Test and update of regionalised ESIF payments 1989-2018

Final report

CCI Contract: 2018CE16BAT115

Open Evidence
May – 2020

The study was executed by Samuele Lo Piano of Open Evidence, Barcelona.

Mr Roman Römisch, Vienna Institute for International Economic Studies, provided support and guidance on the earlier regionalisation of data from 2007-2013 to allow additional data from 2016-2018 to be added.

Daniele Vidoni and John Walsh of the European Commission provided relevant information and advice on the structure of the programmes and of the data bases.

EUROPEAN COMMISSION

Directorate-General for Regional and Urban Policy
Directorate B — Policy Coordination
Unit B.2 — Evaluation and European Semester

Contact: John WALSH

E-mail: REGIO-EVAL@ec.europa.eu

European Commission
B-1049 Brussels

Test and update of regionalised ESIF payments 1989-2018

Final report

CCI Contract: 2018CE16BAT115

Directorate-General for Regional and Urban Policy

2020 EN

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Manuscript completed in May 2020

Luxembourg: Publications Office of the European Union, 2020

ISBN: 978-92-76-19200-8
doi: 10.2776/569313

TABLE OF CONTENTS

1. Abstract

The modelled expenditures elaborated in the context of the previous tender contract[1] have been tested against the incurred member-state expenditures reported by the beneficiaries to the member-state agencies. Two Member States (MSs) were inquired - Italy (IT) and Czech Republic (CZ) - as regards the expenditures incurred under the funding schemes Cohesion Fund (CF) and European Regional Development Fund (ERDF) along the programming period 2007-2013.

Mismatches in the total amounts have been documented as Member-state Authorities (MAs) database also included uncertifed expenditure. These exceeding figures were taken out so as to normalise the modelled and incurred expenditures. The impact of the modelling assumptions on the yearly difference between these figures has been tested by sensitivity analysis, which revealed a primary role for the normalisation assumption. The quality of information of the MAs database should be improved so as to allow for a more robust benchmarking.

This result is also confirmed when examining individual funding programmes for IT in task 2.c. We have then proceeded to estimate expenditures at the NUTS2 granularity level for the programming period 2014-2020 in task 2.c as per the specifics of the model elaborated in the previous tender contract. Initial mismatches were documented in the *Stata* input file for the closure of the European Commission (EC) payments remitted in the programming period 2007-2013 in the context of task 2.a. These were finally overcome, and the regionalisation was finally accomplished under a new n+2 rule for this programming period in task 2.c. Conversely, in task 2.b we have successfully completed the regionalisation of the payments for the following programming period 2014-2020.

Finally, we have detailed in task 3 the granularity level at which the EC payments were remitted to dependent on the funding scheme and member state in the programming period 2014-2020.

Die im Rahmen des vorherigen Ausschreibungsvertrags[1] erarbeiteten modellierten Ausgaben wurden anhand der angefallenen Ausgaben überprüft, die den Agenturen der begünstigten Mitgliedstaaten gemeldet wurden. Zwei Mitgliedstaaten - Italien IT und die Tschechische Republik CZ - wurden im Programmplanungszeitraum 2007-2013 zu den Ausgaben im Rahmen der Finanzierungssysteme Kohäsionsfonds CF und Europäischer Fonds für regionale Entwicklung ERDF untersucht.

Aufgrund der Tatsache, dass die Datenbank der Behörden der Mitgliedstaaten auch nicht zertifizierte Ausgaben enthielt, wurden Abweichungen in den Gesamtbeträgen dokumentiert. Es wurde ein Normalisierungsverfahren angewendet, um den Vergleich zwischen den beiden Datenbanken für modellierte und angefallene Ausgaben zu ermöglichen, indem die übersteigenden Zahlen aus dem zuletzt genannten herausgenommen wurden.

Die Auswirkung der Modellannahmen auf die gemeldete jährliche Differenz zwischen den Zahlen der Datenbanken wurde durch Sensitivitätsanalyse getestet, welche eine primäre Rolle der oben genannten Normalisierung aufdeckte.

[1] CCI 2016CE16BAT081 – Regionalisation of ESIF payments 1989-2015 - https://op.europa.eu/en/publication-detail/-/publication/cd2133d5-1d04-11e8-ac73-01aa75ed71a1/language-en/format-PDF/source-68135162

Die Qualität der Informationen in der MA-Datenbank sollte verbessert werden, um eine wirksame Überarbeitung des Modells zu ermöglichen und den Unterschied zu den dokumentierten Ausgabenmustern zu verringern. Bei der Prüfung der einzelnen Förderprogramme in Task 2.c auf IT wurde dieses Ergebnis bestätigt. Anschließend wurden die Ausgaben auf der NUTS2-Granularitätsstufe für den Programmplanungszeitraum 2014-2020 in Task

2.b gemäß den Einzelheiten des im vorherigen Ausschreibungsvertrag erarbeiteten Modells geschätzt. In Task 2.a wurden Inkongruenzen auch in der Stata-Eingabedatei für den Abschluss der Zahlungen der Europäischen Kommission EC dokumentiert, die im Programmplanungszeitraum 2007-2013 überwiesen wurden. Diese Inkongruenzen waren schließlich überwunden, und die Regionalisierung wurde schließlich unter einer neuen n+2-Regel für diesen Programmplanungszeitraum in Task 2.c durchgeführt.

Les dépenses modélisées établies dans le contexte du précédent contrat d'appel d'offres[1] ont été comparées aux dépenses engagées par les Etats membres et déclarées par les bénéficiaires aux agences des Etats membres. Deux Etats membres – l'Italie (IT) et la République tchèque (CZ) - ont été interrogés sur les dépenses engagées´ au titre des régimes de financement des Fonds de cohésion (CF) et des Fonds européens de développement régional (ERDF) au cours de la période de programmation 2007-2013.

Des asymétries dans les montants totaux ont été relevées, la base de données des autorités des Etats membres (MS) comprenant également des dépenses´ non certifiées. Une procédure de standardisation a été mise en place afin de permettre la comparaison, dans les deux bases de données, des dépenses modélisées et des dépenses encourues, en retirant les chiffres excédentaires de ces dernières. L'impact des hypothèses de modélisation sur la différence annuelle rapportée entre les chiffres des bases de données a été testé par le biais d'une analyse de sensibilité, qui a montré le rôle primordial de la standardisation susmentionnée. La qualité des informations de la base de données des Etats membres devrait être améliorée pour permettre une révision efficace du modèle afin de réduire la différence par rapport aux modèles de dépenses documentés.

Ce résultat est également confirmé lorsque les programmes de financement uniques sont examinés pour l'Italie dans la Tache 2.c. Nous avons ensuite procédé `a l'estimation des dépenses au niveau de granularité NUTS2 pour la période de programmation 2014-2020 dans la Tache 2.c, conformément aux spécificités du modèle élaboré dans l'appel d'offres précédent. Des incompatibilités ont également été documentées dans l'entrée Stata pour la clôture des paiements de la Commission européenne (CE) versés au cours de la période de programmation 2007-2013 dans le cadre de la Tache 2.a. Ces incompatibilités ont finalement été surmontées et la régionalisation a finalement été réalisée sous une nouvelle règle n+2 pour cette période de programmation dans la Tache 2.c. A` l'inverse, dans la Tache 2.b, nous avons achevé avec succès la régionalisation des paiements pour la période de programmation 2014-2020.

Enfin, nous avons détaillé dans la Tache 3, le niveau de granularité auquel les paiements communautaires ont été versés au cours de la période de programmation 2014-2020, en fonction du régime de financement et de l'Etat´ membre.

2. INTRODUCTION - THE ISSUE AT HAND

A general model to infer expenditures from the reimbursement pattern of the European Structural and Investment Funds was proposed in the previous tender contract. The aim of this new contract is to test its robustness against the actual expenditure reported by MAs. To this end, comparisons have been carried out against the Czech and Italian authorities' data bases as regards the ERDF and CF over the programming period 2007-2013 (task 1.a). This analysis could result in plausible figures, partial or major mismatches. On the basis of this outcome, the modelling of expenditures estimates elaborated in the context of the previous tender contract may be readjusted accordingly as per task 1.b.

In task 2, we disaggregate all the payments reported by the EC to MAs at the NUTS2 granularity level. Some of these payments were reported in the EC database at variable levels of granularity: Some items were disaggregated at the NUTS2 or NUTS3 level, while others only at the NUTS1 or even NUTS0 level (only at the member-state level).

Finally, in task 3, we report on the granularity level of disaggregation of the EC payments as regards the programming period 2014-2020 across funding schemes and MSs.

3. TASK1 - TAXONOMY OF CASES

Let us label a generic payment remitted by the European Commission to MAs with R_{EC}. R_{MS} is the annual history of expenditures reported by projects to the MA and E is our modelled expenditure. As per equation (1), the sum of the yearly figures over the entire programming period must be equal so as to assure consistency. r_{ECp}^{i} identifies the EC payment remitted to region p over the year i.

$$\sum_{i=1}^{k+m} r_{EC_p^i} = \sum_{i=1}^{k} r_{MS_p^i} = \sum_{i=1}^{k} e_p^i \tag{1}$$

The EC payments are spread over $k + m$ years, while expenditures are not eligible after the k_{th} year of the programming period.

$$r_{EC_p^1} \leq e_p^1 \tag{2}$$

A general rule is that **reimbursements are always following expenses**, on the basis of which we can try to define time patterns and yearly expenditure figures. A situation of cumulative modelled expenditures smaller than the MAs expenditures would be blatantly wrong and would require the amendment of the model. Let us for instance take the first year of the programming period, the relation between the yearly figures for a generic region p must be as per equation (2).

$$r_{EC,p_1} \leq r_{MS,p_1} \tag{3}$$

Furthermore, r_{MSp}^{1} anticipates rECpi and the relation in equation 3 between these two quantities must hold as well.

$$r_{MS_p^1} \approx e_p^1 \tag{4}$$

Finally, when it comes to the relation between $r_{MS,p}^{1}$ and e_p^1, this latter needs to be very similar to the one recorded in the annual history of expenditures reported to MAs. Significant differences in magnitude would not be plausible. $r_{MS,p}^{1}$ also accounts for invoices sent by local authorities (e.g. municipalities). This condition assures the time lag with e_p^1 is minimal, below the yearly granularity at which these figures have been produced. Therefore, the relation between the two figures should be as per equation 4.

$$\sum_{i=1}^{l} r_{EC,p_i} \leq \sum_{i=1}^{l} r_{MS,p_i} \approx \sum_{i=1}^{l} e_p^i \tag{5}$$

We can also extend these relations to the lth year, with l ≤ k as per equation (5):

$$\mu_{IT} = \sum_{i=1}^{k} \mid r_{IT_p^i} - E_{IT_p^i} \mid \tag{6}$$

Therefore, our workflow will imply firstly to verify the closure relation as per equation 1. MSs figures may not be certified, hence we may need to normalise these as to assure the closure with the EC figures and the modelled expenditures. Once made consistent, the compliance with equation 5 will be tested to flag up potential discrepancies and how one could work towards reduce these discrepancies.

3.1. Task 1.a - Consistency across the MSs databases and the modelled expenditure

For the sake of comparability with the EC figures, the databases have been capped to the year 2016. The closure between the modelled expenditure and the EU payments was verified in the previous tender contract. However, discrepancies have emerged against the Italian database in the context of this study due to the uncertified expenditures included in the latter. The Italian database's figures are 15% higher and this required a normalisation. We have tested how the assumption of a normalisation over a variable number of years impacts on the comparison with the modelled expenditure database. The maximum cumulative distance μ (Figure 1) emerges when the normalisation is performed by taking out the excess figure from the last eligible year only (2016) due to the very large gap in this last year. In this setting, MSs figure for the year 2016 would be negative. Normalising over more years reduces the gap in the cumulative distance from 0.58 to 0.43 when the exceeding payment is taken out over the years (2015-2016) or (2009-2015) in equal yearly proportions.

The cumulative figure of the IT expenditures usually exceeds the modelled expenditure when the normalisation is over the last years of the programming period only (Figure 1 and 2). Conversely, performing the normalisation by taking out the exceeding figure over more years makes this trend more variable, with the two figures moving closely to each other (Figure 3).

CZ shows a comparable trend, with smaller mismatches between the estimated expenditures and the reported MS expenditures. The exchange rate adopted for the comparison is a further factor of uncertainty in this case - possible options are a constant or a yearly variable exchange rate against *euro*. One can see in Figures 4 and 5 below (green vs. orange plot) the impact of this assumption is minimal.

