Package: harmonizer (via r-universe)

August 31, 2024

Type Package

Date 2023-01-25

Title Harmonizing CN8 and PC8 Product Codes

Version 0.3.2

Author Christoph Baumgartner [cre, aut]
(<https://orcid.org/0000-0002-2681-071X>), Stjepan Srhoj [aut]
(<https://orcid.org/0000-0002-7105-8088>), Janette Walde [aut]
(<https://orcid.org/0000-0002-1241-5292>)

Maintainer Christoph Baumgartner <Christoph.Baumgartner@uibk.ac.at>

Description Several functions are provided to harmonize CN8 (Combined Nomenclature 8 digits) and PC8 (Production Communautaire 8 digits) product codes over time and the classification systems HS6 and BEC. Harmonization of CN8 codes are possible by default from 1995 to 2022 and of PC8 from 2001 to 2021, respectively.

License GPL (>= 3)

Imports stats, utils, methods

Encoding UTF-8

LazyLoad yes

Depends R (>= 3.5.0)

RoxygenNote 7.2.3

Suggests rmarkdown, knitr

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2023-01-25 15:50:05 UTC

Contents

`cn8_to_bec` *Concordance list between CN8 and BEC*

Description

Provides a dataframe which contains all CN8 product codes and related BEC codes in a given time period.

Usage

```
cn8_to_bec(b, e, historymatrix = NULL, progress = TRUE)
```

Arguments

`b`	first year of interest
`e`	last year of interest
`historymatrix`	History matrix of CN8 product codes. Provided by history_matrix_cn8().
`progress`	logical, determines whether progress is printed in console or not.

Value

A data frame that contains all CN8 product codes and related BEC and HS6 codes in a given time period. The following table offers an overview of all provided variables.

Variable	Explanation
`CN8`	character; a specific CN8 code
`HS6`	character; provides the HS6 classification of the CN8plus code
`BEC`	character; provides the BEC classification on a high aggregation level (1 digit)
`BEC_agr`	character; provides the BEC classification on a lower aggregation level (up to 3 digits)

Examples

```
cn8_bec <- cn8_to_bec(b = 2008, e = 2010)
```

get_data_directory *Data path for custom data*

Description

Provides the directory where custom data must be stored and the used data (e.g., concordance lists, list of codes) can be edited.

Usage

```
get_data_directory(path = TRUE, open_explorer = FALSE,
                   show_data = NULL)
```

Arguments

path	logical, determines whether the path is printed in the console
open_explorer	logical, determines whether an explorer is opened in addition. Only executable if the directory path does not contain any blanks.
show_data	character string, which must take one of the following values: "CN8", "HS6", "PC8" or "HS6toBEC". All available data in in the given directory is printed in the console. Only executable if the directory path does not contain any blanks.

Value

Returns the path (character), of the directory where custom data must be stored and the used data (e.g., concordance lists, list of codes) can be edited.

Examples

```
get_data_directory()

get_data_directory(path = FALSE, show_data = "CN8")
```

harmonize_cn8 *Harmonization of CN8 product codes*

Description

Provides a dataframe which contains all CN8 product codes and their history in the demanded time period, as well as harmonized CN8plus code, harmonized HS6plus code and BEC classification.

Usage

```
harmonize_cn8(b, e, historymatrix, harmonize.to = "e",
              HS6breaks = c(1992, 1996, 2002, 2007, 2012, 2017),
              progress = TRUE)
```

Arguments

b	first year of interest
e	last year of interest
historymatrix	History matrix of CN8 product codes. Provided by history_matrix_cn8(). By default *NULL*; the function computes the needed harmonized data.
harmonize.to	Defines which year for harmonization is used. It may take the following values: • "e", harmonizes product codes towards year e • "b", harmonizes product codes towards year b
HS6breaks	Vector of years, where HS6 codes were changed. Do not edit, unless additional break is needed.
progress	logical, determines whether progress is printed in console or not.

Value

A data frame that contains all CN8 product codes and their history, harmonized CN8plus codes, harmonized HS6plus codes, and BEC classification. The 'plus-codes' are the main outcome of the function. They provide harmonized information of the product codes, i.e. comparable codes. Every harmonization refers to the last year of interest. The following table offers an overview of all provided variables.

