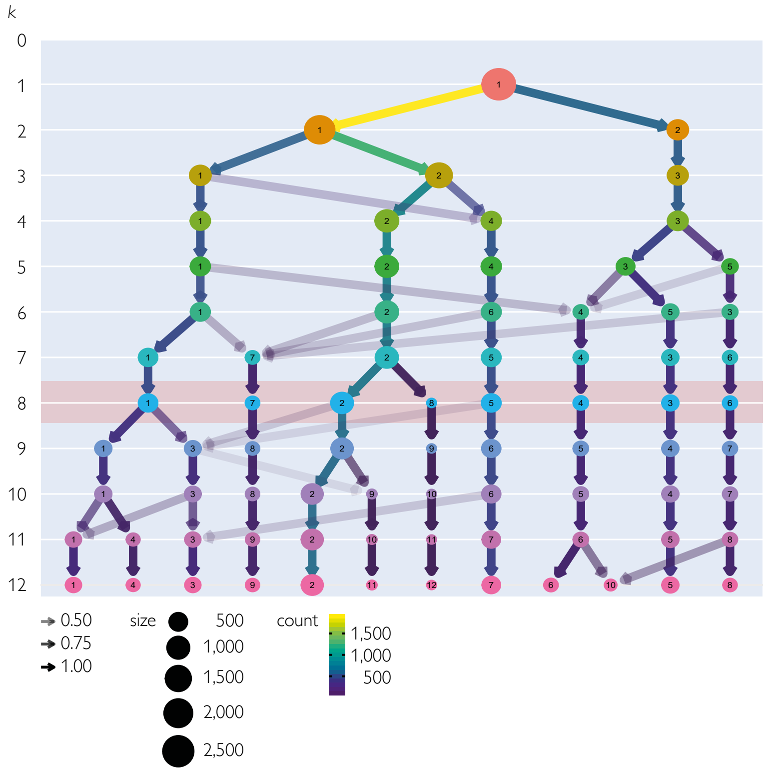
The statistical methods to determine the optimal number of clusters mentioned above typically consider just one cluster at a time. An alternative or rather complementary perspective is to look at how samples move as the number of clusters increases, to gain insights into how homogeneous and (un)stable the clusters are. In the clustering tree shown in chart 3, each line represents the clustering results of the algorithm we applied with a given number of clusters (*k*). The size of the dots reflects the size of each group, while the arrows indicate relevant movements of multinationals to other clusters at the next resolution level. So, at the first node we see the original sample split into cluster 1 consisting of 651 multinationals (yellow arrow) and cluster 2 with 1,904 multinationals (blue arrow). The darker the color of the arrow, the higher the absolute number of observations that move, and vice versa. The degree of transparency of the arrow visualizes the relative

[15] *For details see e.g. Rousseeuw (1987).*

[16] *The highest value is actually at k2, but for analytical reasons a mere two clusters do not provide for adequate granularity.*

Chart 3

Clustering tree at different resolutions of the PAM algorithm



Source: Authors' calculations.

Note: The area boxed in red highlights the resolution of the main results.

importance of observation movements, i.e. the share of group members that move to another group. No transparency indicates that all cluster members move to the group to which the arrow points (e.g. cluster 7 at $k = 9$ and cluster 7 at $k = 10$).[17]

On initial eyeballing, we see one major observation that we can take away from chart 3: as the resolution level k increases, the typical pattern is that a cluster splits up into two. The number of elements moving to the newly built groups from other clusters is fairly limited; or put differently, the remaining clusters stay relatively stable as k rises. This implies that the clustering algorithm is rather robust; otherwise we would see a lot more reshuffling among clusters at each k.

While it has to be borne in mind that the PAM algorithm does not proceed in the iterative, hierarchical way suggested by the clustering tree in chart 3, this tree may, nonetheless, be interpreted as a dynamic decision tree. At each node, we can identify the variable(s) that cause(s) a cluster to split off. How do we do that? At each splitting node, we compute for each variable the standardized difference between the average values of the respective variable in the two subsequent clusters:

$$Diff_v = \frac{\bar{v}_{kC_x}}{SD_v} - \frac{\bar{v}_{kC_y}}{SD_v}$$

where $\bar{v}_{kC_x}$ and $\bar{v}_{kC_y}$ denote the average value of variable v at the clustering level k for, respectively, clusters C_x and C_y, which are the two clusters descending from the same parent cluster at the previous clustering level $k - 1$. Furthermore, SD_v denotes the standard deviation of variable v for the entire dataset. The variable for which $Diff_v$ is highest at a given clustering level k is the variable that causes a new cluster to branch off.

See table 1 for the results and the table rows for the key variables triggering the split, highlighted with the darkest color. For example, when we look at table 1 in combination with chart 3, we see that at clustering level 2 (i.e. $k = 2$), for the split into the two resulting clusters 1 and 2, it is the variable FOREIGN_CONTROL that makes the biggest difference. At the next level ($k = 3$), cluster 1 splits up into

[17] *The order of the clusters was randomly assigned by the PAM algorithm rather than following specific criteria.*

Table 1

Standardized mean value difference between new clusters at subsequent clustering levels

