EU recovery funds shouldn't be medals for battles of the past

Carlo Altomonte, Andrea Coali, Gianmarco I.P. Ottaviano

Technical Annex

1. GVC-related Trade and the External Risk Disruption Indicator (EDRI)

To build our indicator of External Risk Disruption, we consider two measures. First, we look at the domestic value added (DVA) content of total exports in each country-sector pair. We take the total figure and all the DVAs from bilateral flows and we account for the portion of DVA generated either within the EU (including UK and Norway as part of the Single Market), or outside it. Then, we look at the portion of extra-EU import value added (VA) in each country-sector pair. We retrieve the total import-related VA and then look at the VA originated in each bilateral relationship with partner countries. We then aggregate as above in two groups depending on the imports' origin. Thus, we know how much each country-sector pair is "dependent" on Extra-EU flows, both in its imports and exports. We then take a simple average between the two shares and build a regional-level measure by using labor share for the same sector at the regional level as weights. In Section 1.1 we also report an index of overall global value chain exposure of EU regions, to allow for comparisons.

As for data on Gross Exports, GVC-related trade and Imports at the country-sector level, we rely on the WIOD database as in Timmer et al., (2015).[1] The WIOD database contains data back to 2014 for 56 sectors. Out of all these sectors, we consider only those covered by the Eurostat SBS categorization.[2]

Throughout the analysis, we have applied within-region weights, as our aim is to compare the exposure of regions to the shock, regardless of their weight at the EU or country level. The main weights we have employed account for the within-region composition of employment, based on Eurostat SBS data. Using these weights allows us to compare regions without considering

[1] Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R. and de Vries, G. J. (2015), "An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production", *Review of International Economics*., 23: 575–605

[2] Particularly, in our analyses we exclude from the regional-level computations the following NACE codes that are not covered in SBS: A, K, O, P, Q, R, S, T. See infra for an in-depth discussion of the weighting scheme.

their relative importance in the EU framework, thus "normalizing" only with respect to the internal industrial structure, i.e. capturing their 'pure' exposure to the shocks. Using regional weights normalized for European or country-level measures, as it is more common in the trade literature, would give more weight to larger regions. For instance, regions like Lombardy, Catalonia, Madrid, Bavaria, would stand out in the analysis, not necessarily because they have been more exposed to the shock, but because they carry a higher economic weight. Our exercise instead focused on the different industrial structures within each region and points out the associated risks allowing for a comparison across regions. A detailed discussion of our weighting scheme, together with alternatives, is presented in Section 3 of the Technical Annex.

1.1 GVC-related trade measure

We first compute the ratio between GVC-related export value and the total Gross Export value at the country *i* and sector *j* pair:

$$\frac{GVC_{ij}}{GEXP_{ij}}$$

We use the already computed data on GVC-related values in the WIOD tables, exploiting the *icio* STATA routine. The relevant syntax is:

```
icio, exporter(country, all)

icio, exporter(country, all) importer (partner)
```

