Baublatt Manual

Salomon Faure, Dirk Drechsel

August 31, 2012

Contents

1 Introduction

1.1 Swiss construction sector indicator and forecast

The construction investments amount to 10% of the swiss GDP. As the production factors are essentially inside Switzerland, the development in this sector has a substantial impact on the development of other branches. That is why forecasts of construction investment are of significant importance for the construction companies and suppliers.

A swiss construction sector indicator was developed by the KOF based on the monthly publication of the "Baublatt" by the swiss cities ("Gemeinde"), a database that contains all construction requests and permits in Switzerland. Since almost every construction in Switzerland needs a permit, it is natural to think that this database can be used to have an idea of the swiss construction sector development. When a permit is given, the construction begins after several months[1]. Thanks to this time lag, the construction permits can therefore be used to predict the evolution of the construction sector over the next months after the publication of the database. Two indicators are now calculated based on this database:

- The indicator for the entire construction sector,
- The indicator for the housing market.

The time series obtained by summing over the construction costs are then shifted by 8 months for the small projects (whose costs are $<$ CHF 2 million), 10 months for middle projects (whose costs are between CHF 2 million and CHF 100 million) and 11 months for big projects (whose costs are $>$ CHF 100 million). These data are then aggregated quarterly. Afterwards, the annual construction investment data from the BFS are used as reference time series in a temporal disaggregation process (Chow and Lin (1971)). The time shift enables to make forecasts about the evolution of the construction sector over the next 8 months (approximately 2 quarters). Finally, the data is seasonally adjusted (X12 procedure). The inflation in the construction sector in not taken into account in the analysis. In the present documentation, we give a thorough description of how these indicators are calculated in practice and of all details that are taken into account in the aggregation.

Section 2 describes the input data (Baublatt database). Then, the data is read in R (Section 3) and CRB codes are attributed (Section 4). In Section 5, special cases are handled. Section 6 explains the two ways of spreading middle and big projects over time. The different types of aggregations are presented in Section 7. In Section 8, the calculation of the Baublatt indicator is explained. Finally, ideas to improve and develop the aggregation are given in Section 9.

1.2 Annual construction investment data from the BFS

1.2.1 Description

The time series from the BFS are based on data of realized projects only, that are given by the swiss cities ("Gemeinde"). However, the BFS database contains also smaller projects, which are sometimes permitted in a faster procedure after their realization and which are not contained in the Baublatt database. Nevertheless, the major difference between the two time series is that public projects mostly do not request a permit and are therefore not included in the Baublatt database whereas they are included in the BFS database. That is why the Baublatt time series are smaller than the BFS investment time series for the entire construction sector whereas the two time series are much nearer for the housing construction sector.[2]

1.2.2 Update

The annual construction investment data from the BFS are updated every august. This update has to be done before the calculation of the indicators. The procedure is simple and given in Appendix A. Basically, it adds new values for the previous year to the files "annual_data_yyyy_yyyy.csv" and "annual_data_wb_yyyy.csv", for the entire construction sector and the housing sector respectively. These time series will be used afterwards as reference time series in the temporal disaggregation procedure.

[1] Fundamentally, a construction permit is valid for approximately 2 years.

[2] There are less public construction projects proportionally in the housing construction sector.

2 Input data

2.1 Interpretation

The data comes in the format of a file named "KOF_9943D2K_yyyy_mm.csv" where yyyy represents the current year and mm the current month. This file contains mostly projects which were permitted in the current month. It contains, among other, the following parts:

- An object number ("ObjektNr") which characterizes a unique construction project: column A,
- The references of the construction client (BH stands for "Bauherr" and means construction client): columns B:S,
- The references of the architect firm (PL stands for "Planung"): columns T:AL,
- The references of the construction firm (Ing stands for "Ingenieur"): columns AM:BE,
- The references of the construction site: columns BG:BN,
- A short description of the project: column BO,
- The construction stage: column BP,
- The sort of project (1: new construction, 2: extension, 3: transformation, 4: renovation, 5: demolition): column BQ,
- The date of construction start and the date of construction end: columns BR:BS,
- The date of construction request: columns BT,
- The date of construction permit: columns BU,
- The lower and upper bound of construction costs: columns BV:BW,
- The number of buildings: column BX,
- The number of appartments: column BY,
- The number of stories: column BZ,
- The number of basement-levels: column CA,
- The number of car parks: column CB,
- The CRB codes are codes that are used to identify the nature of projects (12 corresponds to "Einfamilienhaus", 14 corresponds to "Mehrfamilienhaus", 39 correponds to "Atelier und Studio", 211 corresponds to "Strassen", etc.). A project has one cost but can have more than one nature: in this case, the project is called mixed project. Mixed projects will cause problems for the aggregation in the housing market. As a matter of fact it can happen that a project contains the construction of a restaurant, a retail store and several houses for example. But the proportion of the costs which are due to the houses are not known. Mixed projects (projects that have more than one CRB code) represent about 80% of total costs. CRB codes are in columns CH:DA,
- The GA codes are codes for the construction materials. They are in columns DB:EY,
- The last columns EZ:FS contain further descriptions of the project.

2.2 Notations

We will use only a small part and potential of the data. For convenience, we give here the notation, we will use throughout the paper.

- An object number will be denoted by $n \in [0, +\infty[$[3],
- The set of all object numbers is denoted by $ObjektNr$,
- The canton where the project n is taking place is denoted by $BaustKanton(n) \in \{ZH, BE, LU, UR, ...\}$,
- The code for the sort of construction project n is denoted by $BauartCode(n) \in \{1, 2, 3, 4, 5\}$,
- The costs of the project n are denoted by $costs(n) \in [0, +\infty[$,
- The number of appartments of the project n is denoted by $AnzWhg(n) \in \mathbb{N}$,
- The month of request date for project n is denoted by $gesuch\_mon(n) \in M := \{1, 2, 3, ..., 12\}$,
- The year of request date for project n is denoted by $gesuch\_jahr(n) \in Y := \{1993, 1994, 1995, ...\}$,
- The set of all projects n whose request date is $(gesuch\_mon, gesuch\_jahr) = t$ is denoted by ges_t,

- The month of permit date for project n is denoted by $bewilligt\_mon(n) \in M$,
- The year of permit date for project n is denoted by $bewilligt\_jahr(n) \in Y$,
- The set of all projects n whose permit date is $(bewilligt\_mon, bewilligt\_jahr) = t$ is denoted by bew_t,

- The first CRB code of project n is denoted by $CRBCode01(n) \in \mathbb{N}$,
- The second CRB code of project n is denoted by $CRBCode02(n) \in \mathbb{N}$,
- The CRB code of project n is denoted by $CRBCode(n) \in \mathbb{N}$.

[3] Originally, n is a positive integer but the special cases will be treated in such a way that two objects with object numbers $n_1 = n + 0.001$ and $n_2 = n + 0.002$ will be created and will replace it. It is then clear that we cannot restrict n to be an integer.

3 Lecture of the data

The data comes in the format of a file named “KOF_9943D2K_yyyy_mm.csv” where yyyy represents the current year and mm the current month. First, this file is treated with the macro available in the file “Macros/MACRO.xls” described in following subsection.

3.1 Erasing columns and sorting CRB codes: “Macros/MACRO.xls”

This macro does 2 things:

- It erases columns that are not necessary for the analysis.
- For each project, it sorts the CRB codes in an increasing manner and deletes duplicated entries in the CRB codes.[4]

3.2 Read the data in *R*: “functions/readdata.r”

After being transformed and saved under “MonatlichDataBaublatt/KOF_9943D2K_yyyy_mm.csv”, the new file is read in *R* with the first script “SCRIPT01.r”, which contains “scripts/01Readdata_AND_CRB_monatlich.r”, which contains “functions/readdata.r”.

- We read the data.

```
data<-read.csv(b, sep=";", header=TRUE, stringsAsFactors = FALSE)
```

  With the last argument “stringsAsFactors”, we avoid using factors as a source of problems and confusion.

- We transform all data, that should contain numbers, into numeric data in order to avoid problems later on. For example, we have for “ObjektNr”

```
data[,"ObjektNr"] <- as.numeric(data[,"ObjektNr"])
```

- We noticed that people sometimes missed the date column and entered the date in the cost column. The conversion of the date by *R* yielded high values for the costs, which had a significant impact on the results. In order to delete these flawed entries, we have erased from the database all of them, whose costs had 10 or more that 10 characters (“2000.0000” has 9 characters for example and would be kept but “01.01.1990” has 10 characters and would be deleted).

```
data <- data[nchar(data$BaukostenVon) < 10,]
```

- We created a new column which is the mean between the lower and upper bound of construction costs. We name this column *baukostends* (*ds* stands for “Durchschnitt” which means “average” in German).

```
data <- cbind(data,"baukostends"=(data$BaukostenVon+data$BaukostenBis)/2)
```

- We erase the data whose permit or request date does not seem realistic. Normally, the entries of the month have a *gesuch_jahr* and *gesuch_mon* which correspond to a date which is before the current date. To handle anormal cases, we erase all entries whose request date is after two years after the current date. We do the same for permit dates.

```
data <- data[data$gesuch_jahr < jahr+2 | is.na(data$gesuch_jahr),]
```

- We keep only the following columns for the analysis.

```
data <- data[,c("ObjektNr","BaustKanton","BauartCode","baukostends","AnzWhg",
"gesuch_mon","gesuch_jahr","bewilligt_mon","bewilligt_jahr","CRBCode01","CRBCode02","Kurzbeschreibung")]
```

[4]As we will see in the next subsections, sorting increasingly the CRB codes has a great impact on the results for the housing market. This will probably be questioned in a new calculation of the Baublatt indicator.

4 Attribution of CRB codes: "functions/crbcodes.r"

The CRB codes are codes that are used to identify the nature of projects (12 corresponds to "Einfamilienhaus", 14 corresponds to "Mehrfamilienhaus", 39 correponds to "Atelier und Studio", 211 corresponds to "Strassen", etc.). A project has one cost but can have more than one nature: in this case, the project is called mixed project. Mixed projects will cause problems for the aggregation in the housing market. As a matter of fact it can happen that a project contains the construction of a restaurant, a retail store and several houses for example. But the proportion of the costs which are due to the houses are not known. Mixed projects (projects that have more than one CRB code) represent about 80% of total costs. In this calculation of the Baublatt indicator, a project can have only one CRB code. Therefore, we have to choose which CRB code to attribute to which project. The CRB codes were sorted in an increasing manner (see section 3.1) and duplicated entries were eliminated. This sorting is questionable and will have a major impact for the results for the housing market. We selected only CRB codes $CRBCode01$ and $CRBCode02$, which are the two smallest CRB codes of a project. CRB codes are then attributed following the subsequently described rule illustrated in figure 3 in the Appendix and in the files "crbcodes.eps" and "crbcodes.jpeg".

Let $n \in ObjektNr$.