			AW	AX	AZ	BA	BD	O	AU	BC	BB	C	D	G	H	P	Y	Z	AA	AB
k	Cx	Cy	Δ SPE	Δ FOREIGN_CONTROL	Δ CESEE_SHARE_ODI	Δ CESEE_SHARE_IDI	Δ BANK	Δ EMP	Δ AGE	Δ IMP_QUOTA	Δ EXP_QUOTA	Δ ODI	Δ IDI	Δ OI_A	Δ OI_L	Δ TURN	Δ FININS_EXP	Δ FININS_IMP	Δ TECH_EXP	Δ TECH_IMP
2	1	2	0.2	2.3	0.7	0.1	0.1	0.2	0.3	0.4	0.0	0.1	0.1	0.1	0.1	0.2	0.1	0.0	0.0	0.1
3	1	2	0.3	0.2	0.1	0.1	0.2	0.1	0.2	1.0	1.9	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.3	0.2
4	2	4	0.4	0.0	0.0	0.2	0.2	0.0	0.4	1.9	0.3	0.1	0.2	0.1	0.1	0.0	0.1	0.1	0.0	0.1
5	3	5	0.0	0.2	2.7	0.1	0.2	0.2	0.5	0.3	0.1	0.0	0.0	0.3	0.2	0.3	0.0	0.0	0.0	0.0
6	4	5	0.0	0.0	0.5	0.2	0.2	0.3	0.4	1.2	2.0	0.1	0.0	0.0	0.0	0.5	0.2	0.0	0.2	0.1
7	1	7	0.1	0.0	2.9	0.1	0.3	0.4	0.4	0.4	1.5	0.1	0.1	0.0	0.1	0.2	0.0	0.1	0.0	0.0
8	2	8	7.1	0.0	0.1	0.3	0.2	0.1	0.9	0.2	0.1	0.8	0.7	0.1	0.1	0.1	0.1	0.1	0.0	0.1
9	1	3	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.8	1.2	0.0	0.1	0.0	0.0	0.1	0.0	0.2	0.0	0.1
10	2	9	0.0	0.1	0.1	6.0	0.1	0.1	0.5	0.3	0.2	0.4	0.5	0.0	0.1	0.0	0.1	0.0	0.1	0.1
11	1	4	0.0	0.0	0.2	0.2	0.0	0.1	0.1	1.9	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1
12	6	10	0.0	0.0	2.1	0.0	0.0	0.1	0.4	0.2	0.4	0.1	0.0	0.0	0.1	0.2	0.1	0.0	0.1	0.0

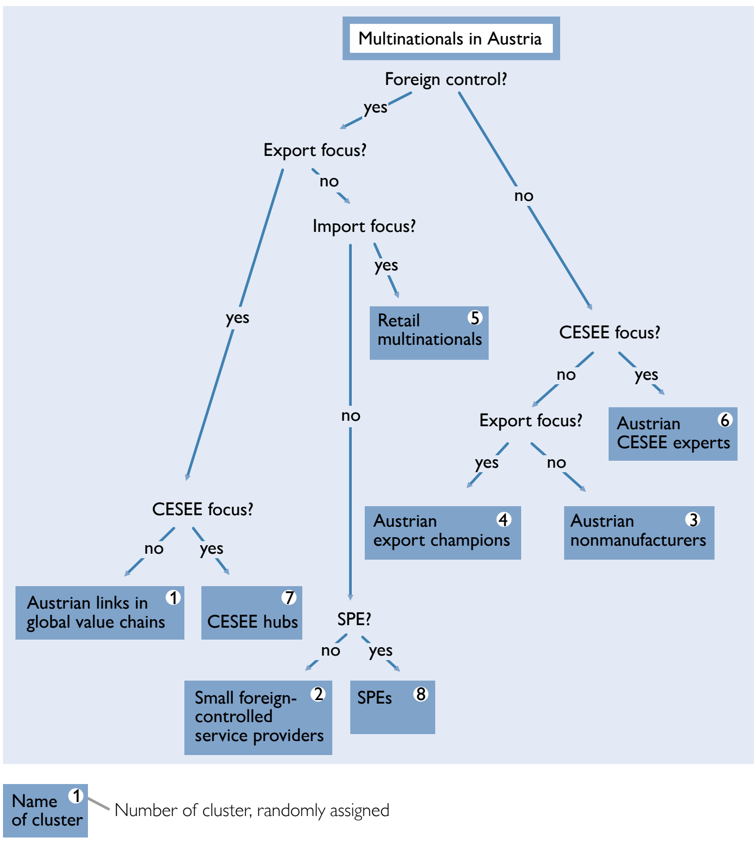
Source: Authors' calculations.

cluster 1 and 2, with the split essentially being driven by the variable EXP_QUOTA (export of goods as a share of turnover). Next (*k* = *4*), variable IMP_QUOTA prompts the division of cluster 2 into clusters 2 and 4. At *k* = *5* a subset of enterprises with significant foreign investment activities in CESEE (variable CESEE_SHARE_ODI) spins off from the cluster of Austrian-controlled multinationals. The export share and CESEE investment focus are the main clustering drivers also at the next two levels. Finally, at *k* = *8*, the variable SPE triggers the split of cluster 2 into clusters 2 and 8.

Chart 4 summarizes the key driving variables at each node by translating the clustering tree into a decision tree where the "branches" represent crucial dimensions for characterizing the different multinationals groups while the "leaves" visualize the clusters. Of the eight identified clusters, three are controlled by domestic companies while five are dominated by entities outside Austria. Due to their key characteristics, which will become more apparent in the detailed