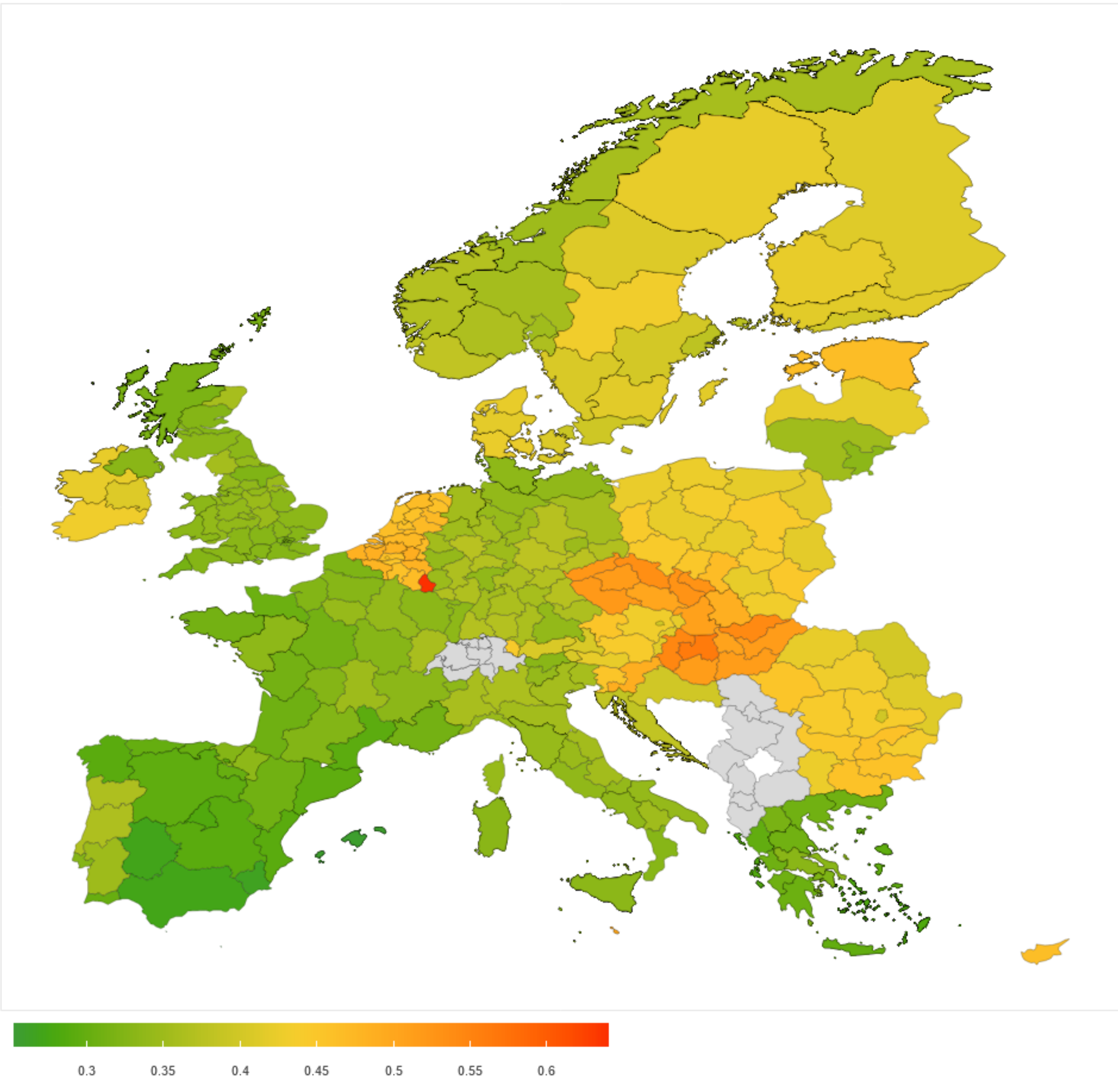
We store all the information in specific excel files, and then use the relevant GVC data retrieved from the decomposition. We then nail down the regional level by creating a within-region weighted average using the regional-employment weights described above. The regional GVC-related trade measure is thus:

$$GVC\_share_{ir} = \sum_j \frac{L_{irj}}{L_{ir}} * (\frac{GVC_{ij}}{GEXP_{ij}})$$

This implies that we have quite some homogeneity by country, since the assumption we make is that, for any given sector, the computed share in a given country will be the same in each of its regions (e.g. if for the automotive sector in Italy the share of GVC-related trade is 40%, the assumption is that such share for the automotive sector will be the same in both Lombardy and Sicily).

Figure A1 plots the NUTS2-level map of GVC-related share of exports. It shows how the most GVC-exposed regions are those in Eastern Europe countries, probably for their strong connection with Western Europe, and those in countries active in more financially related sectors (Netherlands, Luxembourg and Ireland). The least exposed regions seem to be in Southern Spain and Greece.