- If $CRBCode01(n) = 12$ or $CRBCode01(n) = 14$ (the mixed project contains "Einfamilienhäuser" or "Mehrfamilienhäuser"), then
  - If $CRBCode02(n) = 21$ (Kinderhorte und Kindergärten) or $= 127$ (Aussenanlagen, Kinderspielplät) or $= 131$ (Parkhäuser und Einstellhallen) or $= 139$ (Garagen / Fertiggaragen) or $= 174$ (Carports und Abstellplätze) or $= 175$ (Parkplätze) or $= NA$ (single type project), which means that the project contains one of these construction types as second smallest CRB code or the project is a single type project, then
    * If $costs(n) < 2$ or $costs(n) = NA$ (the project has a small size), then
    $$CRBCode(n) = CRBCode01(n) \text{ (12 or 14).}$$
    * If $2 < costs(n) \leq 100$ (the project has a middle size), then
    $$CRBCode(n) = CRBCode01(n) \text{ (12 or 14).}$$
    * If $100 < costs(n)$ (the project has a big size), then
      · If $\frac{AnzWhg(n)}{costs(n)} < \frac{10}{3}$ or $AnzWhg(n) = NA$ (the costs are too big in comparison with the number of appartments (more than CHF 300 000 per appartment), which means that the type $CRBCode02(n)$ has large costs in comparison with the appartments built), then
      $$CRBCode(n) = -1 \text{ (this project will be handled as a special case).}$$
      · If $\frac{AnzWhg(n)}{costs(n)} > \frac{10}{3}$ (the costs are reasonable in comparison with the number of appartments (less than CHF 300 000 per appartment) which means that the type $CRBCode02(n)$ has small costs in comparison with the appartments built), then
      $$CRBCode(n) = CRBCode01(n) \text{ (12 or 14).}$$
  - If $CRBCode02(n) \neq 21$ (Kinderhorte und Kindergärten) and $\neq 127$ (Aussenanlagen, Kinderspielplät) and $\neq 131$ (Parkhäuser und Einstellhallen) and $\neq 139$ (Garagen / Fertiggaragen) and $\neq 174$ (Carports und Abstellplätze) and $\neq 175$ (Parkplätze) and $\neq NA$ (single type project), which means that the project does not contain one of these construction types as second smallest CRB code and the project is not a single type project, then
    * If $costs(n) < 2$ or $costs(n) = NA$ (the project has a small size), then
    $$CRBCode(n) = CRBCode01(n) \text{ (12 or 14).}$$
    * If $2 < costs(n) \leq 100$, then

  - If $\frac{AnzWhg(n)}{costs(n)} < \frac{10}{3}$ or $AnzWhg(n) = NA$ (the costs are too big in comparison with the number of appartments (more than CHF 300 000 per appartment), which means that the type $CRBCode02(n)$ has large costs in comparison with the appartments built), then the housing part is negligible in comparison with the $CRBCode02(n)$ other part:

$$CRBCode(n) = CRBCode02(n).$$

  - If $\frac{AnzWhg(n)}{costs(n)} > \frac{10}{3}$ (the costs are reasonable in comparison with the number of appartments (less that CHF 300 000 per appartment) which means that the type $CRBCode02(n)$ has small costs in comparison with the appartmetns built), then the $CRBCode02$ part is negligible in comparison with the housing part:

$$CRBCode(n) = CRBCode01(n) \text{ (12 or 14)}.$$

* If $costs(n) > 100$ (the project has a large size), then

$$CRBCode(n) = -1 \text{ (this project will be handled as a special case)}.$$

- If $CRBCode01(n) \neq 12$ and $\neq 14$ (the project does not contain an "Einfamilienhaus" or a "Mehrfamilienhaus"),[5] then

$$CRBCode(n) = CRBCode01(n).$$

- If $CRBCode01(n) = NA$ (the project has no CRB code), then
  - If $CRBCode02(n) \neq NA$,[6] then $CRBCode(n) = CRBCode02(n)$,[7]
  - If $CRBCode02(n) = NA$, then $CRBCode(n) = NA$ (this means that the project has no CRB code).

We call special cases the cases where $CRBCode(n) = -1$. These cases are to be handled. The handling procedure is described in the next section.

[5] As 12 and 14 are almost the smallest numbers in the CRB code classification, we can conclude that if $CRBCode01(n)$ which is the smallest of the CRB codes is not 12 and 14, then 12 and 14 do not appear in the CRB codes of the project.

[6] This case should not actually arise because the CRB codes were sorted in the VBA macro.

[7] In the present script, it is -1 but it is clear that $CRBCode02$ has to be attributed, which is made in "SCRIPT02.r". Here, we take the shortcut that $CRBCode02$ is directly attributed.

5 Treatment of special cases: "SCRIPT02.r"

The special cases are cases where the costs cannot be explained alone by either the housing part or by the $CRBCode02$ part. They can only be explained by both at the same time or by an error (if a person has entered CHF 500M instead of .5M thinking that $costs(n)$ were in thousands of CHF instead of millions of CHF).

5.1 Automatic handling

The cases where the costs linked to $CRBCode02$ can be very big (for example the cases for which $CRBCode02$ is a CRB code whose 99.99% *costs* quantile is above 100. We denote by $CRBB$ the set of all CRB codes whose 99.99% *costs* quantile is above 100.) are clear and these cases are treated automatically in the following way.[8] But in order to avoid treating automatically cases where costs are 1000 times too big, we select only those cases where the number of appartments are more than 20 and the costs are less than 500. If the number of apprtments are more than 20, then with an error of 1000, the costs will be approximately $20 * 0.3 * 1000 = 600 > 500$. Most of the time the 1000-errors are for small projects and in these cases, the number of appartments is < 20.

For all object numbers n, if

- $AnzWhg(n) < 20$,
- $costs(n) < 500$,
- $CRBCode02(n) \in CRBB$,

then

- Two objects[9] $n_1 = n + 0.001$ and $n_2 = n + 0.002$ are created and the old one is deleted. The new characteristics of object n_1 (housing part) are
  - $n_1 = n + 0.001$,
  - $costs(n_1) = 0.3AnzWhg(n)$,
  - $CRBCode(n_1) = CRBCode01(n)$ (12 or 14).

  The new characteristics of object n_2 ($CRBCode02$ part) are
  - $n_2 = n + 0.002$,
  - $costs(n_2) = costs(n) - 0.3AnzWhg(n) = costs(n) - costs(n_1)$,
  - $AnzWhg(n_2) = 0$,
  - $CRBCode(n_2) = CRBCode02(n)$.

We have to note that the number of appartments and the total costs are conserved in this way: $costs(n) = costs(n_1) + costs(n_2)$ and $AnzWhg(n) = AnzWhg(n_1) + AnzWhg(n_2)$. The number 2 as a handling code is attributed to the special cases which were handled in this way. The log-file "data/logExceptions.rdata" contains all object numbers of the special cases from the creation of the database with their handling number. The log-file "data/logExceptionsDetailed.rdata" contains all special cases with all their attributes and their handling number from the creation of the database. At each time a special case is treated, the two new objects are added to the database "data/total.rdata", the old object is deleted, the log-files are updated and saved, the database is saved and the object is deleted from the list of all special cases to be treated.

All other special cases, which were not handled automatically need more attention. This is the subject of the following subsection.

[8] See the script "functions/MixedProjectHandlingAnfang.r" for the code.

[9] We use 0.001 to avoid interferences with cases where the two first decimals are used (see subsection 6.1).

5.2 Manual handling

Now, the special cases that were not handled automatically are to be handled manually. At each special case n, its relevant attributes are shown and three propositions are made to the user:

- 1: Cut the costs of n by 1000. In this case, the only change is the following:

$$costs_{new}(n) = \frac{costs(n)}{1000}.$$

- 2: Treat the project as a mixed project in the same way as the automatic way.
- 3: Treat the project differently.

By typing 3, we leave the program. But each change is always saved after a choice is made. This enables the user to start the program but running it normally and have to treat the last object, he or she has dealt with. As a matter of fact, at each time a special case is treated, the two new objects are added to the database "data/total.rdata", the old object is deleted, the log-files are updated and saved, the database is saved and the object is deleted from the list of all special cases to be treated.

6 Splits of the middle and big projects

Middle and big projects can have a very big impact on the monthly results without having any sense because the costs are usually not incurred directly at the beginning of the construction but over a period of time. Therefore, two techniques were developed to deal with this issue. The first one deals with every middle and big project. The second one deals with the middle projects globally and the big projects globally.[10]

6.1 Linear split of the middle and big projects: "scripts/02Aktualisierung_BAU_monatlich.r"

The first technique is a linear split of middle and big projects. Each middle project (whose costs are bigger than 20M) is split over 12 months after the request date or the permit date and each big project (whose costs are bigger than 100M) is split over 24 months after the request date or the permit date. Concretely, a middle project n when handled with respect to the request dates is replaced by 12 similar projects

$$\begin{aligned} n_1 &= n + 0.01, \\ n_2 &= n + 0.02, \\ &..., \\ n_{12} &= n + 0.12, \end{aligned}$$

with the only differences in the attributes that

$$\begin{aligned} costs(n_1) &= \frac{costs(n)}{12}, \\ costs(n_2) &= \frac{costs(n)}{12}, \\ &..., \\ costs(n_{12}) &= \frac{costs(n)}{12}, \end{aligned}$$

and that

$$\begin{aligned} gesuch\_mon(n_1) &= gesuch\_mon(n), \\ gesuch\_mon(n_2) &= ((gesuch\_mon(n) + 1 - 1) \bmod 12) + 1, \\ &..., \\ gesuch\_mon(n_{12}) &= ((gesuch\_mon(n) + 11 - 1) \bmod 12) + 1, \end{aligned}$$

and that[11]

$$\begin{aligned} gesuch\_jahr(n_1) &= gesuch\_jahr(n), \\ gesuch\_jahr(n_2) &= \left\lfloor gesuch\_jahr(n) + \frac{gesuch\_mon(n) + 1 - 1}{12} \right\rfloor, \\ &..., \\ gesuch\_jahr(n_{12}) &= \left\lfloor gesuch\_jahr(n) + \frac{gesuch\_mon(n) + 11 - 1}{12} \right\rfloor. \end{aligned}$$

The same can be written with respect to the permit dates. A big project n when handled with respect to the permit dates is similarly replaced by 24 similar projects

$$\begin{aligned} n_1 &= n + 0.01, \\ n_2 &= n + 0.02, \\ &..., \\ n_{24} &= n + 0.24, \end{aligned}$$

[10]The result when dealing with the middle or big projects individually or globally is clearly the same. But the code when dealing with the projects individually is also clearly more cumbersome.

[11]Recall that $\lfloor x \rfloor$ denotes the largest integer not greater than x.

with the only differences in the attributes that

$$\begin{aligned} costs(n_1) &= \frac{costs(n)}{24}, \\ costs(n_2) &= \frac{costs(n)}{24}, \\ &..., \\ costs(n_{24}) &= \frac{costs(n)}{24}, \end{aligned}$$

and that

$$\begin{aligned} bewilligt\_mon(n_1) &= bewilligt\_mon(n), \\ bewilligt\_mon(n_2) &= ((bewilligt\_mon(n) + 1 - 1) \bmod 12) + 1, \\ &..., \\ bewilligt\_mon(n_{24}) &= ((bewilligt\_mon(n) + 23 - 1) \bmod 12) + 1, \end{aligned}$$

and that

$$\begin{aligned} bewilligt\_jahr(n_1) &= bewilligt\_jahr(n), \\ bewilligt\_jahr(n_2) &= \left\lfloor bewilligt\_jahr(n) + \frac{bewilligt\_mon(n) + 1 - 1}{12} \right\rfloor, \\ &..., \\ bewilligt\_jahr(n_{24}) &= \left\lfloor bewilligt\_jahr(n) + \frac{bewilligt\_mon(n) + 23 - 1}{12} \right\rfloor. \end{aligned}$$

6.2 Degressive split of middle and big projects: "scripts/14Verteilen_quarterly.r"

A degressive split of middle and big projects with a rate[12] $r = 7.5\%$ consists of a global split of the costs in the following way.

- The first period, 7.5% of the costs of the current period are allocated.
- In the second period, 7.5% of the remaining costs are allocated: $7.5\% * 92.5\% = 6.94\%$.
- In the third period, 7.5% of the remaining costs are allocated: $7.5\% * 85.56\% = 6.42\%$.
- This procedure is repeated until the current period.
- This procedure is repeated with the costs of all periods.

Table 1 gives the percentage values that are allocated for initial 100% costs in period 1 and nothing in the other periods.

As this procedure is repeated with the costs of all periods. The degressive split over all periods works in the following way.

