JRC TECHNICAL REPORT

Monitoring the development and integration of EU countries' capital markets

An approach based on composite indicators and cluster analysis

Gucciardi, Gianluca

2019

Joint Research Centre

EUR 29891 EN

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
Name: Gianluca Gucciardi
Address: Via E. Fermi 2749, 21027 Ispra (VA), Italy
Email: gianluca.gucciardi@ec.europa.eu

EU Science Hub
https://ec.europa.eu/jrc

JRC118408

EUR 29891 EN

PDF ISBN 978-92-76-12268-5 ISSN 1831-9424 doi:10.2760/92215

Luxembourg: Publications Office of the European Union, 2019

How to cite this report: Gucciardi Gianluca, Monitoring the development and integration of EU countries' capital markets. An approach based on composite indicators and cluster analysis, EUR 29891 EN, Luxembourg: Publications Office of the European Union, 2019, ISBN 978-92-76-12268-5, doi:10.2760/92215, JRC118408.

Contents

Abstract

This technical report proposes a set of metrics and a methodology to measure the progress that European countries are making towards the development and integration of capital markets. Based on six priorities linked to the achievement of a well-functioning and integrated European capital market, we identify a set of indicators and analyze country performance for the period 2007-2018 by using a composite measure, both in a static and a dynamic environment. In order to account for country specificities and differences in dynamic trajectories, we use robust clustering to identify groups of countries and follow their development through time.

List of figures

List of tables

1. Introduction

In this work, we investigate the progress of Capital Markets' (CMs) development and integration across the 28 EU countries, based on different quantitative approaches.

First, we use composite indicators, synthetic measures summarizing different relevant related dimensions into a single comprehensive number (Nardo et al., 2008). The goal of this approach would be twofold: on one hand, to define and provide a simple-to-interpret (but non-simplistic) tool to measure CMs development and integration between 2007 and 2018; on the other hand, to observe the evolution of this indicator for the whole EU 28 and each single country over the past twelve years, both through static comparisons of ranks emerging from composite indicators, and through dynamic performance assessments based on Data Envelopment Analysis (DEA, Coelli et al., 2005) and Malmquist Index (Färe et al., 1994a; Färe et al., 1994b).

Second, we cluster countries along different CMs-relevant dimensions, adopting a statistically robust technique, i.e. *Robust Trimmed Clustering* model (Garcia-Escudero et al., 2008a). This analysis is conducted with two aims: on one side, we would identify clusters of countries with similar behavior for each year of analysis; on the other side, we would capture country-level patterns across clusters in time, which could be associated to a change in their Capital Markets' development and integration progress.

The remainder of this work is organized as follows. We provide a description of the theoretical framework constituting the rationale for the monitoring of CMs progresses and of the selection of the associated indicators in Section 2. These indicators will underpin all quantitative approaches. We then discuss countries' performance as measured by a static version of the composite measure in Section 3. In section 4, we move to the dynamic performance countries' assessment, based on the Data Envelopment Analysis (DEA), a robust clustering analysis and the Malmquist Index. We then compare our findings with those obtained through similar metrics in literature in Section 5. Lastly, Section 6 concludes. Additional methodological material, including some preliminary analyses conducted to support both methodological approaches, is available in the Technical Annex.

2. Theoretical framework and selection of indicators

In order to measure EU countries progress towards Capital Markets' development and integration, we follow the framework identified by the European Commission within its Capital Markets Union (CMU) Action Plan (European Commission, 2015). Specifically, the European Commission identified six key objectives[1], whose pursuit would have been beneficial for the achievement of a EU single capital market: 1. financing for innovation, start-ups and non-listed companies, 2. making it easier for companies to raise funds on capital markets, 3. promoting investments in long-term, sustainable projects and infrastructure projects, 4. fostering retail and institutional investments, 5. leveraging bank capacity to support the economy, 6. facilitating cross-border investments and promote financial stability.

The first two objectives deal with the introduction of financing sources different from bank loans for companies, at any stage of their life. In particular, the former accounts for the access to alternative finance instruments, even in the pre-IPO phase. On the other side, the latter aims at easing companies' access to public markets and raise of debt and/or equity instruments. Objective 3 describes the need of channeling capitals to infrastructural and sustainable investments, respectively to create new jobs and support a sustainable growth. The fourth objective pushes for increasing market openness to (retail and institutional) investors in order to increase savings, also providing households best options for their pension choices. Objective 5's goal is to strengthen bank capability to keep supporting the economies through financing. This should be enabled by the transfer of risk from the original owner of the loan to investors, freeing banks' capital for new loans. Lastly, the sixth objective includes initiatives towards the establishment of a fully integrated market, with harmonized standards and no capital flows barriers, to increase EU capacity in the global competition.

In this work, we will consider these objectives as essential "pillars"[2] to measure the progress towards CMs development and integration for two main reasons: first, to reduce arbitrariness in the definition of the framework, which is in fact exogenously defined; second, to introduce a monitoring tool in consistency with the framework devised by the policy-maker. Still a certain degree of arbitrariness exists, as more than one indicator can be actually associated to each pillar.

The selection of representative indicators underlying the theoretical framework derives from a mix of different criteria. First, we have looked for a good trade-off between adherence of indicators to their related key objective and their availability on our sample. Consistently, we first tried to adopt indicators presented in European Commission documents (e.g., European Commission, 2016), then we considered previous works measuring the progress of capital markets in EU to enlarge the pool of possible indicators (AFME, 2018; European Central Bank, 2018). Second, we have chosen indicators available on an annual basis (for the period from 2007 to 2018) and for all individual EU-28 countries, with the aim of minimizing as far as possible the imputation of missing data. Third, we have preferred data whose source was public and official, also to favor the replicability of the analysis over time. Nevertheless, these three criteria have been subject to data availability constraints. In particular, we have verified the absence of official statistics with the required granularity for indicators mostly related to pillars 1 (e.g., venture capital and private equity) and 3 (e.g., alternative bonds). Therefore, we have used data from private sources, based on the practical idea that each pillar should be represented by at least one indicator. Finally, if some indicators were not available with the granularity required for our sample, we have opted for comparable alternatives capable of representing the "spirit" underlying the pillar, based on economic literature and policy evidence.

By adopting these criteria, we come to identify 11 indicators: two for the first 5 pillars and 1 for the last one, each available for the period 2007-2018 and for all 28 EU countries. Table 1 synthetically describes the selected

[1] In 2017, the European Commission released a mid-term review of the CMU action plan, introducing a new key objective "strengthening supervision and deepening capital markets in the EU". We do not consider this pillar in the framework discussed in this work, mainly because its introduction in principle could be assessed only for the last two years of our sample (2017-2018).

[2] We will alternatively use both terms with the same meaning later in this work.

indicators - associated to the six pillars - and their contribution to CMs development and integration measurement.

Table 1. Framework of the analysis and CMs metrics underlying indicators.

Pillar	Indicator	What the indicator measures
1. Financing for innovation, start-ups and non-listed companies	**Adoption of Private Equity instrument** Private Equity Volumes / NFCs Loans outstanding	Adoption of alternative finance instruments (Private Equity)
	Adoption of Venture Capital instrument Venture Capital Volumes / NFCs Loans outstanding	Adoption of alternative finance instruments (Venture Capital)
2. Make it easier for companies to raise funds on capital markets	**Relevance of Non-Financial Corporations Debt** NFCs total Debt outstanding / NFCs Loans outstanding	Conversion of loans into Debt Capital Markets instruments
	Relevance of Non-Financial Corporations Equity NFCs Listed Shares Equity outstanding / NFCs Loans outstanding	Conversion of loans into Equity Capital Markets (listed shares) instruments
3. Promote investments in long-term, sustainable projects and infrastructure projects	**Adoption of alternative instruments** Alternative bonds emissions / all bonds emissions	Diffusion of alternative (green, social, sustainable) instruments
	Public-Private Partnerships (PPPs) Volumes of PPP projects (B€)	Investments in long-terms projects
4. Foster retail and institutional investments	**Retail Investments (assets)** (Listed equity shares + Investment equity shares + Bonds + Life Insurance) / GDP [for households]	Availability of pools of savings (households)
	Institutional Investments (assets) (Listed equity shares + Investment equity shares + Bonds + Life Insurance) / GDP [for pension funds and insurances]	Availability of pools of savings (institutional investors)
5. Leverage bank capacity to support the economy	**Covered Bonds** Covered Bonds issuance / Total Economy Loans outstanding	Ability to transform bank loans into capital markets instruments
	Deposit taker capital adequacy Regulatory Tier 1 capital to risk-weighted assets	Financial system soundness
6. Facilitate cross-border investment and promote financial stability	**Cross-border Portfolio Debt and Equity Investments** 100% - Portfolio Debt and Equity Home Bias	EU regional cross-border integration

The adopted indicators are built so that in any case their (positive) growth is associated to a beneficial contribution to Capital Markets' enhancement. Pillar 1 is composed of two different indicators describing the level of adoption of alternative finance instruments, respectively Private Equity and Venture Capital. These indicators are expressed as the ratios of, respectively, Private Equity and Venture Capital volumes on Non-Financial Corporations (NFCs) loans outstanding. Therefore, a raise in this ratio would represent the increasing relevance of alternative instruments with respect to traditional financing. Analogously, two indicators within pillar 2 measure the access to Debt and Equity Capital Markets: "Relevance of NFCs Debt" and "Relevance of NFCs Equity". In these cases, NFCs total debt or listed shares equity outstanding are related to NFCs loans outstanding. Again, an increasing ratio would suggest a larger access to capital markets by NFCs with respects to loans. Two different indicators compose pillar 3. On the one side, the "Adoption of alternative instruments" is built as the ratio between the number of emitted green, social and sustainable bonds[3] on total bonds emissions. The increase of the ratio stands for an ongoing diffusion of alternative bonds. On the other side, investments through Public-Private Partnerships are taken as an absolute value (expressed in B€). Pillar 4 is based on both Retail and Institutional savings taken as fraction of GDP. An increase of these indicators describes a rising availability of pools of savings. Pillar 5 is described by two indicators: on the one hand, "Covered Bonds", taken as ratio of total economy loans outstanding, describes the ability of banks to transfer risks to the market; on the other hand, "Deposit taker capital adequacy" is an IMF indicator which describes financial system soundness, measuring regulatory tier 1 capital to risk-weighted assets[4]. Lastly, pillar 6 measures how the share of foreign investments is evolving. To approximate, we consider the home-bias which provides a measure of how much domestic investments are over-weighted in a given country (Schoenmaker and Bosch, 2008). We take its complement to 100%, to ensure that it contributes positively to the overall metrics. An increase of this ratio would imply more associated equity and debt capital markets within EU, which may lead to benefits in terms of more robust financial stability (Nardo et al., 2017).

The set of selected indicators may comprehensively represent the level of CMs progress. In principle, this work should focus its scope on EU-only integration of capital markets, since the framework is based on the original set up of the CMU Action Plan (European Commission, COM(2015) 468 final) which includes a set of policies specifically dedicated to EU member states. This means that all indicators, when dealing with international capital flows, should be referred to intra-EU cross-country movements only. This would specifically be the case for pillar 6, being directly related to cross-border investments, and partially to pillars 1 and 2, since in principle, EU countries may access to non-EU equity and debt capital markets. In this work, we were able to discriminate between regional (EU) and global (world) cross-country flows for pillar 6, since the granularity of the underlying source (i.e. FinFlows) allowed this analysis. However, we were not able to do the same for the others due to constraints in the selected data sources.

We first investigate the CMs development and integration progress of the European Union adopting a pillar-by-pillar approach. Table 2 shows the average European Union performance across the six pillars. In this work, we will consider the year of the introduction of the CMU Action Plan (2015) as a useful watershed to assess the evolution of CMs development and integration in the EU. Hence, for the sake of synthesis, we represent the average post-CMU action plan performance (2015-2018) and we compare it with the average value for the period (2007-2014) to provide a synthetic view of the CMs pillars' evolution.

One first inspection of the results suggests that on average in the last four years of the sample the average value increased in nine out of eleven indicators compared to the average over the period 2007-2014. On the other hand, negative evolution were focused on pillar 5 (with covered bonds slightly decreasing), and pillar 3 (PPPs more than halved). Moreover, performances are quite heterogeneous across countries, with some of them showing results in contrast to the average. In particular, ~46% of countries experiences decreasing levels in cross-

[3] According to the "Dealogic" definition/categorization.

[4] Missing values were imputed using as reference different sources (i.e., European Central Bank and World Economic Forum), as described in Section A.1 of the Annex.

border portfolio debt and equity investments indicator in the period 2015-2018. Nevertheless, in all the indicators the majority of countries show a performance aligned with the average in terms of its sign.

Even though it provides useful insights, this descriptive approach is limited because it is not possible to correctly evaluate the joint effect of these indicators, and therefore of the CMs enhancement, without being able to correctly aggregate them. In this perspective, in the following section we introduce the concept of composite indicators, as a tool to synthetize and describe a multi-faceted phenomenon.

Table 2. Summary of evolution of key pillars and indicators.

