

## **CODEBOOK**

### **EDB extension variables**

#### **COUNTRIES**

Countries affected by the measure. Notification from the EU list all its member countries, except for measures notified after 2019, which exclude the United Kingdom. This column is useful for merging the dataset with country-level data.

#### **START**

Extracted year for the start of implementation. (Refer to appendix paper for methodology)

#### **END**

Extracted year for the end of implementation. (Refer to appendix paper for methodology)

#### **ISIC**

2-digits ISIC codes roughly corresponding to the sectors listed in the column "Harmonized types of sectors subject to the measure"

#### **MEASURE\_SCORE**

Proxy for the measure strength based on variables in the EDB. Ranges from 0 to 9. Obtained by multiplying BREADTH x DEPTH. Refer to appendix C of the paper for calculation details and description.

#### **BREADTH**

Proxy for the breadth of the measure. Ranges between 0 and 3. Obtained by weighted sum of the sub-components BREADTH\_economic\_sector, BREADTH\_env\_objectives and BREADTH\_env\_keywords. Refer to appendix C of the paper for calculation details and description.

#### **DEPTH**

Proxy for the breadth of the measure. Ranges between 0 and 3. Obtained by summing the sub-components DEPTH\_measures\_variety, DEPTH\_wording and DEPTH\_measures\_instruments. Refer to appendix C of the paper for calculation details and description.

#### **BREADTH\_economic\_sector**

Sub-component of the BREADTH score based on the harmonised sectors affected by the measure.

#### **BREADTH\_env\_objectives**

Sub-component of the BREADTH score based on the number of environmental objectives of the measure.

#### **BREADTH\_env\_keywords**

Sub-component of the BREADTH score based on the number of environmental keywords associated with the measure.

#### **DEPTH\_wording**

Sub-component of the DEPTH score based on the connotation of verbs used in the measure description.

#### **DEPTH\_measures\_variety**

Sub-component of the DEPTH score based on the number of policy instruments used in the measure.

#### **DEPTH\_measures\_instruments**

Sub-component of the DEPTH score based on the type of policy instruments used in the measure.

#### **Tentative HS chapters match**

HS 2-digits codes associated with the measure. For the methodology refer to Appendix B in the paper.

#### **Absolute link strength (L\_tilde)**

A measure of the "confidence" of each matched HS 2-digits code. The higher, the better. In general, confidence is higher when the textual description of the measure are longer, and the matched HS

code overlaps with the harmonised sector in the EDB. Refer to Appendix B in the paper for description of methodology

**Relative link strength ( $L_{\text{bar}}$ )**

Relative "confidence" associated with each HS 2-digits code. This is a measure of the relative linkage with HS code. Obtained by dividing the Absolute link strength ( $L_{\text{tilde}}$ ) of each code, over the total of all codes associated with each measure.

**Reported HS codes**

This column reports all the codes in the column "ICS - HS code", which are unambiguously identified as HS codes.

**Reported ICS codes**

This column reports all the codes in the column "ICS - HS code", which are unambiguously identified as ICS codes.

**Reported ambiguous codes**

This column reports all the codes in the column "ICS - HS code" which are ambiguous (i.e. they could be either HS or ICS codes). When the measure reports only a single code, it is almost always ambiguous because it might have lost a leading or trailing zero. Refer to Appendix B in the paper for a methodology description.

**Similar measures**

List of measures which are similar (measure numbers from column "Nr"). Similarity is computed by comparing the words in the two descriptions.

**Similarity index**

Similarity indices corresponding to the measures listed in "similar measures". The similarity index ranges between 0 and 1. 0 indicates that the two measure description share no words in common. Whereas 1 indicates that the two measures use exactly the same set of words. The calculation is based on the *Jaccard index* of the sets of words from columns "Measure description", "Coverage of measure" and "environment-related objective".