- The first period, 7.5% of the costs of the current period $costs_1$ are allocated: $costs_1^{split} = 7.5\% costs_1$.
- In the second period, 7.5% of the remaining costs of the first period are allocated ($7.5\% * 92.5\% costs_1 = 6.94\% costs_1$) and 7.5% of the costs of the current period are allocated ($7.5\% costs_2$). Finally, the costs allocated to the second period amount to

$$costs_2^{split} = 7.5\% costs_2 + 6.94\% costs_1.$$

[12] The rate of 7.5%, which is now used for the Baublatt indicator, is a value that could be questioned in a new calculation of the indicator. This issue is discussed in the last section of this paper.

Period	Cumulated percentage allocation	Percentage allocation
1	7.50	7.50
2	14.44	6.94
3	20.85	6.42
4	26.79	5.94
5	32.28	5.49
6	37.36	5.08
7	42.06	4.70
8	46.40	4.35
9	50.42	4.02
10	54.14	3.72
11	57.58	3.44
12	60.76	3.18
13	63.71	2.94
14	66.43	2.72
15	68.95	2.52
16	71.27	2.33
17	73.43	2.15
18	75.42	1.99
19	77.27	1.84
20	78.97	1.71
21	80.55	1.58
22	82.01	1.46
23	83.36	1.35
24	84.60	1.25
25	85.76	1.15
26	86.83	1.07
27	87.81	0.99
28	88.73	0.91
29	89.57	0.85
30	90.36	0.78

Table 1: Percentage allocation of the costs for a degressive split of middle and big projects

- In the third period, 7.5% of the remaining costs of the first period are allocated ($7.5\% * 85.56\% costs_1 = 6.42\% costs_1$), 7.5% of the remaining costs of the second period are allocated ($7.5\% * 92.5\% costs_2 = 6.94\% costs_2$) and 7.5% of the costs of the current period are allocated ($7.5\% costs_3$). Finally the costs allocated to the second period amount to

$$costs_3^{split} = 7.5\% costs_3 + 6.94\% costs_2 + 6.42\% costs_1$$

- The procedure is repeated until the current period.

7 Aggregations

7.1 Classifications

Different sorts of aggregation are calculated based on the Baublatt database. The aggregations are usually done for the entire construction sector (all CRB codes) and then for specific CRB codes.[13] The classifications described in the following use some codes, which we will introduce, to interpret the times series resulting from the aggregation over a class of projects. For example, the time series called "BEW_KOEF_NU_HT_NOM" is the time series based on permit dates ("BEW" stands for "Bewilligung") resulting from the aggregation over all public construction projects ("OEF" stands for "öffentlicher Bau") for new construction and transformation projects ("N" stands for "Neubau" and "U" for "Umbau") for the buildings construction sector and the civil engineering sector ("H" stands for "Hochbau" and "T" for "Tiefbau") in nominal terms (this means that the time series was not treated with a degressive split of middle and big projects over time).

7.1.1 Sort of construction classification

For the sorts of construction, the classification is the following one.

- If

  $$BauartCode(n) \in NB := \{1, 2\},$$

  then n is classified as a new construction with code "NB" (which stands for "Neubau"). We denote

  $$NB_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in NB\}.$$

- If

  $$BauartCode(n) \in UB := \{3, 4, 5\},$$

  then n is classified as a transformation with code "UB" (which stands for "Umbau"). We denote

  $$UB_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in UB\}.$$

- If[14]

  $$BauartCode(n) \in [\![1, 5]\!],$$

  then n is not yet classified with code "NU" (which stands for "Neubau, Umbau").

7.1.2 Building/civil engineering classification

For the CRB codes,[15] there are two different classifications. The first one differentiates between building construction projects and civil engineering projects.

- If

  $$CRBCode(n) \in HB := [\![10, 199]\!],$$

  then n is classified as building construction project with code "HB" (which stands for "Hochbau"). We denote

  $$HB_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in HB\}.$$

[13] Remember all the problems linked to the CRB codes and the mixed projects.
[14] Recall that $[\![1, 5]\!]$ is the notation for $\{1, 2, 3, 4, 5\}$.
[15] See the file "BACKUP/Documentation/CRBCodes/CRBCodeDescription.csv" for the signification of the CRB codes.

- If
$$CRBCode(n) \in TB := [\![200, 741]\!],$$
then n is classified as civil engineering project with code "TB" (which stands for "Tiefbau"). We denote
$$TB_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in TB\}.$$

- If
$$CRBCode(n) \in [\![10, 741]\!],$$
then n is not yet classified with code "HT" (which stands for "Hochbau, Tiefbau").

7.1.3 Sectorial classification

Classification This classification differentiates between public construction projects, private construction projects, housing construction projects and industrial construction projects.

- If
$$\begin{aligned} CRBCode(n) \in \quad OEF := \quad & [\![20, 29]\!] \cup [\![51, 53]\!] \cup [\![56, 58]\!] \cup [\![66, 79]\!] \\ & \cup [\![83, 84]\!] \cup [\![90, 109]\!] \cup [\![130, 138]\!] \cup [\![141, 159]\!] \\ & \cup [\![180, 263]\!] \cup [\![265, 312]\!] \cup \{314\} \cup [\![316, 462]\!] \\ & \cup [\![464, 610]\!] \cup [\![614, 741]\!], \end{aligned}$$
then n is classified as a public construction with code "OEF" (which stands for "öffentlicher Bau"). We denote
$$OEF_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in OEF\}.$$

- If
$$\begin{aligned} CRBCode(n) \in \quad PRI := \quad & [\![31, 49]\!] \cup [\![54, 55]\!] \cup [\![59, 65]\!] \cup \{82\} \\ & \cup [\![85, 89]\!] \cup [\![110, 129]\!] \cup \{139\} \cup [\![161, 179]\!] \\ & \cup \{264, 313, 315, 463\} \cup [\![611, 614]\!], \end{aligned}$$
then n is classified as a privat construction project with code "PRI" (which stands for "Privat"). We denote
$$PRI_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in PRI\}.$$

- If
$$CRBCode(n) \in WBT := [\![9, 20]\!],$$
then n is classified as a housing construction project with code "WBT" (which stands for "Wohnbauten"). We denote
$$WBT_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in WBT\}.$$

- If
$$\begin{aligned} CRBCode(n) \in WIB \quad := \quad & [\![31, 39]\!] \cup [\![61, 65]\!] \cup \{47, 54, 59\} \cup [\![111, 118]\!] \\ & \cup \{161\} \cup [\![166, 168]\!], \end{aligned}$$
then n is classifed as an industrial construction project with code "WIB" (which stands for "Wirtschaftsbau"). We denote
$$WIB_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in WIB\}.$$

- If

$$CRBCode(n) \in DIFF := \{0, NA\},$$

then n has no classification with code “DIFF” (which stands for “difference”). We denote

$$DIFF_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in DIFF\}.$$

Subclasses and exceptions There are many subclasses as shown in figure 1. However, some exceptions for the aggregation in these classes arise:[16]

- The aggregated value for the public housing construction projects (code “WBO_NU”) is defined as being 1.8% of the aggregated value for the housing construction projects.
- The aggregated value for the private housing construction projects (code “WBP_NU”) is defined as being 98.2% of the aggregated value for the housing construction projects.
- The aggregated value for the new public housing construction projects (code “WBO_NB”) is defined as being 1.8% of the aggregated value for the new housing construction projects.
- The aggregated value for the new private housing construction projects (code “WBP_NB”) is defined as being 98.2% of the aggregated value for the new housing construction projects.
- The aggregated value for the public housing transformation projects (code “WBO_UB”) is defined as being 1.8% of the aggregated value for the housing transformation projects.
- The aggregated value for the private housing transformation projects (code “WBP_UB”) is defined as being 98.2% of the aggregated value for the housing transformation projects.

These exceptions are described in figure 1.

[16] This proportion seems perhaps a little bit arbitrary and that is why it will be discussed in more detail in the last section.

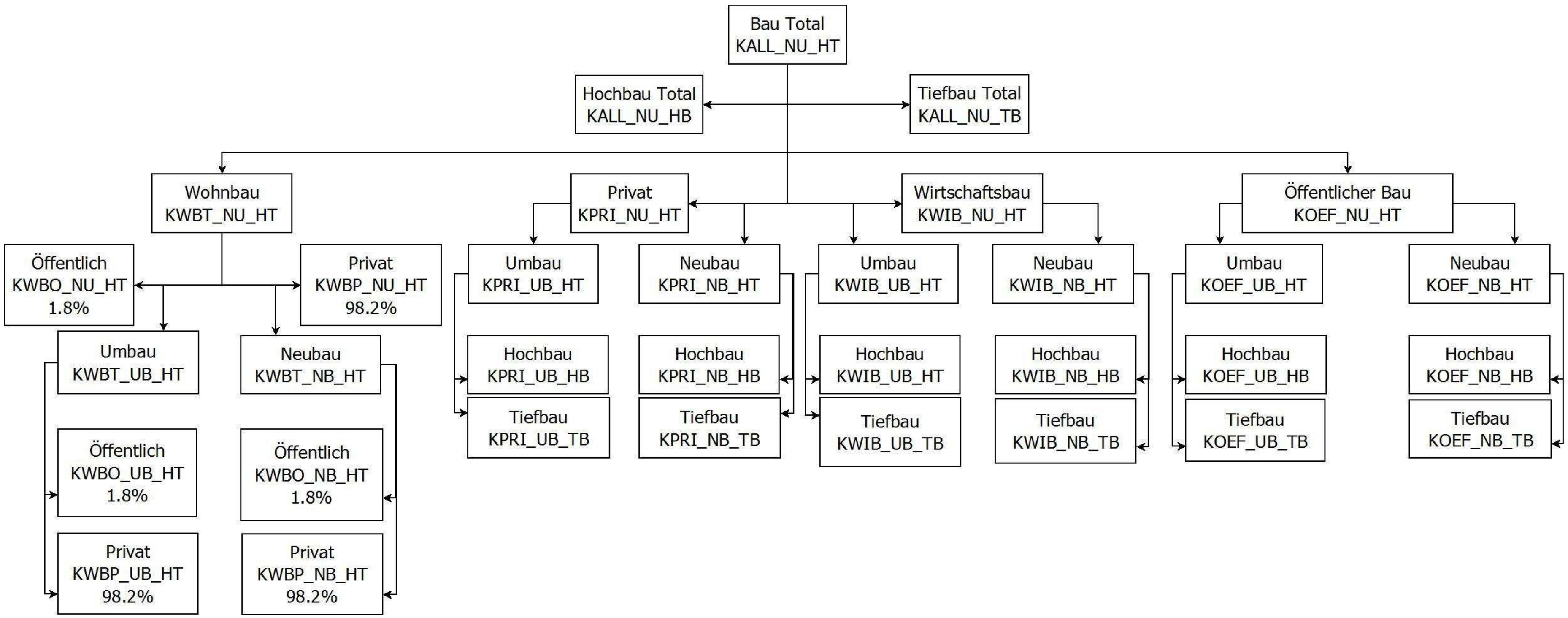

Figure 1: Classifications for the aggregation over the CRB codes and the Bauartcodes

7.1.4 Housing classification

Cantons In the case of the housing construction sector, the aggregation is made regionally over Switzerland and all the 26 cantons

$$\{ \quad ZH, BE, LU, UR, SZ, OW, NW, GL, ZG, FR, SO, BS, BL,\\ SH, AR, AI, SG, GR, AG, TG, TI, VD, VS, NE, GE, JU\}.$$

For a canton KT of this set, we denote

$$KT_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad BaustKanton(n) = KT\}.$$

Housing types Moreover, the housing construction sector differentiates bewteen "Einfamilienhäuser" and "Mehrfamilienhäuser". This classification is introduced in the following.