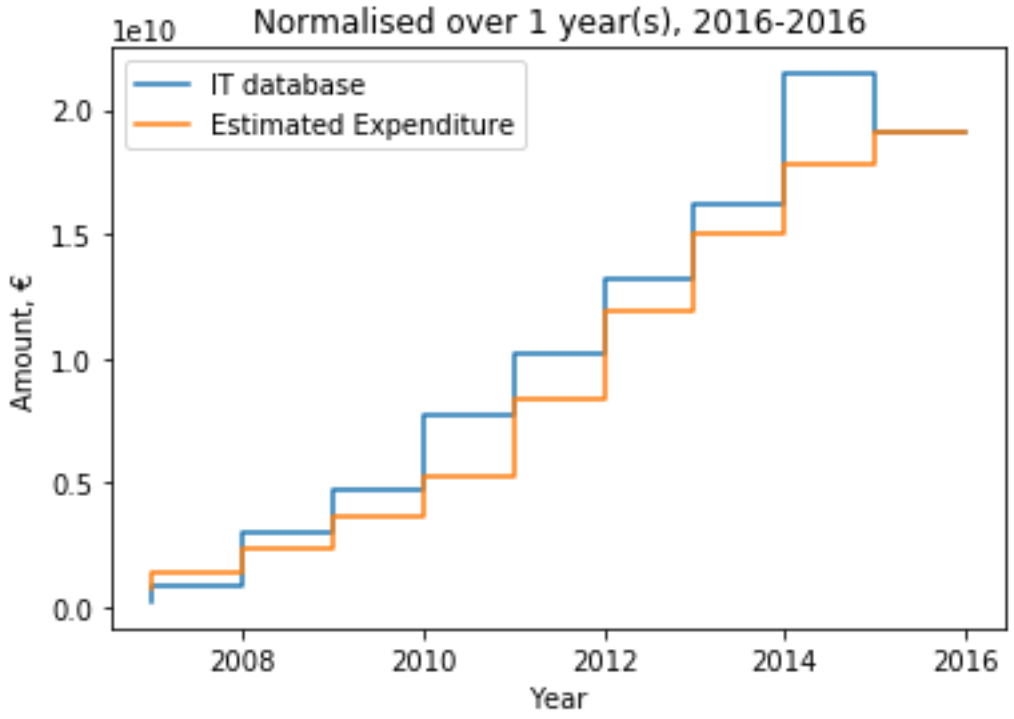

Fig. 1: 2007-2013 ERDF Italian payment dataset against modelled expenditure, normalisation with exceeding figures taken out from 2016

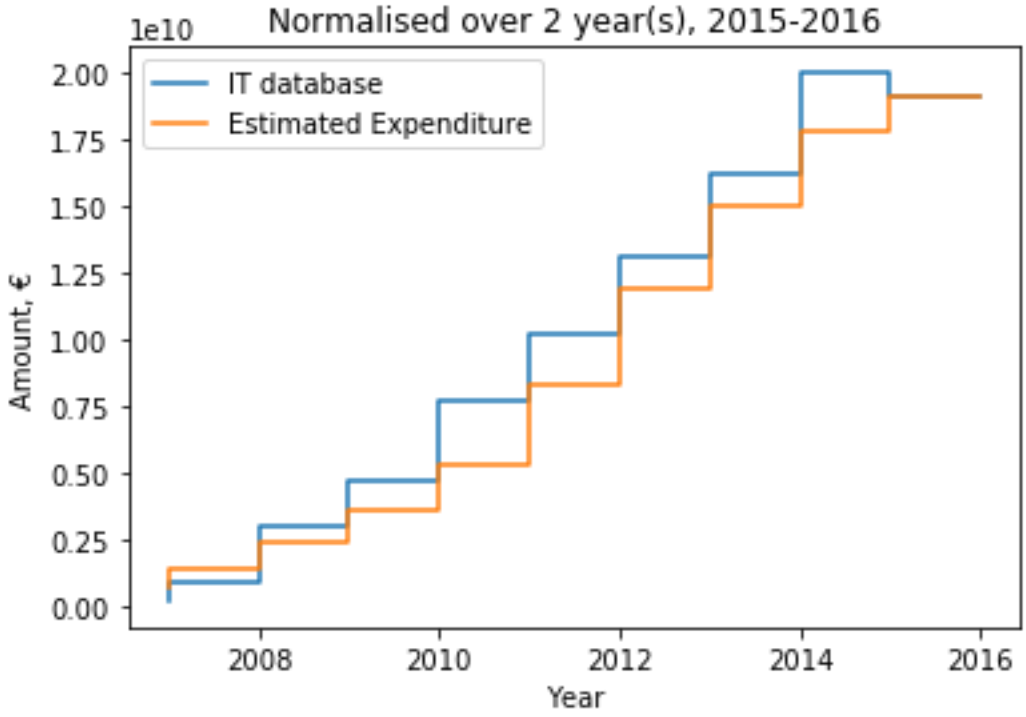

Fig. 2: 2007-2013 ERDF Italian payment dataset against modelled expenditure, normalisation with exceeding figures taken out in equal proportions from 2015, 2016

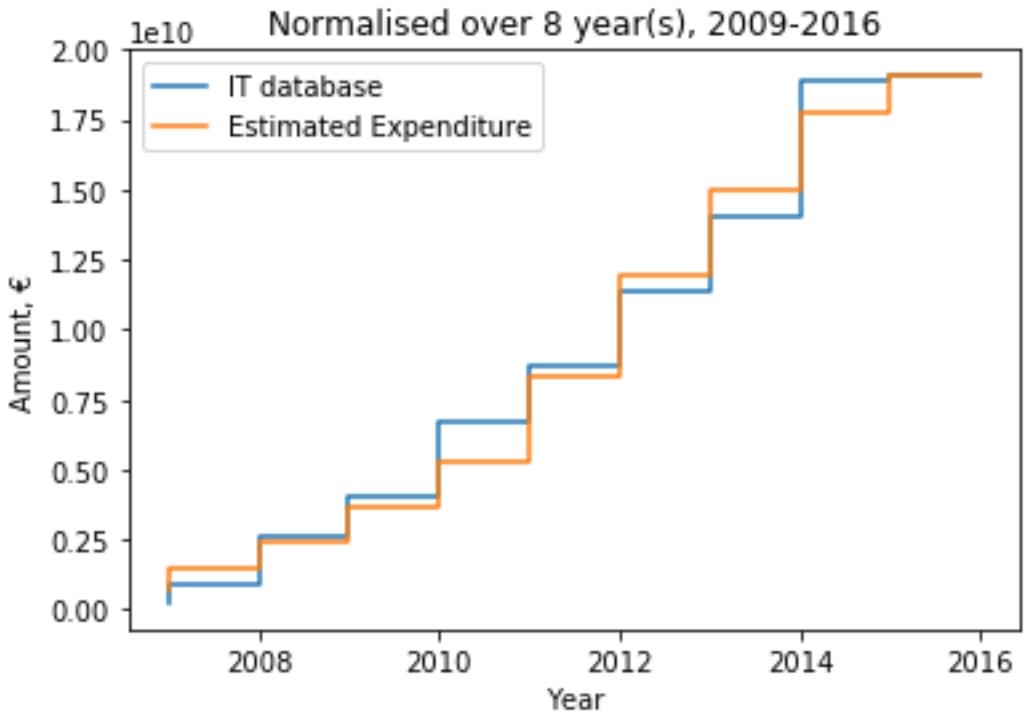

Fig. 3: 2007-2013 ERDF Italian payment dataset against modelled expenditure, normalisation with exceeding figures taken out in equal proportions from 2009-2016

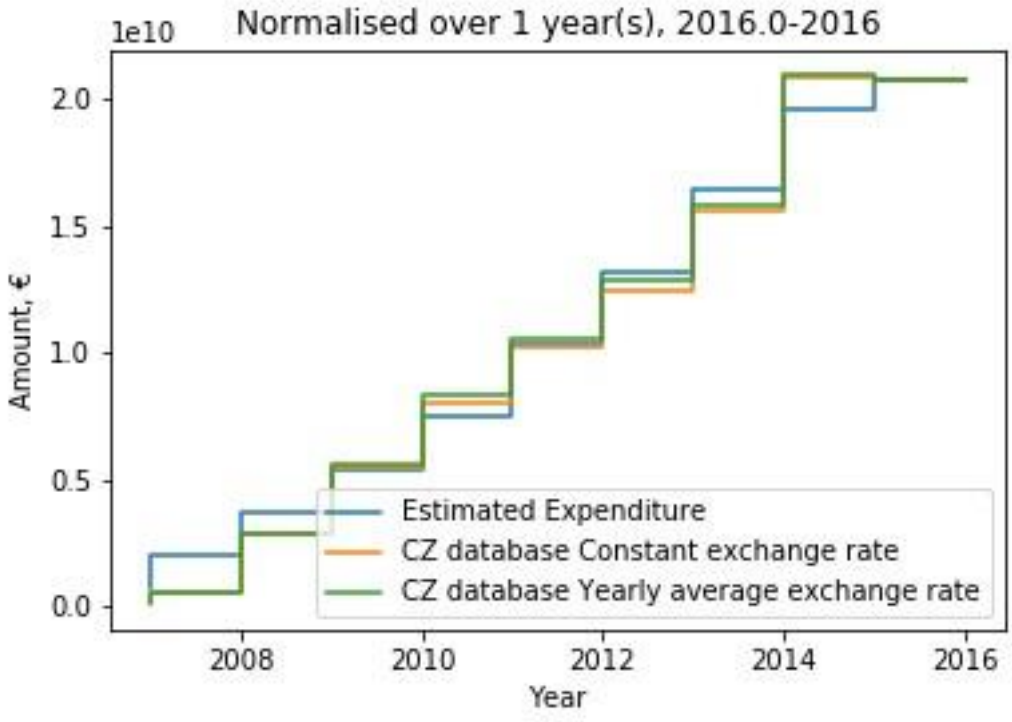

Fig. 4: 2007-2013 sum of CF and ERDF Czech payment data set against modelled expenditure, normalisation with exceeding figures taken out from 2016. Constant exchange rate MA expenditure (orange), yearly variable exchange rate MA expenditure (green), mean estimated expenditure (blue)

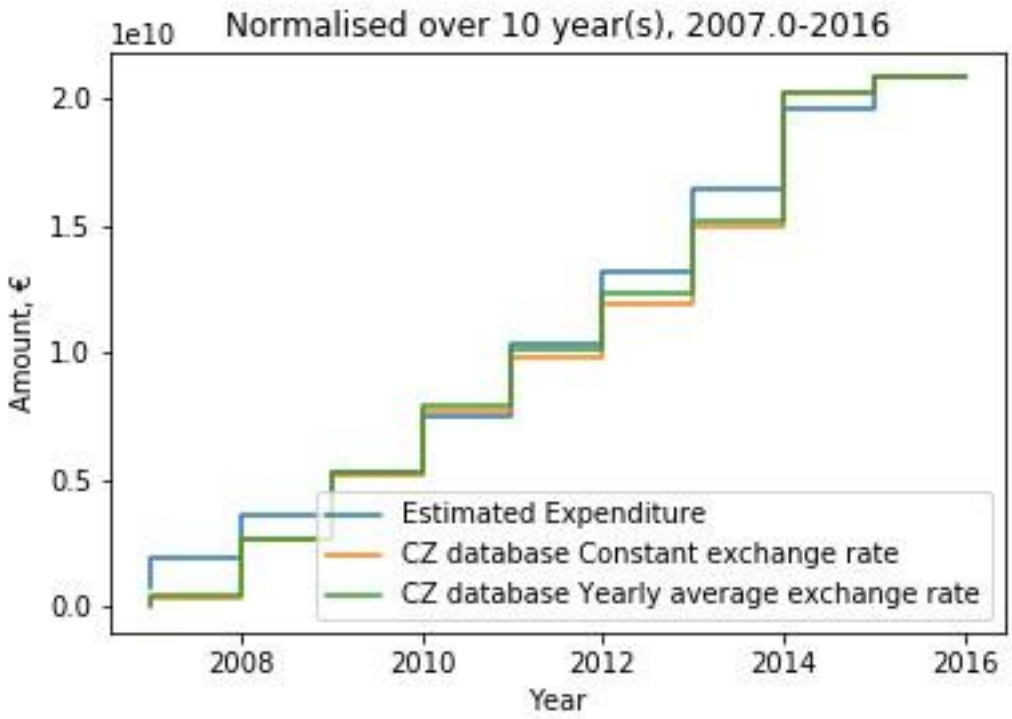

Fig. 5: 2007-2013 sum of CF and ERDF Czech payment dataset against modelled expenditure, normalisation with exceeding figures taken out from 2016. One tenth of it from each year. Constant exchange rate MA expenditure (orange), yearly variable exchange rate MA expenditure (green), mean estimated expenditure (blue)

The mismatch between the normalised MA expenditures reported and the remitted EC payments amount to less than 5% for CZ. The distance between the two patterns is not particularly affected if the normalisation is performed over a variable number of years.

As to identify where the mismatches with the actual incurred expenditures stem from, one can open up the black box of the individual figures composing the distribution. To this end, we have performed this comparison along with the range of adopted assumptions - i) variable number of years from which the exceeding expenditure is taken off; ii) variable exchange rage CZK-EUR. The results are shown in Figures 6 – 25 for a set of 1,000 Monte Carlo random simulations.

One can observe overlapping distributions between the modelled and the incurred expenditures over the central years of the programming period (2010 and 2014 for IT; 2011 and 2014 for CZ). One can perform a more in-depth inquiry on the effect of the modelling parameters by extending this comparison at the NUTS2 level so as to have a better understanding of these trends. In this way, one can evaluate the highest mismatches across regions and years.