Figure A1 - Share of GVC-related trade at NUTS-2 level (weighted by labor share)



1.2 Building the EDR indicator (EDRI)

The first component of the EDRI is the Domestic Value Added (DVA) content of total exports in each country-sector pair. To build it, we follow the prescriptions of Belotti, Borin and Mancini

(2020)[3] and Borin and Mancini (2019), by taking both the total DVA figure and all the exported DVAs from bilateral flows. We then aggregate the bilateral data in two groups: belonging to EU29 (EU27 + GBR + NOR) and extra-EU29 countries (including the fictional *Rest of the World* country in WIOD).

Thus, we can account for the portion of DVA absorbed either within Europe or outside Europe. The country-sector level measure is thus:

$$\frac{DVA_{ij}^{extraEU}}{DVA_{ij}}$$

since we are interested in the exposure to extra-EU flows.

The relevant STATA code used to retrieve DVA information is the same as the one indicated in section 1.1, using DVA-related data rather than GVC-related data.

The second component we consider is the portion of Extra-EU Import Value Added (VA) in each country-sector pair, i.e. the foreign content of exports. We retrieve the total import-related VA from WIOD and then look at the VA originated in each of the WIOD countries, again following Belotti, Borin and Mancini (2020). We then aggregate as above in two groups: EU29 and extra-EU29. The country-sector level measure is thus:

$$\frac{VA\_IMP_{ij}^{extraEU}}{VA\_IMP_{ij}}$$

The relevant STATA code is:

```
icio, importer(country, sector) origin(all) output(va)
```

To build the final "External Disruption Risk Indicator "(EDRI), we take a simple average between the two shares explained above and build a regional-level measure by using labor share for the same sector at the regional level as weights.

Summing the numbers over sectors and by region results in a within-region weighted average by sector of the extra-EU29 trade exposure at the regional level.

[3] Belotti, F., Borin, A. and Mancini, M. (2020). icio: Economic Analysis with Inter-Country Input-Output Tables in Stata. *World Bank: World Development Report 2020 Policy Research Paper*.
Borin, A., and Mancini, M. (2019). Measuring What Matters in Global Value Chains and Value-Added Trade. *Policy Research Working Paper; No. 8804. World Bank.*

$$regio\_exposure_{ir} = \sum_j \frac{L_{irj}}{L_{ir}} * averageExp_{ij}$$

Where

$$averageExp_{ij} = mean(\frac{DVA_{ij}^{extraEU}}{DVA_{ij}}; \frac{VA_{IMP}{}_{ij}^{extraEU}}{VA_{IMP}{}_{ij}})$$

The EDRI has an average value of 0.44, with a standard deviation of 0.071.

2. The Internal Disruption Risk Indicator (IDRI)

To compute the "Internal Disruption Risk Indicator" (IDRI), we rely on the sector-specific risk measure coming from INAIL, the Italian agency for job safety. The IDRI captures the extent to which the economic structure of each region (or country) is going to be affected by the potential disruptions arising from obstacles to the production process

INAIL has classified each narrow sector (ISIC/NACE 2-digit level) in terms of three characteristics:

- *Contact*: the probability of social contact while on the work premises (the index takes value from 0, low probability, e.g. agriculture to 4, high-probability, e.g. nurses);
- *Proximity*: the intrinsic characteristics of the workflow that do not allow for sufficient social-distancing (the index takes value from 0, working alone, to 4, work in strict contact with other people);
- *Aggregation:* the level of contact with other subjects other than the firm's workers (e.g. restaurants, education, also taking value from 1 to 4).

Publicly available data include a combined index that considers the two measures of contact and proximity, plus the aggregation one[4]. In line with the prescriptions of the technical document, we use the aggregation measure to *weight* the first risk measure. Each social aggregation risk class correspond to a weight ranging from 1 to 1.5.

The final risk measure thus considers these two components and is computed at the NACE 2-digit level as:

[4] INAIL Dossier Covid 19 (Italian language. Available at: https://www.inail.it/cs/internet/comunicazione/covid-19-prodotti-informativi.html)

$$risk_{weighted} = risk_{class} * weight_{aggregation}$$

Since some sectors in WIOD are the result of an aggregation of multiple 2-digit sectors, we took the average *risk_weighted* measure for aggregated WIOD sectors.