- If

$$CRBCode(n) \in EF := \{14, 19\},$$

  then n is classified as a single family house with code "EF" (which stands for "Einfamilienhaus"). We denote

$$EF_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in EF\}.$$

- If

$$CRBCode(n) \in MF := \{10, 11, 12, 13, 15\}$$

  then n is classified as a multiple family house with code "MF" (which stands for "Mehrfamilienhaus"). We denote

$$MF_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in MF\}.$$

- If

$$CRBCode(n) \in [\![9, 20]\!]$$

  then n is classified as a housing construction project with code "EM" (which stands for "Einfamilienhaus, Mehrfamilienhaus"). We denote

$$EM_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad CRBCode(n) \in EM\}.$$

7.1.5 Size classification

Importance of a size classification A size classification is important for at least two things.

- It enables to spread the middle and big size projects over time.
- It enables to treat small, middle and big size projects differently with different shifts (respectively 8 months, 10 months and 11 months).

Classification Finally, we use a classification based on the size of the projects as well.

- If

$$costs(n) \in K := [0, 20],$$

  then n is classified as a small project with code "KP" (which stands for "kleines Projekt"). We denote

$$K_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad costs \in K\}.$$

- If

$$costs \in M :=]20, 100]$$

then n is classified as a medium project with code "MP" (which stands for "mittleres Projekt"). We denote

$$M_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad costs(n) \in M\}.$$

- If

$$costs(n) \in]100, +\infty[$$

then n is classified as big project with code "GP" (which stands for "grosses Projekt"). We denote

$$G_{ObjNr} := \{n \in ObjektNr \quad s.t. \quad costs(n) \in G\}.$$

7.2 Aggregation for the entire construction sector with a linear split of middle and big projects

This is done in

- "scripts/03Bau_Auswertung2_quarterly.r" with resulting time series
  - "data/SumBewBau.rdata",
  - "data/SumGesBau.rdata".

We assume here that the database was transformed as explained in section 6.1 i.e. each middle and each big project was split linearly over the 12 or 24 months following the request date or permit date. We use indeed the databases "data/gesuch24database.rdata" (for the split over the request dates) and "data/bewilligt24database.rdata" (for the split over the permit dates).

Now, we create the time series thanks to the previously described classifications (see figure 1). t can denote $(gesuch\_mon, gesuch\_jahr)$ or $(bewilligt\_mon, bewilligt\_jahr)$ (whose codes are respectively "GES" and "BEW", which stand respectively for "Gesuch" (request) and "Bewilligung" (permit)).

As the aggregation over the request dates is similar to the one over the permit dates, we just give the results for the aggregation over the permit dates.[17] For the entire construction sector, we have

$$\begin{aligned}
BEW_KALL_NU_HT_NOM_t &= \sum_{n \in bew_t} costs(n), \\
BEW_KALL_NU_HB_NOM_t &= \sum_{n \in HB_{ObjNr} \cap bew_t} costs(n), \\
BEW_KALL_NU_TB_NOM_t &= \sum_{n \in TB_{ObjNr} \cap bew_t} costs(n).
\end{aligned}$$

For the housing construction sector, we have

$$\begin{aligned}
BEW_KWBT_NU_HT_NOM_t &= \sum_{n \in WB_{ObjNr} \cap bew_t} costs(n), \\
BEW_KWBO_NU_HT_NOM_t &= 1.8\% KWBT_NU_HT_NOM_t, \\
BEW_KWBP_NU_HT_NOM_t &= 98.2\% KWBT_NU_HT_NOM_t.
\end{aligned}$$

For the housing transformation sector, we have

$$\begin{aligned}
BEW_KWBT_UB_HT_NOM_t &= \sum_{n \in WB_{ObjNr} \cap UB_{ObjNr} \cap bew_t} costs(n), \\
BEW_KWBO_UB_HT_NOM_t &= 1.8\% KWBT_UB_HT_NOM_t, \\
BEW_KWBP_UB_HT_NOM_t &= 98.2\% KWBT_UB_HT_NOM_t.
\end{aligned}$$

[17] Recall that NOM means that the time series was not treated with a degressive split of middle and big projects over time.

For the housing building sector, we have

$$
\begin{aligned}
BEW\_KWBT\_NB\_HT\_NOM_t &= \sum_{n \in WB_{ObjNr} \cap NB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KWBO\_NB\_HT\_NOM_t &= 1.8\% KWBT\_NB\_HT\_NOM_t, \\
BEW\_KWBP\_NB\_HT\_NOM_t &= 98.2\% KWBT\_NB\_HT\_NOM_t.
\end{aligned}
$$

For the privat construction sector, we have

$$
\begin{aligned}
BEW\_KPRI\_NU\_HT\_NOM_t &= \sum_{n \in PRI_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KPRI\_UB\_HT\_NOM_t &= \sum_{n \in PRI_{ObjNr} \cap UB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KPRI\_UB\_HB\_NOM_t &= \sum_{n \in PRI_{ObjNr} \cap UB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KPRI\_UB\_TB\_NOM_t &= \sum_{n \in PRI_{ObjNr} \cap UB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KPRI\_NB\_HT\_NOM_t &= \sum_{n \in PRI_{ObjNr} \cap NB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KPRI\_NB\_HB\_NOM_t &= \sum_{n \in PRI_{ObjNr} \cap NB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KPRI\_NB\_TB\_NOM_t &= \sum_{n \in PRI_{ObjNr} \cap NB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n).
\end{aligned}
$$

For the industrial construction sector, we have

$$
\begin{aligned}
BEW\_KWIB\_NU\_HT\_NOM_t &= \sum_{n \in WIB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KWIB\_UB\_HT\_NOM_t &= \sum_{n \in WIB_{ObjNr} \cap UB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KWIB\_UB\_HB\_NOM_t &= \sum_{n \in WIB_{ObjNr} \cap UB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KWIB\_UB\_TB\_NOM_t &= \sum_{n \in WIB_{ObjNr} \cap UB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KWIB\_NB\_HT\_NOM_t &= \sum_{n \in WIB_{ObjNr} \cap NB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KWIB\_NB\_HB\_NOM_t &= \sum_{n \in WIB_{ObjNr} \cap NB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KWIB\_NB\_TB\_NOM_t &= \sum_{n \in WIB_{ObjNr} \cap NB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n).
\end{aligned}
$$

For the public construction sector, we have

$$
\begin{aligned}
BEW\_KOEF\_NU\_HT\_NOM_t &= \sum_{n \in OEF_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KOEF\_UB\_HT\_NOM_t &= \sum_{n \in OEF_{ObjNr} \cap UB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KOEF\_UB\_HB\_NOM_t &= \sum_{n \in OEF_{ObjNr} \cap UB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KOEF\_UB\_TB\_NOM_t &= \sum_{n \in OEF_{ObjNr} \cap UB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KOEF\_NB\_HT\_NOM_t &= \sum_{n \in OEF_{ObjNr} \cap NB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KOEF\_NB\_HB\_NOM_t &= \sum_{n \in OEF_{ObjNr} \cap NB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BEW\_KOEF\_NB\_TB\_NOM_t &= \sum_{n \in OEF_{ObjNr} \cap NB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n).
\end{aligned}
$$

7.3 Aggregation of the number of appartments over the EF, MF or EM

This is done in

- "scripts/08Bau_Wohnungen2ANZEF_quarterly.r" with resulting time series
  - "data/SumBewAnzEF.rdata",
  - "data/SumGesAnzEF.rdata",
- "scripts/09Bau_Wohnungen2ANZEM_quarterly.r" with resulting time series
  - "data/SumBewAnzEM.rdata",
  - "data/SumGesAnzEM.rdata",
- "scripts/10Bau_Wohnungen2ANZMF_quarterly.r" with resulting time series
  - "data/SumBewAnzMF.rdata",
  - "data/SumGesAnzMF.rdata".

For this aggregation, we do not use any split of the middle and big projects. We use the original database "data/total.rdata".

Instead of adding the costs, we add this time the number of appartments (code "ANZ" which stands for "Anzahl Wohnungen") in the aggregation. As the other cases are very similar, we just give the results for the permit dates (BEW) and the single family house (EF).

$$BEW\_WANZ\_CH\_EF = \sum_{n \in EF_{ObjNr} \cap bew_t} AnzWhg(n)$$

and for a canton $KT \in \{ZH, BE, LU, UR, SZ, ...\}$,

$$BEW\_WANZ\_KT\_EF = \sum_{n \in EF_{ObjNr} \cap KT_{ObjNr} \cap bew_t} AnzWhg(n).$$

7.4 Aggregation of the costs over the EF, MF or EM

This is done in

- "scripts/11Bau_Wohnungen2BAUEF_quarterly.r" with resulting time series

– "data/SumBewBauEF.rdata",
– "data/SumGesBauEF.rdata",

- "scripts/12Bau_Wohnungen2BAUEM_quarterly.r" with resulting time series
  – "data/SumBewBauEM.rdata",
  – "data/SumGesBauEM.rdata",
- "scripts/13Bau_Wohnungen2BAUMF_quarterly.r" with resulting time series
  – "data/SumBewBauMF.rdata",
  – "data/SumGesBauMF.rdata".

This is similar to the previous subsection. The only difference is that the aggregation is made by adding the costs (code "KOS" which stands for "Kosten") instead of the number of appartments. Therefore, we just give the results for the permit dates (BEW) and the single family house (EF).

$$BEW\_WKOS\_CH\_EF = \sum_{n \in EF_{ObjNr} \cap bew_t} costs(n)$$

and for a canton $KT \in \{ZH, BE, LU, UR, SZ, ...\}$,

$$BEW\_WKOS\_KT\_EF = \sum_{n \in EF_{ObjNr} \cap KT_{ObjNr} \cap bew_t} costs(n).$$

7.5 Aggregation over the small, middle and big projects

This is done in

- "scripts/04Bau_auswertung2_k_quarterly.r" with resulting time series
  – "data/SumBewBauK.rdata",
  – "data/SumGesBauK.rdata",
- "scripts/05Bau_auswertung2_m_quarterly.r" with resulting time series
  – "data/SumBewBauM.rdata",
  – "data/SumGesBauG.rdata",
- "scripts/06Bau_auswertung2_g_quarterly.r" with resulting time series
  – "data/SumBewBauG.rdata",
  – "data/SumGesBauG.rdata".

In this subsection, we do not use any split of the middle and big projects. We use the original database "data/total.rdata".

We use the size classification and the classification presented in figure 1.