Pillar	Indicator[5]	EU 07-14 avg	EU 15-18 avg	Data source
1	**Adoption of Private Equity instrument** (%)	0.64%	0.93%	Zephyr + Eurostat
	Adoption of Venture Capital instrument (%)	0.02%	0.04%	Zephyr + Eurostat
2	**Relevance of Non-Financial Corporations Debt** (%)	8.3%	9.8%	Eurostat
	Relevance of Non-Financial Corporations Equity (%)	36%	39%	Eurostat
3	**Adoption of alternative instruments** (%)	0.05%	1.59%	Dealogic
	Public-Private Partnerships (PPPs) (B€)	0.64	0.31	European PPP Expertise Centre (EIB)
4	**Retail Investments (assets)** (%)	63%	71%	Eurostat
	Institutional Investments (assets) (%)	47%	56%	Eurostat
5	**Covered Bonds** (%)	1.6%	1.3%	ECBC[6] + Eurostat
	Deposit taker capital adequacy (%)	13%	18%	International Monetary Fund[7]
6	**Cross-border Portfolio Debt and Equity Investments** (%)	23%	24%	FinFlows[8]

5 Figures are calculated taking the average value of the countries' performances by year. Then, these values are averaged respectively for the period 2007-2014 and 2015-2018. More details on the dataset construction and on indicators' descriptive statistics are available in Section A.1 of the Annex.

6 European Covered Bond Council.

7 "Financial Soundness Indicators".

8 Elaborations from Finflows database, jointly developed by DG-ECFIN & DG-JRC groups (Nardo et al., 2017) – portfolio equity and debt figures are averaged by year and country.

3. Composite Indicator approach

One of the possible tools to measure multi-dimensional phenomena is the composite indicator (Floridi et al., 2011). Indeed, this method allows the aggregation of several indicators into a single number and the comparison of aggregated performances of single observed units one with each other and over time (Saisana and Tarantola, 2002). Hence, it could be a useful metrics to investigate the evolution of single countries' relative performance in terms of CMs development and integration progress in our sample.

At the same time, it is known in the composite indicator literature that the relative performance of countries is subject to the "arbitrary" approach according to which the composite is built (Valdes, 2018). Specifically, each phase of the indicator construction (i.e. raw data normalization, normalized data weighting and its final aggregation) implicitly carries a source of arbitrariness and, consequently, of variability of the final composite expression (Ebert and Welsch, 2004; Böhringer and Jochem, 2007).

In order to account for the level of arbitrariness, following Saisana and Munda (2008), Munda and Saisana (2011), Floridi et al. (2011), in this work we adopt a "non-simplistic" approach for composite indicator construction (Luzzati and Gucciardi, 2015). The intuition behind this methodology is to directly propose a "plausible" rank, rather than a single one based on a specific *ad hoc* combination of (normalization, weighting and aggregation) techniques. Hence, following this approach we generated 32 yearly different rankings for the period 2007-2018, emerging from the combination of five normalization, four weighting and two aggregation techniques[9]. We produced 11 different "experiments"[10], each keeping fixed one technique and changing the others. Hence, five experiments come from normalization, four from weighting and two from aggregation. For each of these *scenarii*, we calculate the average rank, by country and year. The minimum and the maximum values among these averages can vary between 1 to 28 (even if in practice some ranks are not filled due to the average in the experiments) and represent the delimiters of the range of possible ranks. Lastly, to determine the "plausible" rank, we take the average of the ranks of the 11 experiments, again by country and year.

3.1 Evolution of CMs composite indicator rank

We first provide the results for 2018 performance versus the previous year, as shown in Figure 1.

This map depicts with greenish colors countries showing a positive growth of their ranking in 2018 with respect to 2017, with reddish colors countries with decreasing rankings and in yellow countries having the same rank in both years. It emerges that ~43% of the EU countries have decreased their relative position in the ranking, while ~36% improved and the remainder maintained their rank. We do not find differences in the distribution of countries decreasing their rank, since they are evenly divided between the highest and the lowest part of the ranking. Nevertheless, 80% of countries improving their relative rank were in the lower part of the ranking in 2017.

Hence, overall, these results provide first evidence of a substantially stable evolution of our CMs metrics comparing 2017 and 2018, with an average light worsening of relative performances in the last year of the sample, and most of improvements concentrated on low-ranked countries. Building on this approach, we are able to provide a first grouping of countries based on 2017 ranking and their growth performance in 2018, as shown in Figure 2.

[9] See Sections A.2 to A.4 in Annex for all details on the different techniques and their combinations.

[10] See Section A.5 in Annex as reference for the adopted "experimental" set-up.

Figure 1. Evolution of CMs composite indicator rank (2018 vs 2017) by EU country.

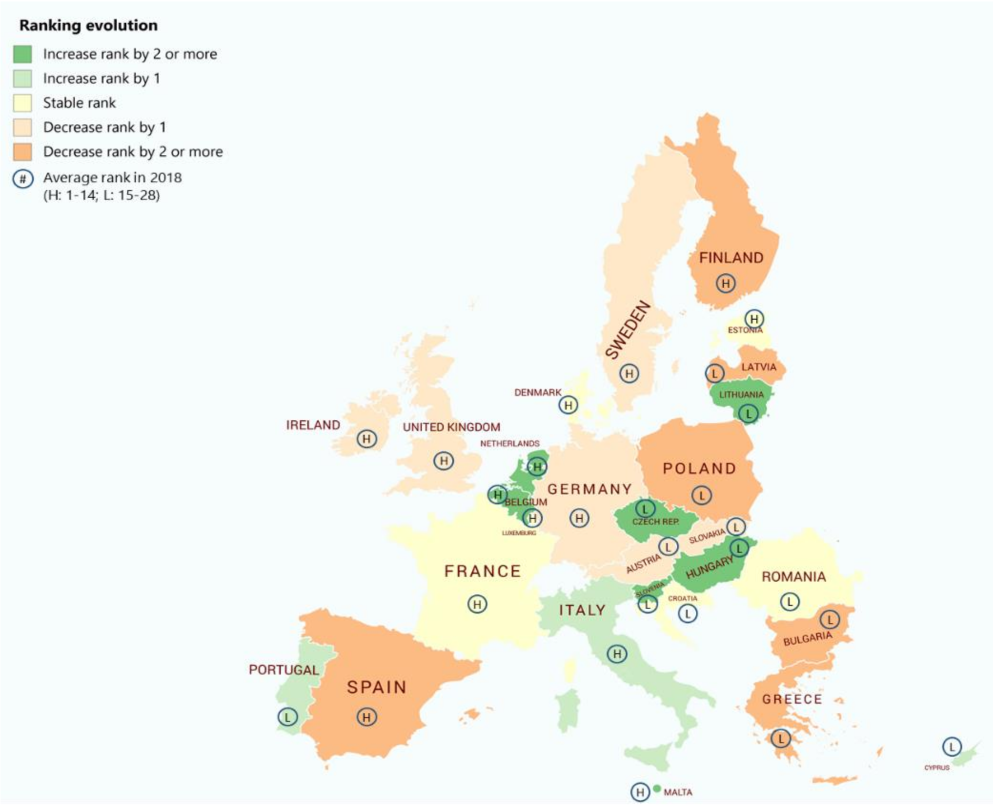
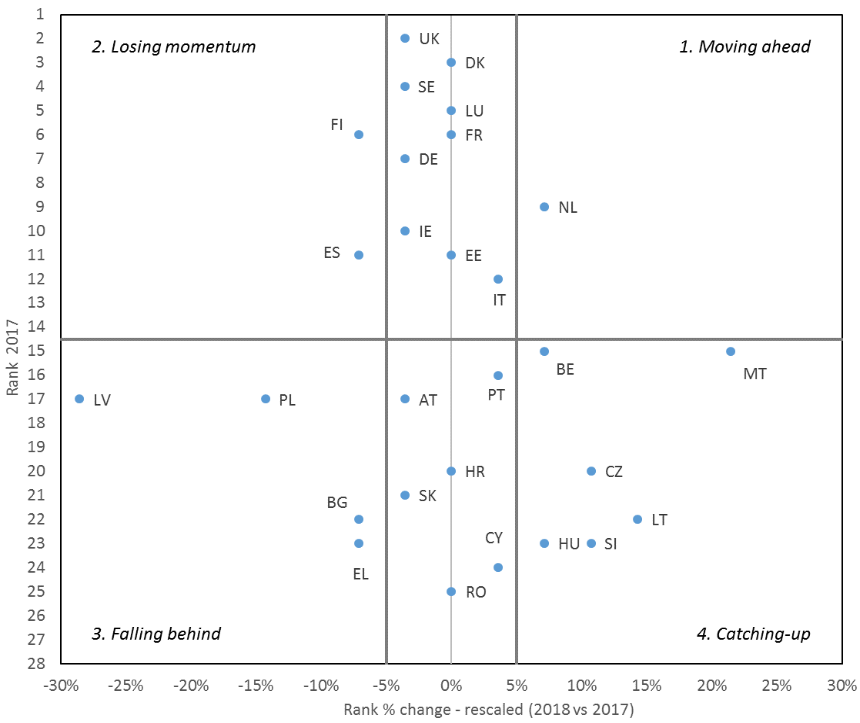


Figure 2. CMs performance matrix (2018 vs 2017) by EU country[11].



11 The taxonomy adopted for the definition of the quadrants in Fig. 2 and Fig. 4 is mediated from Fagerber and Shrodec (2005).

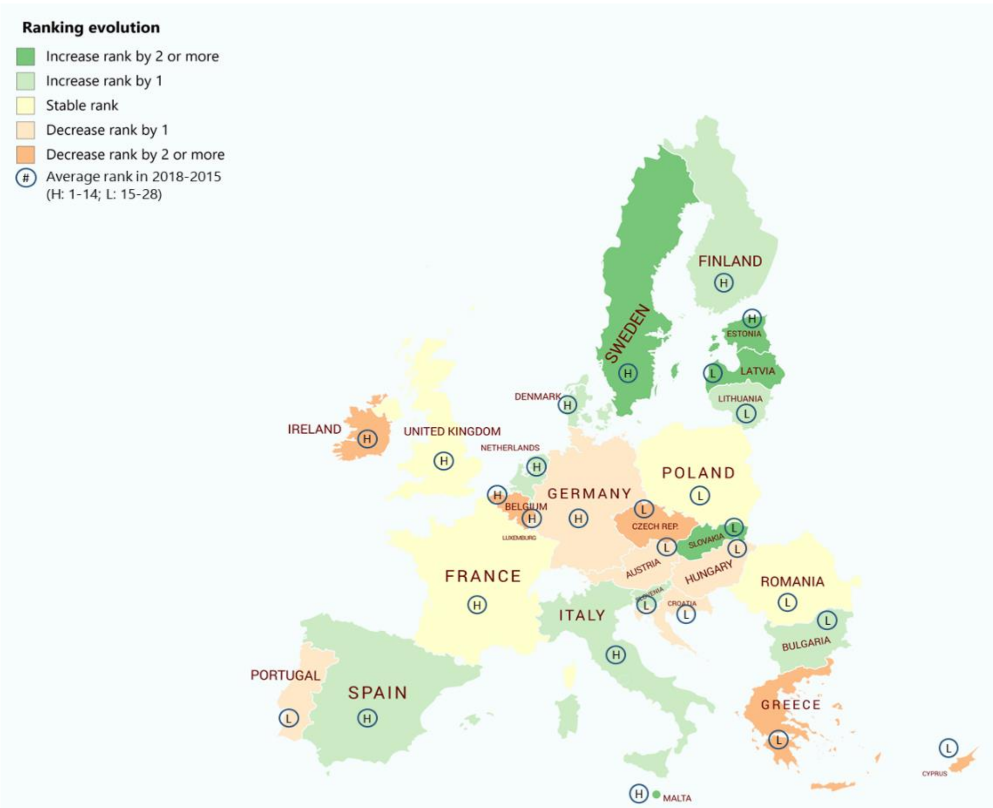
Two axes divide the area into four quadrants. The first, along the vertical axis, divides countries between those with a promising rank (first half of the ranking) and those with room for improvements (second half). The second, along the horizontal axis, clusters countries between those increasing (right side) and decreasing (left side) their relative rank. In order to avoid rigid interpretation of the results, we also introduced a "buffer" area between -5% and +5% within which country performance can be considered as substantially stable over time.

According to this representation, four clusters emerge. First, "moving ahead" cluster in the top right quadrant shows consolidating top level performances, with higher than 5% increases from 2017 to 2018. On the other side of the matrix, four "falling behind" countries are losing positions in their ranking, already standing in the bottom 50% of 2017 ranking. Two countries are "losing momentum", since they still keep a good position, but with significant relative ranking decrease versus the previous year. Conversely, six countries are "catching-up" since they obtained a significant increase in terms of ranking which is allowing them to converge towards more robust performances. All the other countries show a substantially stable performance in the last two years of our sample.

Overall, a larger relative mobility emerges for lower levels of the ranking, with ten countries showing significant changes in their position. At the same time, we find more countries moving from the bottom to the top part of the ranking, rather than the opposite, confirming the insights emerged from Figure 1.

In order to investigate whether a country has experienced a significant change in CMs metrics over the analyzed period, we now compare its average rank in the pre-CMU Action Plan period (2007-2014) with the average rank after the CMU Action Plan introduction (2015-2018), as shown in Fig. 3.

Figure 3. Evolution of CMs composite indicator rank (2015-2018 vs 2007-2014) by EU country.