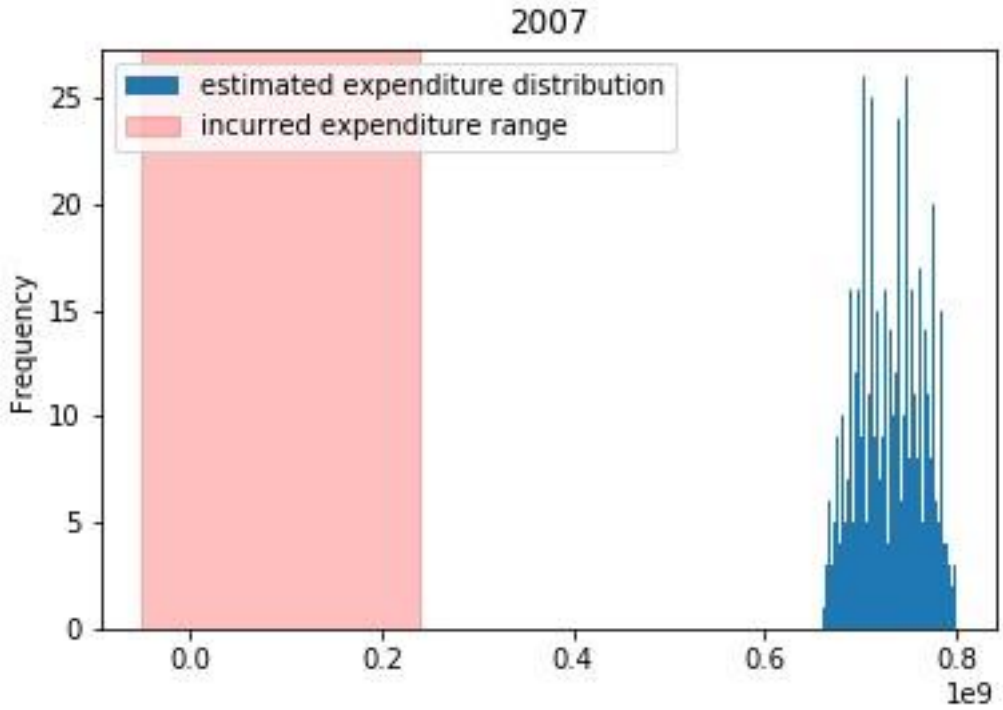

Fig. 6: IT - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

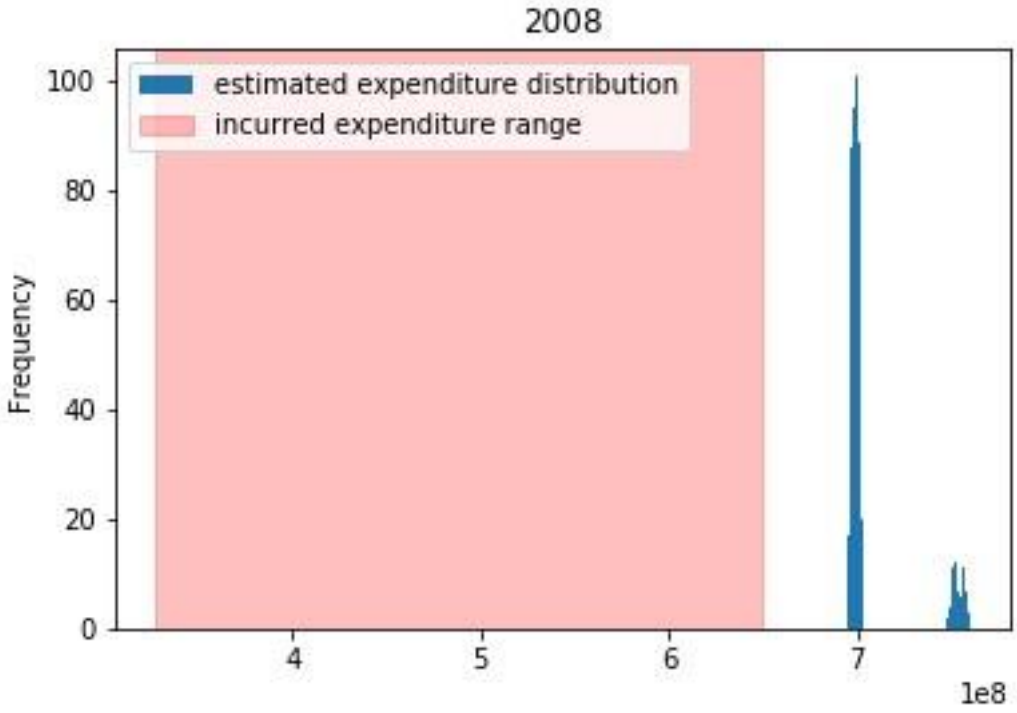

Fig. 7: IT - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

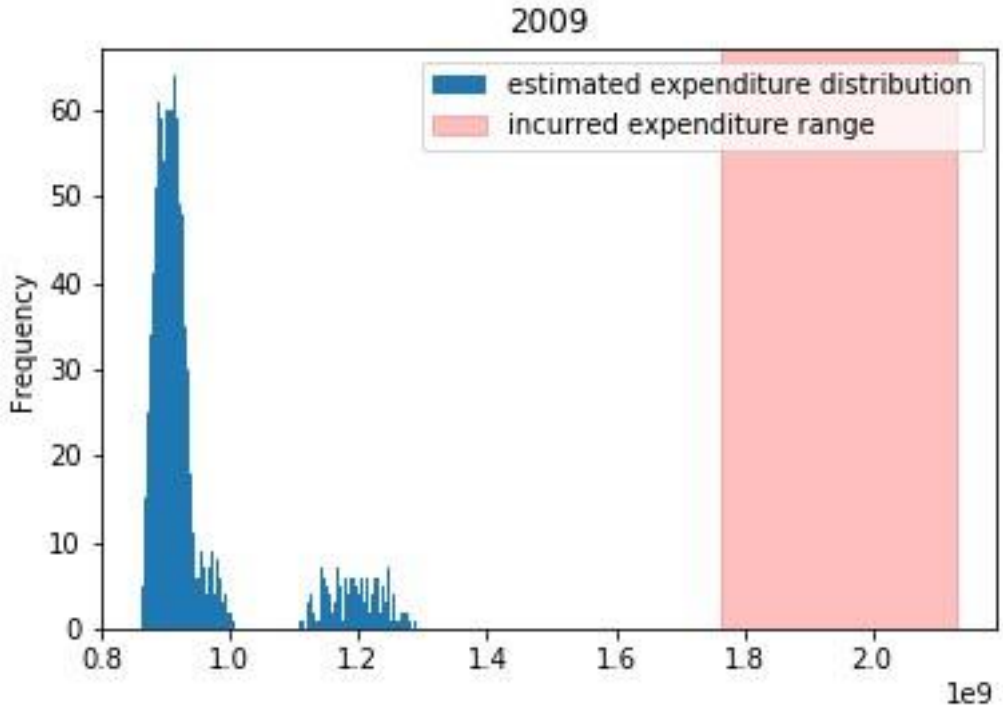

Fig. 8: IT - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

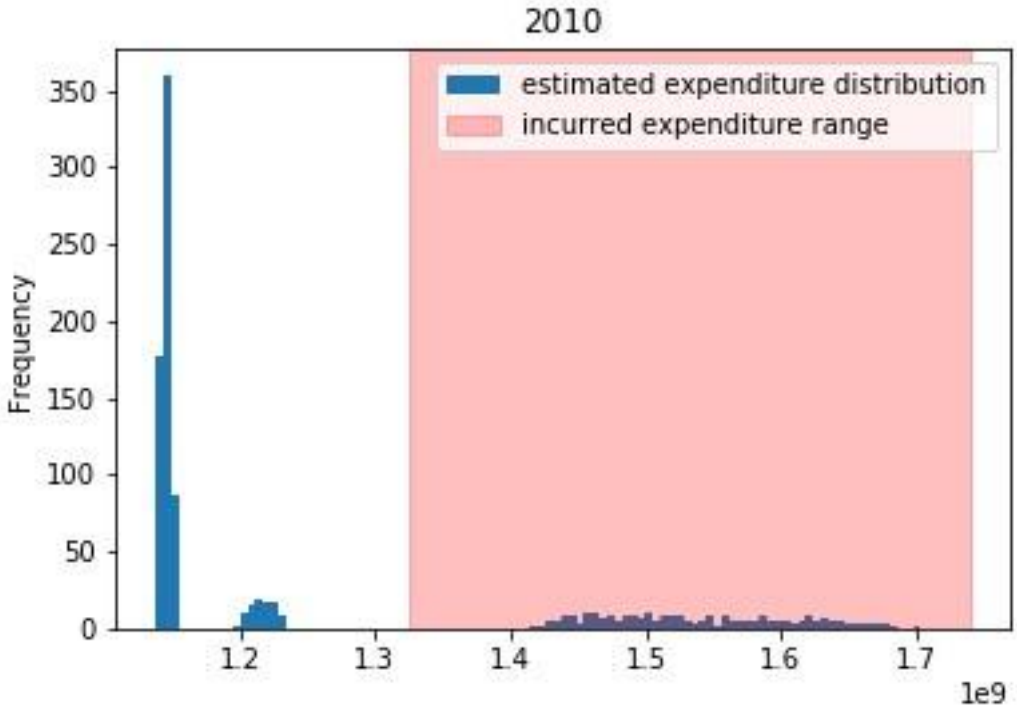

Fig. 9: IT - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

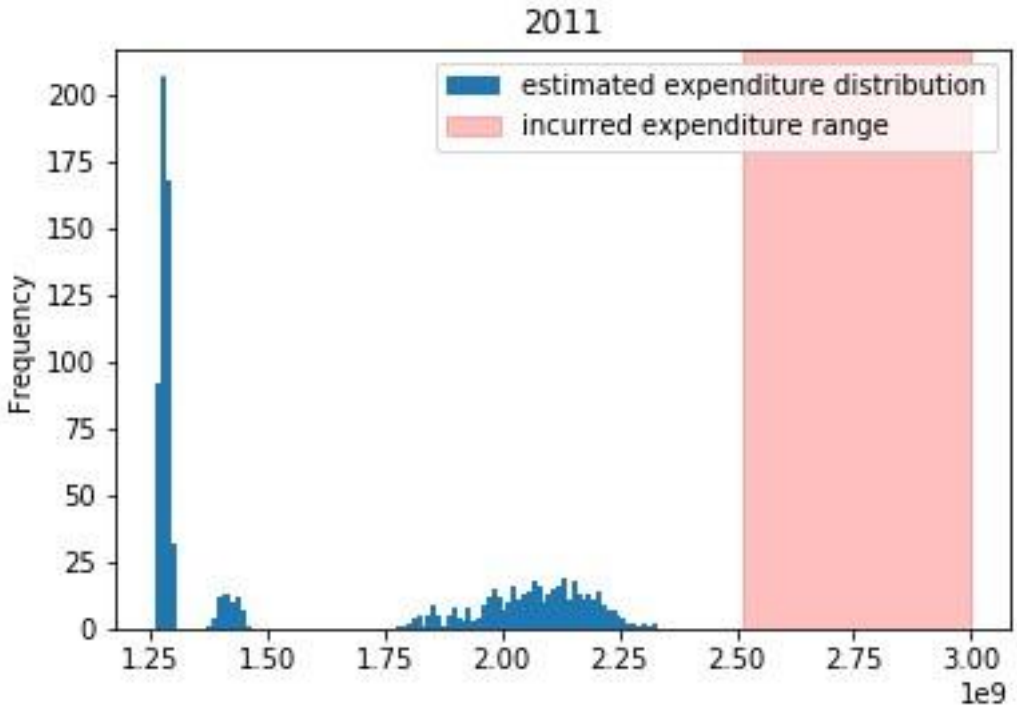

Fig. 10: IT - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

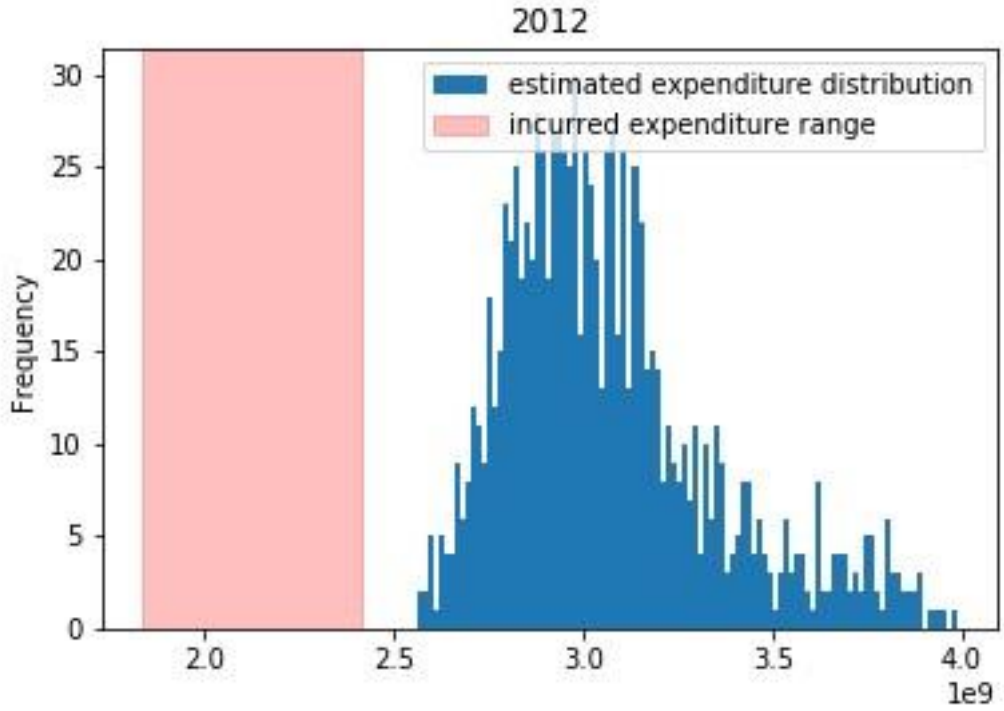

Fig. 11: IT - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

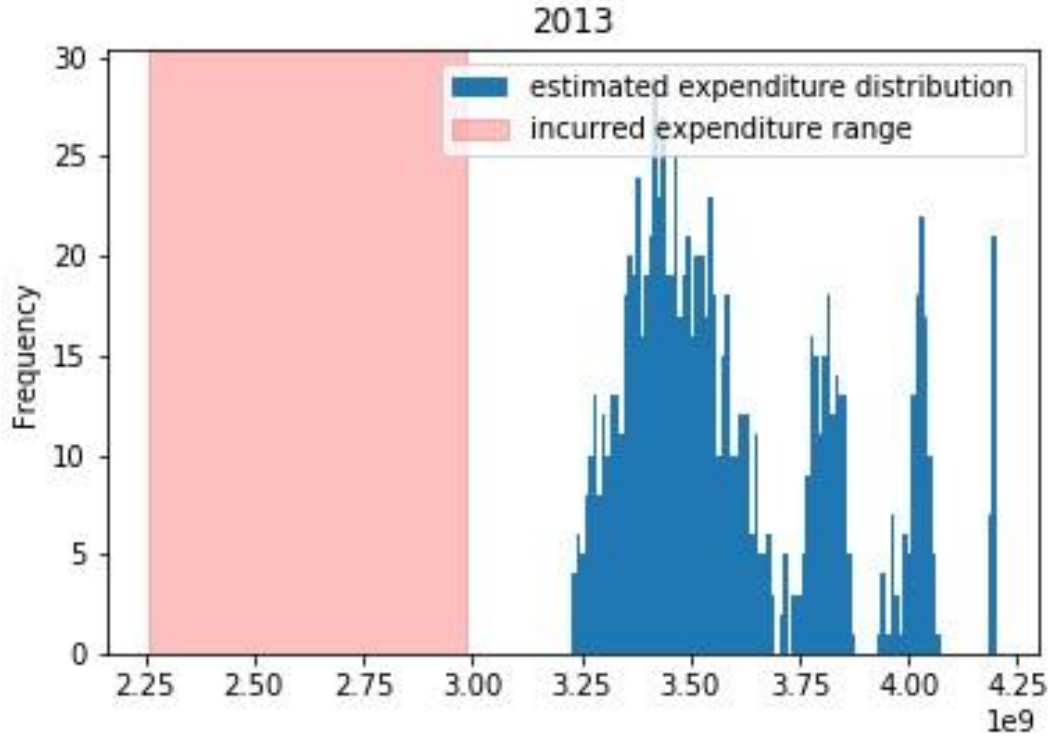

Fig. 12: IT - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

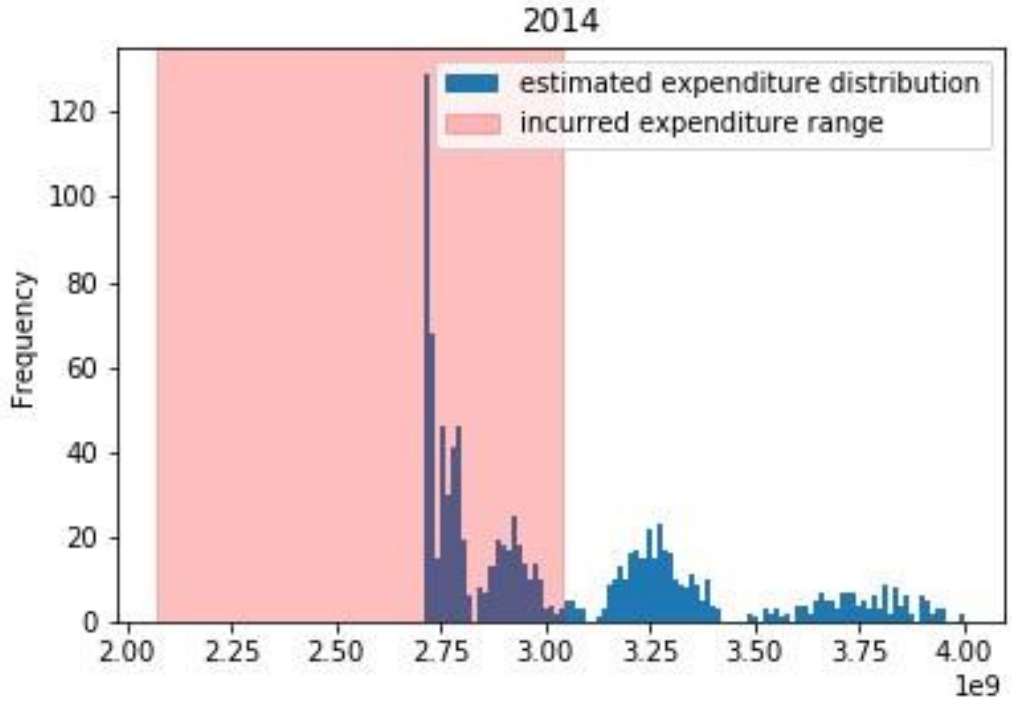

Fig. 13: IT - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

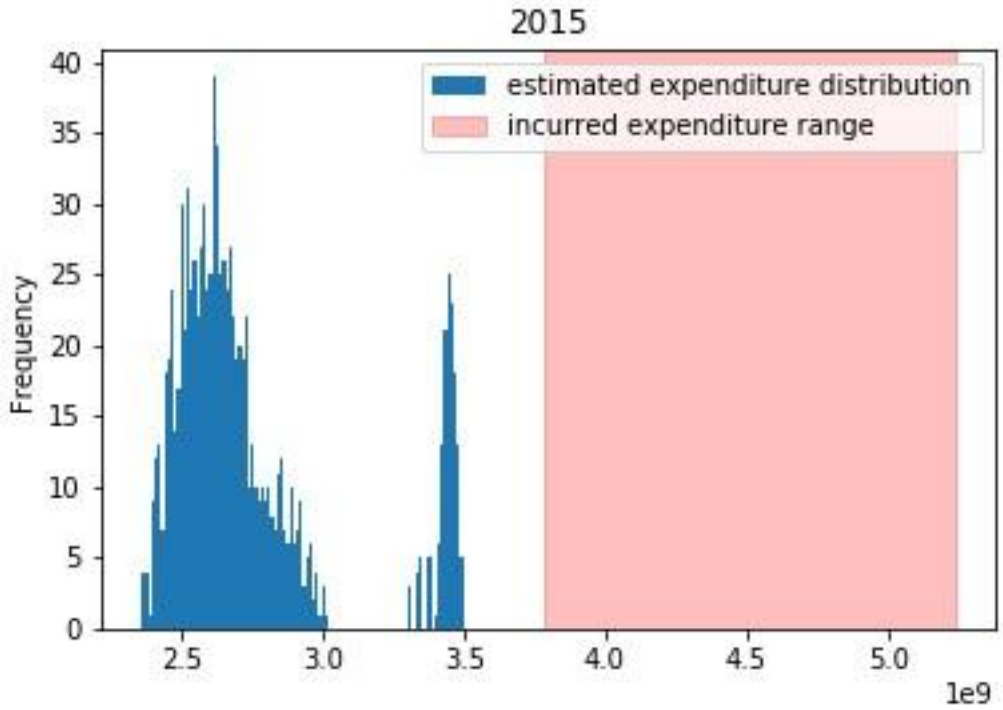

Fig. 14: IT - Distribution of the modelled expenditure. Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