To compute the NUTS-2 level risk, as for the EDRI we have computed a weighted average by sectors (*j*) of the risk at the regional level *r*:

$$risk_{ir} = \sum_{j} \frac{L_{irj}}{L_{ir}} * riskweighted_j$$

The resulting indicator has finally be rescaled according to the EU-level distribution, resulting in an index ranging from 0 to 1, with an average of 0.34 and a standard deviation of 0.16.

We run a comparative check using the aggregation measure only as a proxy for social risk, obtaining similar results (the two indicators are correlated at 0.9). We also run a second robustness check where we take the maximum value within aggregated WIOD sectors rather than averages: results are qualitatively similar, and the two measures are highly correlated (r = 0.96).

For robustness, we add here the EU-regions map showing the distribution of the aggregation component only, always using SBS weights (Figure A2).

3. Weighting Scheme

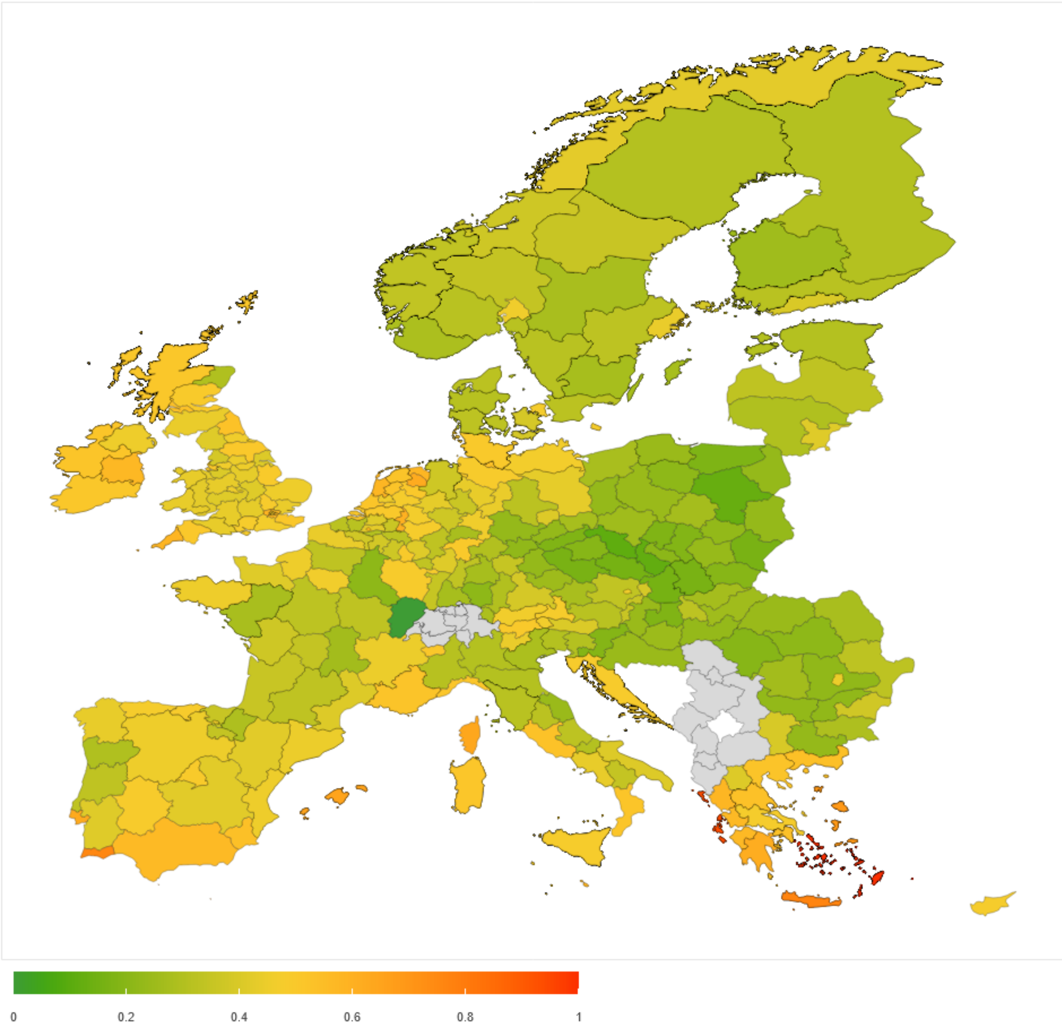
The within-region weights reproduce the industrial structure of each NUTS-2 region by exploiting the regional-level Eurostat Structural Business Statistics database (database code: *sbs_r_nuts06_r2*). We thus consider only the NACE rev.2 sectors covered by SBS and use the relevant employment figures to re-create the industrial composition.