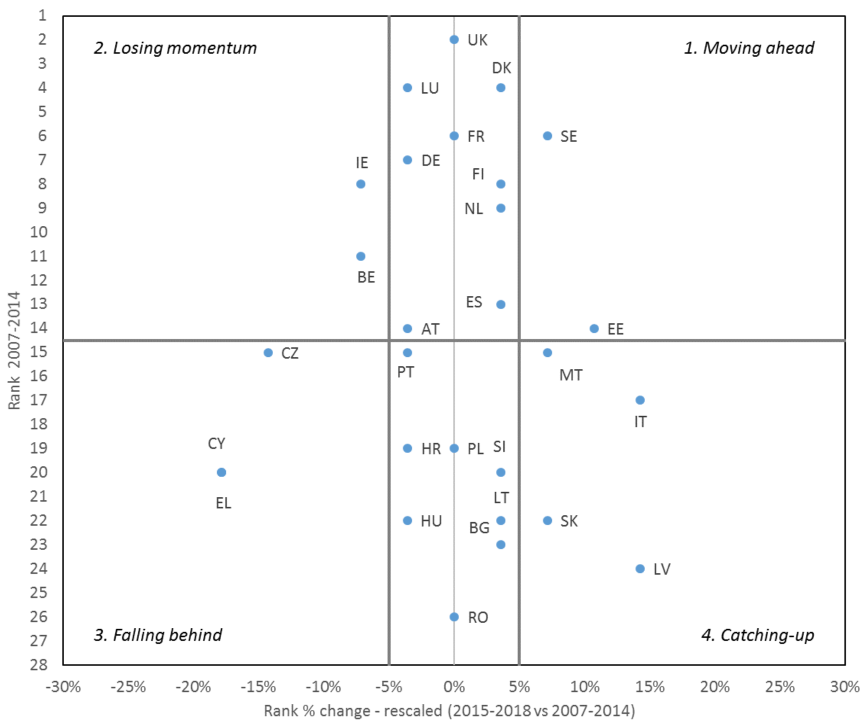


This map depicts the evolution of countries' average rank in the post-CMU Action Plan introduction phase with respect to the pre-CMU Action Plan phase. As in Fig. 1, greenish colors show a positive growth of their average rank, while reddish colors describe lower ranks and yellow stable ones in both periods.

Looking at the phenomenon from a long-term perspective, an average increase of relative ranks emerges. Indeed, ~60% of countries have – at least – not decreased their position in the ranking. Nevertheless, differently from Figure 1, countries belonging to the bottom part of the ranking in the period 2007-2014 represent only approximately 50% of the total of increasing countries (vs 80% when comparing 2018 with 2017). This finding suggests that the improvement in relative ranks is more evenly distributed along the entire ranking.

We now group countries based on the same approach of Fig. 2, but with reference to the post/pre CMU Action Plan setting.

Figure 4. CMs performance matrix (2015-2018 vs 2007-2014) by EU country.



When looking at the post-CMU period, Figure 4 confirms that, while variability in rank % changes emerges along the whole ranking, the most significant deviations (>|5|%) are concentrated in the lower part, with four countries in the “catching-up” quadrant and three in the “falling-behind” (out of fifteen). On the other side, countries in the upper part of the ranking are more stable in their rank (~70% of the total). This result suggests that countries belonging to the lower part of the ranking appear to be more sensible to changes in underlying indicators, rather than those in the top positions, which instead show a rather consolidated performance.

Finally, comparing countries behaviors emerging in Fig. 2 and 4, a light slowdown in CMs metrics relative average performance in the last year of the sample is confirmed. Indeed, the share of countries increasing their ranks in the 2018 vs 2017 setting is lower than the same share in the 2015-2018 vs 2007-2014 setting (~46% vs ~36% of the total).

All the results so far show that process of convergence towards higher average levels of CMs metrics is not completed and is still in progress. In particular, we found an overall average relative improvement since 2015, with a light slowdown in the last year of the sample (2018). This insight comes as expected, since on the one hand, the Action Plan is a progressive program including actions to be implemented throughout the period 2015-2019, on the other hand, many of the dimensions underlying the composite indicator need a congruous time to positively respond to the implemented policies.

In order to investigate the underlying reasons for the evolution of countries' relative performance, in the next section we decompose the phenomenon at the pillar level.

3.2 Capital Markets metrics' performance by pillar

One of the main advantages of composite indicators is to provide a synthetic representation of complex phenomena. Nevertheless, the decomposition of the final scores into its components might be useful to understand which role each underlying indicator or pillar plays in the determination of the final ranking. With this in mind, we investigate pillars' deviations from the overall Capital Markets metrics, building a country ranking for each of the six pillars[12], and comparing them to the plausible ranking emerging from the overall composite indicator. In other terms, we rank countries based on the six objectives underlying the CMU Action Plan to see whether the resulting rankings are (on average) more or less aligned to the plausible ranking obtained from the overall definition of CMs metrics, and which pillars are compensating each other.

We take 2018 as reference year for this analysis. Interestingly, the 2018 rankings based on the individual pillars do not return the same countries' rank emerging from the plausible ranking (depicted in Fig. 1). Specifically, comparing ranks at country level it emerges that pillar 4 (foster retail and institutional investments) shows the most consistent behavior with the plausible ranking one. Conversely, pillars determining larger variability are 5 and 6. This variability indicates that, for example, countries placed in the first part of the plausible ranking are instead penalized when ranked according to pillars 5 (leverage bank capacity to support the economy) and 6 (facilitate cross-border investment and promote financial stability) only. This result could be relevant from a policy perspective because it may highlight countries' vulnerable areas, which may be focus of actions to further enhance CMs development and integration.

Looking at the pillars with the largest difference when comparing their ranks with the plausible one (i.e. 5 and 6), some of the countries with the most relevant gaps are not in the Euro Area, which could act as facilitator for cross-border investment having no risk of currency, according to our interpretation.

All in all, two main results emerge from these findings. First, despite we find evidence suggesting that the fostering of retail and institutional investments (i.e. pillar 4) seems to be the best single proxy for investigating at a glance the CMs performance, the definition of a metrics measuring Capital Market development and integration is inherently multi-faceted, and cannot be fully synthetized by a single pillar without losing relevant information provided by the other remaining pillars. Second, based on the evidence that countries' performances are not homogeneous across pillars and that some (i.e. 5 and 6) determine larger variabilities in rankings, policy-makers/researchers might prioritize their efforts on them to accelerate CMs development and integration process.

4 Dynamic performance assessment

So far, we have analyzed countries' performance through their relative position within yearly rankings in the time-span 2007-2018. Despite allowing comparison among countries along the period, this kind of analysis is still intrinsically static, because rankings are built on a yearly basis and any improvement (worsening) in terms of a country's rank can only provide information on its relative position. In this section, we introduce a set of tools capable of providing two further relevant pieces of information: on one side, whether the country is improving its performance with respect to a common benchmark; on the other side, whether the whole EU-28 group performance is increasing (decreasing) in time. We adopt two complementary methodologies to analyze the evolution of country performances over time: the Data Envelopment Analysis (Farrel, 1957; Charnes et al., 1978) and the Malmquist Index (Malmquist, 1953; Caves et al., 1982; Färe et al., 1994a; Färe et al., 1994b). Moreover, for the sake of robustness, we compare DEA results with those emerging from the robust cluster analysis.

---

[12] See Section A.6 in Annex for technical details concerning the construction of the "by-pillar" composite indicators.

4.1 Data Envelopment Analysis (DEA)

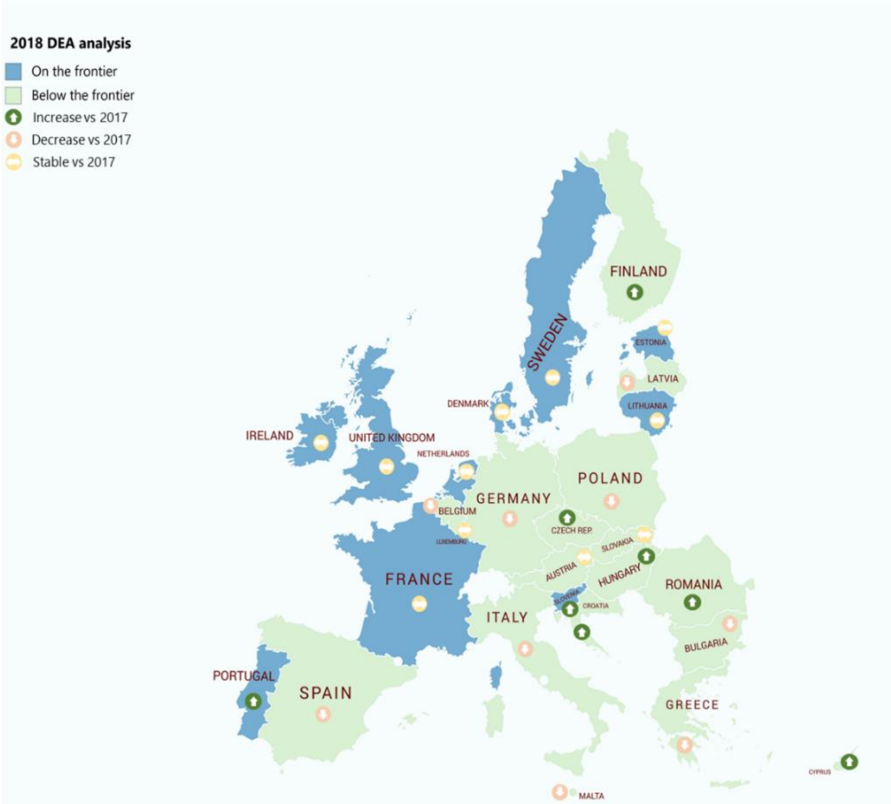
DEA original scope was the measurement of efficiency of productive "decision making units" (usually firms), based on inputs and outputs, in a non-parametric approach within which the functional form of the production function was not *a priori* known. In broader terms, the goal of this kind of analysis is to find, for each analyzed unit, the most reduced set of feasible inputs to determine the highest possible output (Coelli et al., 2005). Consequently, units may then be ranked based on their positioning towards the efficiency frontier: if their efficiency cannot be improved given their inputs, units are on the frontier and get the highest possible score (i.e. 1), while if it can be improved adopting a feasible different combination of inputs, their score is represented as a fraction of 1, based on the distance from the frontier (Charnes et al., 1978).

Interestingly, this approach has also influenced the composite indicator construction literature, basically for two main affinities. First, DEA returns a ranking of units based on some input and output variables, similarly as composite indicators providing a score (and therefore a ranking) based on the aggregation of sub-indicators. Second, DEA allows to obtain this ranking without defining a rigid functional form in the construction of the indicator. This is particularly relevant with regards to the weighting system, to the extent that, by construction, DEA may assign weights to dimensions in order to guarantee the best possible results to the units, given their dimensions. This effect, also known as "Benefit of the Doubt" (BoD) in the related literature (Melyn and Moesen, 1991), is particularly relevant because reduces the level of arbitrariness in the construction of the indicator.

Hence, following Cherchye et al. (2007) DEA estimation approach, on our framework units coincide with the 28 EU countries, inputs are constructed as an equal-to-one dummy variable, and outputs are the 11 indicators[13] underlying the CMs metrics, while the reference periods are yearly changes in the time-span between 2007 and 2018. We first focus on 2018 results, in comparison to 2017, as shown in Fig. 5.

[13] Since some of our indicators contain zeros, to avoid the DEA estimate being biased, we have adopted a transformation of the indicators based on Bowlin (1998), which adds a very low scalar (in our case $1E^{\wedge -16}$) in case of zeros. The same transformation applies for the Malmquist Index Analysis.

Figure 5. Countries on the 2018 DEA frontier vs 2017.



In 2018, ~40% of the countries managed to reach the CMs metrics' frontier (in blue in Fig. 5), being ~46% in 2017, even though still higher than the average value for the whole sample (i.e. ~37%). Interestingly, while four moved away from the frontier from 2017 to 2018, two countries reached it in the same period. At the same time, more than 50% of countries placed below the frontier in 2017, still managed to improve their relative performance, getting nearer or reaching it in 2018. Moreover, the average distance to the frontier calculated as the arithmetic mean of the distance to the frontier by country is substantially stable in 2017 and 2018 (~87%).

Hence, overall these results confirm those emerging from the composite indicator analysis: increasing performance since 2015, with a light slowdown in the last year of the sample. Indeed, while on average the EU experienced a substantially stable performance in Capital Markets development and integration progress from 2017 to 2018, a light decrease of number of countries belonging to the frontier emerged in the same period.

We now replicate the analysis comparing the post- with the pre-CMU Action Plan average performances. In particular, we calculate the average number of countries laying on the frontier in the period 2007-2014 vs 2015-2018, as shown in Tab. 3.

Table 3. Evolution of the number of countries on the frontier.

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Countries on the frontier (#)	12	13	7	14	9	6	9	10	10	11	13	11
Average Ratio (%)	36%								40%			

Two main results emerge. First, the ratio of countries moving to the frontier has increased on average after 2015. Indeed, in the post-CMU Action Plan period ~40% of countries obtained a score equal to 1, while only ~36% in 2007-2014. Therefore, this result represents a further signal for the convergence of EU countries towards the enhancement of Capital Markets development and integration. Second, this methodology also allows to associate the effects of the European Sovereign Debt Crisis with the DEA CMs metrics evolution. Indeed, the ratio of countries on the frontier reached an average of ~34% in that period (2010-2013), decreasing from the ~38% of the previous period (2007-2009), then increasing again since 2014 and stabilizing, as already mentioned, in the post-CMU period approximately at the pre-crisis level.

4.2 Robust Cluster Analysis

In order to provide further supporting evidence to these findings, following Cariboni et al. (2015) we also perform a robust cluster analysis, with the aim of further investigating dynamic trajectories of countries towards EU Capital Markets development and integration. Specifically, we adopt the *Robust Trimmed Clustering* model, developed by Garcia-Escudero et al. (2008a; 2008b), for its flexibility and suitability with our investigation's needs. The details are reported in Section C of the Annex for the sake of synthesis. Indeed, this methodology allows to take advantage of three relevant features. First, it is not necessary to choose *a priori* the proposed number of clusters, as the most adequate number emerges from the basis of statistical criterion (e.g. the Bayesian Information Criterion) which allows the choice of the most "informative" model, effectively reducing a-priori assumption (Fraley and Raftery, 1998). Second, this technique allows to build more flexible clusters, also allowing "non-spherical" shapes, differently from standard methodologies as the *k-means* (Garcia-Escudero et al., 2008a). Third, this technique extracts from the data a number of outliers, which can be excluded from the general clustering estimation to avoid biases (Rocke and Woodruff, 1996).