Chart 4

Cluster decision tree



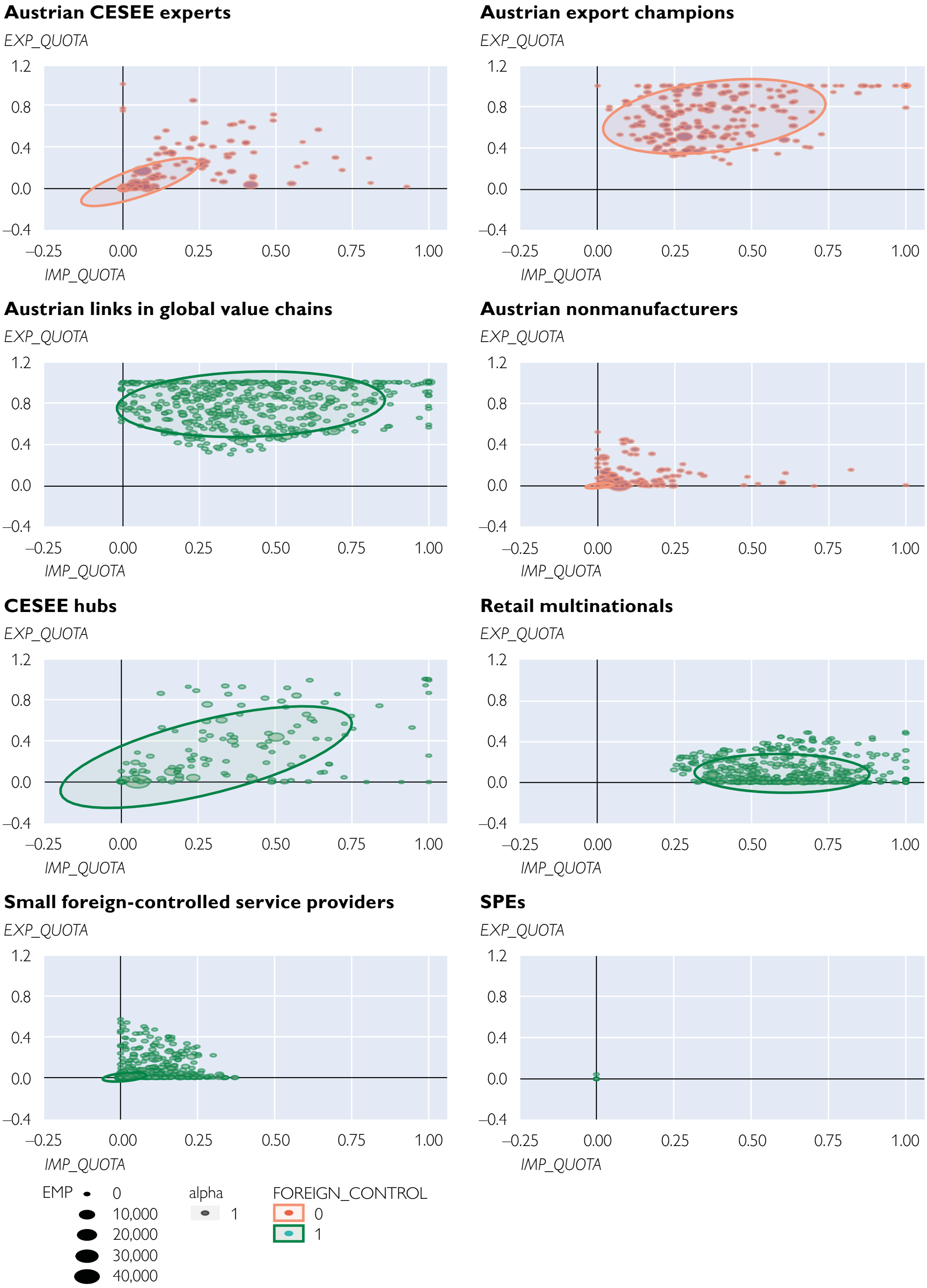
Source: Authors' calculations.

analysis below, we label the eight clusters as follows: (C1) Austrian links in global value chains, (C2) small foreign-controlled services providers, (C3) Austrian nonmanufacturers, (C4) Austrian export champions, (C5) retail multinationals, (C6) Austrian (-controlled) CESEE experts in contrast to (C7) (foreign-controlled) CESEE hubs and (C8) SPEs. The decision tree shows that the most relevant variables for the partitioning algorithm were SPE, FOREIGN_CONTROL, CESEE_SHARE_ODI, IMP_QUOTA and EXP_QUOTA.

By way of example, we can demonstrate the important role that the variables IMP_QUOTA and EXP_QUOTA played for cluster-building, yet from another perspective. The scatter plot (chart 5), which plots import against export shares for the eight clusters, shows a clear concentration of multinationals in the upper half of the quadrant (i.e. high EXP_QUOTA) for the clusters labeled "Austrian links in global value chains" (C1) and "Austrian export champions" (C4). In line with the decision tree, the only definite concentration of high IMPORT_QUOTA values was calculated for "retail multinationals" (C5). The multinational groups "small foreign-controlled service providers" (C2), "Austrian nonmanufacturers" (C3) and "SPEs" (C8) show small or no values for IMP_QUOTA and EXP_QUOTA. The clusters "Austrian CESEE experts" (C6) and "CESEE hubs" (C7) do not exhibit a clear pattern of distribution. The reason for this random distribution is that the multinationals clustered into these groups are to a high degree defined by other variables (especially FOREIGN_CONTROL and CESEE_SHARE_ODI).

Chart 5

Import and export shares by cluster and foreign control



Source: Authors' calculations.

To deepen the picture of how the clusters of multinationals differ from each other, we look at the main underlying economic activity[18] (chart 6) and find that the identified clusters are rather homogeneous with respect to the industry breakdown (as represented by MNE_NACE, which was *no* input variable for the cluster analysis). Two clusters stand out with manufacturing as the predominant economic activity: "Austrian links in global value chains" (59%) and "Austrian export champions" (73%). The other economic sectors of these two clusters show a similar distribution, the main distinguishing feature being foreign vs. domestic control. In the "trade multinationals" cluster, 80% of the group members are classified as trade companies. In the "SPEs" cluster, activities like manufacturing and trade are absent almost by definition ("up to five employees" in the IMF's definition of SPEs), leaving the vast majority of SPEs to be classified in the service sector or as financial companies. The other clusters show no clear indication of a dominating economic activity within a cluster.