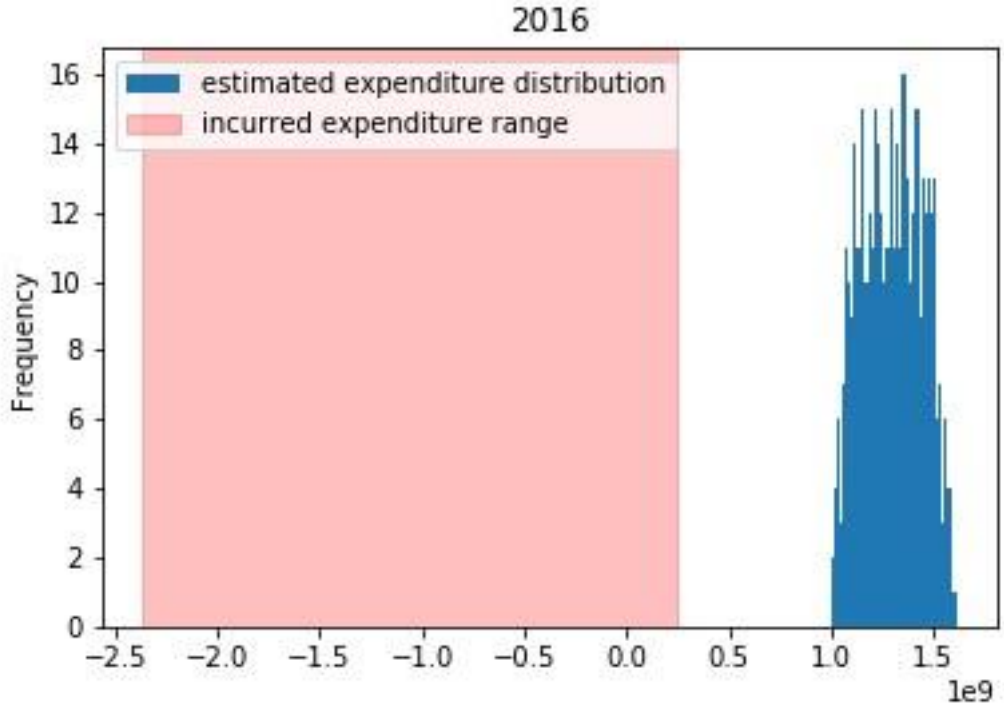

Fig. 15: IT - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

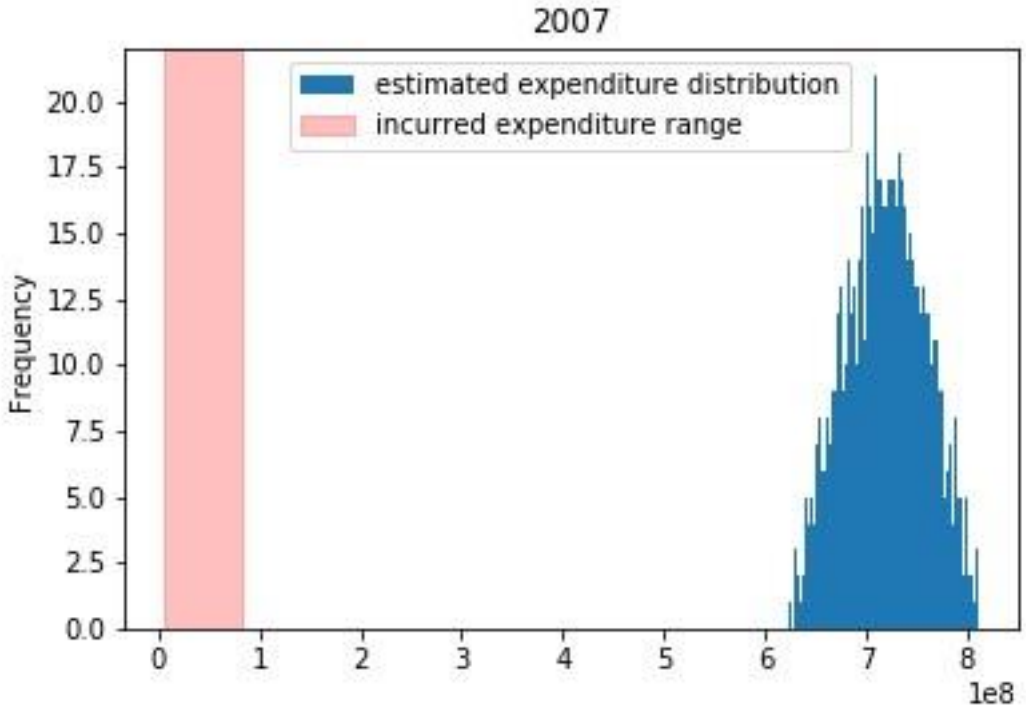

Fig. 16: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

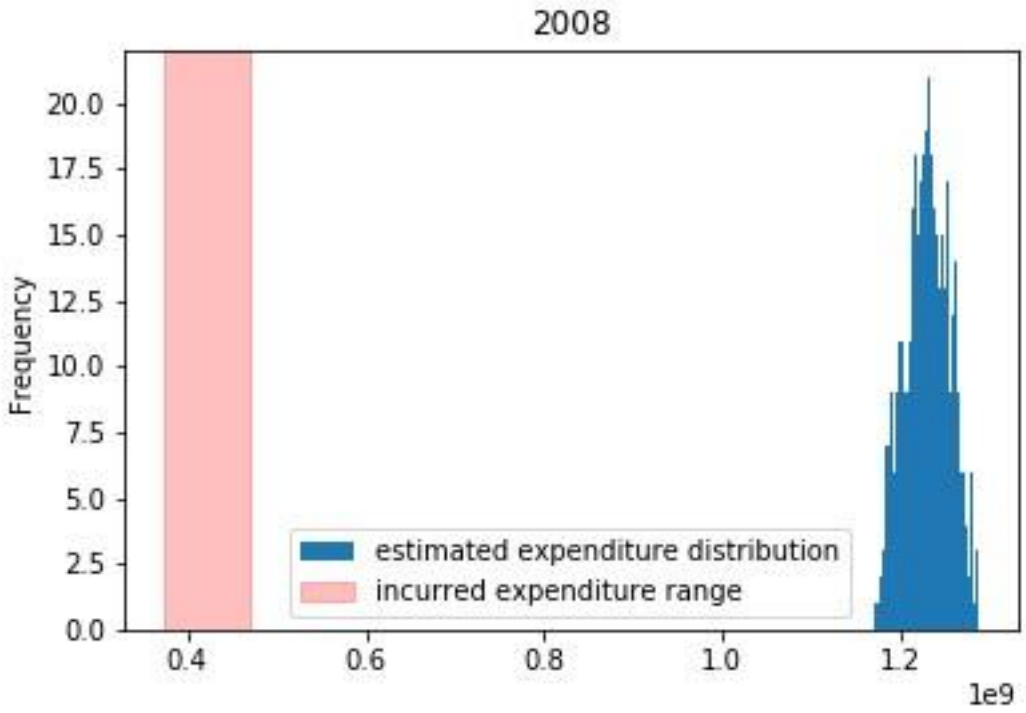

Fig. 17: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

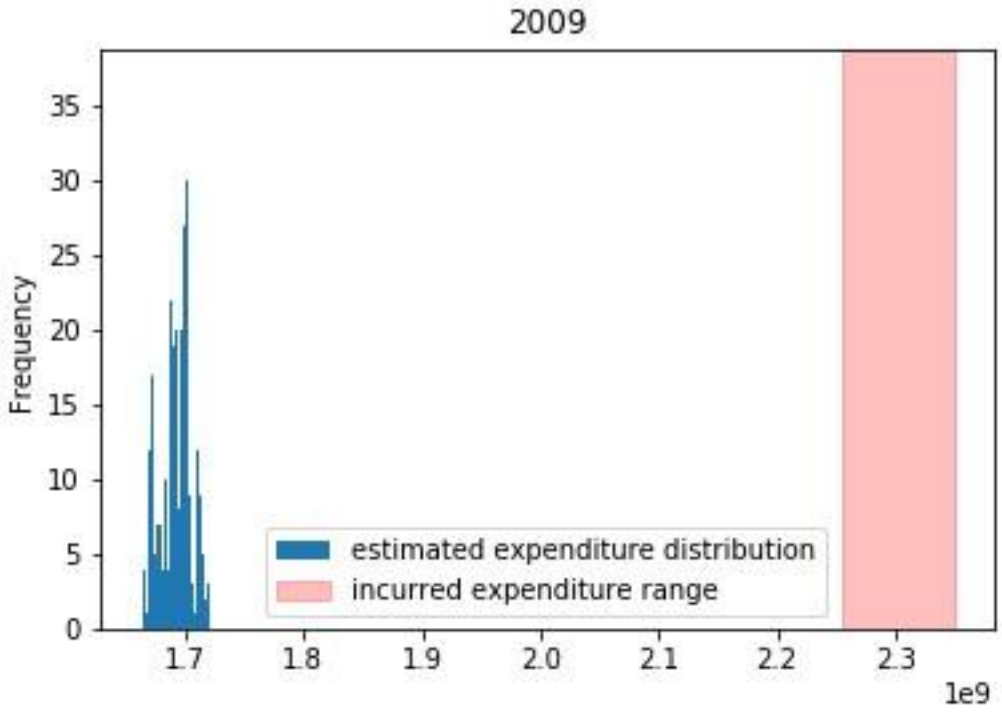

Fig. 18: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

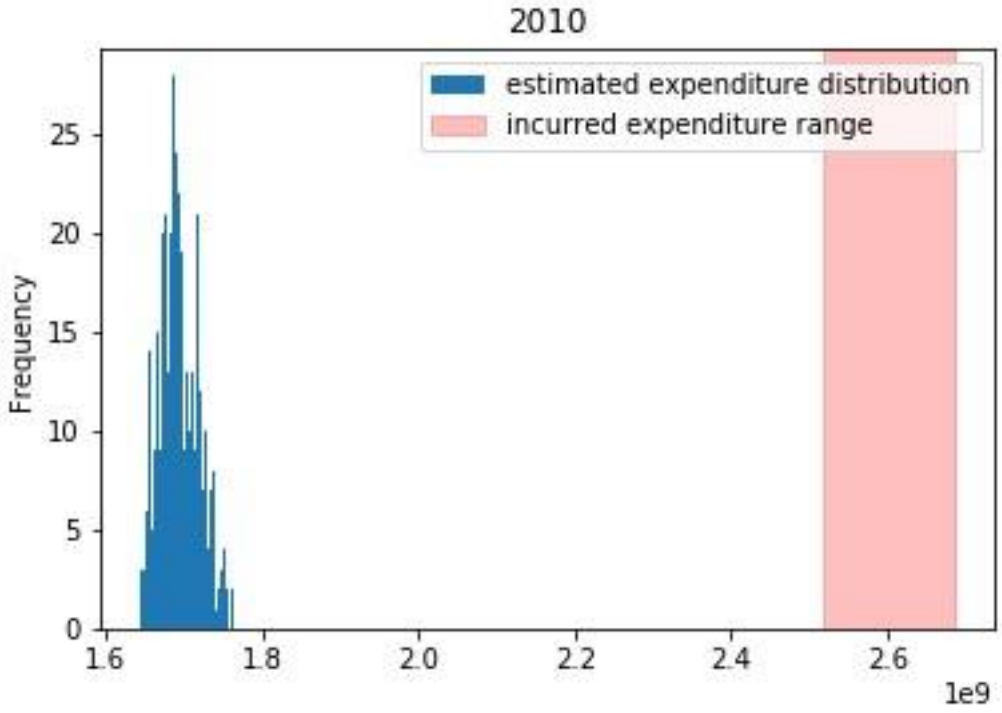

Fig. 19: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

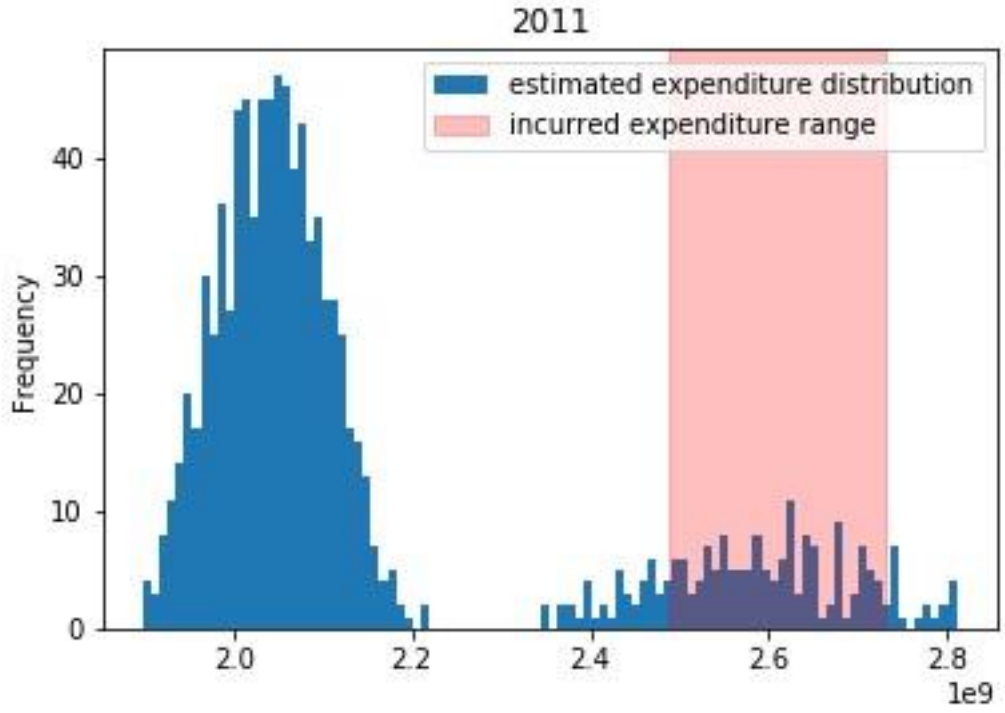

Fig. 20: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

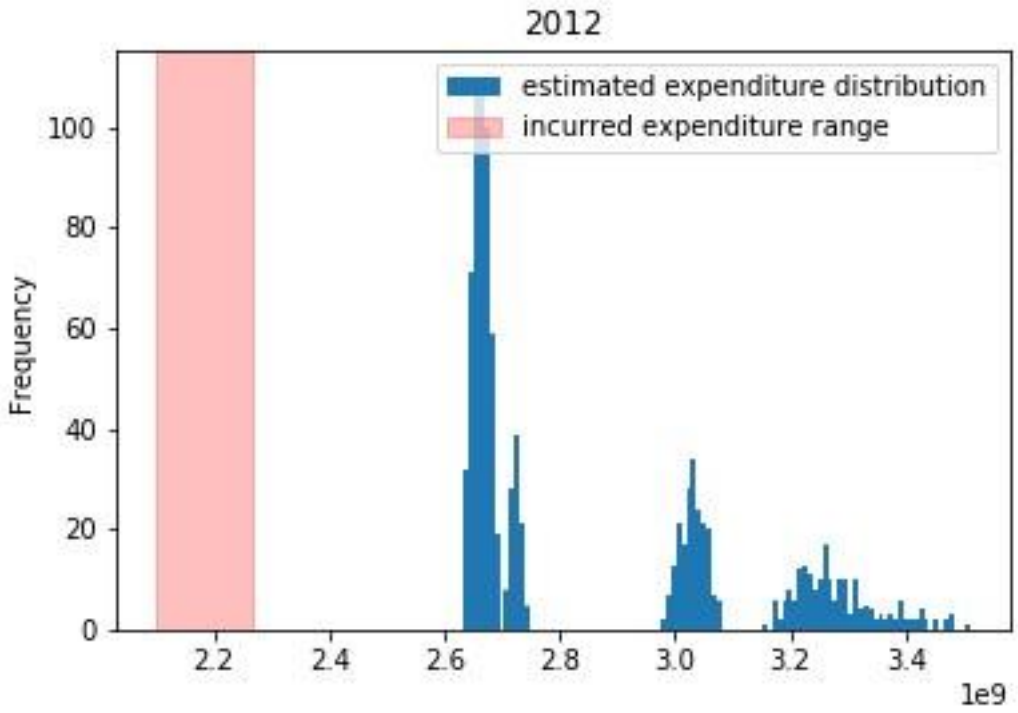

Fig. 21: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

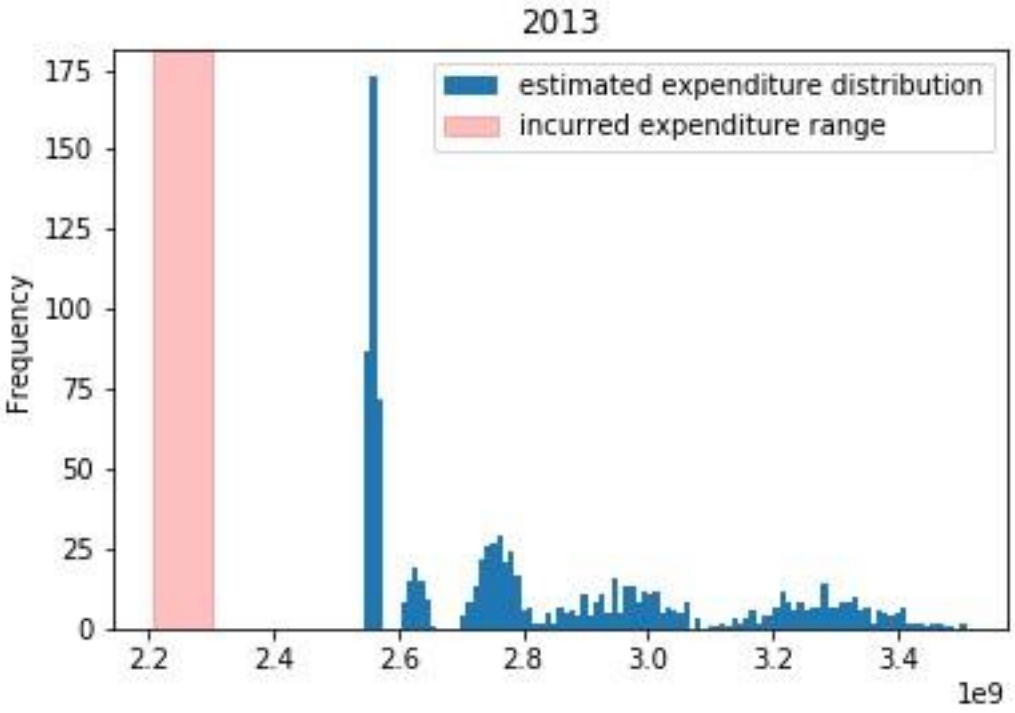

Fig. 22: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

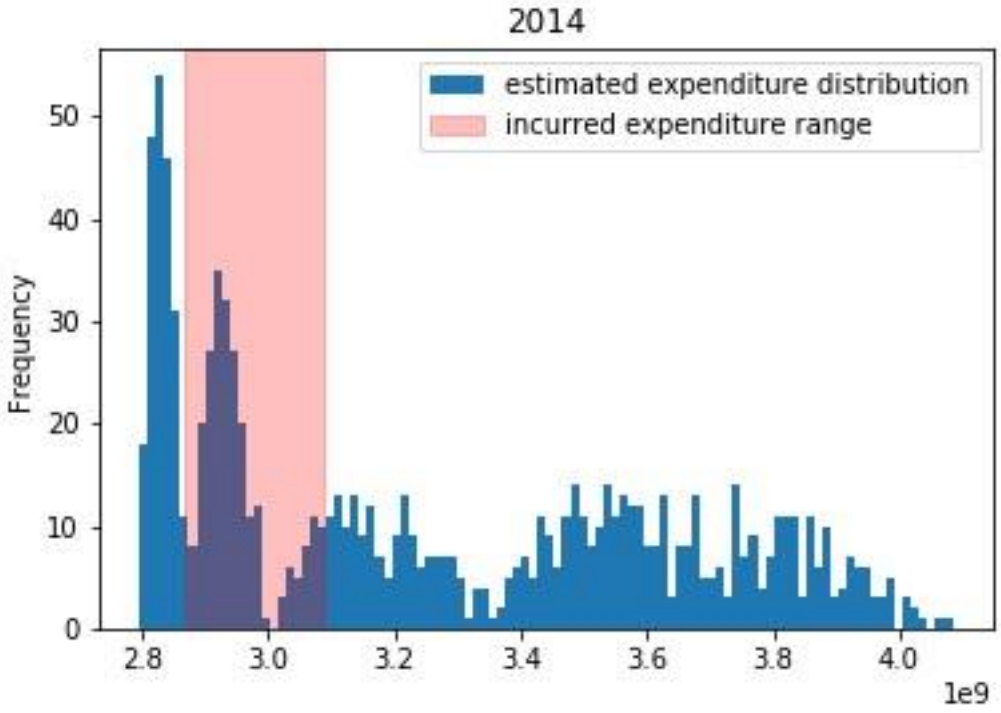

Fig. 23: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

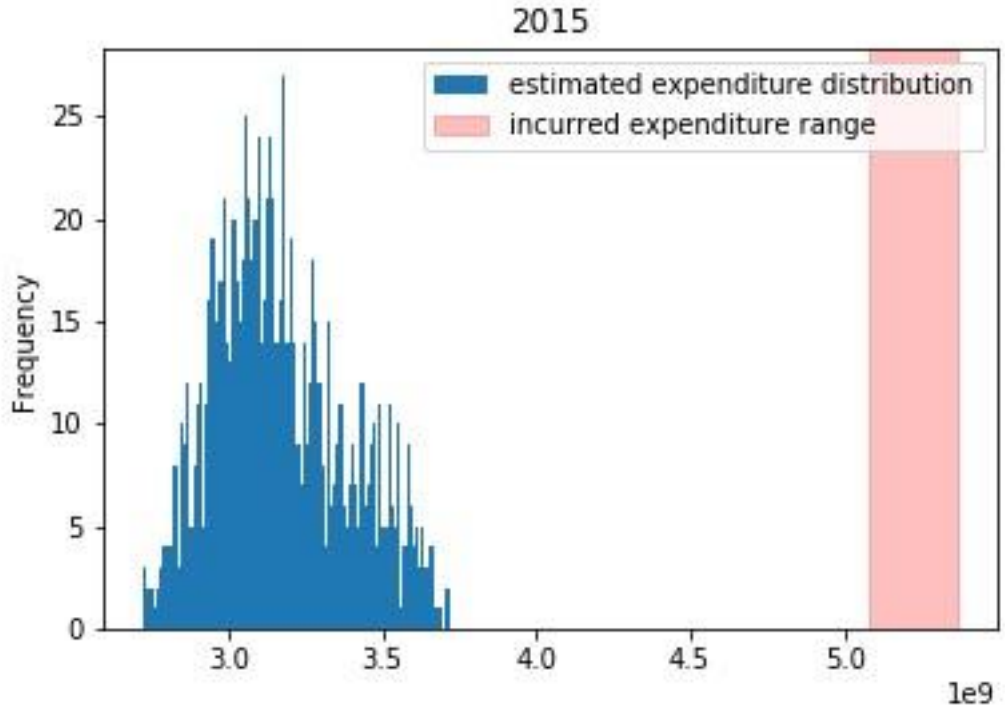

Fig. 24: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

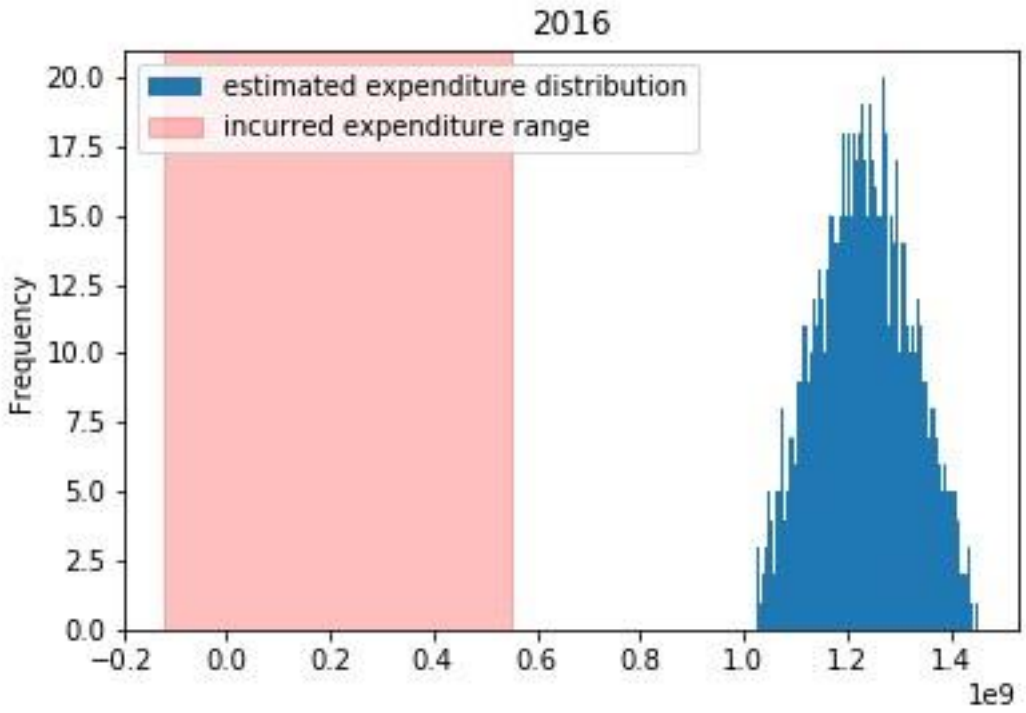

Fig. 25: CZ - Distribution of the modelled expenditure (blue) against the range of incurred expenditure (pink). Frequency of occurrence in the Monte Carlo simulations against yearly expenditure (€).

3.1.1. Comparison at the NUTS2 level

The number of figures produced in the context of this inquiry is too high to be uploaded in this section. However, they can be accessed to and reproduced by running the code on the *GitHub* repository. The relevant notebooks are

CZ_NUTS2_DistributionsVsExpenditures_GitHub

and

IT-NUTS2_DistributionsVsExpenditures_GitHub.

Breaking down the figures at the regional level further complicates the analysis: several reported expenditures for Italy are not clearly attributed at the NUTS2 level. They are rather allocated at the member-state level (*Ambito nazionale* as per the datasheet terminology), or to multiple regions (*Pluri*) or even to other states (Stato Estero). The monetary amount and the number of these allocations is small but tangible (2% and 1% of the total; 3% and 9%; less than 1% according to these metrics, respectively).

This poses the challenge of how mapping these figures onto regional allocations. Some of these schemes hint to a specific regional attribution (e.g. POR CONV FESR SICILIA) for some expenditures reported at the member state level. One can then test how this assumption would impact on the final result. The same applies to the multiple-region attributions: Several hypotheses can be tested on how to split these figures onto individual regions, such as distributing them proportionally to the regional attributions over the whole programming period or to the specific year they have been filed under. As regards specific expenditure items that make reference to a precise NUTS2, one may attribute for instance 50% of them to the region is made reference to and the remainder may be split over the other regions as per the same criterion above.