As the other cases are similar, we just present the results of the aggregation for the permit dates and the middle size projects. For the entire construction sector, we have

$$\begin{aligned}
BMP_KALL_NU_HT_NOM_t &= \sum_{n \in M_{ObjNr} \cap bew_t} costs(n), \\
BMP_KALL_NU_HB_NOM_t &= \sum_{n \in M_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BMP_KALL_NU_HT_NOM_t &= \sum_{n \in M_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n).
\end{aligned}$$

For the housing construction sector, we have

$$
\begin{aligned}
BMP\_KWBT\_NU\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KWBO\_NU\_HT\_NOM_t &= 1.8\% KWBT\_NU\_HT\_NOM_t, \\
BMP\_KWBP\_NU\_HT\_NOM_t &= 98.2\% KWBT\_NU\_HT\_NOM_t.
\end{aligned}
$$

For the housing transformation sector, we have

$$
\begin{aligned}
BMP\_KWBT\_UB\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WB_{ObjNr} \cap UB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KWBO\_UB\_HT\_NOM_t &= 1.8\% KWBT\_UB\_HT\_NOM_t, \\
BMP\_KWBP\_UB\_HT\_NOM_t &= 98.2\% KWBT\_UB\_HT\_NOM_t.
\end{aligned}
$$

For the housing building sector, we have

$$
\begin{aligned}
BMP\_KWBT\_NB\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WB_{ObjNr} \cap NB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KWBO\_NB\_HT\_NOM_t &= 1.8\% KWBT\_NB\_HT\_NOM_t, \\
BMP\_KWBP\_NB\_HT\_NOM_t &= 98.2\% KWBT\_NB\_HT\_NOM_t.
\end{aligned}
$$

For the privat construction sector, we have

$$
\begin{aligned}
BMP\_KPRI\_NU\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap PRI_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KPRI\_UB\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap PRI_{ObjNr} \cap UB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KPRI\_UB\_HB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap PRI_{ObjNr} \cap UB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KPRI\_UB\_TB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap PRI_{ObjNr} \cap UB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KPRI\_NB\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap PRI_{ObjNr} \cap NB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KPRI\_NB\_HB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap PRI_{ObjNr} \cap NB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KPRI\_NB\_TB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap PRI_{ObjNr} \cap NB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n).
\end{aligned}
$$

For the industrial construction sector, we have

$$
\begin{aligned}
BMP\_KWIB\_NU\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WIB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KWIB\_UB\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WIB_{ObjNr} \cap UB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KWIB\_UB\_HB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WIB_{ObjNr} \cap UB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KWIB\_UB\_TB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WIB_{ObjNr} \cap UB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KWIB\_NB\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WIB_{ObjNr} \cap NB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KWIB\_NB\_HB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WIB_{ObjNr} \cap NB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KWIB\_NB\_TB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap WIB_{ObjNr} \cap NB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n).
\end{aligned}
$$

For the public construction sector, we have

$$
\begin{aligned}
BMP\_KOEF\_NU\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap OEF_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KOEF\_UB\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap OEF_{ObjNr} \cap UB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KOEF\_UB\_HB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap OEF_{ObjNr} \cap UB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KOEF\_UB\_TB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap OEF_{ObjNr} \cap UB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KOEF\_NB\_HT\_NOM_t &= \sum_{n \in M_{ObjNr} \cap OEF_{ObjNr} \cap NB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KOEF\_NB\_HB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap OEF_{ObjNr} \cap NB_{ObjNr} \cap HB_{ObjNr} \cap bew_t} costs(n), \\
BMP\_KOEF\_NB\_TB\_NOM_t &= \sum_{n \in M_{ObjNr} \cap OEF_{ObjNr} \cap NB_{ObjNr} \cap TB_{ObjNr} \cap bew_t} costs(n).
\end{aligned}
$$

8 Baublatt indicator

8.1 Aggregation type and split

The split used for the calculation of the Baublatt indicator is the degressive split.

The aggregation is the one described in section 7.5, i.e. the aggregation over small, middle and big projects and over the classification presented in figure 1. The computed time series are therefore

$$\begin{aligned} &BKP\_KALL\_NU\_HT\_NOM_t, \\ &BKP\_KOEF\_NU\_HT\_NOM_t, \\ &\ldots \\ &BMP\_KALL\_NU\_HT\_NOM_t, \\ &\ldots \\ &BGP\_KALL\_NU\_HT\_NOM_t, \\ &\ldots \end{aligned}$$

8.2 Relevant Time Series

The relevant time series for the Baublatt indicator are

$$\begin{aligned} &BKP\_KALL\_NU\_HT\_NOM_t, \\ &BKP\_KWBT\_NU\_HT\_NOM_t, \\ &BMP\_KWBT\_NU\_HT\_NOM_t, \\ &BGP\_KWBT\_NU\_HT\_NOM_t. \end{aligned}$$

8.3 Shifts

The data from the Baublatt is a forecast of the sales of the construction companies. As a matter of fact the construction client has first to request a permit. Between the request and the permit (if any), there is a time lag[18] in order for the authority to check the construction. When the authority delivers the construction permit, the construction client has to find a construction company[19].

The requests outnumber the final constructions because not every request is accepted. Moreover, the time lag between the request and the construction is larger than the time lag between the permit and the construction, which leads to more uncertainty for the requests than for the permits. Therefore, taking the requests in the aggregation for the Baublatt indicator is perhaps not the best idea to capture the real construction sales. This is why only the permit dates are now used in the aggregation for the calculation of the Baublatt indicator.

The shift between the permit dates and the construction is fixed at 8 months for small projects ($costs \leq$ CHF 20M), 10 months for middle size projects (CHF 20M $< costs \leq$ CHF 100M) and 11 months for big projects (CHF 100M $\leq costs$). These numbers are the results of the optimization fit of the temporal disaggregation (see subsection 8.6).

Concretely, the shift has the following effect on the relevant time series.

$$\begin{aligned} BKP\_KALL\_NU\_HT\_NOM\_8_t &:= BKP\_KALL\_NU\_HT\_NOM\_8_{t+8}, \\ BKP\_KWBT\_NU\_HT\_NOM\_8_t &:= BKP\_KWBT\_NU\_HT\_NOM_{t+8}, \\ BMP\_KWBT\_NU\_HT\_NOM\_10_t &:= BMP\_KWBT\_NU\_HT\_NOM_{t+10}, \\ BGP\_KWBT\_NU\_HT\_NOM\_11_t &:= BGP\_KWBT\_NU\_HT\_NOM_{t+11}. \end{aligned}$$

[18] The time lag is comprised between 3 and 6 months mostly.

[19] Fundamentally, the construction permit is valid for approximately 2 years.

8.4 Two indicators, two time series

The two time series that will be used in the Baublatt indicator are the following ones.

$$
\begin{aligned}
&BKP\_KALL\_NU\_HT\_NOM\_8_t, \\
BKMGP\_10\_11\_KWBT\_NU\_HT\_NOM\_8_t \quad &:= \quad BKP\_KWBT\_NU\_HT\_NOM\_8_t \\
&\quad +BMP\_KWBT\_NU\_HT\_NOM\_10_t \\
&\quad +BGP\_KWBT\_NU\_HT\_NOM\_11_t.
\end{aligned}
$$

The first one will be used as an indicator for the entire construction sector and the second one will be used for the housing market.

8.5 Quarterly aggregation: "scripts/15Quarterly_Sampling_quarterly.r"

The relevant time series are aggregated quarterly. This means that a new variable *bewilligt_qua* is created ant that the time series is temporally aggregated as described in the following. The new values of the time series are

$$
\begin{aligned}
TS^{qua}_{bewilligt\_jahr,bewilligt\_qua=1} &= TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=1} + TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=2} \\
&\quad +TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=3} \\
TS^{qua}_{bewilligt\_jahr,bewilligt\_qua=2} &= TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=4} + TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=5} \\
&\quad +TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=6} \\
TS^{qua}_{bewilligt\_jahr,bewilligt\_qua=3} &= TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=7} + TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=8} \\
&\quad +TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=9} \\
TS^{qua}_{bewilligt\_jahr,bewilligt\_qua=4} &= TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=10} + TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=11} \\
&\quad +TS^{mon}_{bewilligt\_jahr,bewilligt\_mon=12}.
\end{aligned}
$$

8.6 Temporal disaggregation: "scripts/21Quartalisierung.r"

The yearly values from the BFS for the construction investments (entire construction sector and housing construction sector) are quarterly disaggregated with the time series obtained with the Baublatt database. This means that the yearly aggregated values of the disaggregated time series will be the values from the BFS and that the variations of the disaggregated time series will follow the variations of the Baublatt time series.

This temporal disaggregation was performed with different shifts. The best fit of the temporal disaggregation gave the best shifts for small, middle and big projects. These shifts were fixed and used afterwards.

Using the temporal disaggregation regression, we can obtain a natural forecast of the construction investments, which is made possible by the shifts found.

The method used for the temporal disaggregation is a method developed by **Chow and Lin (1971)**. We briefely present here the main ideas (for more detail see the file "BACKUP/documentation/temporaldisaggregation/ChowLin_1971.pdf").

Let us denote the yearly values for the entire construction sector[20] from the BFS by y_a (dim $N \times 1$[21]) and the quarterly values by y (dim $4N \times 1$), which we want to estimate. The quarterly values from the Baublatt are denoted by x (dim $4N \times 1$). Now suppose that quarterly y can be predicted using a multiple linear regression:

$$y = X\beta + u$$

where X is a $(4N \times 2)$ matrix of 2 predictor variables: a constant (a vector (1,...,1) of dim $4N \times 1$) and the time series $x = BKP\_KALL\_NU\_HT\_NOM\_8_t$ (dim $4N \times 1$), u is a $(4N \times 1)$ random vector with

[20] The analysis for the housing market is similar with the time series $BKMGP\_10\_11\_KWBT\_NU\_HT\_NOM\_8_t$.

[21] We denote by N the number of years from 1993 to the last date when BFS data are available.

zero mean and covariance matrix V (dim $4N \times 4N$). Using subscript a to denote annual figures, the previous equation can be converted to a regression of annual aggregates:

$$y_a = Cy = CX\beta + Cu = X_a\beta + u_a$$

where C is the matrix (dim $N \times 4N$) that converts $4N$ quarterly observations into N annual observations:

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & . & . & . & . & . & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & . & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & . & . & . & . & . & . & . & . & . & 1 & 1 & 1 & 1 \end{bmatrix}$$

We observe that

$$y_a = X_a\beta + u_a$$

can be estimated. This way, we obtain $\hat{\beta}$ and the predicted residuals $\hat{u}_a = y_a - X_a\hat{\beta}$. Calculus gives us the Chow-Lin best linear unbiased predictor (disaggregate) of y:

$$\hat{y} = X\hat{\beta} + VC'(CVC')^{-1}\hat{u}_a. \tag{1}$$

The first term on the RHS of this equation gives the predicted quarterly y based on observed quarterly X and estimated β from annual totals. What the second term does is to allocate annual residuals $\hat{u}_a$ to the four quarters of the year such that the annual sum of the interpolated values equal the observed value y_a. The variations in the Chow Lin procedures are about the choice of V. In the present analysis, it is supposed that u follows an **AR(1)** process:

$$u_t = \rho u_{t-1} + e_t$$

where e_t is a white noise and $|\rho| < 1$. In this case, V takes a form that can be calculated thanks to an estimation of ρ_a from the regression

$$(u_a)_t = \rho_a(u_a)_{t-1} + (e_a)_t$$

using the predicted errors u_a obtained previously. With the estimation of V, equation 1 gives the estimation of y, that we were looking for. In order to reproduce exactly the results from ECOTRIM (used in the previous version of the Baublatt indicator calculation), we use the same covariance matrix as in ECOTRIM.

In R, we use the package "tempdisagg", whose documentation is in directory "BACKUP/documentation/TemporalDisaggregation/R_package_tempdisagg.pdf". The package to be installed on the computer is in directory "BACKUP/R_packages/tempdisagg_0.13.zip".

8.7 Seasonal adjustment: "scripts/22X12.r"

8.7.1 Justification

There are strong seasonal variations in the time series and we want to get this effect out of them. For example, in december, almost no permit is given due to Christmas and the construction sector is also subject to high fluctuations (in winter, the construction activity is substantially reduced). That is why seasonal adjustment is performed.

8.7.2 Procedure

The procedure now used in R is the X12 with the multiplicative mode. It gives very similar results to the results obtained thanks to the previous procedure (previously done in FAME). The idea of the seasonal adjustment is the following one. We write the time series as

$$TS = T * S * I$$

where T is the trend, S the seasonal component and I the irregular component. Using low-pass filters, we want to have only the trend and the irregular component:

$$\frac{TS}{S} = T * I.$$

In R, we use the package "x12", whose documentation is in directory "BACKUP/documentation/x12/R_package_x12.pdf". The package to be installed on the computer is in directory "BACKUP/R_packages/x12_1.0-3.zip". Furthermore, the X12 program has to be installed. Its setup file is in directory "BACKUP/x12/winx12.zip".

9 Ideas to improve and develop the aggregation

The Baublatt indicator rose some doubts about its validity, when it predicted in the first and second quarter of 2012 that the construction sector would shrink during the rest of the year. This is why the Baublatt indicator was entirely programmed in R and documented in order to be understood by new people at the KOF and then improved.