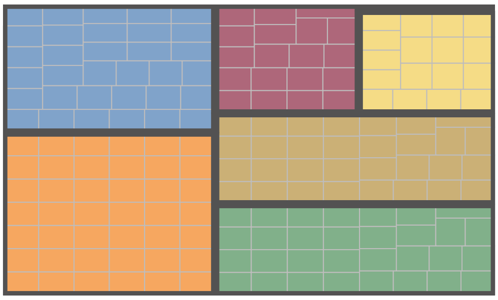
[18] *Variable MNE_NACE: For details concerning grouping and calculation, see section 1.*

Chart 6

Economic activity by cluster

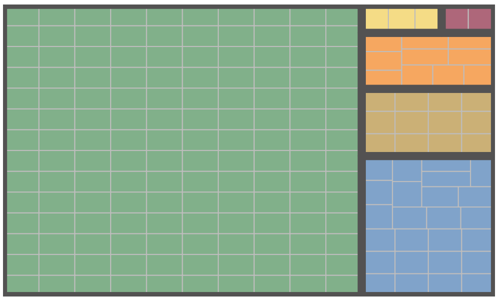
Austrian CESEE experts

1 square = 1 multinational enterprise



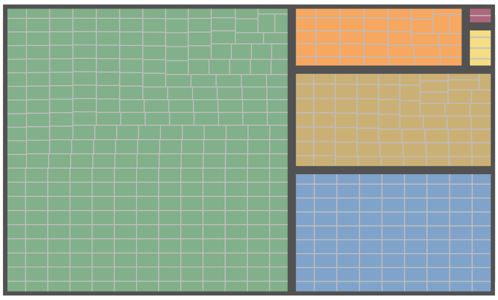
Austrian export champions

1 square = 1 multinational enterprise



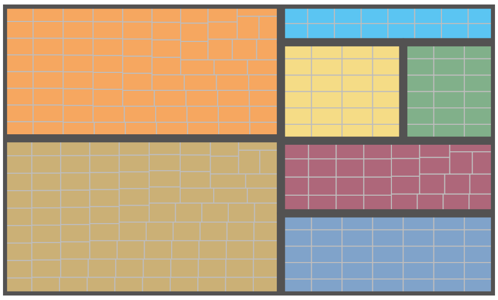
Austrian links in global value chains

1 square = 1 multinational enterprise



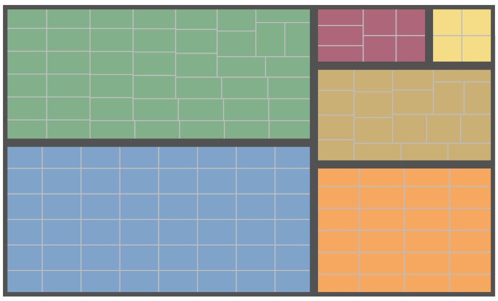
Austrian nonmanufacturers

1 square = 1 multinational enterprise



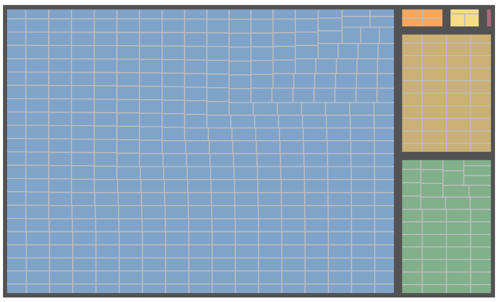
CESEE hubs

1 square = 1 multinational enterprise



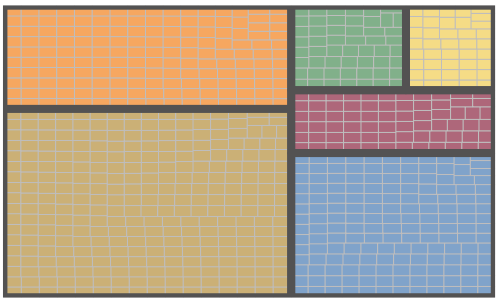
Retail multinationals

1 square = 1 multinational enterprise



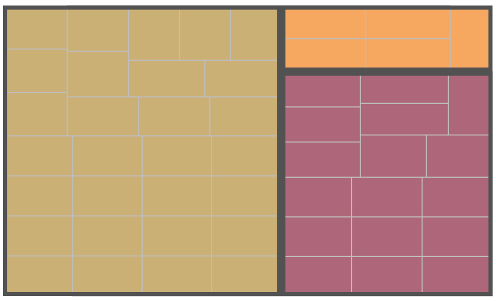
Small foreign-controlled service providers

1 square = 1 multinational enterprise



SPEs

1 square = 1 multinational enterprise



Trade
Services
Resources, energy, construction
Activities of households
Professional, scientific and technical activities
Financial and insurance activities
Manufacturing

Source: Authors' calculations.

3.2 Cluster results in detail

In lieu of a descriptive summary, see chart 7 for an at-a-glance overview.

3.2.1 "Austrian links in global value chains"

Consisting of 455 members, this comparatively large group is characterized by an export ratio of 78%. Since all companies are under foreign control, it can be assumed that investors clustered into this group above all seek to benefit from Austria's high levels of productivity, enabled by a highly skilled workforce, good infrastructure and a favorable geographical location. The average import ratio (44%) indicates that a significant part of the value added is produced in Austria. Examples of major companies (in terms of employment) in this group are household names such as BMW, MAGNA, FACC, BOSCH and NOVARTIS.

While the median number of employees (113) is neither exceptionally high nor low compared to other clusters, the median age (23) indicates a comparatively young set of multinationals. The median turnover (EUR 38 million) is also in the mid range of the total population. These figures are well in line with the large share of medium-sized[19] enterprises (58%) in this cluster. Other distinguishing features include the comparatively high share of external trade in services (service exports: 75%, service imports: 72%). The dominant economic activity is manufacturing (59%), followed by trade (18%) and services (14%). SPEs aside, this is the most globalized group in terms of the location of ultimate investors: 26% of

Chart 7

Cluster overview

Source: Authors' calculations.

Note: MNEs = multinational enterprises.

[19] *https://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en.*

them are located outside Europe, with the Americas contributing strongly with 16%.