The maximum distance between the figures is also attained in this regional-level comparison when the exceeding budget is taken out from a single or all the years of eligible expenses. Not surprisingly, the highest yearly distance is on the last year when the normalisation is performed by taking out the excess from only this year. The maximum distance in negative values, i.e. European expenditures anticipating the invoicing to the member state, occurs if the normalisation is performed over ten years.

Trentino Alto-Adige is disaggregated in the two provinces of Trento and Bolzano in the European database. Conversely, a unique regional figure is presented in the IT database. Hence, we aggregated the figures for the two provinces at the regional level so as to produce a consistent comparison.

No visible pattern emerges when analysing how the distance calculated maps onto the series of μ_p coefficients: The regions having the highest distance are not those with the highest value of the coefficient of regional specificity.

The Czech database does not have any ambiguous attribution unlike the Italian one. This simplifies the process of breaking down the figures at the NUTS2 level. The effect of the variable exchange rate seems to be practically negligible on the observed distance also in this case. When disaggregated at the regional level, the excess figure varies its sign across years. Curiously enough, the payments reported at the EU level are larger than the reported expenditures in the CZ database for some years. The normalised distances are inferior to the Italian database and the trend observed when taking out the exceeding figure over a variable number of years is increasingly/decreasingly monotonously.

As regards the μ coefficients, the conclusion we have drawn for Italy also apply to CZ: A precise mapping onto the μ_p figures is to be ruled out in this case as well. Analysing the distribution trends across years reveals interesting trends with a general pattern for consecutive years: when the distribution of the estimated expenditures is larger than the interval of the incurred expenditure in the year $i+1$, it is typically the reverse in the preceding year i. In the next section, we will scrutinise the model robustness through sensitivity analysis, which may also offer insights on how to reduce the reported distance between modelled and actual incurred expenditures

3.2. Task 1.b - Modelling robustness through sensitivity analysis