Our estimation leads to the selection of a model with three outlier countries and two clusters for each year of the sample. Among the outliers, we find some countries showing extremely positive performances compared to the average in one or more indicators[14]. This is especially the case of the UK (with top performance in many areas) and Denmark (particularly strong in the indicator relating to the adoption of Covered Bonds), as well as Sweden and Luxembourg. On the other hand, a few other countries show outlying performance in only some years of the sample, so we do not consider them as strict outliers. Concerning the two clusters, we categorize them into "standard" and "top-performer", where the latter is defined as the group of countries for which most of the components underlying the cluster estimation show a better performance[15]. Interestingly, the robust cluster analysis confirms the insight on pillars 5 and 6's limited correlation with overall CMs metrics (see paragraph 3.2), since a good performance in components accounting for, respectively, deposit taker capital adequacy and cross-border investments, and covered bonds is associated with the membership in the "top-performer" cluster only, respectively, in 25% and 50% of occurrences (i.e. three and six out of eleven years).

Investigating countries' patterns in time, it is possible to notice that the ratio of countries moving to better performances is increasing since 2015. Indeed, in the post-CMU Action Plan period ~28% of countries on average belongs to the "top performer" cluster, while the same ratio for the period 2007-2014 accounted for only ~18%. Moreover, if we isolate the period interested by the European Debt Crisis (2010-2013) it emerges that the ratio of countries belonging to the "top performer" cluster is again lower (~10%) than the post-CMU Action plan period, also decreasing with respect to previous period in the sample (2007-2009).

Hence, overall, these results seem to confirm the findings obtained through the DEA analysis, providing robust evidence on the closer adherence to Capital Markets development and integration progress.

[14] As detailed in Section C of the Annex, we should more precisely refer to components rather than indicators, since the cluster analysis is performed on the four components emerging after a PCA on the original set of 11 indicators.

[15] At least three out of four components emerging from the PCA and adopted for the clustering. In case of a draw (2 vs 2 components), we picked as "top performer" the cluster showing the largest median value of components' centroids location.

4.3 Measuring the evolution in time: Malmquist Index

In their seminal work, Caves et al. (1982) proposed the DEA theoretical adoption within a dynamic context. Since Färe et al. (1994a) and Färe et al. (1994b), several works have adopted this methodology essentially to study productivity changes over time. More recently, its application to composite indicators has renewed and extended the interest in this useful tool, particularly in the social (Bernini et al., 2013; Carboni and Russo, 2015; Peiró-Palomino and Picazo-Tadeo, 2018) and environmental (Kortelainen, 2008; Wang, 2015; Wang et al., 2016; Wang, 2019) fields.

The common intuition behind these works is to extend the DEA approach, to measure the evolution of units' performances between two periods computing the ratio of the distance between each point and a common best practice. Indeed, DEA cannot be used to compare the evolution of units' performances in time, because its scores are inherently referred to different frontiers. Conversely, the Malmquist Productivity Index (MPI) allows to compare the performance change between two observations in time, computing the ratio between the distances of these observations to the same common benchmark (i.e. the frontier). In other terms, the MPI measures the evolution (from period *t* to period *t+1*) of the distance of each country's performances to the frontier in the same period (at period *t* or *t+1*)[16]. Hence, within this setting, it is possible to compare both the performance of a country in time and with the performance of other countries (Bosetti et al., 2007).

Another interesting feature of the MPI is the possibility of decomposing the index into two factors, usually called "Efficiency" and "Technical"[17] effects (Wang, 2015), whose product is by construction equal to the MPI. While these two definitions – and their interpretation – directly come from the productivity literature, they are adaptable to our setting. Intuitively, a country can reach the frontier (i.e. the benchmark performance) through two (combinable) effects. First, it can simply improve its own performance (efficiency effect). Second, it may result nearer to the frontier because the frontier itself moves (technical effect). In our setting, the efficiency effect would therefore measure the "idiosyncratic" Capital Markets development and integration progress of each country, once the overall trend of the market (technical effect) has been taken into account.

More formally, the first factor can be defined as the ratio of the CMs relative performance at time *t* and *t+1* with respect to the frontier at the same period. In our case, it can be interpreted as the share of Capital Markets performance evolution inherently due to the individual country's behavior (we call it "idiosyncratic" effect). In other terms, this effect describes countries' dynamic behavior within a DEA-like setting. In practice, an increase in this factor (i.e. a >1 ratio) represents a convergence towards the frontier, while a decrease (i.e. a <1 ratio) a divergence from it. On the other side, the second factor is defined as the geometric mean of two ratios: 1. the distance of a country's performance at time *t* towards the frontier taken first at time *t* and then *t+1*; 2. the distance of a country's performance at time *t+1* towards the frontier taken first at time *t* and then *t+1*. In our setting, we can think of it as the share of Capital Markets development and integration evolution due to the global common best practice trends (we call it "adherence to global trends" effect). In practice, an increase in this factor (i.e. a >1 ratio) signals that the movement of the frontier from *t* to *t+1* is favorable for the country determining an improvement of the CMs metrics performance.

Hence, we compute the MPI for the CMs metrics (MCM) for our sample, i.e. taking the 28 EU countries as units of analysis within the 2007-2018 period. We first discuss the Malmquist Index for the last year in the sample in order to investigate how countries' performance in 2018 has approached (or moved away from) a common benchmark in time.

[16] To avoid arbitrariness, it is usually calculated towards both frontiers (*t* and *t+1*) and, then, the final MPI is obtained as their geometric average. See Section B.1 in Annex for all technical details on the MPI for CMs metrics construction.

[17] See Section B.2 in Annex for technical details on factors decomposition and their definition.

Table 4. Evolution of Malmquist CMs Index and factor decomposition (yearly average – 2017-2018).

	Overall effect	"Idiosyncratic" effect	"Adherence to global trends" effect
Malmquist CMs Index (EU 28 avg)	+2.5%	-0.3%	+2.8%

First, we should note that in our sample countries almost evenly experience increases or decreases in this index from 2017 to 2018. Second, as shown in Table 4 the geometric average of the 2017-2018 index for the 28 EU countries performances suggests a slight (+2.5%) increase. However, this result comes as the combination of a substantially stable "idiosyncratic" factor (consistently with previously presented DEA results) and a slightly growing (+2.8%) "adherence to global trends" factor. This implies that, on average, during the last year country-specific efforts towards Capital Markets development and integration progresses have not lead to increasing results in CMs metrics, while, conversely, countries have benefitted from the openness to global markets.

We now investigate the evolution of CMs Malmquist performances, comparing the post- and the pre-CMU Action Plan results. Therefore, to do that, we calculate the yearly index by country, and then we take the geometric average of those performances respectively for the periods 2007-2014 and 2015-2018. Finally, we compute the geometric average of these results by country to obtain the average EU-28 performance for the two periods, as shown in Tab. 5 together with the decomposition of the two factors.

Table 5. Evolution of Malmquist CMs Index and factor decomposition (yearly average – 2007-2014, 2015-2018, 2007-2018).

	Overall effect	"Idiosyncratic" effect	"Adherence to global trends" effect
Overall period	+4.4%	-0.02%	+4.5%
2015-2018	+4.5%	+2.2%	+2.2%
2007-2014	+4.3%	-1.3%	+5.8%

According to this synthetic index, on average the 28 EU countries have experienced an increasing yearly performance (+4.4%) since 2007. Moreover, we do not find significant differences in yearly performances when comparing the pre- and post- CMU Action Plan introduction (+4.3% vs +4.5%, respectively). Interestingly, the overall effect seems to be driven in the whole sample by the "adherence to global trends" factor (+4.5%), suggesting that countries' increasing performance in CMs metrics is based on their openness to global markets, more than their country-specific policy efforts. Nevertheless, its contribution appears to be more crucial in the pre-CMU period (+5.8%), being also able to more than compensate a decreasing level of the idiosyncratic factor (-1.3%), in consistency with our DEA results. Finally, since 2015 the two factors have provided an even contribution to the average Malmquist CMs growth, with both factors showing a 2.2% yearly raise.

We now look at Fig. 6 to further investigate the link between the decomposition of the performances into the two factors and the overall evolution of the index, leveraging on geographical heterogeneity.

Figure 6. Openness of capital markets: global trend vs idiosyncratic effects (2007-2018).

This graph combines two kinds of information. First, the color attributed to countries describes the 2007-2018 overall MCM index. Therefore, countries in green significantly (>5%) increased their performance, those in red significantly (<-5%) decreased it, while those in blue showed a more stable performance. Second, countries are scattered along two dimensions, representing the "idiosyncratic" and "adherence to global trends" factors. Looking at Fig. 6 we find higher dispersion in the "adherence to global trends" with respect to the "idiosyncratic" effect.

The interpretation of this finding is twofold. First, global openness to markets seems to matter more than country-specific extra-efforts to determine Capital Markets enhancements. Indeed, for instance we observe that the eight countries showing a significant overall increasing performance (in green in Fig. 6) have also experienced a growing performance of the "adherence global trends" factor (in seven cases >+5%). In other terms, global markets openness seems to be more relevant in predicting a country's overall positive growth in Capital Markets development and integration, rather than its further specific efforts to adhere to dedicated policies. Indeed, all countries showing a significant growth in of the "adherence of global trends" factor obtained an overall growth of the index (and larger than 5% in 70% of the occurrences). Second, a significant growth in the "idiosyncratic" factor itself hardly seems to be the most determinant predictor of the overall increasing performance in capital markets integration (~12% of the occurrences). In other terms, once the global markets trends have been taken into account, the country-specific efforts (possibly including the actions in place in adherence to CMU Action Plan) are not sufficiently effective to catch-up in Capital Markets development and integration according to our metrics.

Overall, two main findings emerge. On one side, EU countries yearly growth in Malmquist CMs metrics is fully explained by their participation to global markets. In other terms, when the global trend is taken into account, country-specific "idiosyncratic" efforts (as, e.g., EU policies application) do not seem to play a role towards Capital Markets development and integration. On the other side, while this latter result is fully applicable also when looking at the 2007-2014 period, since 2015 the global trends have explained only half of the overall growth, with "idiosyncratic" measures emerging as increasingly relevant to sustain the previous yearly growth pace.

5 Comparison with similar metrics in literature

To our knowledge, this is the first attempt in literature aimed at monitoring the development and integration of EU countries' Capital Markets through the use of composite indicators and cluster analysis. Nevertheless, we are aware of other works focusing their analysis on the adoption of composite indicators to measure relatively similar concepts, as economic integration (König and Ohr, 2013), globalization (AT Kearney/Foreign Policy Globalization Index, 2002) and financial integration (Baele et al., 2004).

A more recent and similar to our case is the European Central Bank periodic report presenting a composite indicator measure of Financial Integration in Europe (European Central Bank, 2018). Specifically, the ECB produces two metrics, a "price-based" composite - built with price-based indicators measuring four segments, i.e. the money, bonds, equity and banking markets - and a "quantity-based" composite - aggregating data on cross-border holdings for different asset classes. Since we focus on volumes and not on prices, our composite indicator may be in principle comparable only with the latter.

However, some differences emerge between ECB's and our composite indicator, in terms of scope, granularity and methodology. The first and most important concerns the scope of the analysis. Indeed, ECB measures financial integration and not Capital Markets' development and integration progress. This difference is clear when looking at the indicators underlying the composite. Specifically, ECB uses the share of cross-border lending among Monetary Financial Institutions (MFIs), MFIs' and investment funds' shares of cross-border holdings of debt securities, and MFIs' and investment funds' cross-border holdings of equity. Conversely, we adopt eleven indicators, directly attributed to the six key objectives defined by the policy maker as being relevant for CMs development in the EU. Some of these indicators are not linked to financial integration (e.g., the adoption of alternative bonds). Moreover, our approach allows to decompose the metrics by pillar and to more easily compare performances' evolutions with undertaken policy actions. The second difference is that ECB indicator is measured on the Euro area, while our composite is at the EU-28 level, consistently with the CMU policy. Third, our indicator is also available at the individual country level, while ECB proposes an aggregated composite for the EMU. Our approach allows a geographical comparison of relative performance across countries. Fourth, ECB adopts a unique combination of normalization, weighting and aggregation techniques, while in this work we adopt a robust approach, taking into account 32 different methodological combinations and to highlight any differences in the emerging results. Fifth, the 2018 ECB version of the indicators is presented comparing data as of end of first quarter of each year, while we use end of year figures.

All these differences do not allow for a direct comparison of the two results. Nevertheless, some common patterns seem to emerge. Indeed, according to the ECB quantitative indicator, within the time-span between late nineties and 2017, financial integration has reached its peak between 2005 and 2008, then decreased until 2013, and more recently (2015-2017) increased again, despite not coming back to the pre-crisis levels. This result seems to be in line with our DEA analysis – also confirmed by the robust clustering methodology – which underlines increasing performances in the post-CMU Action Plan period, though not fully compensating the decrease occurred in the previous years.

6 Conclusive remarks

The goal of this work was to provide a measure of the EU-28 countries progress towards Capital Markets' (CMs) development and integration. First of all, we defined the perimeter of the analysis adopting as theoretical framework the Capital Markets Union (CMU) Action Plan introduced by the European Commission in 2015, since it is based on six key objectives to be pursued in order to enhance CMs development and integration across Europe.