We create a within-region weight as:

$$shareL_{irj} = \frac{L_{irj}}{L_{ir}}$$

Where *i* indicates the country, *r* the NUTS-2 region and *j* the industrial sector as in SBS. *L* stands for the number of persons employed from Eurostat SBS. We use data from 2017, i.e. the last available year in the database.

Figure A2 - Aggregation Risk (re-scaled), NUTS-2 level



Thus, we consider the share of people employed in sector *j* in region *r* of country *i* over the total number of people employed in the same region *r* in the covered SBS sectors.

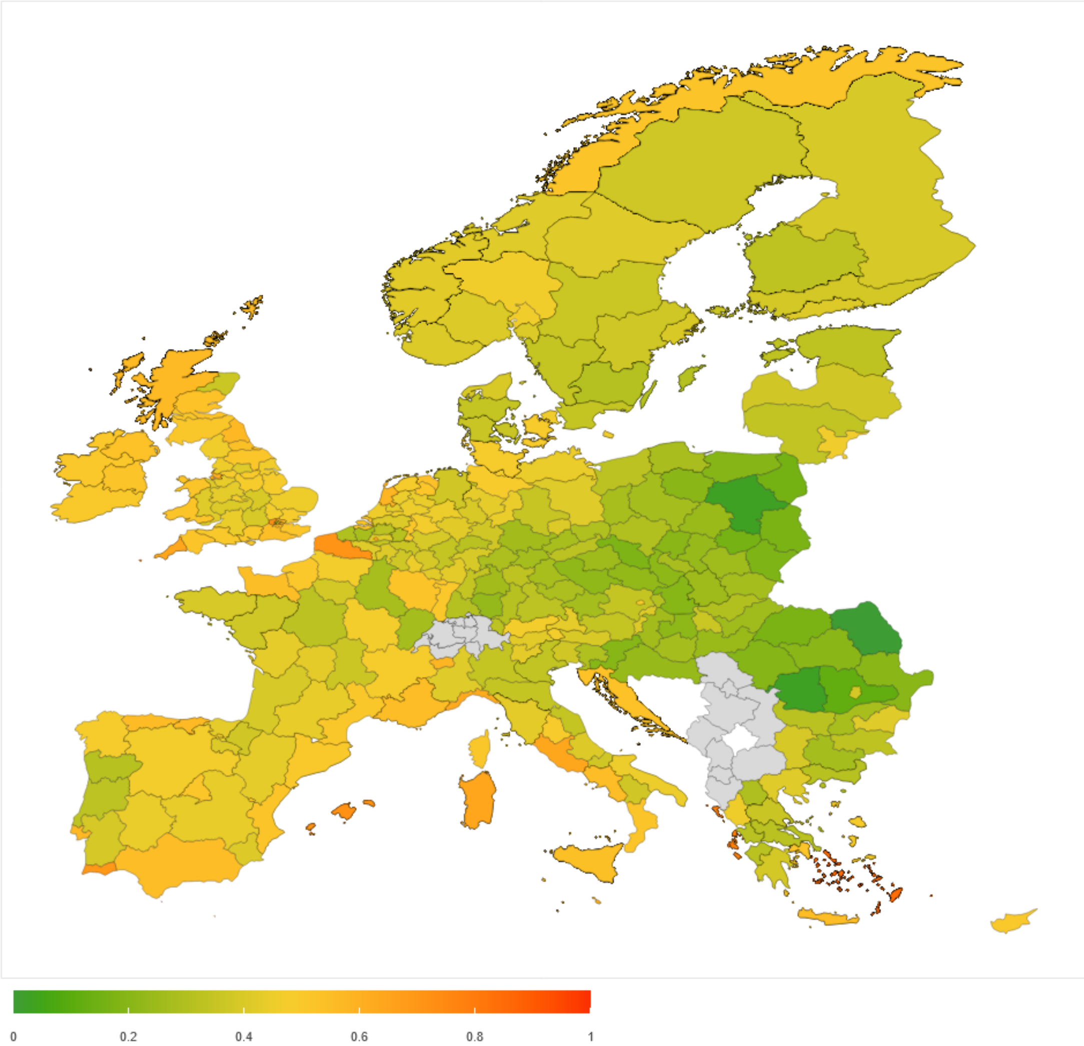
This weighting scheme allows us to understand which sectors are the most important (in employment terms) within a region. The rationale for using within-region weights is that we want to compare regions regardless of their weight at the EU level. These numbers thus accounts for the within-region composition of the employment in SBS sectors. Using EU or country-level weights, would have given higher weights to regions where labor is concentrated (e.g. Lombardy, Cataluna, Madrid, Bayern etc…) and would account for a different type of comparison.