The output variable μ_p measures the cumulative relative distance between the modelled expenditures and the incurred MSs expenditures. Should the variance of this output be primarily affected by the MSs expenditure standardisation parameters, one would need to prioritise improving the quality of the information presented in this dataset rather than updating the model to reduce μ_p. Conversely, a primary importance of our modelling parameters' uncertainty would point towards possible points of intervention to improve the model developed and fill the gaps encountered.

As to ensure a consistent comparison, several edits to the modelled expenditure are to be introduced. These include mainly two edits as per the previous sections: i) assess the Italian region Trentino Alto-Adige (TAA) all together in lieu of its two autonomous provinces; ii) assess the aggregated CF + ERDF funds as regards CZ. The list of the other modelling hypotheses is found below.

The comparison can result into a dynamic and iterative process: several rounds of model edits followed by new comparisons against the actual incurred expenses. All with the aim to reduce μ_p towards the theoretical lower boundary of zero.

- Normalisation: i) variable number of years from which the exceeding figure gets taken out - *Excess years taken out MS*; ia) constant share across years or ib) variable share across years, halving each year backwards *Excess share years taken out MS*; ii) the exceeding figure gets taken out from the reported yearly maximum expenditure only.
- Exchange rate: ia) constant exchange rate across the whole programming period; ib) variable yearly

3.2.1. Uncertainty analysis of the generated distributions

2^{13} (8,192) Monte Carlo simulations were run on quasi-random samples extracted from the distributions of the six uncertain parameters. The result was a distribution for the outcome variable μ_p including 2^{13} samples. The statistical features of these outcome distributions can be inferred by analysing histogram charts: few region showed a normal-like distributions Molise (ITF2) (Figure 26) a skewed one (Calabria (ITF6), Prague (CZ01) Figures 27 and 28, respectively) and multi-modal distributions (Toscana (ITI1), Severozapad (CZ04) (Figures 29 and 30, respectively). More figures can be accessed to by running the *Jupyter notebook* made available from the quoted *GitHub* repository (see "Supporting Material" below).

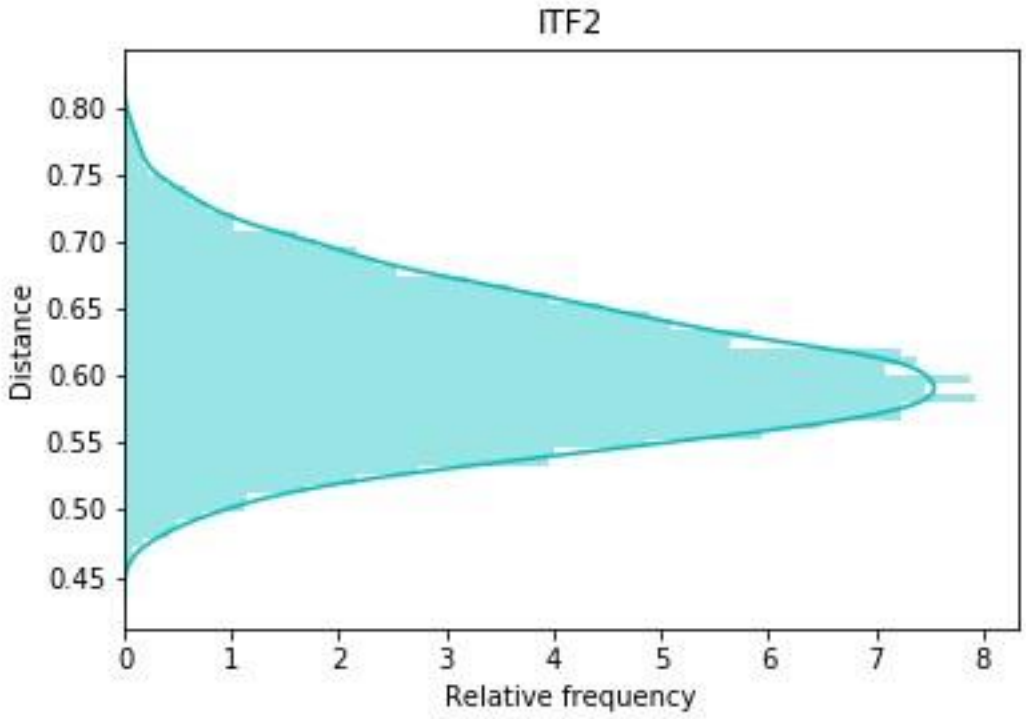

Fig. 26: ITF2, Molise - Distribution of the cumulative distance μ_p between the modelled expenditures and the incurred MSs expenditures

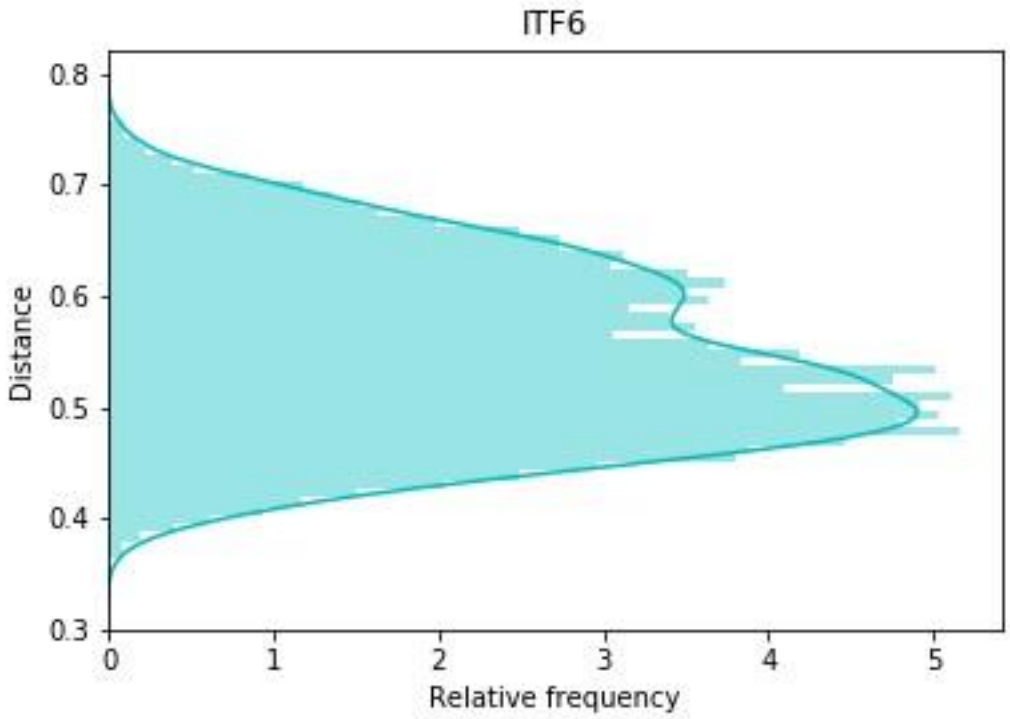

Fig. 27: ITF6, Calabria - Distribution of the cumulative distance μ_p between the modelled expenditures and the incurred MSs expenditures

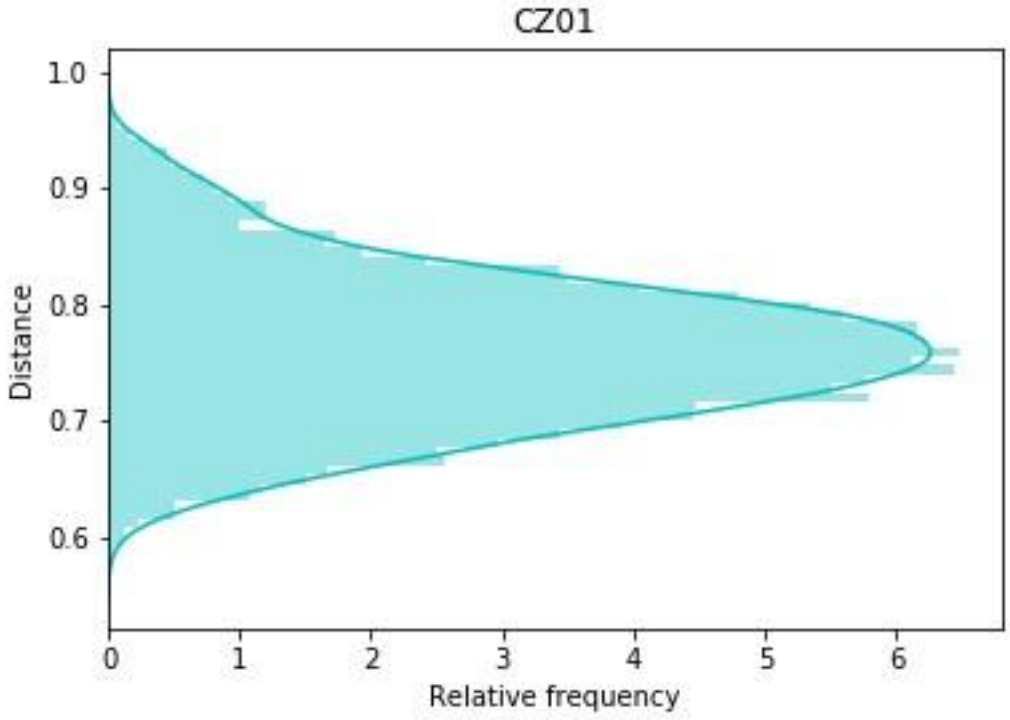

Fig. 28: CZ01, Prague - Distribution of the cumulative distance μ_p between the modelled expenditures and the incurred MSs expenditures