The homogeneity of clusters measured by intra-cluster variance is in the mid range compared to other groups,[20] with intra-cluster variance being rather high for import share and turnover in particular. At the same time, employment and outward FDI are homogenous. Cluster stability is low, as more detailed resolutions of the clustering tree (see chart 3) show two additional upcoming splits of this group until resolution 12 *(k = 2)*.

3.2.2 "Small foreign-controlled service providers"

Small foreign-controlled service providers are the largest cluster by far, comprising 782 multinationals that are almost exclusively in the nonmanufacturing business, above all in the service sector (38%), followed by "professional, scientific and technical activities"[21] (21%) and trade (20%). Homogeneity within this cluster is rather low, not least because of size. The cluster contains many small enterprises (51%), some larger multinationals (14%) and a significant share of enterprises without any employees in Austria (31%). Examples include companies like BAWAG, VAMED and HOFER. What they have in common with other group members is basically the fact that they are under foreign control and have a rather low export ratio.

The cluster's median turnover is among the lowest (EUR 7 million), as is the median number of employees (17). Only a small fraction of the companies grouped into this cluster trade in services, and the average share of imports (5%) and exports (4%) is very low. Essentially, this group of foreign-controlled multinationals in Austria serves the domestic market for services (e.g. hotel industry, catering, car rentals, transport services, financial services), goods (a wide range of industries, e.g. food, office equipment, opticians, petrol stations) and professional business services (holding companies). A close "neighboring" group is the "SPEs" cluster, which exists as a distinct group only from resolution *k = 8* downward (see chart 3 for details). The main regions of origin of FDI investors in this cluster are Western Europe (74%), particularly Germany (36% of all cases). This group also contains the highest concentration of CESEE investors, with a small subgroup of the CESEE-controlled multinationals branching off at resolution 10 *(k = 10)*. Thus, cluster stability is neither high nor low.

3.2.3 "Austrian nonmanufacturers"

This is the "residual" cluster under Austrian control, consisting of only 651 multinationals compared with 1,904 multinationals in the five foreign-controlled clusters. This cluster is dominated by mostly small providers of various services

[20] *We compute homogeneity H_c of a cluster C in the following way: For each variable $v \in \{1, \dots, 20\}$ and a given cluster $C \in \{1, \dots, 8\}$ the intra-cluster standard deviation $SD(v_C)$ is calculated. For each variable v the standard deviations are then ranked across clusters in descending order and assigned a corresponding position value (rank), i.e. $Rv_C = f(SD(v_C))$ such that for two different clusters $i \neq j$, where $i, j \in \{1, \dots, 8\}$, if $SD(v_i) < SD(v_j)$ then $Rv_i > Rv_j$ and $Rv_C \in \{1, \dots, 8\}$. The overall score for a cluster is then computed as the sum of the cluster ranks across all variables $H_C = \sum_{v=1}^{20} Rv_C$. Hence, the higher H_C the more homogenous the cluster C. "SPEs" and "retail multinationals" achieve best results by a large margin (109 and 97, respectively) and are thus attributed the homogeneity label "high." Values between 70 and 80 were classified as "medium," cases below 70 as "low."*

[21] *A large part of the holdings with nil employment are classified in this sector.*

and surprisingly robust, with rather small differences between resolution level 12 *(k = 12)* and level *k = 8* (see chart 3).

The multinationals in this cluster are mainly engaged in services (31%) and professional business services (in the form of holding companies; 26%). More than half of all private individuals included in the population were clustered into this group,[22] which is one explanation for the low median number of employees (9), low turnover (EUR 4 million) and low degree of foreign trade activity (export share: 4%, import share: 5%). More than half (58%) of the companies in this cluster qualify as small multinationals, but there are also some widely known larger enterprises, such as SPAR (HOLDAG Bet. GmbH), FLUGHAFEN WIEN, PORR and RAIFFEISEN-HOLDING NÖ/W. Essentially, this cluster contains Austrian enterprises in the nonmanufacturing business and individuals engaged in outward FDI without a specific CESEE focus.

3.2.4 "Austrian export champions"

This cluster, encompassing 192 multinationals, features many of Austria's very large multinationals, thus accounting for the highest values with regard to many variables, such as the median number of employees (305) and median turnover (EUR 424 million) and the average amount of service exports (EUR 115 million) and service imports (EUR 111 million). The vast majority of multinationals in this group is in the manufacturing business (73%), followed with a huge gap by trade companies (13%). There is a clear export focus (72% export share), although many of the multinationals seem to be integrated in global value chains (import share: 41%). Recourse to international financial markets is strong in this group, as 21% of the multinationals in this cluster are known to be counterparties in cross-border financial derivatives contracts or issuers of bonds held by foreign investors. Every tenth member of this cluster also performs a cash-pooling function.[23]

Many of the multinationals in this cluster are household names in Austria, e.g. OMV, ANDRITZ, KTM, VERBUND, VOESTALPINE, LENZING and RED BULL. With a comparatively low CESEE outward FDI ratio (20%), investment targets and markets are spread globally. Another explanation for this rather low value could be that large multinationals with a very strong CESEE focus were clustered into the group of "Austrian CESEE experts." While homogeneity is low because of outliers in many variables, stability is high (this cluster remains broadly unchanged until cluster resolution *k = 12*).

3.2.5 "Retail multinationals"

The second-largest cluster, consisting of 472 multinationals, is a very homogeneous group of foreign-controlled multinationals serving retail markets of all kinds in Austria. Two-thirds of the companies are trade businesses, with the second-largest sector (manufacturing) accounting for just 10% of the cluster population. The composition of the cluster remains broadly stable for resolutions from *k = 4* down to *k = 12*.