We also tried to merge the SBS data with relevant information from Business Demography (BD) for retrieving data on sectors not covered by SBS. Results are qualitatively similar for countries where data is available. However, we decided to keep weights based on SBS since BD (database: *bd_enace2_r3*) does not cover all EU countries and information is less complete (for instance, data for many Netherlands regions is not available in recent years and German data is not available at all). Thus, to allow the most robust and cleaner cross-regional comparison, we decided to rely on the most complete weights as possible.

Adopting this approach however limits our analysis in the following way. SBS sectors do not cover sectors such as Agriculture (NACE code A), Public Administration (O), Education (P), Human Health and Social Work Activities (Q), Arts and Recreation (R) and a bunch of other aggregated sectors. This has surely an effect on the estimates we produce since we are not considering the full spectrum of employment. Moreover, we might over/under weight our results depending on the full composition of labor in each region. At the same time, however, data for these sectors are often offered at a too aggregated level (e.g. BD data offers data aggregated for sectors R plus S, and P plus Q omitting public administration) compared to the higher detail of other sectors. and thus even using these broader data might result in distortions. Public administrative jobs are also difficult to classify and might group employees operating in very different segments of the economy at the same time.

In any case we have run a robustness exercise where we exploit employment data from the Regional Eurostat data on Employment (database: *lfst_r_lfe2en2*) to recover some of the sectors not covered by SBS. This database contains data on employees' numbers for aggregated NACE sectors (for instance, it provides aggregated employment figures for the sectors R-U). The larger aggregation of sectors in the weighting measure implies that assumptions have to be made when aggregating risk measures that are instead captured at a higher level of detail. Figure A3 shows the re-scaled IDR indicator using original weights complemented with this new data for the SBS missing sectors.
The regional ranking is very much like the one we obtained by using the SBS weights and it is clear how most affected regions are still those that we identified in the main section of this research.