9.1 Error treatment

We have to keep in mind that errors can be made by people who enter the data into the Baublatt database. We could develop further programs to search for errors and handle them. We list the errors we found:[22]

- The typical error is the one for which costs are 1000 times bigger than they should be.
- It can happen that people skip a column and write the date in the costs column. The date is then read as a simple number by the computer.
- The project can be given CRB codes that are not appropriate.[23]

There may be other errors that were not seen or handled. This can therefore be a good way to improve the quality of the Baublatt database.

9.2 Missing entries

There are omissions in the public construction sector: not all public constructions need an official permit. This is a major problem and it explains partly the difference of magnitude between the BFS time series and the Baublatt time series for the entire construction sector.[24] The number of public projects added stays now constant in the regression, which may be a bad assumption over time (increase of construction expenditure) and especially in periods of crises, when the state tries to stimulate the economic situation with additional expenditure.

9.3 Realization of construction permits

That the realization rate of what was permitted is constant over time is a strong assumption of the Baublatt indicator. That is why it has to be questioned. As a matter of fact, we can imagine for example that there may be an incentive for the construction client to renege on his/her right to construct if he/she thinks that the interest rate will rise in the mid-term. To account for this fact, we could include into the temporal disaggregation regression the interest rate as proposed in subsection 9.8 of this section. It is also plausible to think that this realization rate is correlated with the work in hand and the work already contracted but not delivered. For example, if the work in hand and the work already contracted but not delivered is at an all-time high, this means that the construction sector is in a boom and that the Baublatt indicator has to be adjusted (increased) in order to take into account this effect.

9.4 Lags between permit and realization

The assumption that the lags between permit and realization (8 months for small projects, 10 for middle projects and 11 for big projects) is constant over time and equal to these values may clearly be questioned.

[22] The two first errors produced large differences with the results obtained with SAS even on the entire construction indicator and therefore are of high significance.

[23] Looking at the database gives many examples.

[24] The public projects that are missing are essentially civil engineering projects. But the part of building construction projects is however not negligible.

9.5 Mixed projects

9.5.1 Attribution of CRB codes

We recall that the CRB codes are sorted increasingly for each project and then the attribution of CRB codes is made as described in section 4. This attribution has been made very primitive due to the presence of mixed projects in the Baublatt database. We take the following concrete example to illustrate this fact.

- $n = 99594615$
- $BaustKanton(n)$ =ZH
- $costs(n) = 80$
- $AnzWhg(n) = 100$
- $bewilligt\_jahr(n) = 2001$ and $bewilligt\_mon(n) = 12$
- $CRBCode(n) = 38$
- $CRBCode01(n) = 12$ (Mehrfamilienhäuser)
- $CRBCode02(n) = 38$ (Betriebs- und Gewerbebauten)
- $CRBCode03 = 61$ (Ladenbauten)
- $CRBCode04 = 64$ (Bürobauten, erhöhte Anforderungen)
- $CRBCode05 = 111$ (Restaurationsbetriebe)
- $CRBCode06 = 131$ (Einstellgaragen und Parkhäuser)
- $CRBCode07 = 166$ (Kino-, Diskothek- und Saalbauten)
- $CRBCode08 = 174$ (Carport)
- Kurzbeschreibung: Neubau Mehrfamilienhäuser mit Büros, Läden, Kinos, Bar und UN-Garage

As 100 appartments are built in Zurich, we would expect to have something like CHF 30M in Zurich for appartments. But with the actual CRB code attribution, this object will be only seen as a "Betriebs- und Gewerbebauten" and nothing will appear in the aggregation over the housing construction in canton Zurich.[25]

With this example, it is also clear that the attribution of CRB codes is very arbitrary because the costs for small CRB codes will be automatically too big due to the mixed project and the costs for high CRB codes will be automatically too small due to the mixed project as well.

What, we think for the moment, can be done to improve the CRB code attribution is summarized in the following points.

- We could split big projects in different projects with different costs, which would add to the initial costs, more systematically (at least do it systematically for mixed projects where there are appartments).
- Systematically, we should review manually the middle and big projects whose costs are bigger than CHF 20M because it is clear that big projects are normally mixed projects. In average, there would be monthly 15 projects to review, which can be done in 10 minutes with a good R program, which would handle these cases.

[25] For example, there are more than 9000 projects representing more than CHF 10 000M, which contain in the first two CRB codes one civil engineering project ("Tiefbau") and one building construction project ("Hochbau"). Therefore, we cannot conclude that the time series differentiating between civil engineering project and building construction project are right.

- We could develop a procedure to better capture the housing sector (other CRB codes than between 10 and 19 are housing construction projects)[26].

Generally, we can read further the database to get familiar with it and gain more ideas to deal with the very important issue of CRB code attribution.

In the present situation and due to the problems posed by the relatively arbitrary CRB code attribution, we think that the time series based on such classifications consist of values that can be hardly credible.

9.5.2 Split of mixed projects

The split of mixed projects could be more systematically done as explained in the previous subsection. Furthermore, allocating CHF 300 000 per appartment may not be an absolute value. There is no justification for this value. Moreover, the prices of appartments is continually rising over time and maintaining this value at CHF 300 000 is therefore very artificial and not convincing. The prices of appartments in Switzerland have substantially increased over the last 20 years.

9.6 Degressive split of middle and big projects

The rate of 7.5% in the degressive split of middle and big projects explained in section 6.2, which is now used for the Baublatt indicator, is a value that could be questioned in a new calculation of the indicator. As a matter of fact it can be easily understood that a different rate must be used for middle and big projects. For example, a project with costs CHF 20M lasts shorter than a project with costs CHF 2000M. Therefore, the rate for big projects must be smaller than for middle projects. Furthermore, there is no justification at all for the value $r = 7.5\%$. Finally, the rate may not be constant over time as technological progress makes the construction industry faster.

Another track would be to think of other ways to split middle and big projects.

9.7 Aggregating over other entities as Wohnbau or Gesamtbau

When the CRB code attribution is improved or for objects that are very often single nature projects, we will be able to credibly estimate other entities (aggregate over specific CRB codes).

9.8 Other factors in the temporal disaggregation regression

Explaining the investment data with the Baublatt series may not be sufficient. We could use here other time series, which have an influence on the construction investments and are not explained by the construction permits, and include them in the X matrix from section 8.6. An hint for this approach is that the residuals in the temporal disaggregation regression have blown up the two last years, which let think that the Baublatt time series are not sufficient anymore to explain the BFS investment data. We could use for example the following time series.[27]

- Capacity utilization
- The work in hand and the work already contracted
- The public construction investment[28]

9.9 Percentage of private and public housing construction projects

In the housing construction sector, it is assumed in section 7.1.3 that a fix percentage (1.8%) of the aggregated costs for the housing construction sector consists of public projects and another fix percentage (98.2%) of the aggregated costs for the housing construction sector consists of private projects. The same proportions are conserved for new construction projects and for transformation projects. This justification for this proportion is not clear and we do not see why it could not change over time. Therefore, this constitutes a track of understanding and improvement for the Baublatt aggregations.

[26] See for example CRB codes 171 to 174.
[27] These time series are obtained from the business tendency survey of the construction sector.
[28] This would account for the fact that the Baublatt database has very few public projects.

10 New estimation of the Baublatt indicator

10.1 Introduction

The first idea of the new estimation of the Baublatt indicator is to question the assumption that the lags between permit and construction are fixed (8 months for small projects, 10 months for middle projects and 11 months for big projects). Furthermore, the clear deviation of the Baublatt forecasts from the BFS forecasts has casted doubt on the assumption that the realization rate of projects is constant over time. These two assumptions that we wanted to question have lead us to the idea that we could use a distribution of the costs of a project after the delivery of its construction permit. This distribution would therefore take into account the lags between construction permit and realization and the time span during which the costs of the project are spread. Using different distributions, we would obtain different Baublatt indicators. Then, we choose the distribution that maximizes the fit of the regression of the Baublatt indicator obtained with the distribution on the BFS investment data. Finally, we adjust the values obtained in this way with the variable representing the work in hand and already contracted but not delivered. Therefore, this way, we tackle subsections 9.2, 9.3, 9.4, 9.6 and 9.8.

We concentrate mostly on the entire construction sector and on the housing construction sector in order to avoid trouble with the CRB codes and the mixed projects. We will just present in detail the procedure for the entire construction sector for simplicity. The procedure for the housing construction sector is rapidly described in subsection 10.7.

10.2 Spread of the costs after the delivery of a permit

10.2.1 Log-normal distribution

We concentrate on log-normal distributions because the computational power requested is reasonable (there are only two parameters μ and σ to vary) and because we think that this distribution may well explain the distribution of the costs after the delivery of a permit. We denote the log-normal distribution function with parameters μ and σ by $f_{\mu,\sigma}(\cdot)$ and the log-normal cumulative distribution function with parameters μ and σ by $F_{\mu,\sigma}(\cdot)$.

10.2.2 Discretization of the log-normal distribution

We have to discretize the log-normal distribution in order to have values for each month. For this we have to define a value denoted by *tolerance* that measures how much of the mass will be cut by the discretization in the tail. We take the values denoted by $\hat{w}_t$ with $t = 1, ..., I$ where

$$\begin{aligned} \hat{w}_t &:= \int_{t-1}^{t} f_{\mu,\sigma}(x)dx \\ &= F_{\mu,\sigma}(t) - F_{\mu,\sigma}(t-1) \end{aligned}$$

and I is the smallest integer such that

$$F_{\mu,\sigma}(I) \geq 1 - tolerance.$$

We denote by w_t for $t = 1, ..., I$ the discrete distribution, which is normalized to 1 in the following way:

$$w_t := \frac{\hat{w}_t}{\sum_{t=1}^{I} \hat{w}_t}.$$

It is clear that the mean and the standard deviation of the discrete distribution w_t will not be exactly identical to the mean and the standard deviation of the continuous distribution $f_{\mu,\sigma}(\cdot)$.

10.2.3 Principe

The spread of the costs using the discrete distribution is done mathematically in the following way.

- The first period, w_1 of the costs of the current period $costs_1$ are allocated: $costs_1^{spread} = w_1 costs_1$.

- In the second period, w_2 of the costs of the first period are allocated ($w_2 costs_1$) and w_1 of the costs of the current period are allocated ($w_1 costs_2$). Finally, the costs allocated to the second period amount to

$$costs_2^{spread} = w_1 costs_2 + w_2 costs_1.$$

- In the third period, w_3 of the costs of the first period are allocated ($w_3 costs_1$), w_2 of the costs of the second period are allocated ($w_2 costs_2$) and w_1 of the costs of the current period are allocated ($w_1 costs_3$). Finally the costs allocated to the second period amount to

$$costs_3^{spread} = w_1 costs_3 + w_2 costs_2 + w_3 costs_1.$$

- The procedure is repeated until the current period.

Therefore, the costs of the current period t is given in the following way:

$$costs_t^{spread} = \sum_{i=1}^{t} w_{t-i+1} costs_i.$$

10.2.4 Other distributions

Other distributions that could be used for this are the exponential distribution or the gamma distribution. We prefered the log-normal distribution because we found the form of this distribution more plausible.

10.2.5 Duration

We limit us to distributions with a duration of maximum 72 months in order to avoid better fits due to fewer values. Therefore, each distribution with duration greater than 72 months is systematically not considered in the optimization process described in the subsequent subsection.

10.3 Optimization

10.3.1 Three log-normal distributions: small (K), middle (M) and big (G) projects

We think that the small projects denoted by K (projects of costs $<$ CHF 20 million), the middle size projects denoted by M (projects of costs $<$ CHF 100 million and $>$ CHF 20 million) and the big projects denoted by G (projects of costs $>$ CHF 100 million) have different distributions of costs spread over time.