[22] *Private individuals and foundations exist as multinationals in this paper if they hold outward foreign direct investments but no shares of domestic companies.*

[23] *"Cash pooling" is a position on reporting templates for the balance of payments and the international investment position.*

The main characteristics are a high share of imports (61%) combined with a low share of exports (11%). Most companies are neither small nor large, and homogeneity within the group is high. The median number of employees is 52, median turnover is EUR 28 million. Typical representatives of this cluster are companies like H&M, IBM, MAN, EDUSCHO, DEICHMANN, MIELE, NEW YORKER or ZARA. Typically, they do not hold outward FDI; this is the case only for 27 out of 472 companies in this cluster. The group heads of these multinationals are overwhelmingly located in Western Europe (80%; Germany: 42%).

3.2.6 "Austrian CESEE experts"

This cluster encompasses 174 Austrian-controlled multinationals with a dedicated CESEE focus in their outward FDI. Homogeneity is rather low because the cluster comprises enterprise groups of all sizes and industries, including some major Austrian banks (ERSTE GROUP, RBI, OBERBANK), large multinationals from other sectors, e.g. EVN, STRABAG, UNIQA, WIENER STAEDTISCHE, POST, XXXLUTZ, but also a number of lesser-known smaller CESEE experts. About a third of the companies clustered into this group employ up to 10 people.

In this cluster, 94% of all outward FDI is invested in CESEE countries on average. The industry mix is highly balanced, led by professional business services (i.e. holding companies; 24%) and followed by trade (19%), services (18%) and manufacturing (18%). This cluster stands out with regard to the variables "BANK" (7%) and "capital market participation" (25%), which are likely to be correlated since all banks in this cluster are active on international financial markets. With a median age of 35 years, this cluster contains the most mature of all companies. Moreover, the cluster is highly stable; the partitioning algorithm would build almost identical clusters when forced to build nine, ten or eleven groups. Only at $k = 10$ would some companies (mainly the manufacturing companies) break off to form a new cluster.

3.2.7 "CESEE hubs"

The second-smallest cluster (144 multinationals) is defined mainly by two characteristics: foreign control and outward FDI focus on CESEE countries. To some degree, there are similarities with the "Austrian CESEE experts" cluster, given the outward FDI focus on CESEE and the lack of a clear emphasis on a specific economic activity. The Austrian multinationals in this cluster serve exclusively as a hub to the CESEE region, with a minimum of managing and administrative personnel in Austria. In terms of economic activity, the single-biggest homogeneous group of multinationals in this cluster provides "professional business services" (i.e. holding companies; 17% of all multinationals in this group). The other 83% obviously are in some sort of production, trade or service business in Austria. The share of exports (29%) and imports (33%) is significantly higher than in the "Austrian CESEE experts" cluster (14% and 13%, respectively).

Household names in this cluster are REWE, TELEKOM AUSTRIA, SIEMENS and ALLIANZ. Homogeneity is also comparatively low given the broad mix of companies (companies of all sizes and industries) as in the "neighboring" cluster 6 above. Stability is high, with no further split occurring at least until $k = 12$.

3.2.8 "SPEs"

According to the IMF's definition:

"An SPE resident in an economy, is a formally registered and/or incorporated legal entity recognized as an institutional unit, with no or little employment up to maximum of five employees, no or little physical presence, and no or little physical production in the host economy." [24]

Additional characteristics include foreign control and an exceptionally high degree of cross-border assets and liabilities.

As was to be expected, this cluster is highly homogeneous, with 51 out of the 52 multinationals in the dataset marked as SPE having been clustered into this group. The median number of employees in this cluster is zero, as is turnover and foreign trade. The PAM algorithm conducts a "SPE split" at $k = 12$. The closest "neighboring" group is "small foreign-controlled service providers," which builds a common cluster with SPEs at $k = 7$.

4 Interpretation, conclusion and further research

This paper sheds some light on the relevance and increasing role of multinational enterprises active in Austria. To this effect we built a comprehensive dataset of multinationals in Austria and then clustered them into groups according to their key characteristics, using the partitioning-around-medoids algorithm.

The analysis delivers eight meaningfully interpretable groups of multinationals, three of which are Austrian-controlled, with the other five clusters being in foreign hands. There is a significant difference in complexity between these two segments. The Austrian-controlled units are characterized by a high degree of stability from a relatively early partitioning stage *(k6)*, with no further splits occurring until $k = 12,$ except for the branching-off of a CESEE group containing larger, export-orientated manufacturing companies. As the foreign-controlled units are more heterogeneous, they tend to be subject to splits between $k = 7$ and $k = 11$. Looking ahead, further experimentation in input parameters could verify the stability of our results at these cluster resolutions.

Additionally, one must bear in mind that we conducted a one-off exercise based on 2017 data only. An obvious next step could be to perform a similar analysis with 2018 data or time series with historic data to be able to assess the stability of the clusters are over time. Furthermore, we did not investigate all available variables, and there might be additional relevant data sources that could be linked to the multinationals database (e.g. R&D expenditure). Another rewarding question could be a more detailed investigation of "neighboring" clusters, e.g. "Austrian CESEE experts" and "CESEE hubs," to be able to establish the effect foreign control has had over time compared with domestic control. Likewise, intra-cluster consistency could be the subject of further research.

Finally, future research may look at the impact of various policy measures ranging from taxes to labor market policies for different types of multinationals, which would then allow for more effective and efficient policymaking. Last but not least, once the proposed taxonomy of environmentally sustainable activities has been completed, another avenue of future research might focus on identifying the characteristics and drivers of multinationals' "green" activities.

[24] *See www.imf.org/external/pubs/ft/bop/2018/pdf/18-03.pdf (page 19).*

References

Budiaji, W. and F. Leisch. 2019. Simple K-medoids partitioning algorithm for mixed variable data. In: Algorithms. 12(9). 177.

Baser, P. and J. Saini. 2013. A comparative analysis of various clustering techniques used for very large datasets. In: International Journal of Computer Science and Communication Networks. 3(5). 271–275.