Due to the multi-faceted nature of CMU Action Plan, we adopted the composite indicator as the most appropriate tool to synthetize complex phenomena in a (set of) scores and related rankings. To reduce arbitrariness in its construction, we built a set of composite indicators, organized in "experiments", according to the different adoptable techniques. We then obtained a “plausible” rank as the average value of the experiments for each country and year of the sample. We analyzed changes in countries’ plausible ranks. First insights suggested that convergence towards higher level of CMs metrics is still in progress, with increasing average results after 2015, and a light slowdown in 2018 vs 2017.

In order to investigate which of the rationales underlying counties’ performance, we compared their ranks, respectively built on the six key objectives, and the overall CMs metrics. It emerged that countries’ performances across the key objective is not homogenous, and that some of them (in particular related to soundness of banks, and to cross-border investment facilitation) show peculiarities compared to the overall metrics. This result, even confirmed by the clustering analysis, allows to investigate policy implications in the CMs development and integration area.

We then moved to a dynamic assessment, based on Data Envelopment Analysis (DEA) and Malmquist Index Analysis, in order to investigate the evolution of countries’ performance with respect to a yearly benchmark. DEA analysis showed an increasing average number of countries belonging to the CMs metrics frontier after 2015, in comparison to the pre-Action Plan period. Moreover, this methodology allowed to highlight countries deteriorating performance during the European Sovereign Debt crisis (2010-2013). We obtained similar results through the robust clustering analysis, which showed that the ratio of countries moving to the “top-performer” cluster from the “standard” one is increasing in recent years, after a reduction during the European crisis. These results are also substantially in line with ECB (2018) findings, obtained through the construction of a Financial Integration composite indicator.

Lastly, the Malmquist Index allowed to investigate the evolution of countries’ performances from 2007 to 2018. We found an increasing average yearly Malmquist CMs metrics performance for the 28 EU countries in the whole sample, with no significant differences between the post and pre CMU Action Plan 2015 launch. Moreover, we were able to account for the factors underlying this growth looking at the whole time-span. It emerged that it was mainly driven by countries’ adherence to open global Capital Markets, especially in the period 2007-2014, rather than “idiosyncratic” factors (as also suggested by the DEA), which, conversely, seem to start being more relevant since 2015. Finally, focusing on countries experiencing significant growth in Malmquist CMs metrics, it emerges that the global markets openness is decisive in predicting EU countries’ CMs enhancements, rather than their country-specific policy actions. Hence, a further effort on the implementation of European and national policies in this area could be useful to improve the overall CMs development and integration in Europe.

Overall, all our findings suggest that CMs development and integration progress has started, though the process of convergence towards higher levels is not completed and is still ongoing. This result is not surprising since the Action Plan was still ongoing at the end of 2018, and that the dimensions underlying our metrics may not immediately respond to the identified policies. Nevertheless, our metrics seem to be mainly associated to EU countries average adherence to the world-wide trend towards increasing openness of global capital markets. At the same time, the implementation of (further) policy actions is becoming more relevant and, possibly, necessary to consolidate this growth in time.

References

Ülengin, B., Ülengin, F., & Güvenç, Ü. (2001). A multidimensional approach to urban quality of life: The case of Istanbul. *European Journal of Operational Research*, *130*(2), 361-374.

AFME (2018). Capital Markets Union. Measuring progress and planning for success. https://www.afme.eu/Portals/0/globalassets/downloads/publications/afme-cmu-kpi-report-4.pdf

A.T. Kearney/Foreign Policy (2002) Globalization Index (Washington, DC: Carnegie Endowment for International Peace).

Baele, L., Ferrando, A., Hördahl, P., Krylova, E., & Monnet, C. (2004). Measuring European financial integration. *Oxford review of economic policy*, *20*(4), 509-530.

Benczúr, P., Karagiannis, S., & Kvedaras, V. (2018). Finance and economic growth: financing structure and non-linear impact. *Journal of Macroeconomics*.

Beliakov, G., Pradera, A., & Calvo, T. (2007). *Aggregation functions: a guide for practitioners* (Vol. 221). Heidelberg: Springer.

Bernini, C., Guizzardi, A., & Angelini, G. (2013). DEA-like model and common weights approach for the construction of a subjective community well-being indicator. *Social indicators research*, *114*(2), 405-424.

Billor, N., Hadi, A. S., & Velleman, P. F. (2000). BACON: blocked adaptive computationally efficient outlier nominators. *Computational statistics & data analysis*, *34*(3), 279-298.

Böhringer, C., & Jochem, P. E. (2007). Measuring the immeasurable—A survey of sustainability indices. *Ecological economics*, *63*(1), 1-8.

Bosetti, V., Cassinelli, M., & Lanza, A. (2007). Benchmarking in tourism destinations; keeping in mind the sustainable paradigm. In *Advances in modern tourism research* (pp. 165-180). Physica-Verlag HD.

Bowlin, W. F. (1998). Measuring performance: An introduction to data envelopment analysis (DEA). *The Journal of Cost Analysis*, *15*(2), 3-27.

Bullen, P. S. (2013). *Handbook of means and their inequalities* (Vol. 560). Springer Science & Business Media.

Carboni, O. A., & Russu, P. (2015). Assessing regional wellbeing in Italy: An application of Malmquist–DEA and self-organizing map neural clustering. *Social indicators research*, *122*(3), 677-700.

Cariboni, J., Pagano, A., Perrotta, D., & Torti, F. (2015). Robust clustering of EU banking data. In *Advances in Statistical Models for Data Analysis* (pp. 17-25). Springer, Cham.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: Journal of the Econometric Society*, 1393-1414.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, *2*(6), 429-444.

Cherchye, L., Lovell, C. K., Moesen, W., & Van Puyenbroeck, T. (2007). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review*, *51*(3), 749-779.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.

Cohen-Addad, V., Kanade, V., Mallmann-Trenn, F., & Mathieu, C. (2019). Hierarchical clustering: Objective functions and algorithms. *Journal of the ACM (JACM)*, *66*(4), 26.

Cournède, B., & Denk, O. (2015). *Finance and economic growth in OECD and G20 countries* (No. 1223). OECD Publishing.

Dick-Nielsen, J., Gyntelberg, J., & Sangill, T. (2012). Liquidity in government versus covered bond markets.

Ebert, U., & Welsch, H. (2004). Meaningful environmental indices: a social choice approach. *Journal of Environmental Economics and Management*, *47*(2), 270-283.

European Central Bank (2018). Financial Integration in Europe. https://www.ecb.europa.eu/pub/pdf/fie/ecb.financialintegrationineurope201805.en.pdf

European Commission (2004). Economic Sentiment Indicator, DG ECFIN, Brussels, http://europa.eu.int/comm/economy_finance/index_en.htm

European Commission (2015). *Economic Analysis Accompanying the Action Plan on Building a Capital Markets Union*, Brussels, 30.9.2015, SWD (2015)

European Commission (2016). *Commission Staff Working Document: European Financial Stability and Integration Review (EFSIR): A focus on Capital Markets Union*

Fagerberg, J., & Srholec, M. (2005). Catching up: what are the critical factors for success. *Vienna, UNIDO, Background paper for the UNIDO Industrial Development Report*.

Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1994a). Productivity developments in Swedish hospitals: a Malmquist output index approach. In *Data envelopment analysis: theory, methodology, and applications* (pp. 253-272). Springer, Dordrecht.

Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994b). Productivity growth, technical progress, and efficiency change in industrialized countries. *The American economic review*, 66-83.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, *120*(3), 253-281.

Floridi, M., Pagni, S., Falorni, S., & Luzzati, T. (2011). An exercise in composite indicators construction: Assessing the sustainability of Italian regions. *Ecological economics*, *70*(8), 1440-1447.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The computer journal*, *41*(8), 578-588.

Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., & Wu, J. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological indicators*, *81*, 491-502.

García-Escudero, L. Á., Gordaliza, A., Matrán, C., & Mayo-Iscar, A. (2008a). A general trimming approach to robust cluster analysis. *The Annals of Statistics*, *36*(3), 1324-1345.

García-Escudero, L. Á., Gordaliza Ramos, A., Matrán Bea, C., & Mayo Iscar, A. (2008b). The TCLUST approach to robust cluster analysis.

García-Escudero, L. Á., Gordaliza, A., Matrán, C., & Mayo-Iscar, A. (2010). A review of robust clustering methods. *Advances in Data Analysis and Classification*, *4*(2-3), 89-109.

Goedkoop, M., & Spriensma, R. (2001). The Eco-indicator 99: A damage oriented method for life cycle assessment. *Methodology report*, *2*.

Grabisch, M., Marichal, J. L., Mesiar, R., & Pap, E. (2009). *Aggregation functions* (Vol. 127). Cambridge University Press.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, *141*(1), 61-94.

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

Jolliffe, I. T. (1973). Discarding variables in a principal component analysis. II: Real data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *22*(1), 21-31.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, *20*(1), 141-151.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, *1*(6), 90-95.

König, J., & Ohr, R. (2013). Different Efforts in E uropean Economic Integration: Implications of the EU I ndex. *JCMS: Journal of Common Market Studies*, *51*(6), 1074-1090.

Kortelainen, M. (2008). Dynamic environmental performance analysis: a Malmquist index approach. *Ecological Economics*, *64*(4), 701-715.

Langfield, S., & Pagano, M. (2016). Bank bias in Europe: effects on systemic risk and growth. *Economic Policy*, *31*(85), 51-106.

Lee, K. R., Leem, B. I., Lee, C. W., Thales, S., & Lee, C. (2011). Malmquist Productivity Index using DEA frontier in Stata. *URL: https://pdfs. semanticscholar. org/847f/0d96a2b8e5581ba45e32406ec16a8858839c. pdf*.

Luzzati, T., & Gucciardi, G. (2015). A non-simplistic approach to composite indicators and rankings: an illustration by comparing the sustainability of the EU Countries. *Ecological economics*, *113*, 25-38.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadistica y de Investigacion Operativa*, *4*(2), 209-242.

Melyn, W., & Moesen, W. (1991). Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. *Public economics research papers*, 1-24.

Munda, G., & Nardo, M. (2005). Non-compensatory composite indicators for ranking countries: A defensible setting. *EUR Report, EUR*, *21833*.

Munda, G., & Saisana, M. (2011). Methodological considerations on regional sustainability assessment based on multicriteria and sensitivity analysis. *Regional Studies*, *45*(2), 261-276.

Nardo, M., Ndacyayisenga, N., Pagano, A., & Zeugner, S. (2017). Finflows: a database for bilateral financial investment stocks and flows. *European Commission JRC Technical Reports*.

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Tools for composite indicators building. *European Comission, Ispra*, *15*, 19-20.

Nardo, M., Saisana, M. S., & Saltelli, A. A., & Tarantola, S. (2008). Handbook on constructing composite indicators: Methodology and user guide.

Parker, J. D. E. (1991). *Environmental reporting and environmental indices* (Doctoral dissertation, University of Cambridge).

Peiró-Palomino, J., & Picazo-Tadeo, A. J. (2018). OECD: One or many? Ranking countries with a composite well-being indicator. *Social Indicators Research*, *139*(3), 847-869.

Pollesch, N., & Dale, V. H. (2015). Applications of aggregation theory to sustainability assessment. *Ecological Economics*, *114*, 117-127.

Reddy, C. K., & Vinzamuri, B. (2018). A survey of partitional and hierarchical clustering algorithms. In *Data Clustering* (pp. 87-110). Chapman and Hall/CRC.

Riani, M., Perrotta, D., & Torti, F. (2012). FSDA: A MATLAB toolbox for robust analysis and interactive data exploration. *Chemometrics and Intelligent Laboratory Systems*, *116*, 17-32.

Rocke, D. M., & Woodruff, D. L. (1996). Identification of outliers in multivariate data. *Journal of the American Statistical Association*, *91*(435), 1047-1061.

Rokach, L., & Maimon, O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook* (pp. 321-352). Springer, Boston, MA.

Saisana, M. (2012). A do-it-yourself guide in Excel for composite indicator development. *Joint Research Centre, Italy: European Commission*.

Saisana, M., & Munda, G. (2008). Knowledge Economy: measures and drivers. *EUR Reports*, *23486*.

Saisana, M., & Tarantola, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development* (p. 214). Ispra, Italy: European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, Technological and Economic Risk Management Unit.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C. T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664-681.

Schoenmaker, D., & Bosch, T. (2008). Is the home bias in equities and bonds declining in Europe?. *Investment Management and Financial Innovations*, *5*, 90-102.

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2007). Development of composite sustainability performance index for steel industry. *Ecological Indicators*, *7*(3), 565-588.

UNDP (1990). Human Development Report 1990. Oxford University Press, New York.

Valdés, J. (2018). Arbitrariness in multidimensional energy security indicators. *Ecological Economics*, *145*, 263-273.

Wang, D. D. (2019). Performance assessment of major global cities by DEA and Malmquist index analysis. *Computers, Environment and Urban Systems*, *77*, 101365.

Wang, H. (2015). A generalized MCDA–DEA (multi-criterion decision analysis–data envelopment analysis) approach to construct slacks-based composite indicator. *Energy*, *80*, 114-122.

Wang, H., Zhou, P., & Wang, Q. (2016). Constructing slacks-based composite indicator of sustainable energy development for China: A meta-frontier nonparametric approach. *Energy*, *101*, 218-228.

Weber, S. (2010). bacon: An effective way to detect outliers in multivariate data using Stata (and Mata). *The Stata Journal*, *10*(3), 331-338.

Yeheyis, M., Hewage, K., Alam, M. S., Eskicioglu, C., & Sadiq, R. (2013). An overview of construction and demolition waste management in Canada: a lifecycle analysis approach to sustainability. *Clean Technologies and Environmental Policy*, *15*(1), 81-91.