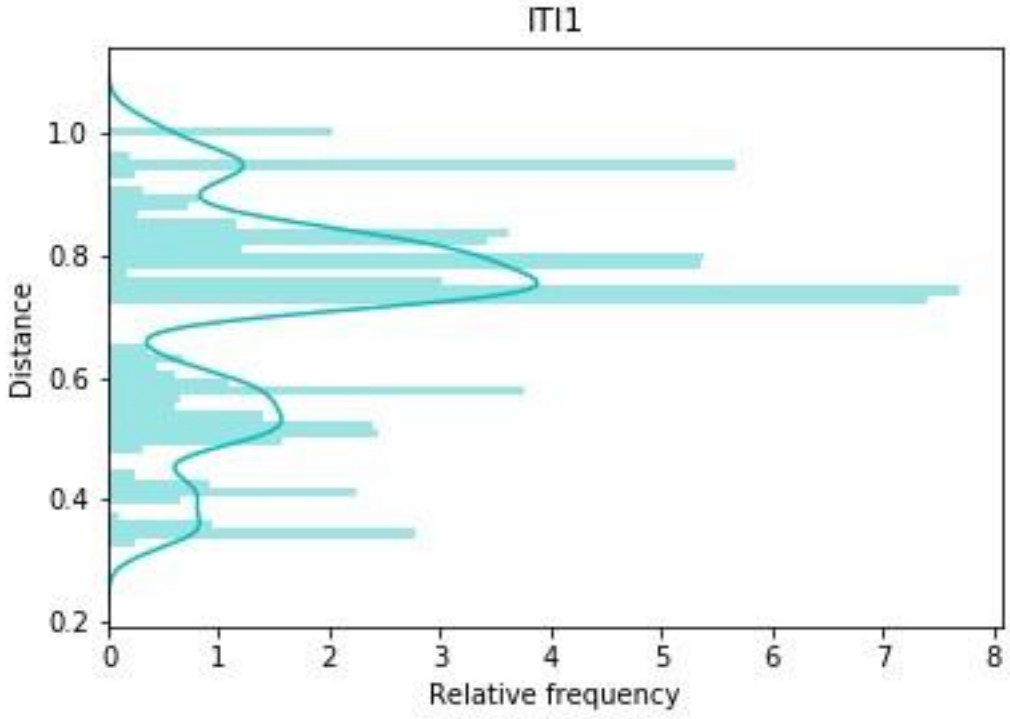

Fig. 29: ITI1, Toscana - Distribution of the cumulative distance μ_p between the modelled expenditures and the incurred MSs expenditures

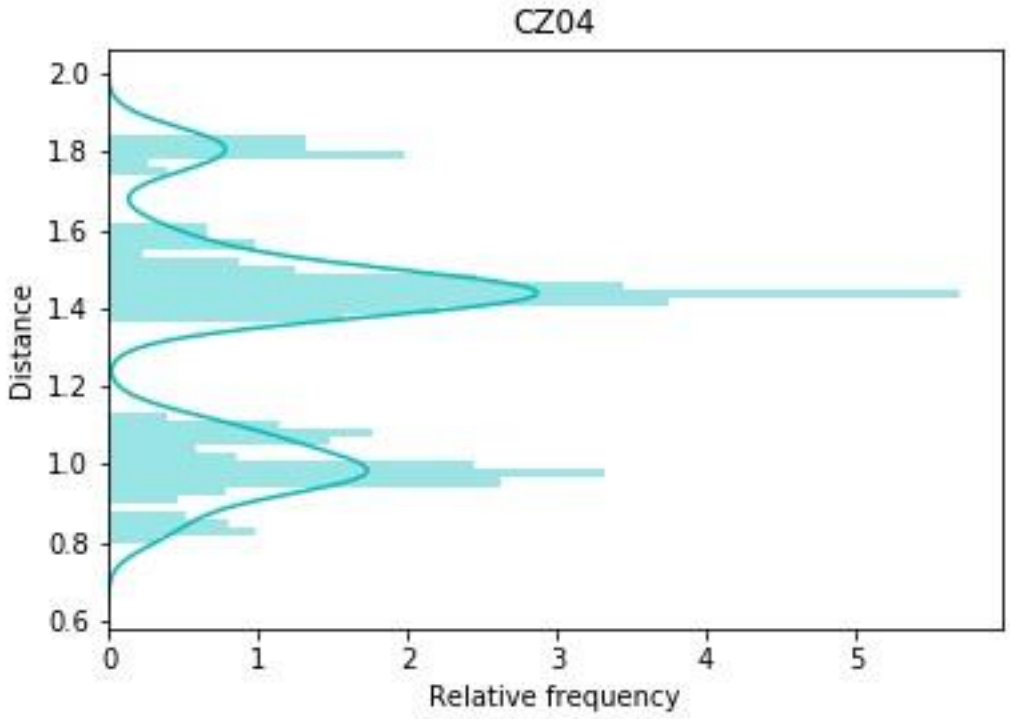

Fig. 30: CZ04, Severozápad - Distribution of the cumulative distance μ_p between the modelled expenditures and the incurred MSs expenditures

3.2.2. Global sensitivity analysis

Let us now try to apportion the output uncertainty to the input parameters through sensitivity analysis. To this end, we calculated variance-based sensitivity indices. These quantities inform on the share of the output variance each parameter's uncertainty is responsible for, either alone (first-order indices) or also through interactions with the other parameters' (total order indices). First-order indices are typically labelled with the letter **S**, whereas total order with the letter **T**.

We have run 1,000 replicas of our calculation as to strengthen the analysis by generating confidence intervals for the indices computed. Unfortunately, the complicatedness of the output distributions resulted in large confidence intervals for first-order sensitivity indices for several of the regions.

A sample of the sensitivity-indices whisker-box plots is reported below in Figures 31 and 32 for Jihozapad (CZ03), ITI1 and Sicilia (ITG1) respectively.

In general, Czech regions have narrower confidence interval than Italian ones, which denote better estimates for the former. The only exception is CZ01. For most of the region, the most influential parameters are the number of years from which the exceeding MSs expenditure is taken out from and the backwards attributing yearly scheme. For some regions, *Residual years attributed _model* plays also a tangible role on the output uncertainty. This applies especially to Italian regions such as Campania (ITF3), ITF6 and ITG1 - (Figure 33). Finally, for a couple of regions even the other modelling parameters φ_{max} and φ_{min} have non-negligible indices, this applies especially to CZ01, for which the situation is less clear.

Hence, the uncertainty of μ_p is totally affected by the uncertainty of the standardisation parameters of the MSs database as regards CZ. In this setting it is not possible to improve the model of estimated expenditures unless this source of uncertainty can be reduced by adopting different modelling hypotheses.

As regards Italy, one can inquire the effect of *Residual years _attributed _model* on μ_p. Were a clear trend discernible, one may apply *Monte Carlo filtering* to select a sub-range of values for this parameter so as to reduce the reported distance. We may test this option for the regions which show the most promising trend. The results are reported in the scatter plots (Figures 34 -36) for ITF3, ITF6 and ITG1, respectively.

While the effect on the output distribution is tangible, the trend across years is not monotonous. Hence, reducing or increasing the number of years allowed in the model would not necessary result in the reduction of μ_p.

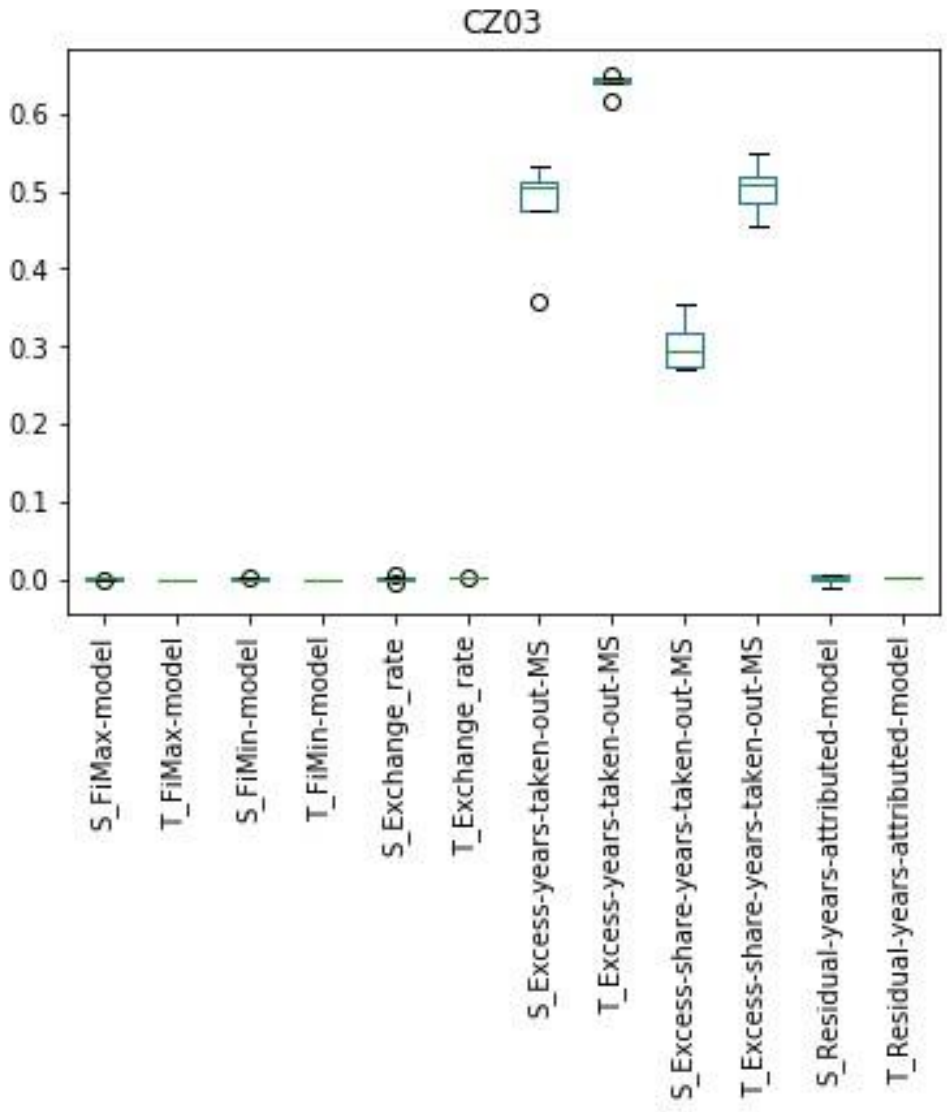

Fig. 31: CZ03, Jihozápad - Whisker-box plots of the first-order and total sensitivity indices of the six parameters inquired

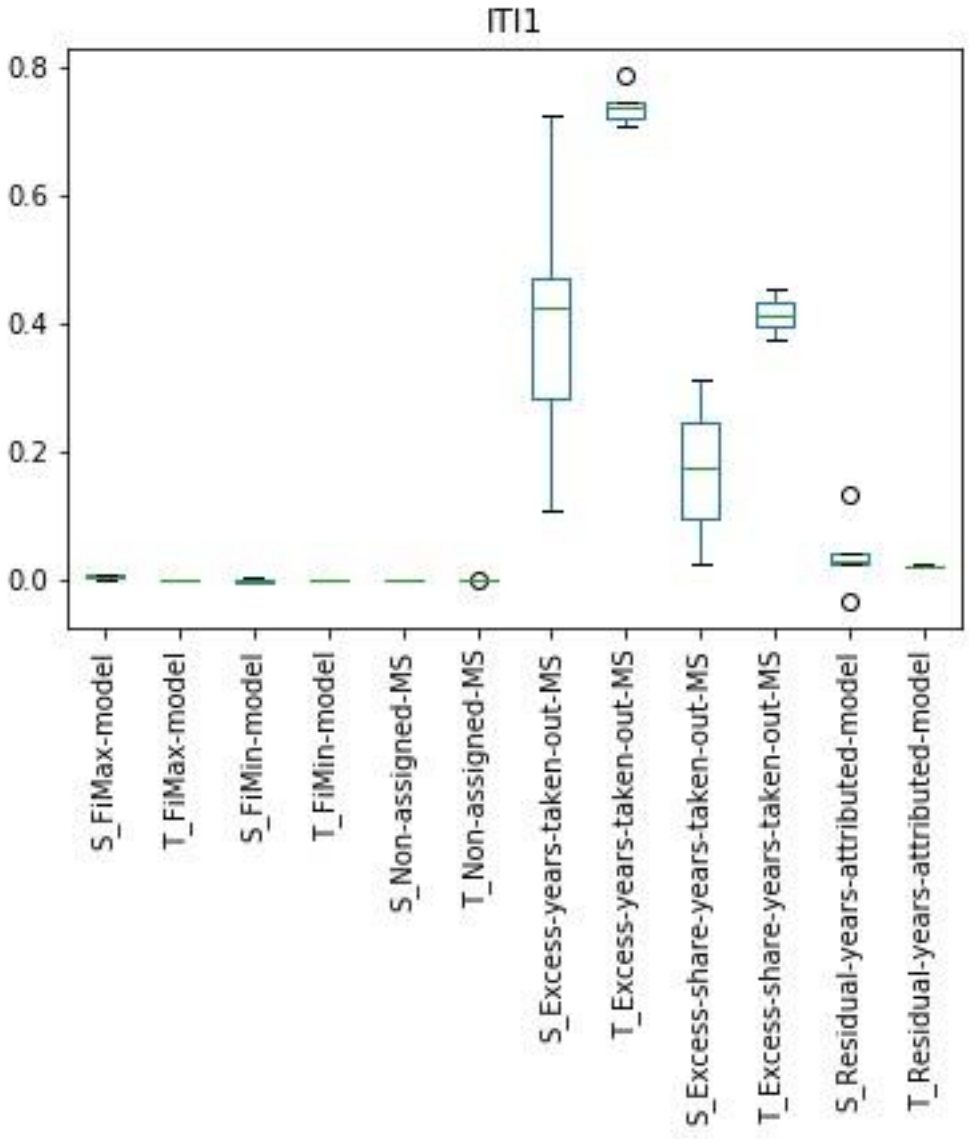

Fig. 32: ITI1, Tuscany - Whisker-box plots of the first-order and total sensitivity indices of the six parameters inquired