Variable	Explanation
CN8_xxxx	character; a specific CN8 code in a given year
CN8plus	character; the harmonization code for CN8, which refers to the last/first year of the time period
HS6plus	character; the harmonization code of HS6, which refers to the last/first year of the time period
BEC	character; provides the BEC classification at a high aggregation level (1 digit)
BEC_agr	character; provides the BEC classification at a lower aggregation level (up to 3 digits)
SNA	character; provides information if the code is classified as consumption, capital or intermediate good in SNA
flag	numeric; integer from 0 to 3; 1 indicates that this code remained the same in notation over the whole time period
flagyear	numeric; indicates the first year in which the flag was set

Examples

```
myharmonization <- harmonize_cn8(b = 2008, e = 2010)

mydata <- history_matrix_cn8(b = 2016, e = 2018)
myharmonization <- harmonize_cn8(b = 2016, e = 2018,
                                 historymatrix = mydata)
```

---

harmonize_pc8 *Harmonization of PC8 product codes*

Description

Provides a dataframe which contains all PC8 product codes and their history in the demanded time period, as well as harmonized PC8plus code, harmonized HS6plus code and BEC classification.

Usage

```
harmonize_pc8(b, e, historymatrix = NULL, harmonize.to = "e",
              HS6breaks = c(1992, 1996, 2002, 2007, 2012, 2017),
              progress = TRUE)
```

Arguments

b	first year of interest
e	last year of interest
historymatrix	History matrix of PC8 product codes. Provided by history_matrix_pc8(). By default *NULL*; the function computes the needed harmonized data.
harmonize.to	Defines which year for harmonization is used. It may take the following values: • "e", harmonizes product codes towards year e • "b", harmonizes product codes towards year b
HS6breaks	Vector of years, where HS6 codes where changed.
progress	logical, determines whether progress is printed in console or not.

Value

A data frame that contains all PC8 product codes and their history, harmonized PC8plus codes, harmonized HS6plus codes, and BEC classification. The 'plus-codes' are the main outcome of the function. They provide harmonized information of the product codes, i.e. comparable codes. Every harmonization refers to the last year of interest. The following table offers an overview of all provided variables.

Variable	Explanation
PC8_xxxx	character; a specific PC8 code in a given year
PC8plus	character; the harmonization code for PC8, which refers to the last/first year of the time period
HS6plus	character; the harmonization code of HS6, which refers to the last/first year of the time period
BEC	character; provides the BEC classification at a high aggregation level (1 digit)
BEC_agr	character; provides the BEC classification at a lower aggregation level (up to 3 digits)
SNA	character; provides information if the code is classified as consumption, capital or intermediate good in BEC
flag	numeric; integer from 0 to 3; 1 indicates that this code remained the same in notation over the whole time period
flagyear	numeric; indicates the first year in which the flag was set

Examples

```
myharmonization <- harmonize_pc8(b = 2009, e = 2011)

mydata <- history_matrix_pc8(b = 2015, e = 2017)
myharmonization <- harmonize_pc8(b = 2015, e = 2017,
                                 historymatrix = mydata)
```

`history_matrix_cn8` *History matrix of CN8 product codes*

Description

Provides a dataframe which contains all CN8 product codes and their history in a given time period.

Usage

```
history_matrix_cn8(b, e, c1 = 1988, c2 = 2022,
                   progress = TRUE)
```

Arguments

`b`	first year of interest
`e`	last year of interest
`c1`	first year of the concordance list
`c2`	last year of the concordance list
`progress`	logical, determines whether progress is printed in console or not.

Value

A data frame that contains all CN8 product codes and their history over time for the demanded time period. This dataset is the basis for the main function `harmonize_cn8()` and can be obtained therewith as well. The following table offers an overview of all provided variables.

Variable	Explanation
`CN8_xxxx`	character; a specific CN8 code in a given year
`flag`	numeric; integer from 0 to 3; 1 indicates that this code remained the same in notation over the whole time period
`flagyear`	numeric; indicates the first year in which the flag was set

Examples

```
history <- history_matrix_cn8(b = 2008, e = 2010)
```

history_matrix_pc8 *History matrix of PC8 product codes*

Description

Provides a dataframe which contains all PC8 product codes and their history in a given time period.

Usage

```
history_matrix_pc8(b, e, progress = TRUE)
```

Arguments

b	first year of interest
e	last year of interest
progress	logical, determines whether progress is printed in console or not.

Value

A data frame that contains all PC8 product codes and their history over time for the demanded time period. This dataset is the basis for the main function `harmonize_pc8()` and can be obtained therewith as well. The following table offers an overview of all provided variables.

Variable	Explanation
PC8_xxxx	character; a specific PC8 code in a given year
flag	numeric; integer from 0 to 3; 1 indicates that this code remained the same in notation over the whole time period
flagyear	numeric; indicates the first year in which the flag was set

Examples

```
history <-  history_matrix_pc8(b = 2008, e = 2010)
```

pc8_to_bec *concordance list between PC8 and BEC*

Description

Provides a dataframe which contains all PC8 product codes and related BEC codes in the demanded time period.

Usage

```
pc8_to_bec(b, e, historymatrix = NULL, progress = TRUE)
```

Arguments

b	first year of interest
e	last year of interest
historymatrix	History matrix of PC8 product codes. Provided by history_matrix_pc8().
progress	logical, determines whether progress is printed in console or not.