Technical Annex

A Composite Indicator Construction

A.1 Dataset construction and multivariate analysis

The original dataset is a panel composed of 3,696 observations[18], deriving from 11 indicators taken for the 28 EU countries in 12 years (2007-2018).

A first inspection of the raw data was conducted to check for the presence of possible outlying observations. In particular, since we want to detect extreme country behaviors across different dimensions of the analyzed phenomenon and years, following Billor et al. (2000) we looked for multivariate outliers based on Mahalanobis distance and found seven potential outliers[19]. We then further inspected these observations at the univariate level. It emerged that all seven cases were strongly affected by the behavior of a single indicator (i.e. adoption of alternative instruments), showing an extremely high performance with respect to the panel, as empirically confirmed by a high kurtosis (>3.5) and skewness (>2) of the indicator across countries in the four affected years. Hence, following Saisana (2012) we opted for Winsorization, i.e. we substituted the seven affected values with maximum value of the same indicator across all years and countries. Finally, a new multivariate outlier analysis was conducted on the reviewed dataset, now suggesting the absence of multidimensional outliers. The resulting dataset was used for all the analyses conducted in this work. Table A.1 shows descriptive statistics for all indicators.

18 23 observations (<1% of the total) were manually imputed to avoid missing observations. Most of them (i.e. 20) were relative to the IMF "Deposit taker capital adequacy" indicator. In this case, our imputing strategy was to linearly interpolate missing observations using two similar indicators, first from European Central Bank ("Tier 1 ratio") and, if also missing for the same country and year, from World Economic Forum ("Soundness of Banks" survey data) for known observations. Similarly, we imputed "Institutional Investments" for Czech Republic in 2017 and 2018 linearly interpolating missing observations, using known observations from our "Retail Investments" indicator. Finally, we imputed "Cross-border Portfolio Debt and Equity Investments" for Luxembourg in 2007 as the average value taken for the remaining period (2008-2018).

19 Specifically, Latvia (2015-2017), Estonia (2015), Lithuania (2017-2018) and Slovenia (2018) - we used "bacon" STATA routine (Weber, 2010) to conduct this analysis.

Table A.1. Descriptive statistics for the periods 2007-2014 and 2015-2018.

Indicator	2007-2014			2015-2018		
	Average	Min	Max	Average	Min	Max
Adoption of Private Equity instrument *(%)*	0.64%	0%	5.70%	0.93%	0%	7.64%
Adoption of Venture Capital instrument *(%)*	0.02%	0%	0.38%	0.04%	0%	0.25%
Relevance of Non-Financial Corporations Debt *(%)*	8.3%	0%	31.2%	9.8%	0%	31.0%
Relevance of Non-Financial Corporations Equity *(%)*	36%	1%	148%	39%	2%	127%
Adoption of alternative instruments *(%)*	0.05%	0%	2.56%	1.59%	0%	8.79%
Public-Private Partnerships (PPPs) *(B€)*	0.64	0	11.38	0.31	0	4.20
Retail Investments (assets) *(%)*	63%	3%	225%	71%	9%	247%
Institutional Investments (assets) (%)	47%	1%	239%	56%	5%	252%
Covered Bonds *(%)*	1.6%	0%	25.0%	1.3%	0%	20.7%
Deposit taker capital adequacy *(%)*	13%	6%	35%	18%	11%	31%
Cross-border Portfolio Debt and Equity Investments *(%)*	23%	1%	79%	24%	2%	71%

We then analyzed the correlation between indicators in order to check for collinearity and avoid redundancy. The idea would be to exclude (or to differently weigh) indicators in case they are persistently correlated over time, in particular if correlation occurs within the same dimension. Being our sample structured as a panel, we performed the correlation analysis between indicators for each analyzed year (2007-2018) on z-score normalized data. The definition of a threshold for the identification of relevant correlation is not uniquely determined. Nevertheless, following Nardo et al. (2008) we decided to be conservative in our approach, in order to avoid losing too much information, *ex ante* excluding indicators. Hence, based on Saisana (2012) we decided to evaluate for further investigations only those cases with a level of correlation significant and higher than 0.92. This case does not occur in our sample for any pairwise combination of indicators and years of analysis[20]. For the sake of robustness, we have also checked the result setting the threshold at a lower level (i.e. 0.8). In this case, correlation emerges in less than 1% of the occurrences (i.e. 2 cases on 660 possible combinations), in one case within the same dimension (Private Equity vs Venture Capital in 2012) and once across different dimensions (Institutional Investments vs Private Equity in 2010). Hence, we can conclude that in our sample a reduction of the number of indicators is not necessary, considering that no persistent correlation among indicators over time emerges.

[20] We also replicated the analysis adopting different normalization methodologies (min-max, distance from the leader and distance from the mean, see paragraph 3.1), with no significant differences.

A.2 Normalization techniques

The first element of heterogeneity in composite indicators counstruction derives from normalization, that is from the different techniques allowing the transformation of raw data - expressed in different units - into homogeneous units which can be correctly aggregated at a later stage. Building on Nardo et al. (2005), Table A.2 synthetically describes the different techniques adopted in this work.

Table A.2. Normalization techniques.

Name	Rule
Rank	*Each country gets a decreasing score (from 28 to 1) depending on its position in each indicator ranking*
Z-score	$I_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc=\mu c}^{t^*}}{\sigma_{qc=\mu c}^{t}}$
Min-Max	$I_{qc}^{t} = \frac{x_{qc}^{t} - \min_{c} x_{c}^{t^*}}{\max_{c} x_{c}^{t^*} - \min_{c} x_{c}^{t^*}}$
Distance from the best performance	$I_{qc}^{t} = \frac{x_{qc}^{t}}{\max_{c} x_{c}^{t^*}}$
Distance from the average performance	$I_{qc}^{t} = \frac{x_{qc}^{t}}{x_{qc=\mu c}^{t^*}}$

I_{qc}^{t} is the normalised indicator for indicator *q*, country *c*, and year *t*. t^* is the reference year, μ indicates the arithmetic mean, σ indicates the standard deviation, and *min* and *max* the minimum and the maximum values.

Hence, we adopt five methodologies: the ranking, the standardization (z-score), the rescaling (min-max), and the distance from a reference point (both from the average and from the best performance). These techniques can be applied in the presence of cross-sectional observations, therefore in case of countries observed in a single time period (usually a reference year). However, they can also be adjusted to describe phenomena that affect more countries over time (e.g. in a panel setting). Dealing with panel data, we need to normalize data for ranking and z-score with a year-by-year approach, while we adopt a single reference year for the entire panel[21] for rescaling and distance from reference point techniques.

A.3 Weighting systems

Coming to the weighting system and the aggregation, to minimize the level of arbitrariness in their definition we refer to the approach of Gan et al. (2017), who collected a large number of possible techniques and propose a structured scheme to define which of these should be used according to research needs. Concerning the weighting system, we exclude from the list of adoptable techniques those based on public[22] or experts[23] opinions, since similar weights are not currently publicly available and any new collection goes beyond the scope of this work. Among the remaining methods, we are interested in identifying those allowing comparability of rankings over time and space. Hence, we end up using four different techniques, two basic ones (Equal Weighting by Indicator and Equal Weighting by Pillar) and two statistic-based ones (Principal Component Analysis / Factor Analysis and Regression Analysis).

More specifically, the Equal Weighting by Indicator (EWI) approach attributes the same weight to all the indicators, while the Equal Weighting by Pillar (EWP) one attributes the same weight to the six pillars, and within

21 We adopted 2007 (i.e., the first available year) for all indicators, with the exception of adoption of alternative instruments for which we used 2018 (i.e., the last available year) because alternative bonds emissions started in 2012.

22 Public Opinion (Parker, 1991).

23 Budget Allocation (Goedkoop and Spriensma, 2001), Analytic hierarchy process (Singh et al., 2007) and Conjoint analysis (Ülengin et al., 2001), Balance of Opinions (European Commission, 2004).

them indicators are weighted according to the EWI method. Although intuitive, easily replicable, and therefore adopted for the construction of many well-known indicators such as the HDI (UNDP, 1990), these two modalities still imply an arbitrary choice on the relevance of single underlying indicators and, consequently, on the overall definition of the composite indicator. The two statistic-based methodologies partially overcome this constraint by making weights directly emerge from the data. In particular, the Principal Component Analysis / Factor Analysis (PCA-FA) reduces the dimensionality of the original dataset, without a significant loss of information, to a subset of relevant factors. Once factors are estimated, identified on the basis of the eigenvalues and rotated, it is possible to construct the weights to be attributed to each indicator as the scaled-to-one value of the factor loading square, in proportion to the variance explained by each factor (Nardo et al., 2008). Despite some limitations of interpretability of the results due to a lower degree of explicability of the meaning of the single components (Gan et al., 2017) and to the need to analyze indicators with a good level of correlation (Nardo et al., 2008), this methodology still allows to reduce the risk of double weighting, otherwise implicit in the EWI and EWP techniques (Yeheyis et al., 2013). The regression analysis allows the estimation of weights through a linear model, in our case on a fixed-effects panel regression, in which the regressors are the indicators making up the composite indicator, and the dependent variable is a proxy of the benefit associated with a high level of the composite indicator (Nardo et al., 2008). This method has the advantage of being purely statistical, without data manipulation and restriction, although it is based on the dependent variable choice which strongly influences results (Nardo et al., 2005). In our case, we have used as dependent variable the GDP annual growth rate, since the achievement of a real capital market in the EU is related to a greater mobility of capital (European Commission, 2016; European Commission, 2015), aimed at reducing financial instability through the development of the financial system, thus favoring economic growth[24]. The weighting system is then expressed as the (rescaled-to-one) absolute value of the statistically significant coefficients estimated through the regression.

A.4 Aggregation

In the final phase of indicator construction, we aggregate normalized and weighted data. Based on the Gan et al. (2017) categorization, three aggregation macro-methods[25] are possible: 1. additive, 2. geometric, 3. non-compensatory. Table A.3 briefly represents the three methods.

[24] For further details on relevant literature related to the positive impact of financial system development on economic growth and, in particular, to the need of diversification of the composition of finance among banking and capital markets to achieve growth see, for instance, Cournède and Denk (2015), Langfield and Pagano (2016) and Benczúr et al. (2018).

[25] Greco et al. (2019) identify a further category named "Mixed Strategies" which we do not consider in this work, basically being combinations / hybridizations of the main three categories.

Table A.3. Aggregation techniques.

Name	Rule	Type
Linear (additive) average	$CI_c^t = \sum_{q=1}^{Q} w_q I_{qc}^t$	Compensatory
Geometric average	$CI_c^t = \prod_{q=1}^{Q} {x^t}_{qc}^{w_c}$	Partially compensatory
Multi-criteria	*Condorcet-Kemeny-Young-Levenglick (CKYL) ranking procedure (Munda and Nardo, 2005)*	Non-compensatory

CI_c is the composite indicator for country *c*, with $\sum_{q=1}^{Q} w_q = 1$ and $0 \le w_q \le 1$, for all q = 1, ..., Q and c = 1, ..., C, where w_q is the weight and I_{qc} is the value of the country normalized indicator.

The first method is rather advantageous from the computational point of view as it can be obtained from an arithmetic (weighted) average of the normalized indicators. This simple approach would also imply the immediate practical advantage, that is the possibility of individually isolating, even *ex post*, the marginal contribution of each indicator to the overall composite indicator score, in order to draw descriptive conclusions. However, this interpretation is subject to the validity of the condition of mutually preferentially independence between the indicators, i.e. a very strict requirement that implies that trade-offs between two indicators are independent from all the other possible trade-offs of the dataset (Nardo et al., 2008). At the same time, the additive method is by definition compensatory (Greco et al., 2019), therefore allowing good (bad) performances in certain indicators to compensate for bad (good) performances in others. Therefore, overall, it is preferable not to exclusively adopt this technique in the case of relevant interactions between indicators (Gan et al., 2017). The geometric method represents a middle way between a compensatory and a non-compensatory approach towards the aggregation issue. In exchange for a slightly more complex calculation process (typically the weighted geometric mean), this method of aggregation is useful to reduce the compensability of poor performance in some indicators by high values in others due to the "geometric-arithmetic means inequality" (Beliakov et al., 2007; Bullen, 2013). Although the compensation is not constant as in the linear method and is lower for lower levels of the indicators, the geometric method is therefore not fully non-compensatory (Greco et al., 2019). Moreover, the assumptions related to mutually preferentially independence are still valid also in this case (Nardo et al., 2008). Finally, the third method is fully non-compensatory, being based on two main principles: on the one hand, its goal is to generate an order among the performances of the analyzed countries, thus ultimately constructing their own rankings (Munda and Nardo, 2005); on the other hand all the dimensions of the analyzed phenomenon must contribute separately to the definition of the phenomenon as a whole without compensations (Grabisch et al., 2009; Pollesch and Dale, 2015). Since the non-compensability of the results comes at the cost of a greater computational effort, especially in the presence of many analyzed countries, several indicators (Gan et al., 2019) and combinations of normalization and weighting techniques, in this work we stick to the first two aggregation methodologies.

A.5 Rankings and Experiments set-up

Overall, we produced 32 different basic rankings based on the feasible combinations of normalization, weighting and aggregation methodologies, as shown in Table A.4. The rankings are generated for any year of the sample (2007-2018).

Table A.4. Composite indicators rankings.