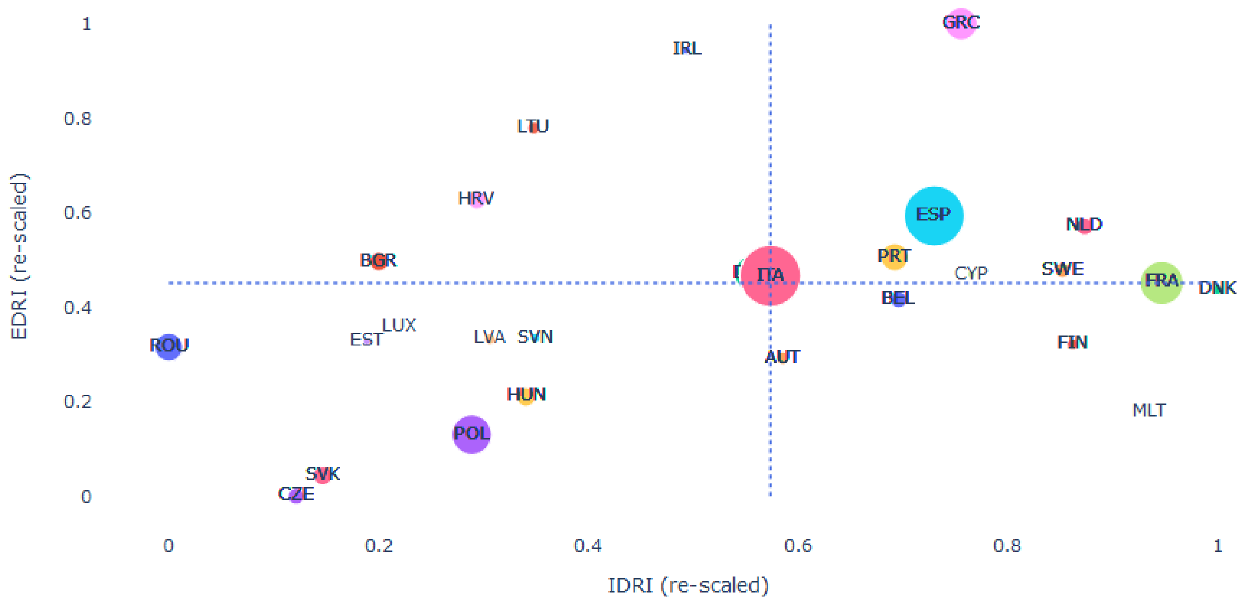
Figure A3 - IDRI with alternative weights



We have thus chosen to use SBS data only, as the latter require less assumptions to be made when sectoral aggregation has to be done, and does not merge data coming from different sources and of different types (e.g. persons employed and employees). Results seem to be robust in any case.

We also tried using a different weighting scheme at the country level. We rely on Value Added data from the WIOD database, thus exploiting information for all the available sectors, and weight each sector by its contribution to the overall country Value Added. Figure A4 below shows a replication of Figure 4 using such weighting scheme (bubble size is still the allocation share based on EC estimates).

Figure A4 – Country Combined Risk (VA weights)



The picture is different from Figure 4, albeit not strikingly, as countries move along the axes but appear to be in the same quadrants. The higher weight of the risk measure for northern EU countries is due to their higher sectoral share of health and social care sectors, that are classified as high risk in our dataset. Nevertheless, this exercise shows how it is crucial to choose the most appropriate weighting scheme if these indicators are used in any allocation exercise. Moreover, it is important to stress how aggregating at the country level ignores information coming from the regional heterogeneity, which is something that should be taken into account in such a context.

To ease replications and to let researchers propose and try other weighting schemes, we provide all the country-level data we used in the *github* repository linked in the last section. Researchers are thus able to nail down the regional dimension using the weights they consider to be better suited for this exercise, also exploiting different aggregation levels or more detailed data if available.

4. *COVID-related data*

We retrieved data on COVID contagions up to 29th May 2020 at the NUTS-2 level exploiting different sources, such as websites and databases published by national health authorities. We were able to retrieve regional data for 225 out of 296 European regions. While in the main article we reported measures for the relative number of contagions, i.e. the proportion of COVID cases to the total regional population, we show in Figure A5 the regional map with absolute levels of COVID contagions.

Figure A5 - Absolute Number of COVID cases (re-scaled, up to 29/05)