First, we find the Baublatt time series representing the aggregated costs of small, middle and big projects. We proceed an optimization in order to find the parameters of the three distributions. We use the fact that setting the mean denoted by E and standard deviation denoted by SD of the log-normal distribution determines completely the log-normal distribution. Indeed, we have the following expressions giving the parameters μ and σ in function of E and SD.[29]

$$\begin{aligned} \mu &= \ln(E) - \frac{1}{2}\ln\left(1 + \left(\frac{SD}{E}\right)^2\right), \\ \sigma &= \ln\left(1 + \left(\frac{SD}{E}\right)^2\right). \end{aligned}$$

Let P_K, P_M and P_G denote the set of possible means E and standard deviations SD for the different distributions. We take discrete sets in order to be able to loop over all coefficients. We denote for $S \in \{K, M, G\}$,

$$P_S = \{(\mu, \sigma) \in \{\mu_S^1, ..., \mu_S^{n_S}\} \times \{\sigma_S^1, ..., \sigma_S^{m_S}\}\}.$$

[29]Using the mean is not very relevant. An idea to be developed would be to use the median instead of the mean.

We use linear specifications for the sets $\{\mu_S^1, ..., \mu_S^{n_S}\}$ and $\{\sigma_S^1, ..., \sigma_S^{m_S}\}$:

$$\mu_S^i = (i-1)\frac{\mu_S^{n_S} - \mu_S^1}{n_S - 1} + \mu_S^1$$
$$\sigma_S^i = (i-1)\frac{\sigma_S^{m_S} - \sigma_S^1}{m_S - 1} + \sigma_S^1.$$

Looping over all coefficients is an algorithm with complexity $O(m_K * m_M * m_G * n_K * n_M * n_G)$. If these numbers are equal to n, we find a complexity of $O(n^6)$ which can very big even for small values of n ($30^6 = 729$ million for example).

10.3.2 The regression

Using these distributions, we can spread the Baublatt values of the aggregated time series of the small, middle and big projects over time as explained in subsection 10.2.3. We obtain time series denoted by TSS_S for $S \in \{K, M, G\}$. We use these time series as explanatory variables in the following regression:

$$BFS\_DATA = \alpha + \beta_K TSS_K + \beta_M TSS_M + \beta_G TSS_G + u$$

where BFS_DATA is the BFS time series for construction investments. The coefficients β_S represent the realization rates of the different projects for each size. The coefficient α contains the public projects that do not need official construction permits. As this coefficient α is constant over time, the variations over time of the costs of public projects that are not included in the Baublatt database is reported in the β coefficients and in the error term.

10.3.3 Bayesian method for Model Comparison

The regressions obtained with different means and standard deviations are then compared assuming that means and standard deviations constitute models. The comparison of the different models is made as in the book of Gary Koop: Bayesian Econometrics (2003) calculating marginal likelihood and odds ratios.[30]

The algorithm for finding the best model (E_{max}, SD_{max}) is the following one. We denote by $lmarglik(E_n, SD_n)$ the log-marginal likelihood of the model (E_n, SD_n).

1. Calculate the log-marginal likelihood $lmarglik(E_n, SD_n)$ for the model (E_n, SD_n).
2. Compare $lmarglik(E_n, SD_n)$ and $lmarglik(E_{max}, SD_{max})$.
   - If $lmarglik(E_n, SD_n) > lmarglik(E_{max}, SD_{max})$, redefine

     $$\begin{aligned} E_{max} &:= E_n \\ SD_{max} &:= SD_n \\ lmarglik(E_{max}, SD_{max}) &:= lmarglik(E_n, SD_n) \end{aligned}$$

     and repeat the algorithm with the next model (E_{n+1}, SD_{n+1}).
   - Otherwise, repeat the algorithm with the next model (E_{n+1}, SD_{n+1}).

10.4 Vorlauf

We define the time span of the forecast called "Vorlauf" (difference between the end date of forecast and the current date)[31] as being the largest integer $Vorlauf$ such that

$$\sum_{i=1}^{Vorlauf} w_i \leq 0.5$$

when this integer is more than 8 months. If this integer is less than 8 months, we redefine $Vorlauf := 8$ months. This sum, $\sum_{i=1}^{Vorlauf} w_i$, is the mass of values that are not known for the last observation of the forecast.

[30] Part to be completed. Include the plot with the noninformative prior.
[31] This is also the number of forecasted values.

10.5 Forecast

Using the Baublatt aggregated costs values until the current date, we can have spread values until the current date + I months[32]. We limit us at $Vorlauf$ months as explained in subsection 10.4.

The problem is now that the mass of the values obtained in the forecast is incomplete. We have to adjust for that. We do it as explained in the following. Let us denote by $costs$ the time series of the Baublatt aggregated values without use of the distribution (no spread) and by $costs^{spread}$ the time series of the Baublatt aggregated values with use of the distribution. Let us denote by $costs^{spread}_{FORECAST}$ the time series of the Baublatt aggregated values with use of the distribution and with complete forecast values ($costs^{spread}$ is identical to $costs^{spread}_{FORECAST}$ until the current date and the values of $costs^{spread}_{FORECAST}$ afterwards are what we think being a good forecast). Let us denote by t the current date expressed in months. Notice that we have values for the $costs$ time series until t and values for the $costs^{spread}$ time series until $t + Vorlauf$.

- For the months $t + 1$ just after the current month, the mass of values that is missing is equal to $w_1 costs_{t+1}$. But as we do not know $costs_{t+1}$, we adjust by assuming that w_1 of the current date value is missing. Therefore, as we think that $costs^{spread} = (1 - w_1) costs^{spread}_{FORECAST}$, we have

$$costs^{spread}_{FORECAST} = \frac{costs^{spread}}{1 - w_1}.$$

- For the months $t + n$ where $n \in [\![1, Vorlauf]\!]$, the mass of values that is missing is equal to $w_1 costs_{t+n} + w_2 costs_{t+n-1} + ... + w_n costs_{t+1}$. But as we do not know $costs_{t+n}$ for $n \in [\![1, Vorlauf]\!]$, we adjust by assuming that $w_1 + w_2 + ... + w_n$ of the current date value is missing. Therefore, as we think that $costs^{spread} = (1 - w_1 - w_2 - ... - w_n) costs^{spread}_{FORECAST}$, we have

$$costs^{spread}_{FORECAST} = \frac{costs^{spread}}{1 - w_1 - w_2 - ... - w_n}.$$

10.6 Adjustment of the fitted values with the work in hand and already contracted

10.6.1 Transformation of the variable representing the work in hand and already contracted

The variable representing the work in hand and already contracted ("Auftragsbestand") is a variable denoted by z from a survey aggregation. This variable is large when the situation is good and small when the situation is bad. As the values are expressed monthly and we want to use yearly values, we first assume that the yearly value in year j is the average of the 12 monthly values in year j.

We want to convert the work in hand and already contracted into a variable that will contain corrective information for the Baublatt fitted values. For this, we need a transformation of the variable. We take an affine transformation of the type $f(z_{old}) = a z_{old} + b = z_{new}$ with the following two restrictions that uniquely determines a and b for a given parameter par, which measures the intensity of the corrections of z.

$$\begin{aligned} \mathbb{E}[z_{new}] = 1 &= a\mathbb{E}[z_{old}] + b, \\ par = \max(z_{new}) - \min(z_{new}) &= a(\max(z_{old}) - \min(z_{old})). \end{aligned}$$

The first condition means that in average, the new variable has no influence on the Baublatt fitted values. The second condition is a measure of the intensity of the corrections of z. We denote now by z the transformed time series.

10.6.2 Principe

The principe of the adjustment of the fitted values with the work in hand and already contracted is based on the idea developed in subsection 9.3. We want to convert this time series into a time series

[32] Recall that I is the duration of the discrete distribution.

that can be used to adjust the realization rate to the current economic situation. The Baublatt fitted values found in subsections 10.3.2 and 10.6.3 are multiplied by z giving a new time series representing "Auftragsbestand"-adjusted Baublatt fitted values. The result is denoted by $T\hat{S}S^{adjusted}$ and given in the following equation.

$$T\hat{S}S^{adjusted} = \hat{\alpha} + z_t\hat{\beta}_K TSS_{K,t} + z_t\hat{\beta}_M TSS_{M,t} + z_t\hat{\beta}_G TSS_{G,t}.$$

10.6.3 The three scenarios

We do not have a value of z for the year of the forecast. Therefore, we must generate a value that could be credible. For this, we have developed three scenarios: a good, bad and neutral case.

The good case is the case for which the value of the work in hand and already contracted has increased. The growth is taken as the average of the positive growths observed since the beginning of the time series. Then the value is transformed as previously described. Its transformation is denoted by $z_{FORECAST}$. It is multiplied by the Baublatt forecast $T\hat{S}S$ in order to find the "Auftragsbestand"-adjusted Baublatt forecast $T\hat{S}S^{adjusted}_{FORECAST}$.

$$T\hat{S}S^{adjusted}_{FORECAST} = T\hat{S}S z_{FORECAST}.$$

The bad case corresponds to the case when the work in hand and already contracted has decreased by the rate equal to the average of the negative growth rates. The neutral case assumes a growth rate equal to 0.

10.7 The housing construction sector

We can do exactly the same thing for the housing construction sector separating the single family houses from the multiple family houses ("Einfamilienhäuser" from the "Mehrfamilienhäuser"). In this case, we have to take the BFS time series for the housing construction sector and the "Auftragsbestand" time series for the building construction sector. The analysis runs otherwise in the same way.

11 Appendix

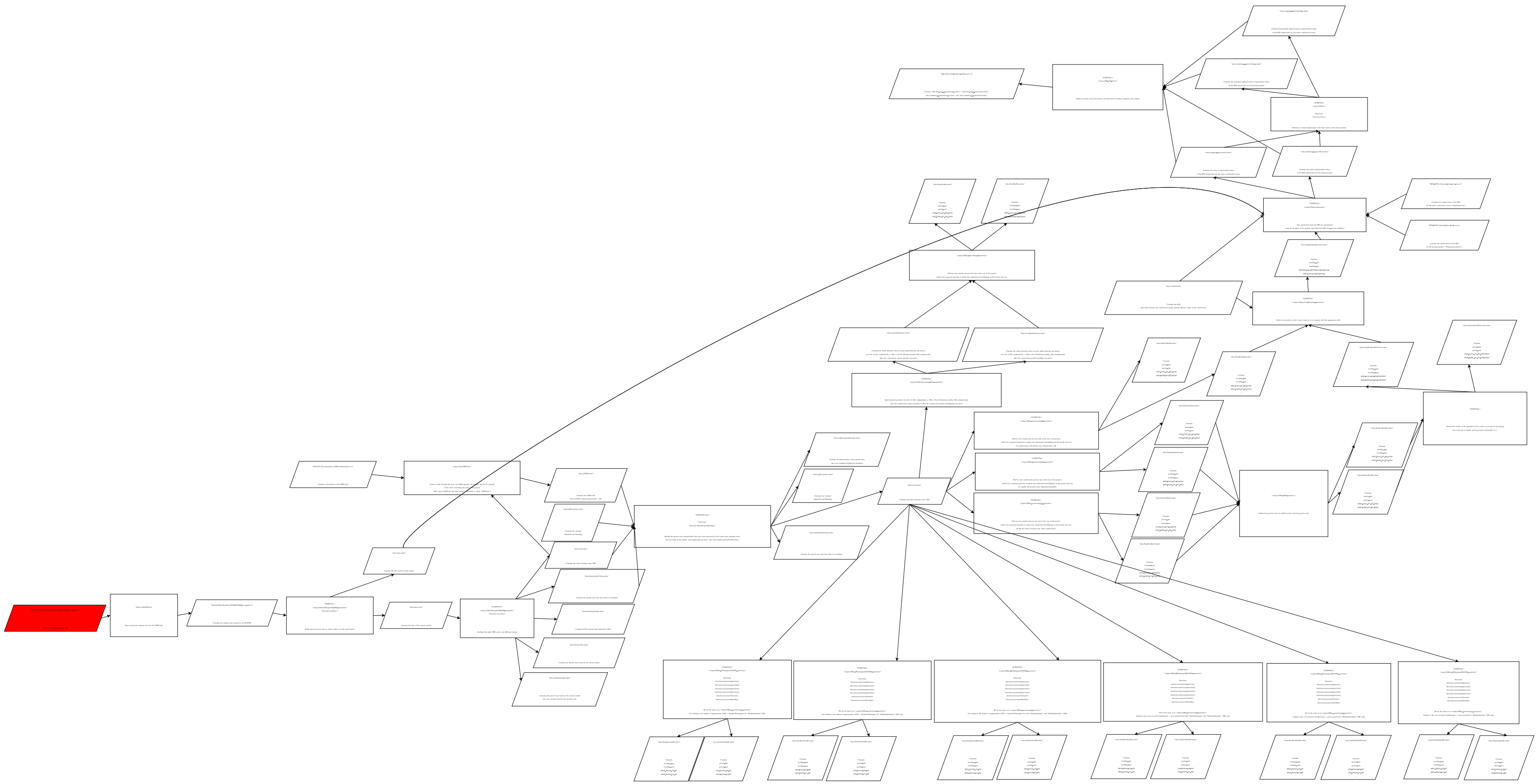

Figure 2: Data flow of the Baublatt indicator: see files Datenfluss.eps or Datenfluss.jpeg