Rankings	Normalization	Weighting	Aggregation
1	Rank	EWI	Linear
2	Rank	EWP	Linear
3	Rank	PCA	Linear
4	Rank	RA	Linear
5	z-score	EWI	Linear
6	z-score	EWP	Linear
7	z-score	PCA	Linear
8	z-score	RA	Linear
9	Min-Max	EWI	Linear
10	Min-Max	EWP	Linear
11	Min-Max	PCA	Linear
12	Min-Max	RA	Linear
13	Distance from the average perf.	EWI	Linear
14	Distance from the average perf.	EWP	Linear
15	Distance from the average perf.	PCA	Linear
16	Distance from the average perf.	RA	Linear
17	Distance from the best performance	EWI	Linear
18	Distance from the best performance	EWP	Linear
19	Distance from the best performance	PCA	Linear
20	Distance from the best performance	RA	Linear
21	Rank	EWI	Geometric
22	Rank	EWP	Geometric
23	Rank	PCA	Geometric
24	Rank	RA	Geometric
25	Distance from the average perf.	EWI	Geometric
26	Distance from the average perf.	EWP	Geometric
27	Distance from the average perf.	PCA	Geometric
28	Distance from the average perf.	RA	Geometric
29	Distance from the best performance	EWI	Geometric
30	Distance from the best performance	EWP	Geometric
31	Distance from the best performance	PCA	Geometric
32	Distance from the best performance	RA	Geometric

We then produce 11 different experiments, focusing each of them on one technique, so that five experiments come from normalization, four from weighting and two from aggregation. Specifically, we started focusing with composite indicators built using “rank” normalization techniques (experiment 1), then “z-score” (experiment 2), “min-max” (experiment 3), “distance from the average performance” (experiment 4) and “distance from the best performance” (experiment 5). Then, we moved to composite built using “EWI” weighting (experiment 6), “EWP” (experiment 7), “PCA” (experiment 8) and “RA” (experiment 9). Last, two last set of experiments are based on “linear” (experiment 10) and “geometric” (experiment 11) aggregations. For each of the experiments, we calculate the average rank by country and year. Finally, we take the average of the ranks of the 11 experiments in order to get the final “plausible” rank (by country and year).

A.6 Countries' performance by pillar

In order to compare countries' performances by pillar with the overall CMs metrics, we investigate the differences of countries' CMs composite indicator plausible ranking respectively with the six rankings computed at the pillar level. To do that, we replicate the construction of one composite indicator for each pillar, for the sake of simplicity adopting just one methodological approach (z-score normalization, Equal Weighting by Indicator[26] and linear aggregation), whose methodology is chosen since it is the one that more closely resembles[27] our 2018 plausible ranking. Then, we compare each of them to the plausible ranking obtained for 2018 for the overall CMs composite indicator. If the difference between one pillar's rank and the plausible rank belongs to the interval [-1;1], then we consider it as substantially negligible, therefore we interpret pillar's performance as aligned with the overall one. In other terms, taking the pillar as a proxy, it could be a "good guess" of the overall CMs rank. Conversely, if the same difference is larger (lower) than one, we conclude that the analyzed pillar is under(lower)-estimating the overall rank. Hence, if the number of "good guesses" of each pillar is limited, then we conclude that it provides larger variability in the CMs ranking definition. For the sake of robustness, we also adopt the sum of square metrics to measure the differences in rankings.

Overall, in our sample, the fifth and the sixth are the most distant pillar to the overall metrics, showing respectively the highest sum of squares the lowest share of "good guesses". On the other side, pillar 4 resembles the most the overall CMs metrics, showing the highest proportion of "good guesses" of the plausible rank and the lowest sum of square.

B DEA-Malmquist

B.1 Malmquist Index on the CMs metrics

Following Coelli et al. (2005), we define for the Malmquist Productivity Index for the CMs metrics, i.e. the MCM:

$$MCM_t^{t+1} = \left[\frac{CM^t(t+1)}{CM^t(t)} * \frac{CM^{t+1}(t+1)}{CM^{t+1}(t)}\right]^{1/2}, \qquad (1)$$

where t denotes time period. Looking at the first factor, $CM^t(t+1)$ represents the CMs indicator for a unit analyzed in year $t+1$ adopting the best practice frontier in year t, while $CM^t(t)$ measures the same unit in year t versus the same best practice. Moving to the second factor, $CM^{t+1}(t+1)$ represents the CMs indicator for a unit observed in year $t+1$, in this case referred to the best practice frontier in the same year, and $CM^{t+1}(t)$ describes the same unit in year $t+1$ with respect to the best practice in year t. Therefore, the two ratios measure the evolution of unit between two consecutive years, *vis-à-vis* two different frontiers (respectively, t and $t+1$ for the former and the latter ratios). While both factors of the multiplication taken as stand-alone elements can correctly represent *per se* a measure for the Malmquist index, the first as a Laspeyres and the second as a Paasche, the final MCM is taken as their geometric mean (as in Eq. 1), to avoid an arbitrary selection of the base year. When MCM > 1 we find a positive growth of the CMs metrics from year t to year $t+1$ for the observed unit, while the opposite occurs if MCM < 1. In our setting, units coincide the 28 EU countries.

[26] We a priori excluded Equal Weighting by Pillar, Principal Components Analysis and Regression Analysis system of weights since each pillar is composed at most by two indicators.

[27] We looked at the lowest sum of squares among rankings.

B.2 Decomposition of Malmquist Productivity Index in "Efficiency" and "Technical" factors

One of the advantages of the Malmquist Index Approach is that it allows to discriminate changes in overall performance between "technical" and "efficiency" changes (Wang, 2015). Indeed, the index in Eq. 1 may be decomposed as follows:

$$MCM_t^{t+1} = \frac{CM^{t+1}(t+1)}{CM^t(t)} * \left[\frac{CM^t(t+1)}{CM^{t+1}(t+1)} * \frac{CM^t(t)}{CM^{t+1}(t)}\right]^{1/2}, \quad (2)$$

where

$$MCME_t^{t+1} = \frac{CM^{t+1}(t+1)}{CM^t(t)} \quad (3)$$

represents the efficiency change and

$$MCMT_t^{t+1} = \left[\frac{CM^t(t+1)}{CM^{t+1}(t+1)} * \frac{CM^t(t)}{CM^{t+1}(t)}\right]^{1/2} \quad (4)$$

denotes the technical change.

Being the expression in Eq. 3 the ratio between two consecutive years of the MCM, it directly measures the relative evolution between year *t* and *t+1* of any analyzed country performance in terms of CMs enhancements. At the same time, this ratio allows to describe whether a unit is improving (or worsening) its performance relatively to the best practice, when moving from any *t* to *t+1*. Specifically, when a unit shows a MCME > 1, the unit is "catching-up" the best practice at a certain rate. Otherwise, when MCME <1 the unit is moving away from the frontier. On the other hand, the expression in Eq. 4 denotes how the best practice changes from any year *t* to *t+1*. Indeed, when MCMT > 1 the best practice frontier is still improving, while it is worsening when MCMT < 1. In other terms, this ratio becomes a proxy for overall CMs "technical" development for the whole sample, since it describes how the best performance evolves in time.

As also mentioned in paragraph 4.2, in this work we call "idiosyncratic" the efficiency effect, and "adherence to global trends" the technical factor, in consistency with the research setting of this work, and units coincide with the 28 EU countries.

For the DEA/Malmquist estimation we used the STATA "malmq" package (Lee et al., 2011).

C Cluster Analysis

Cluster analysis is a set of algorithms useful to classify objects like geographical units, with the aim of reducing the dimensionality of the data-set, exploiting the similarities between different individuals (Kodinariya and Makwana, 2013). In this work, we aim at clustering the 28 European Union countries along the CMs metrics dimensions underlying the six key objectives emerging from the CMU Action Plan. The goal of our analysis is to use robust cluster techniques to identify groups of countries with similar behaviors within our 2007-2018 sample, with the aim of capturing country-level, and EU-level, patterns across clusters in time.

Several clustering methodologies have been adopted in literature, due to a certain degree of ambiguity in the definition of the cluster itself (Rokach and Maimon, 2005), and also based on the different features that each analysis should meet (Saxena et al., 2017). One of the most common classification deals with the nature of the algorithm used to separate the space, and conducts to two main typologies: first, hierarchical clustering according to which a hierarchy of partitions is constructed, characterized by a decreasing number of groups, typically depicted through a dendrogram, in which the grouping / division steps of the groups are represented (Cohen-Addad et al., 2019); second, partitional (or non-hierarchical) clustering, according to which the membership of a unit in a group is defined based on its distance from a representative point of the cluster, thus having prefixed the number of groups of the resulting partition (Reddy and Vinzamuri, 2018).

The choice of the set-up of our clustering analysis was guided by three main principles. First, we do not rely on an *ex ante* theory which may suggest an expected number of clusters. Therefore, we need a statistical tool that can clearly state which model specification can be considered the most informative, e.g. that can provide the Bayesian information criterion (BIC) for each model specification (Fraley and Raftery, 1998). Second, we look for a tool which may relax the assumption of spherical clustering groups, allowing for a wider flexibility of the covariance structure towards an ellipsoidal form (Garcia-Escudero et al., 2008a). Third, we want to potentially exclude from our clusterization a number of outliers, whose presence may otherwise bias the overall result (Rocke and Woodruff, 1996).

The first and the second requirement are typically addressed in literature through Gaussian mixture models. This model-based clustering approach, usually estimated using the Expectation Maximization (EM) algorithm, associates the observations to groups incorporating information about the covariance structure of the data, and providing, for any of the observations, the probability of belonging to each of possible clusters (Fraley and Raftery, 1998). This model also allows the restriction of the relative variability within clusters, through the introduction of constraints on the value of the ratio between the maximum and minimum eigenvalues of the covariance matrices (Garcia-Escudero et al., 2008a), thus fulfilling our second requirement. Nevertheless, mixture models do not take into consideration the presence of outliers, i.e. outlying observations are included within the clusters together with the others with a possible bias in the groups' definition and interpretation. One possibility to overcome this issue could be to *a priori* exclude the "usual suspects" from the computation, based on descriptive analyses and on scholar's sensitivity, though with the trivial drawbacks of arbitrariness of the choice and of reduced flexibility in the replication and update of the results. Hence, similarly to Cariboni et al. (2015), we opted for the adoption of *Robust Trimmed Clustering* model (Garcia-Escudero et al., 2008a; Garcia-Escudero et al., 2010). Indeed, building on mixture models, this approach automatically allows the possibility of making a share of outlying observations unassigned, *de facto* treating them as a different (excluded) cluster (Garcia-Escudero et al., 2008b). Hence, in this way, also our third requirement is fulfilled.

Our clusterization is based on the same eleven indicators underlying our CMs (composite indicator) metrics definition. To run the estimation of the clusters using robust techniques, we used the MATLAB function *tclust* developed within the FSDA project[28] (Riani et al., 2012). However, this algorithm cannot manage the estimation of more than two clusters given the number of analyzed countries[29], hence we first perform a PCA to get a reduced number of components to be included in the clusterization, obtaining 4 components.

The analysis of PCA results is shown in Tab. C.1.

[28] Jointly owned by the European Commission and the University of Parma (see, for reference, https://fsda.jrc.ec.europa.eu and http://rosa.unipr.it/fsda.html)

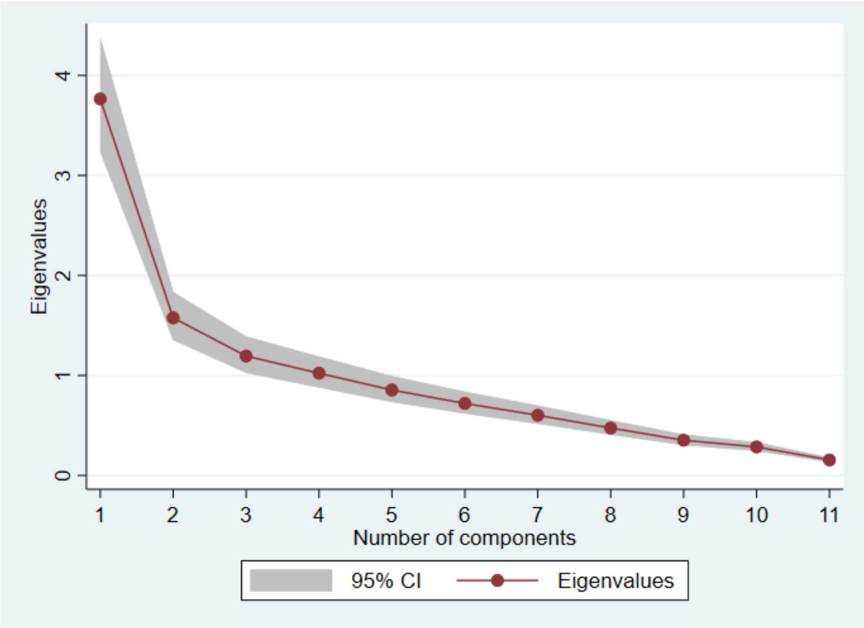
[29] TCLUST starts by estimating k ellipsoids based on $i + 1$ observations (where i is the number of indicators) chosen at random for each group. In our case, therefore, since the dataset has 28 countries and 11 indicators:
for k = 2: $(i + 1) * k = 12 * 2 = 24$
for k = 3: $(i + 1) * k = 12 * 3 = 36$
hence, TCLUST would fail for any k > 2.

Table C.1. Results of Principal Component Analysis conducted on the eleven indicators underlying the CMs metrics.