Value

A data frame that contains all PC8 product codes and related BEC and HS6 codes in a given time period. The following table offers an overview of all provided variables.

Variable	Explanation
PC8	character; a specific PC8 code
HS6	character; provides the HS6 classification of the PC8plus code
BEC	character; provides the BEC classification on a high aggregation level (1 digit)
BEC_agr	character; provides the BEC classification on a lower aggregation level (up to 3 digits)

Examples

```
pc8_bec <- pc8_to_bec(b = 2008, e = 2010)
```

---

utilize_cn8 *A possible utilization of harmonized CN8 products codes*

Description

Provide an application of the data frames obtained by the main function, harmonize_cn8. To use these additional functions, data on firm-level is required, which is data that is not provided by the package.

Usage

```
utilize_cn8(b, e, firm_data, harmonized_data = NULL,
            progress = TRUE, output = "merged.firm.data",
            value = FALSE, base = "CN8")
```

Arguments

b	first year of interest
e	last year of interest
firm_data	Data on firm level which must provide the following columns: "firmID", "year" and "CN8".
harmonized_data	Harmonized data of CN8 product codes. Provided by harmonize_cn8(). By default *NULL*; the function computes the needed harmonized data.

progress	logical, determines whether progress is printed in console or not.
output	Defines which dataframe is returned. It may take the following values: • "product.changes", returns all changed CN8 product codes per firm per year (see description of (a) below) • "merged.firm.data", returns entered firm data, extended by harmonized data (see description of (b) below) • "all", returns both dataframes as a list
value	logical, determines whether value is calculated for same/new/dropped products. Only possible if data contains a column: "value". Value may contain different quantities (e.g. sales [Euro] or weight [kg]).
base	Defines which plus-codes are used as a base for calculating added/dropped/same products and their corresponding values. It may take the following values: • "CN8", uses CN8plus codes for computation. • "HS6", uses HS6plus codes for computation.

Value

Provides two possible data frames:

(a) One dataframe that contains all changed CN8 product codes per firm per year. In more detail, this means how many products remained the same, were added or dropped - the value of the same/added/dropped products - how many products were produced by a certain firm in a given year, and how many products were produced in the year after. As a base of this computation CN8plus codes or HS6plus codes can be used.

(b) One dataframe that is based on the entered firm data. The entered firm data is extended by harmonized data (that is "CN8plus", "flag", "flagyear", "HS6plus", "BEC", "BEC_agr", "SNA_basic_class").

Table that summarizes the output, described by the notation (a) above:

Variable	Explanation
firmID	character; specific code that describes a firm over the years (this code does not change over t
period_UL	character; lower limit of the time period
period	character; time period in which the product was produced
gap	numeric; indicating if the time period is greater than one (i.e. upper limit - lower limit > 1)
same_products	numeric; number of products that were produced in both years (i.e. remained in the product
value_same_products	numeric; value of products that were produced in both years (i.e. remained in the product po
new_products	numeric; number of added products in the upper limit of the time period (i.e. added to the pr
value_new_products	numeric; value of added products in the upper limit of the time period (i.e. added to the prod
dropped_products	numeric; number of dropped products in the upper limit of the time period (i.e. removed of t
value_dropped_products	numeric; value of dropped products in the upper limit of the time period (i.e. removed of the
nbr_of_products_period_LL	numeric; number of all products produced in the lower limit of the time period (i.e. entire pr
nbr_of_products_period_UL	numeric; number of all products produced in the upper limit of the time period (i.e. entire pr

Table that summarizes the output, described by the notation (b) above:

Variable	Explanation
firmID	character; specific code that describes a firm over the years (this code does not change over time, provided by use
year	numeric; year in which the firm produced a product (provided by user)
CN8	character; CN8 code of firm product (provided by user)
(value)	numeric; value of the corresponding product code (may be provided by user)
...	character; additional columns from original firm data (provided by user)
CN8plus	character; final harmonization, which refers to the last year of the time period
flag	numeric; integer from 0 to 3; 1 indicates that this code remained the same in notation over the whole time period
flagyear	numeric; indicates the first year in which the flag was set
HS6	character; provides the HS6 classification of the CN8plus code
HS6plus	character; also adjusts for the change lists of HS6
BEC	character; provides the BEC classification on a high aggregated level (1 digit)
BEC_agr	character; provides the BEC classification on a less aggregated level (up to 3 digits)
SNA	character; provides information if the code is classified as consumption, capital or intermediate good in BEC

Examples

```
sampledata <- read.table(paste0(system.file("extdata", package = "harmonizer"),
                         "/sampledata/cn8sample.txt"), sep = ";",
                         header = TRUE, colClasses = "character")

newdata <- utilize_cn8(b = 2008, e = 2010, firm_data = sampledata)

newdata <- utilize_cn8(b = 2008, e = 2010, firm_data = sampledata,
                       output = "all")
changes <- newdata[[1]]
merged_data <- newdata[[2]]
```

---

utilize_pc8 *A possible utilization of harmonized PC8 products codes*

Description

Provide an application of the data frames obtained by the main function, harmonize_pc8. To use these additional functions, data on firm-level is required, which is data that is not provided by the package.