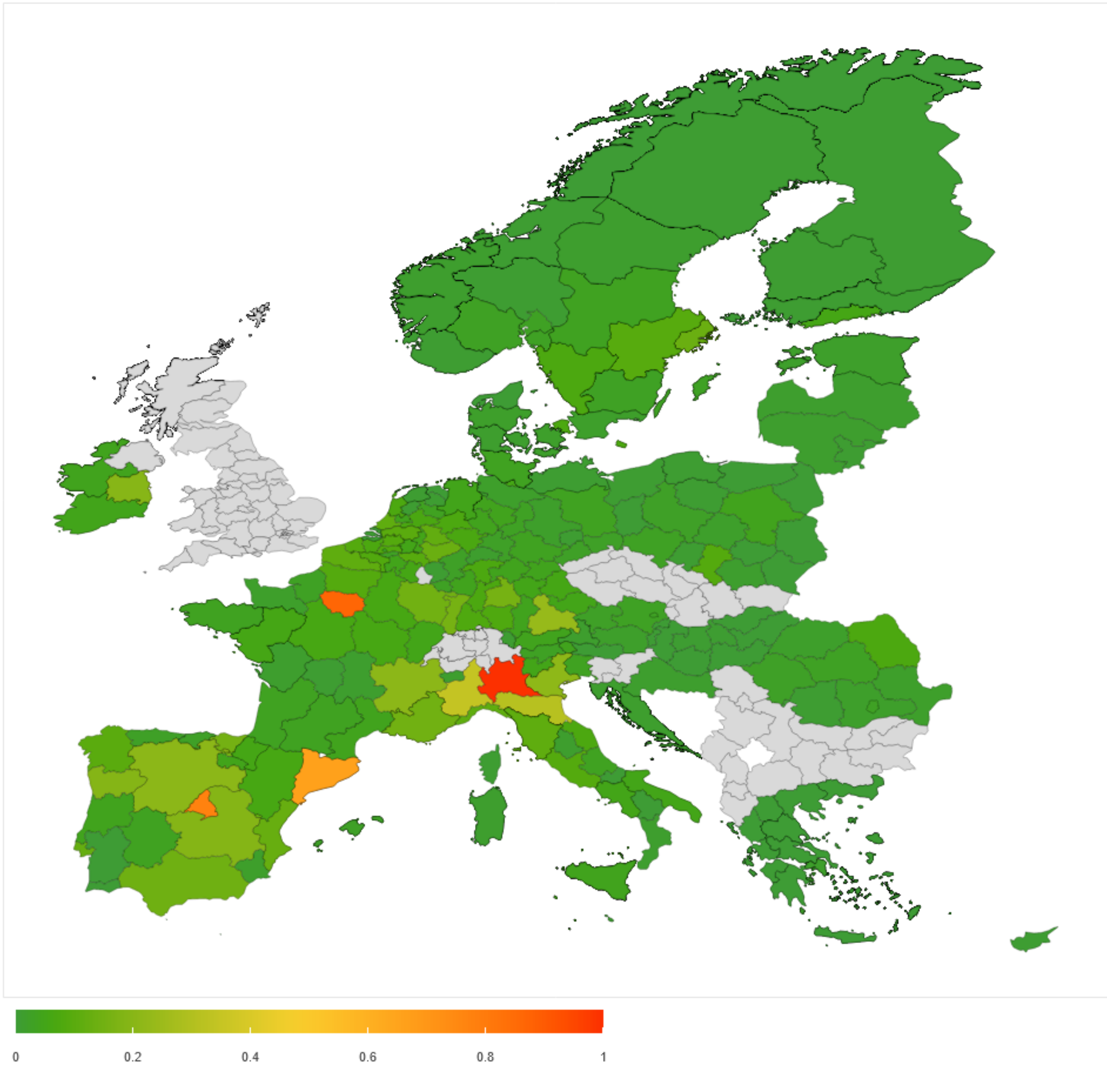


Figure A5 - Absolute Number of COVID cases (re-scaled, up to 29/05/2020)

5. Additional Information

Throughout the article, we considered as EU countries the EU27 together with Norway and United Kingdom (i.e. EU29). Thus, extra-EU trade does not contain trade values from these two countries. Similarly, median levels showed in graphs and re-scaling of indicators are based on such EU29 distribution.

However, re-scaled numbers are not affected by the inclusion or removal of such countries, given that the standard 0-1 rescaling formula for each variable is:

$$x_{rescaled} = (x - \min_x)/(\max_x - \min_x)$$

6. Data Availability

To allow for replications or leverage on the indicators we have built, all the data used for our computations is freely available on *github*.

In the same repository, it is also possible to find an interactive version of the bubble chart in Figure 3 with all the EU regions for which we were able to retrieve contagions' data.

To access the data please visit: [*https://github.com/andreacoali/COVID_back_to_future*](https://github.com/andreacoali/COVID_back_to_future)