Component	Eigenvalue	Difference	Proportion	Cumulative
1	3.76519	2.18898	0.3423	0.3423
2	1.57621	0.382	0.1433	0.4856
3	1.19421	0.171796	0.1086	0.5941
4	1.02241	0.168239	0.0929	0.6871
5	0.854173	0.134396	0.0777	0.7647
6	0.719777	0.118188	0.0654	0.8302
7	0.601589	0.127309	0.0547	0.8849
8	0.47428	0.12223	0.0431	0.928
9	0.35205	0.0667047	0.032	0.96
10	0.285345	0.130576	0.0259	0.9859
11	0.154769	.	0.0141	1

The choice of the number of components is quite straightforward, based on both the Kaiser (1960) and Jolliffe (1973) criteria, since both the number of components with larger than one eigenvalues (also considering their confidence interval as shown in Fig. C.1) and explaining approximately 70% of total variance are equal to four.

Figure C.1. Scree plot of eigenvalues (and related confidence interval) after PCA.



Despite the five identified components are not always dominated by a single indicator, it is possible to identify a prevalence for each of them. Specifically, we first consider factors with loadings larger than |0.5|, and in case of absence, larger than |0.4|.

Table C.2. PCA: factor loadings (>|0.5| or, if none, >|0.4|).

Indicator	Comp. 1	Comp. 2	Comp. 3	Comp. 4
Adoption of Private Equity instrument	-	-	-	-
Adoption of Venture Capital instrument	-	-	-	-
Relevance of NFCs Debt	-	-	-	-
Relevance of NFCs Equity	42%	-	-	-
Adoption of alternative instruments	-	-	-	87%
Public-Private Partnerships (PPPs)	-	-	-	-
Retail Investments (assets)	42%	-	-	-
Institutional Investments (assets)	42%	-	-	-
Covered Bonds	-	-	75%	-
Deposit taker capital adequacy	-	50%	-	-
Cross-border PD and PE Investments	-	63%	-	-

Therefore, as shown in Table C.2, the first component is guided by the two measures of the pillar 4 indicators (retail and institutional investments) and one of pillar 2 (relevance of NFCs equity), the second by cross-border portfolio debt and equity investments and deposit taker capital adequacy, the third by the covered bonds indicator, and the fourth by the adoption of alternative instruments.

In order to estimate our clusters, our analysis relies on the specification of three parameters: the supposed number of clusters (k), the share of outliers to be excluded (α), and the restriction factor on the covariance matrix shape (*RF*). For our purposes, considering the dimension of the sample (i.e., 28 countries), we adopt two possible values for k (2 and 3), and three for α (3.6%, 7.1% and 10.7%, respectively accounting for the exclusion of one, two or three countries from the sample). Moreover, we consider six different levels of restriction factors, i.e. 1 (equal to the standard *k-means* clustering approach (Hartigan and Wong, 1979)), 5, 50, 100, 200 and 500. Hence, based on the three parameters setting, overall we estimate 36 different cluster analyses. We then pick the best specification based on the monitoring of BIC criterion. The estimation is performed, by year, for the whole period of the analysis (2007-2018). This would allow to investigate: first, whether the best specification is constant or not in time and, second, the CMs metrics performance of countries, through the dynamic analysis of their belonging to the same or different clusters.

Figure C.2. BIC scores for 36 different combinations of cluster numbers (K=2, 3), α (A=1, 2, 3) and restriction factors (RF= 1, 5, 50, 100, 200, 500) by year.

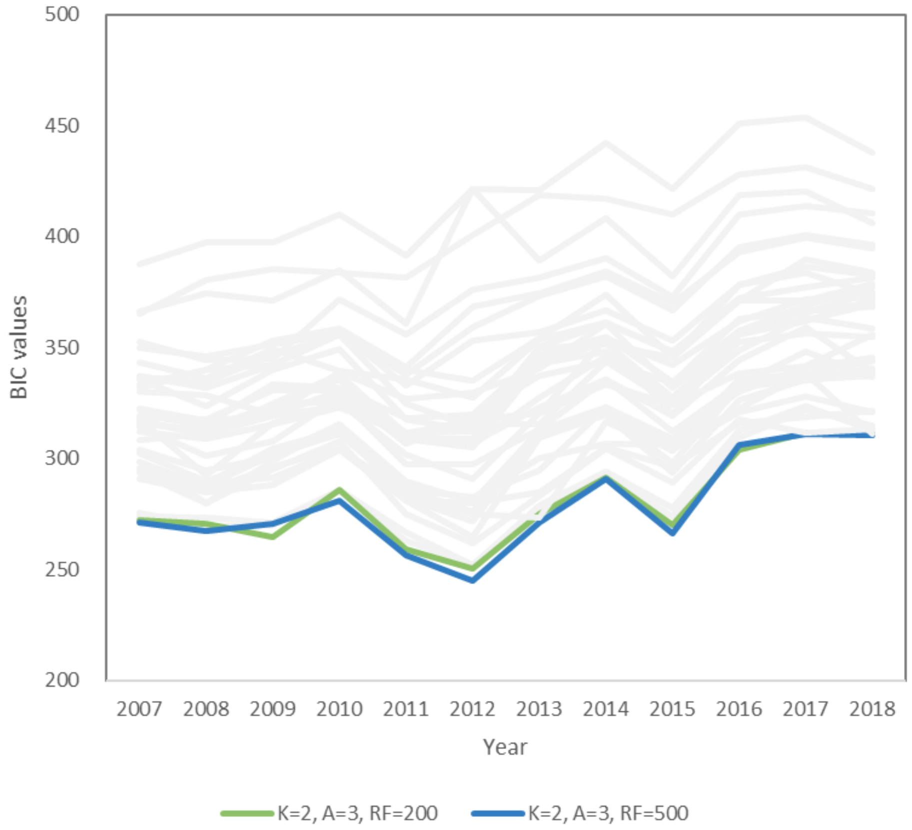
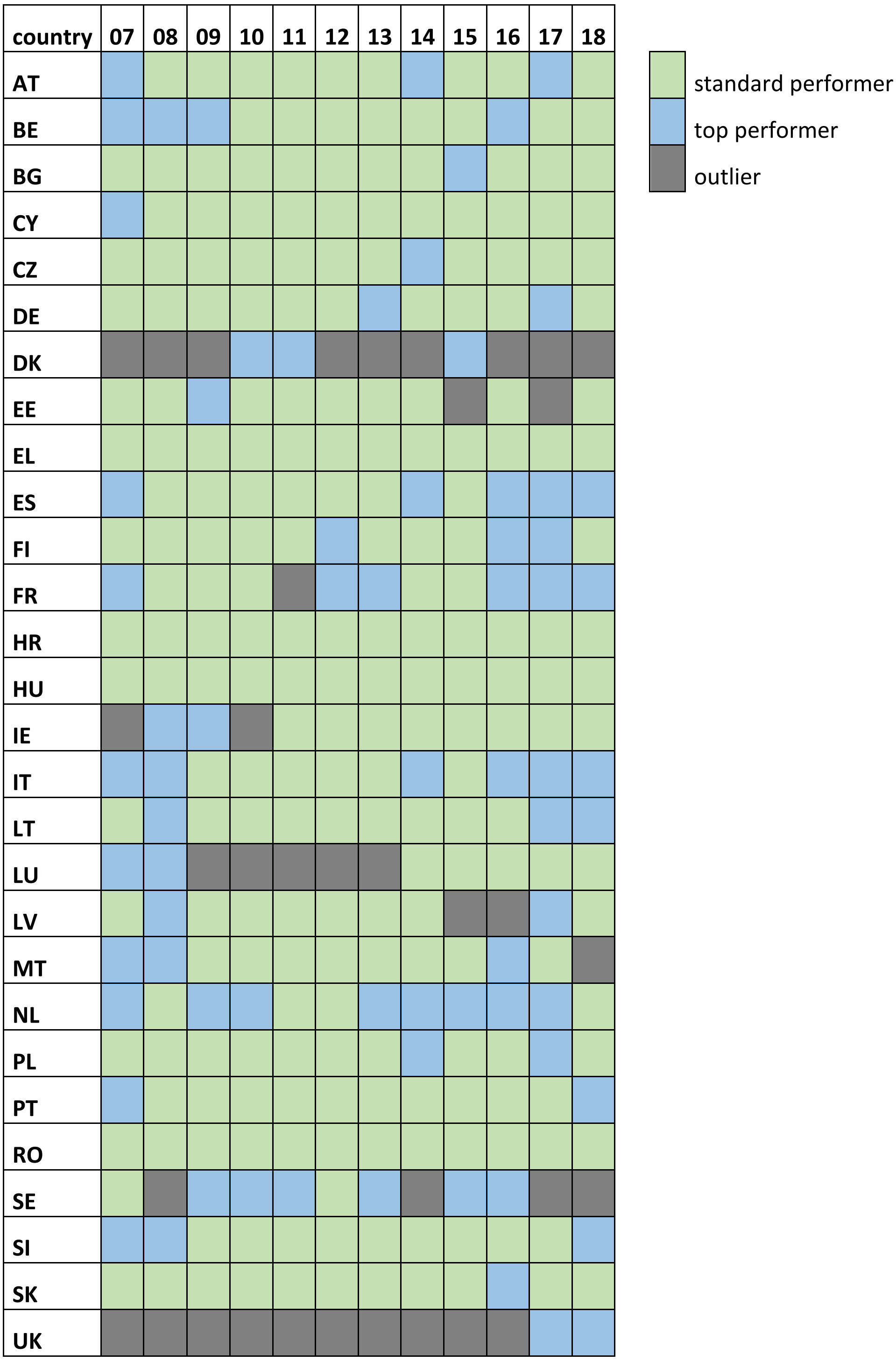


Figure C.2 depicts the 36 different estimated robust clustering trimmed models. The vertical axis describes the level of BIC for each model, while the horizontal axis provides its temporal evolution in the period 2007-2018. We represented with the same color (grey) those models that do not ever provide for any year of the sample the minimum BIC of the panel; otherwise, we adopt a different color for the other two models. From the analysis of the BICs, it emerges a substantial stability of the features of the most informative models. Indeed, for the whole sample, we should choose two clusters (*k*=2) and three outliers (*α*=10.7%*)*. Thus, trimming is relevant for the specification of the model. At the same time, the restriction factor should be higher than 100, and, in particular, equal to 500 for most of years (2007-2008, 2010-2015, 2017-2018), and 200 for 2009 and 2016, even though, even in this cases, the difference with the 100 and 500 versions is quite low. This means that an ellipsoidal form is more suitable for the grouping of our data, rather than the standard *k-means* spherical structure. Fig. C.3 shows the results of the robust clustering trimmed models.

Figure C.3. Evolution of clusters by country and years (2007-2018).



The trimming has the main effect of excluding from the clusterization countries with usual top performances, e.g. the UK (in ten out of twelve years), which typically shows the largely best performance in several dimensions among which retail and institutional investments, and Denmark (nine years), which always shows an outlying performance in the Covered Bond to Loans dimension, being the Danish covered bond market one of the oldest and most sophisticated in Europe (Dick-Nielsen et al., 2012). In addition, Luxembourg and Sweden show significant outlying performances (respectively, five and four years), even though not in most of the time-span. We can consider these four countries as "strict outliers" due to their mostly outlying behavior in the analyzed period. Our models also identify other seven countries showing isolated performance as outliers (i.e. one or two years in the whole sample). Despite their exclusion from the analysis may be relevant in the year by year interpretation of results, we consider them more as “episodic outliers”, rather than strict ones, with their overall performances being affected by single specific components.

Moving to the rest of the analysis, two clusters emerge: one composed of countries whose performance dominates within at least three out of the four components (“top performer”), and one with less positive performances (“standard performer”)[30]. We should highlight that the components that are determinant for the definition of the “top performer” group are not homogeneous and constant in time, and depend on the estimation of the yearly clustering.

This means that, for instance, in 2007 countries belonging to the “top performer” clusters show common good or very good performance according to components 1, 2 and 4 (and only fair performances in component 3), while countries assigned to the same cluster in 2018 are top performers in components 1, and 4. Nevertheless, some components seem to be more frequently associated to top performances. Indeed, component 1 and 4 are decisive for defining the definition of the “top performer” cluster in eleven out of twelve years. These two components sum up all the contents of pillar 4 (foster retail and institutional investments) and seem to be mostly associated with a good overall performance in CMs enhancement, together with pillars 2 and 3, which are partially covered by components 1 and 4. On the other side, a good performance in components 2 and 3, summing up all the contents of pillars 5 (leverage bank capacity to support the economy) and 6 (facilitate cross-border investment and promote financial stability), is respectively associated with membership to the “top performer” cluster only in 25% and 50% of cases.

As far as concern countries patterns, interestingly some of them show a quite stable performance in the whole period. Specifically, four countries are always clustered among the “standard performers” (Greece, Croatia, Hungary, and Romania), and other ten are predominantly part of them, since they belong to this group in minimum ~80% of the years (Bulgaria, Cyprus, Czech Republic, Germany, Ireland, Estonia, Latvia, Poland, Portugal and Slovakia). On the other side, besides from Denmark, Luxembourg, Sweden and the UK, three countries show a consolidate story of “top performers”: Netherlands (8 years), Italy and France (6 years). The remaining seven countries experienced sporadic changes in their pattern, with some of them belonging to the “top performers” cluster in the first (e.g., Belgium, Malta and Slovenia) or last years (e.g., Finland, France and Spain) of the sample.

To conclude, overall it is possible to notice that the ratio of countries moving to better performances is increasing after 2015. Indeed, in the post-CMU Action Plan period ~28% of countries on average belongs to the “top performer” cluster, while the same ratio for the period 2007-2014 accounted for only ~18%, accounting for a relevant (tough not statistically significant) ~9pp difference between the two periods.

[30] Since we have an even number of components, in case none of the clusters obtains its absolute majority (i.e. 3), the cluster with the higher median of centroids locations location value is taken as “top performer” and the other as “standard performer”.

KJ-NA-29891-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/92215

ISBN 978-92-76-12268-5