De Backer, K., S. Miroudot and D. Rigo. 2019. Multinational enterprises in the global economy: Heavily discussed, hardly measured. https://voxeu.org/article/multinational-enterprises-global-economy.

Fraiman, R., A. Justel and M. Svarc. 2009. Selection of variables for cluster analysis and classification rules. In: Journal of the American Statistical Association. 1294–1303.

Gower, J. C. 1971. A general coefficient of similarity and some of its properties. In: Biometrics 27. 857–874.

Jain, A. K. and R. C. Dubes. 1988. Algorithms for clustering data. Prentice-Hall.

Jin X. and J. Han. 2017. K-Medoids Clustering. In: Sammut C. and G. I. Webb (eds.). Encyclopedia of Machine Learning and Data Mining. Springer.

Kaufman, L. and P. J. Rousseeuw. 1987. Clustering by means of medoids. In: Dodge, Y. (ed.). Statistical Data Analysis Based on the L1-Norm and Related Methods. 405–416.

Kim, I. S. and H. V. Milner. 2019. Multinational Corporations and their Influence Through Lobbying on Foreign Policy. Princeton University Working Paper.

Mirkin, B. 2011. Choosing the number of clusters. In: WIREs Data Mining Knowledge Discovery 1(3). 252–260.

Nguyen, T. T., D. Dinh., S. Sriboonchitta and V. Huynh. 2019. A method for k-means-like clustering of categorical data. In: Journal of Ambient Intelligence and Humanized Computing.

OECD. 2015. Measuring International Investment by Multinational Enterprises. Implementation of the OECD's Benchmark Definition of Foreign Direct Investment, 4th edition. www.oecd.org/corporate/FDI-BMD4-brochure.pdf

OECD. 2018. Multinational enterprises in the global economy. Heavily debated but hardly measured. www.oecd.org/industry/ind/MNEs-in-the-global-economy-policy-note.pdf

Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. In: Journal of Computational and Applied Mathematics 20. 53–65.

Sambandam, R. 2003. Cluster analysis gets complicated. In: Marketing Research 15(1).

Tryfos, P. 1998. Methods for business analysis and forecasting: Text & Cases, Chapter 15. Wiley.

Annex

Cluster analysis

When it comes to extracting intrinsic but unobserved information and structures from large datasets, there are various data-mining techniques to choose from. We are dealing with mixed-type data, and we are looking for an algorithm with which to identify distinct types of multinationals. Unlike in supervised machine-learning techniques, we do not impose any previously known category labels on the data which would denote their a-priori partition. Therefore, we picked cluster analysis from the unsupervised-learning toolbox to unveil clusters in a way that observations are similar within groups with respect to variables of interest, while the groups themselves stand apart from one another (Tryfos, 1998).

Similarly, cluster analysis can be performed with a plethora of different methods classifiable across multiple dimensions (see e.g. Baser and Saini, 2013). In general, different clustering methods lead to different outcomes. Cluster analysis is an explorative technique, meaning that the best approach is highly subjective, and it should be one that is practical to handle and that delivers the "best" – i.e. meaningful, useful and well interpretable – results for the analyst.

Having said that, important decisions need to be made with respect to the clustering method and its parameters in the course of the data exploration process. One of the choices to be made is between a hierarchical or a partitional approach. The hierarchical approach is based on a nested sequence, proceeding either bottom-up – i.e. starting from as many clusters as there are observations and finishing with a single cluster comprising all observations (agglomerative approach) – or vice versa (divisive approach) (see e.g. Jain and Dubes, 1988).

In contrast to hierarchical methods, partitional clustering algorithms generate single partitions of the data into mutually disjoint subsets the number of which – conventionally denoted as *k* – needs to be specified by the researcher ex ante. Essentially, these algorithms first assign each data point to one of *k* clusters and then reshuffle the observations across clusters until each observation has the smallest distance to the center of the cluster. The methods differ, inter alia, in the way the "center" is defined and in the distance metric.

In general, hierarchical clustering methods are rather useful for smaller datasets. Moreover, different parameter specifications tend to produce very different outcomes, as was the case with our dataset. Having experimented with various hierarchical clustering methods, we ultimately opted for a partitional clustering method.

A very popular partitional approach is the so-called k-means method, well known for its efficiency in clustering large datasets. However, one of its key features is the fact that it uses arithmetic data "means" (so-called centroids) as the center of the clusters. The upshot is that this method typically works only on numerical variables.[25] In addition, its results are sensitive to outliers and noise in the data (Budiaji and Leisch, 2019). Since our dataset is a mixture of numerical and categorical variables and contains a number of outliers, the k-means method was not an option. The most common alternative for mixed-variable datasets is the "partitioning around medoids" (PAM) algorithm. It can be considered a more robust and universal algorithm than the k-means, not only because it can handle mixed data but also because it is less sensitive to outliers (Jin and Han, 2017). Rather than using "centroids," this method uses "medoids," which are not computed statistical means but actual data points from the dataset representative of each cluster.

It follows from the description of the PAM algorithm in section 2 that one of the key inputs for the algorithm is some distance metric. If the dataset contains purely numerical variables, different distance measures can be applied directly to the raw dataset just as with k-means. However, in case of mixed data, the distance between observations needs to be computed beforehand and provided as input to the algorithm as a distance matrix. A common option to compute distances for mixed data sets is the Gower-distance matrix (Gower, 1971). It uses an appropriate distance metric for each variable type, i.e. Manhattan for continuous and Dice for categorical datapoints, which is subsequently scaled to fall between 0 and 1. Then,

[25] *Extensions of the k-means method to mixed and categorical data have been developed in the literature. For an example see e.g. Nguyen et al. (2019).*

a linear combination using user-specified weights is calculated to create the final distance matrix.