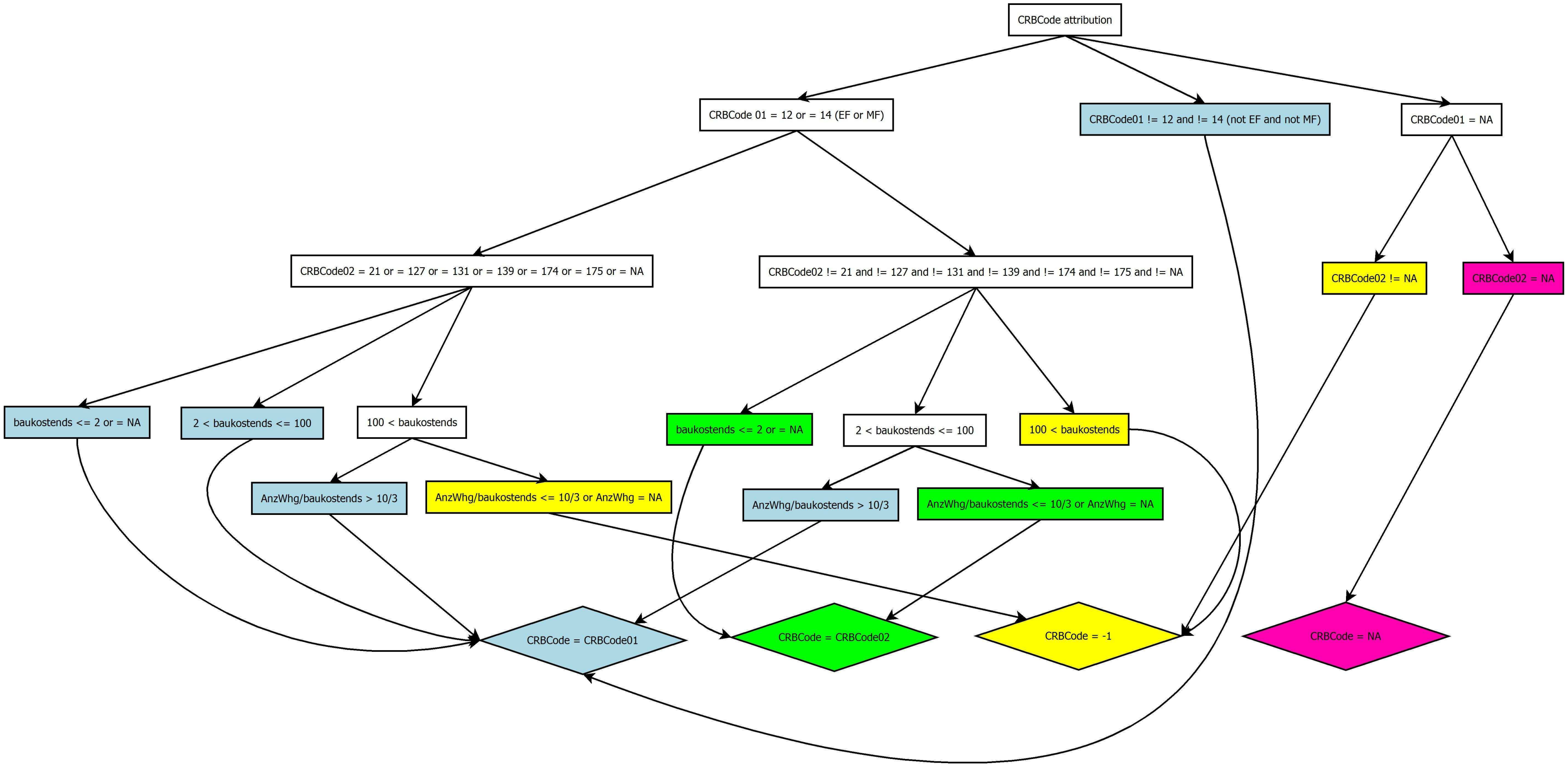

Figure 3: CRB code attribution: see files crbcodes.eps or crbcodes.jpeg

12 List of Notation

Symbol	***Meaning***
n	Object number of a project
$ObjektNr$	Set of the object numbers n of all projects
$BaustKanton(n)$	The canton where the project n is taking place
$BauartCode(n)$	The code for the sort of construction project n
$costs(n)$	The costs of the project n
$AnzWhg(n)$	The number of appartments of the project n
$gesuch\_mon(n)$	The month of request date for project n
$gesuch\_jahr(n)$	The year of request date for project n
ges_t	The set of all projects n whose request date is $(gesuch\_mon, gesuch\_jahr) = t$
$bewilligt\_mon(n)$	The month of permit date for project n
$bewilligt\_jahr(n)$	The year of permit date for project n
bew_t	The set of all projects n whose permit date is $(bewilligt\_mon, bewilligt\_jahr) = t$
$CRBCode01(n)$	The first CRB code of project n
$CRBCode02(n)$	The second CRB code of project n
$CRBCode(n)$	The CRB code of project n
NB	The set of CRB codes that correspond to new construction projects. $NB = \{1, 2\}$
UB	The set of CRB codes that correspond to transformation projects. $UB = \{3, 4, 5\}$
HB	The set of CRB codes that correspond to building construction projects. $HB = \llbracket 10, 199 \rrbracket$
TB	The set of CRB codes that correspond to civil engineering projects. $TB = \llbracket 200, 741 \rrbracket$
HT	The set of CRB codes that correspond to either building construction projects or civil engineering projects. $HT = \llbracket 10, 741 \rrbracket$
OEF	The set of CRB codes that correspond to public construction projects $OEF = \llbracket 20, 29\rrbracket \cup \llbracket 51, 53\rrbracket \cup \llbracket 56, 58\rrbracket \cup \llbracket 66, 79\rrbracket \cup \llbracket 83, 84\rrbracket \cup \llbracket 90, 109\rrbracket \cup \llbracket 130, 138\rrbracket \cup \llbracket 141, 159\rrbracket \cup \llbracket 180, 263\rrbracket \cup \llbracket 265, 312\rrbracket \cup \{314\} \cup \llbracket 316, 462\rrbracket \cup \llbracket 464, 610\rrbracket \cup \llbracket 614, 741\rrbracket$
PRI	The set of CRB codes that correspond to private construction projects. $PRI := \llbracket 31, 49\rrbracket \cup \llbracket 54, 55\rrbracket \cup \llbracket 59, 65\rrbracket \cup \{82\} \cup \llbracket 85, 89\rrbracket \cup \llbracket 110, 129\rrbracket \cup \{139\} \cup \llbracket 161, 179\rrbracket \cup \{264, 313, 315, 463\} \cup \llbracket 611, 614\rrbracket$
WBT	The set of CRB codes that correspond to housing construction projects. $WBT = \llbracket 9, 20\rrbracket$
WIB	The set of CRB codes that correspond to industrial construction project $WIB = \llbracket 31, 39\rrbracket \cup \llbracket 61, 65\rrbracket \cup \{47, 54, 59\} \cup \llbracket 111, 118\rrbracket \cup \{161\} \cup \llbracket 166, 168\rrbracket$
$DIFF$	The set of CRB codes that correspond to not classified projects. $DIFF = \{0, NA\}$
KT	A canton in Switzerland
EF	The set of CRB codes that correspond to single family houses. $EF = \{14, 19\}$
MF	The set of CRB codes that correspond to multiple family houses. $MF = \{10, 11, 12, 13, 15\}$
EM	The set of CRB codes that correspond to single family houses or multiple family houses
K	The set of costs that correspond to small projects. $K = [0, 20]$
M	The set of costs that correspond to middle size projects. $M =]20, 100]$
G	The set of costs that correspond to big projects. $G =]100, +\infty]$
TS	A time series
T	The trend component of a time series
S	The seasonal variations of a time series
I	The irregular component of a time series
N	The number of years form 1993 to the last date when BFS data are available
y_a	The yearly values for the entire construction sector from the BFS
y	The quarterly values for the entire construction sector from the BFS, which we want to estimate in the Chow-Lin procedure
x	The quarterly values from the Baublatt
x_a	The annual values from the Baublatt
X	The matrix of predictor variables. $X = ((1)_{4N\times 1}, x)$
u	A random vector with zero mean and covariance matrix V used in the regression of quarterly data
V	Covariance matrix of the random vector u

u_a	The residualy of the regression of annual data
C	The matrix that converts quarterly observations into annual observations
β	The parameter of the Chow-Lin regression
ρ	The parameter value in the autoregression of the residuals u_t
e_t	A white noise in the autoregression of the residuals u_t
$(e_a)_t$	A white noise in the regression of the residuals $(u_a)_t$
ρ_a	The parameter value in the autoregression of the residuals $(u_a)_t$
r	Rate in the degressive split of middle and big projects

13 Code signification

Symbol	***Meaning***
BEW	“Bewilligung” means permit
GES	“Gesuch” means request
NB	“Neubau” means new construction
UB	“Umbau” means transformation project
NU	“Neubau, Umbau” means new construction or transformation project
HB	“Hochbau” means building construction
TB	“Tiefbau” means civil engineering project
HT	“Hochbau, Tiefbau” means building construction or civil engineering project
DIFF	“Difference” means project with no CRB code classification
ALL	“ALL” means all projects independently of the sectorial classification
OEF	“öffentlicher Bau” means public construction
PRI	“privat” means private
WIB	“Wirtschaftsbau” means industrial construction
WBT	“Wohnbauten” means housing construction
WBO	“Wohnbauten, öffentlich” means public housing construction
WBP	“Wohnbauten, privat” means private housing construction
EF	“Einfamilienhäuser” means single family houses
MF	“Mehrfamilienhäuser” means multiple family houses
EM	“Einfamilienhäuser, Mehrfamilienhäuser” means single or multiple family houses
KP	“kleines Projekt” means small project
MP	“mittleres Projekt” means middle size project
GP	“grosses Projekt” means big size project
ANZ	“Anzahl Wohnungen” means number of appartments
KOS	“Kosten” means costs
ZH	Zürich
BE	Bern
LU	Lucerne
UR	Uri
SZ	Schwyz
OW	Obwalden
NW	Nidwalden
GL	Glarus
ZG	Zug
FR	Fribourg
SO	Solothurn
BS	Basel-Stadt
BL	Basel-Landschaft
SH	Schaffhausen
AI	Appenzell Innerrhoden
AR	Appenzell Ausserrhoden
SG	Saint Gallen
GR	Graubünden

AG	Aargau
TG	Thurgau
TI	Ticino
VD	Vaud
VS	Valais
NE	Neuchâtel
GE	Geneva
JU	Jura