Usage

```
utilize_pc8(b, e, firm_data, harmonized_data = NULL,
            progress = TRUE, output = "merged.firm.data",
            value = FALSE, base = "PC8")
```

Arguments

b	first year of interest
e	last year of interest

`firm_data` Data on firm level which must provide the following columns: "firmID", "year" and "PC8".

`harmonized_data` Harmonized data of PC8 product codes. Provided by harmonize_pc8(). By default *NULL*; the function computes the needed harmonized data.

`progress` logical, determines whether progress is printed in console or not.

`output` Defines which dataframe is returned. It may take the following values:

- "product.changes", returns all changed PC8 product codes per firm per year (see description of (a) below)
- "merged.firm.data", returns entered firm data, extended by harmonized data (see description of (b) below)
- "all", returns both dataframes as a list

`value` logical, determines whether value is calculated for same/new/dropped products. Only possible if data contains a column: "value". Value may contain different quantities (e.g. sales [Euro] or weight [kg]).

`base` Defines which plus-codes are used as a base for calculating added/dropped/same products and their corresponding values. It may take the following values:

- "PC8", uses CN8plus codes for computation.
- "HS6", uses HS6plus codes for computation.

Value

Provides two possible data frames:

(a) One dataframe that contains all changed PC8 product codes per firm per year. In more detail, this means how many products remained the same, were added or dropped - the value of the same/added/dropped products - how many products were produced by a certain firm in a given year, and how many products were produced in the year after. As a base of this computation PC8plus codes or HS6plus codes can be used.

(b) One dataframe that is based on the entered firm data. The entered firm data is extended by harmonized data (that is "PC8plus", "flag", "flagyear", "HS6plus", "BEC", "BEC_agr", "SNA_basic_class").

Table that summarizes the output, described by the notation (a) above:

Variable	Explanation
`firmID`	character; specific code that describes a firm over the years (this code does not change over t
`period_UL`	character; lower limit of the time period
`period`	character; time period in which the product was produced
`gap`	numeric; indicating if the time period is greater than one (i.e. upper limit - lower limit > 1)
`same_products`	numeric; number of products that were produced in both years (i.e. remained in the product
`value_same_products`	numeric; value of products that were produced in both years (i.e. remained in the product pc
`new_products`	numeric; number of added products in the upper limit of the time period (i.e. added to the pr
`value_new_products`	numeric; value of added products in the upper limit of the time period (i.e. added to the prod
`dropped_products`	numeric; number of dropped products in the upper limit of the time period (i.e. removed of t
`value_dropped_products`	numeric; value of dropped products in the upper limit of the time period (i.e. removed of the

nbr_of_products_period_LL	numeric; number of all products produced in the lower limit of the time period (i.e. entire pr
nbr_of_products_period_UL	numeric; number of all products produced in the upper limit of the time period (i.e. entire pr

Table that summarizes the output, described by the notation (b) above:

Variable	Explanation
firmID	character; specific code that describes a firm over the years (this code does not change over time, provided by use
year	numeric; year in which the firm produced a product (provided by user)
PC8	character; PC8 code of firm product (provided by user)
(value)	numeric; value of the corresponding product code (may be provided by user)
...	character; additional columns from original firm data (provided by user)
PC8plus	character; final harmonization, which refers to the last year of the time period
flag	numeric; integer from 0 to 3; 1 indicates that this code remained the same in notation over the whole time period
flagyear	numeric; indicates the first year in which the flag was set
HS6	character; provides the HS6 classification of the PC8plus code
HS6plus	character; also adjusts for the change lists of HS6
BEC	character; provides the BEC classification on a high aggregated level (1 digit)
BEC_agr	character; provides the BEC classification on a less aggregated level (up to 3 digits)
SNA	character; provides information if the code is classified as consumption, capital or intermediate good in BEC

Examples

```
sampledata <- read.table(paste0(system.file("extdata", package = "harmonizer"),
                          "/sampledata/pc8sample.txt"), sep = ";",
                          header = TRUE , colClasses = "character")

newdata <- utilize_pc8(b = 2011, e = 2013, firm_data = sampledata)

newdata <- utilize_pc8(b = 2011, e = 2013, firm_data = sampledata,
                       output = "all")
changes <- newdata[[1]]
merged_data <- newdata[[2]]
```

Index