Table A1

Variables and weights for PAM clustering

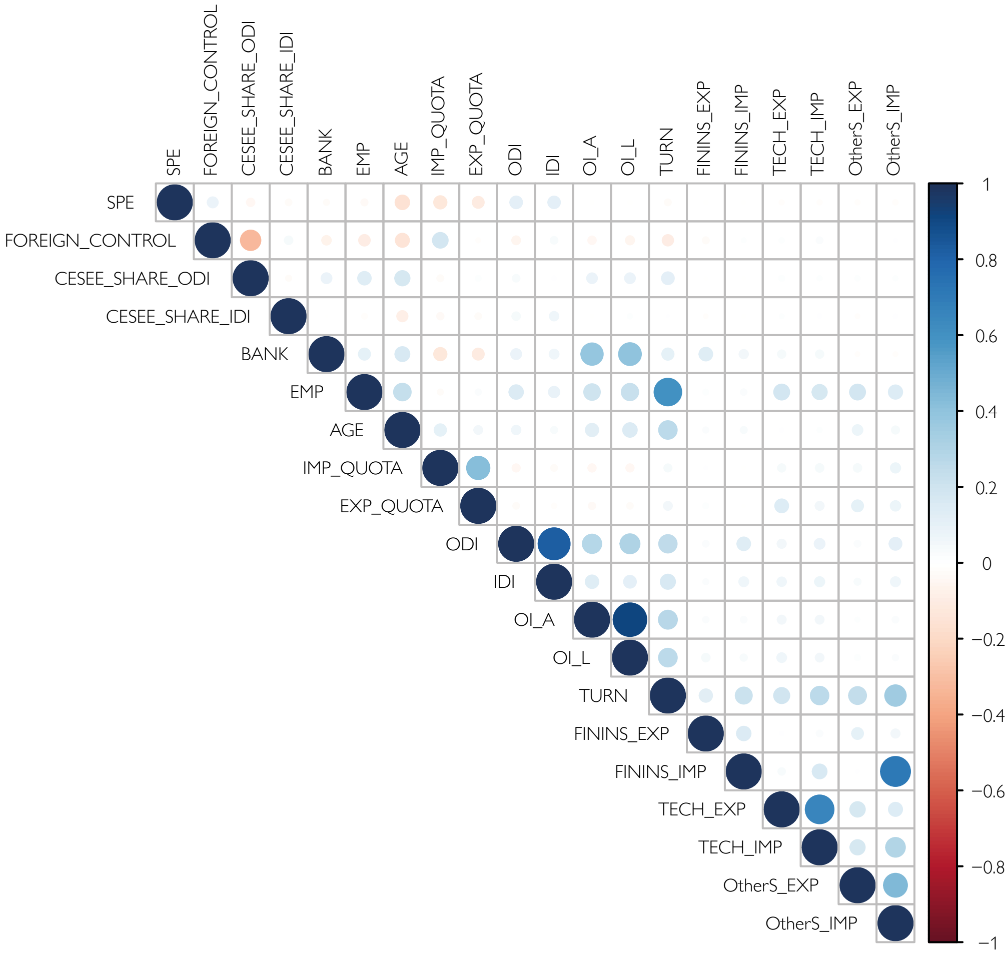
Variable	Name	Type	Values	Description	Weight for clustering in %
SPE	Special purpose entity	Nominal	1 = "yes" 0 = "no"	One or more units are classified as a special purpose entity	8
FOREIGN_CONTROL	Foreign control	Nominal	1 = "yes" 0 = "no"	One or more units are controlled by a nonresident	8
CESEE_SHARE_ODI	CESEE share of outward FDI	Interval	0-100%	Outward FDI in CESEE as a share of total outward FDI	8
CESEE_SHARE_IDI	CESEE share of inward FDI	Interval	0-100%	Inward FDI in CESEE as a share of total inward FDI	8
BANK	Banking license	Nominal	1 = "yes" 0 = "no"	One unit is classified in ESA sector 1220A	8
EMP	Employees	Interval	$\mathbb{N}$	Total number of employees (all units)	8
AGE	Age	Ordinal	$\mathbb{N}$	Age of the oldest unit	8
IMP_QUOTA	Import share	Interval	0-100%	Import of goods divided by turnover	8
EXP_QUOTA	Export share	Interval	0-100%	Export of goods divided by turnover	8
ODI	Outward FDI	Interval	$\mathbb{Z}$	Outward FDI (extended direction principle)	2
IDI	Inward FDI	Interval	$\mathbb{Z}$	Inward FDI (extended direction principle)	2
OI_A	Other investment assets	Interval	$\mathbb{N}$	Other investment assets (BOP/IIP concept)	2
OI_L	Other investment liabilities	Interval	$\mathbb{N}$	Other investment liabilities (BOP/IIP concept)	2
TURN	Turnover	Interval	$\mathbb{N}$	Turnover as reported in structural business statistics	8
FININS_EXP	Insurance and financial services exports	Interval	$\mathbb{N}$	Insurance and financial services exports as reported for ITSS	2
FININS_IMP	Insurance and financial services imports	Interval	$\mathbb{N}$	Insurance and financial services imports as reported for ITSS	2
TECH_EXP	Technical services exports	Interval	$\mathbb{N}$	Technical services exports as reported for ITSS	2
TECH_IMP	Technical services imports	Interval	$\mathbb{N}$	Technical services imports as reported for ITSS	2
OtherS_EXP	Other services exports	Interval	$\mathbb{N}$	Other services exports as reported for ITSS	2
OtherS_IMP	Other services imports	Interval	$\mathbb{N}$	Other services imports as reported for ITSS	2

Source: OeNB, Statistics Austria, Authors' calculations.

Note: $\mathbb{N}$ = positive integers; $\mathbb{Z}$ = positive or negative integers; BOP/IIP = balance of payments/international investment position; ITSS = international trade in services statistics.

Chart A1

Correlation of multinationals variables



Source: Authors' calculations.