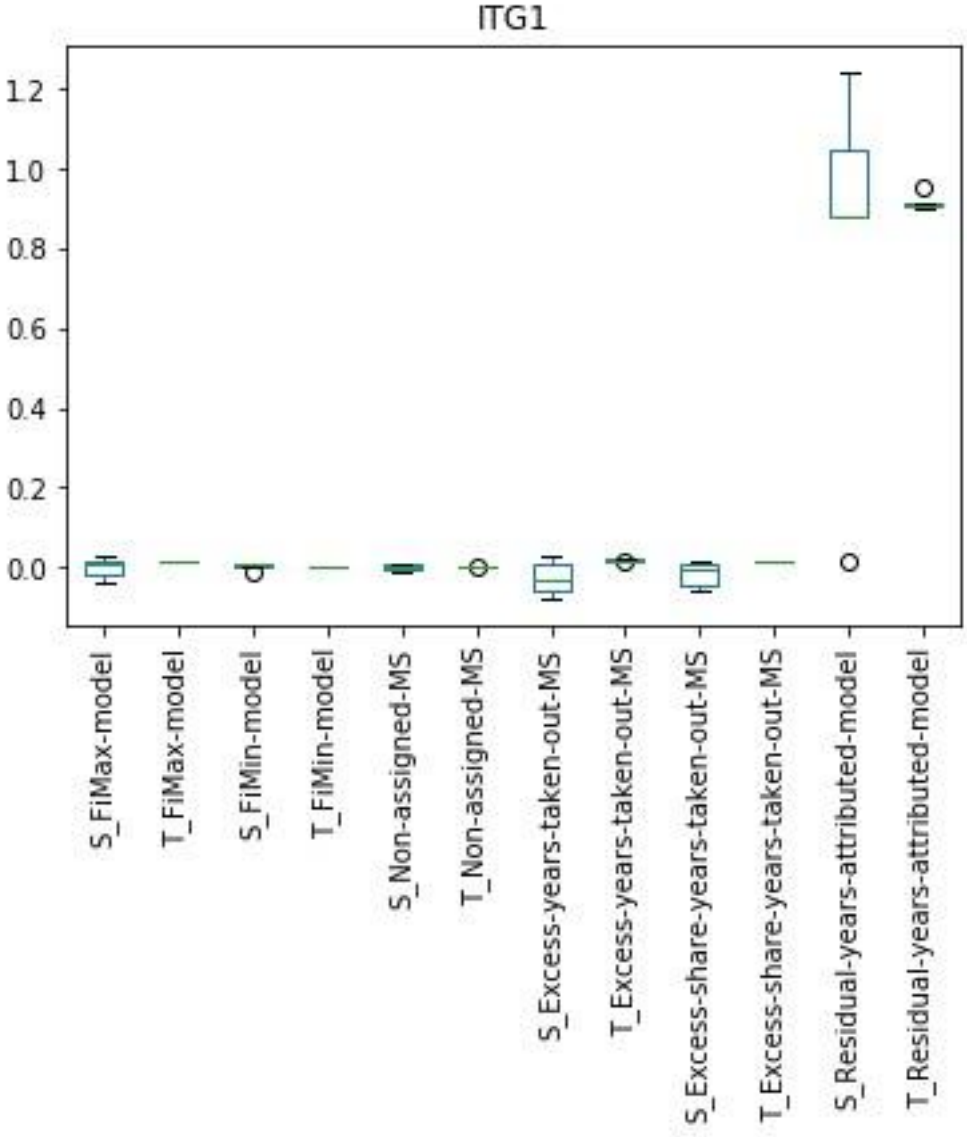

Fig. 33: ITG1, Sicily - Whisker-box plots of the first-order and total sensitivity indices of the six parameters inquired

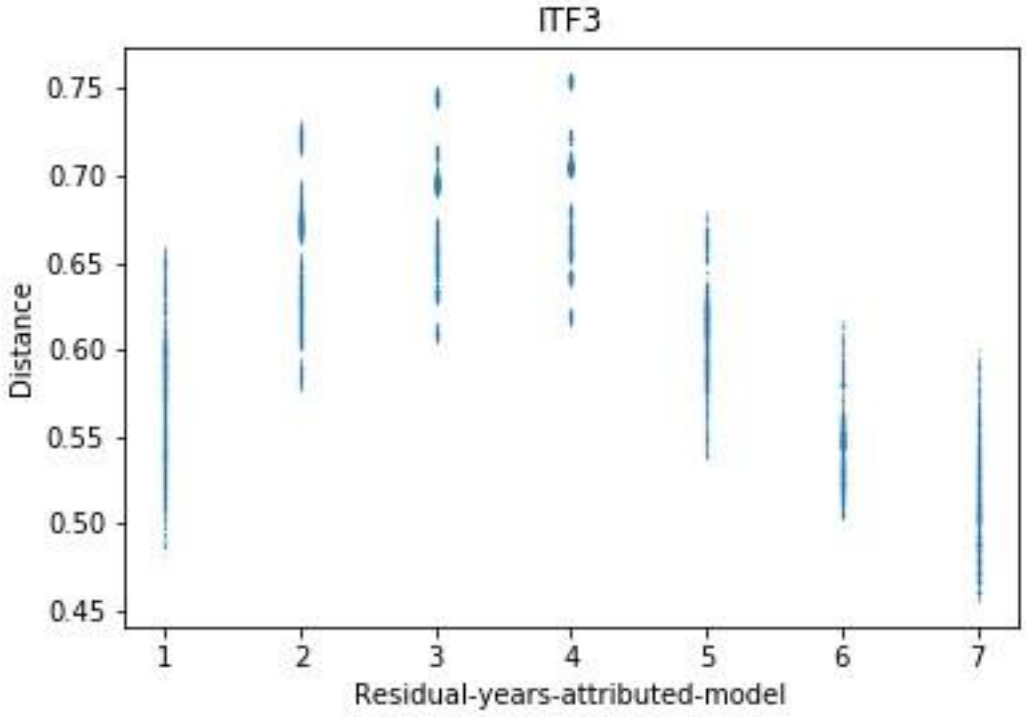

Fig. 34: ITF3, Campania - Scatter plot of μ_p distribution against the number of years of modelled expenditure for the residual of the EC 2016 payment

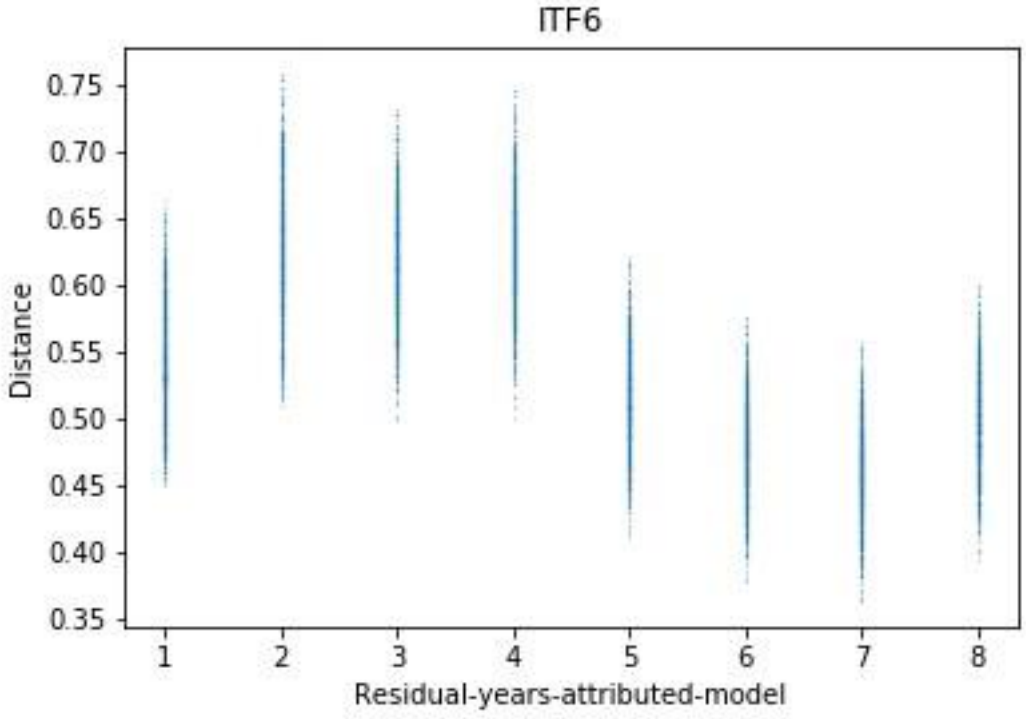

Fig. 35: ITF6, Calabria - Scatter plot of μ_p distribution against the number of years of modelled expenditure for the residual of the EC 2016 payment

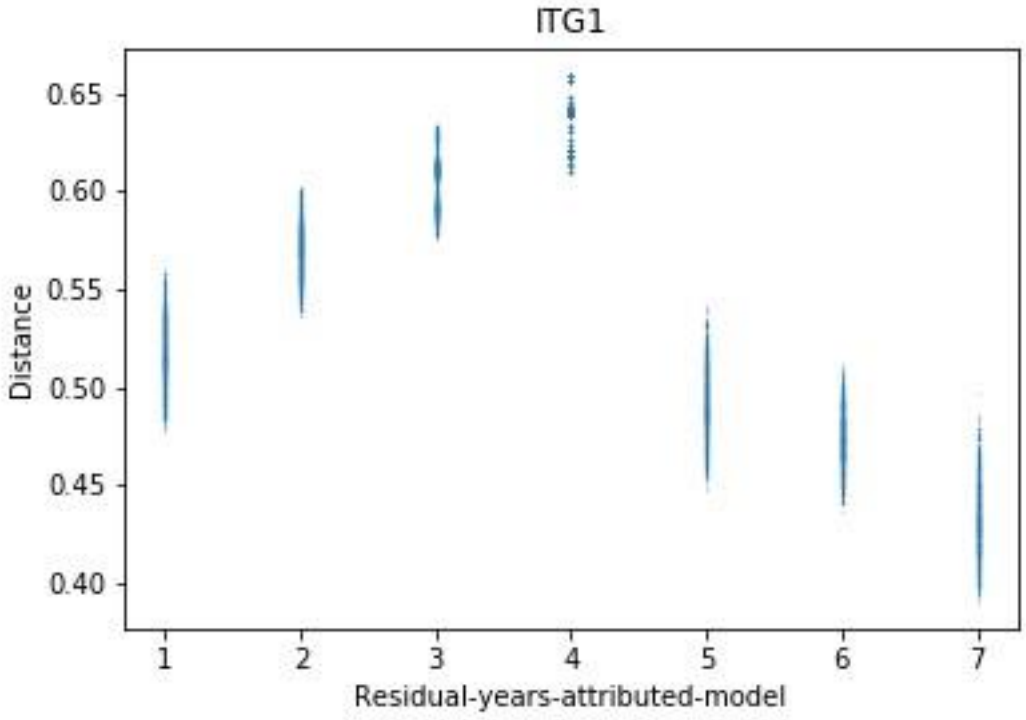

Fig. 36: ITG1, Sicily - Scatter plot of μ_p distribution against the number of years of modelled expenditure for the residual of the EC 2016 payment.

3.3. Conclusion on the comparison with MS data bases

Given the mismatch between the expenditures reported by MSs authorities and the reported EC reimbursed payments, it has not possible to update the algorithm for estimating expenditures. The original algorithm is therefore retained.

4. TASK 2 - UPDATE OF THE 1989-2015 DATABASE

4.1. Task 2.a - Update of the existing database of yearly EC payments at NUTS2 level with the information of the closure payment of the 2007-2013 programming period (2016-2018)

We have documented mismatches between the *Stata* input file and the output file we were provided with in the previous tender contract. The output figures were larger than those of the input database. We have finally managed to realign these figures and complete the regionalisation of the payments, which also included the most up-to-date payments remitted in the context of the 2007-2013 programming period. The last-eligible-year of expenditure rule was amended to a *n+2* scheme for the programming period 2007-2013, so that the last admissible year was 2015.

4.2. Task 2.b - Regionalisation at the NUTS2 level of the payments related to the 2014-2020 operational programme

DG REGIO provided us with the following elements to accomplish the regionalisation of these figures:

- An EC payment table updated on a daily basis: https://cohesiondata.ec.europa.eu/d/gayr-92qh
- A categorisation table with cumulative yearly figures detailed by the specific programme title, priority and dimension: https://cohesiondata.ec.europa.eu/d/3kkx-ekfq
- A look-up table with the mapping of the unique programme CCI codes onto the NUTS2 areas for cohesion policy funded programmes: https://cohesiondata.ec.europa.eu/d/466c-pqi8
- and a similar lookup table for EAFRD programmes: https://cohesiondata.ec.europa.eu/d/t6h3-7956

We firstly excluded the funding scheme beyond our analysis such as the *Interreg programmes*. Potential sources of double counting have also been filtered out, this included specific fund items such as *IPEA-contribution from ERDF, YEI ESF Matching Component, IPAE-contribution from ERDF*.

Some figures were reported as per the 2006, 2013 or 2016 Nomenclature of Territorial Units for Statistics (NUTS) revisions, which required the harmonisation to the 2010 nomenclature.

We have assessed the *Net Interim Payments* figures, from which yearly data could be easily retrieved by subtracting consecutive yearly cumulative figures.

The function of the look-up table was to allow a mapping onto the NUTS areas of the EC payments remitted.

4.2.1. Mapping the look-up table codes onto the programmes

The issue with the mapping from the lookup table was the lack of organisation at different granularity levels. All the codes were reported in the same column and efforts were requested to allocate this information according to its granularity level. This is an essential point to complete the regionalisation of the payments.

To this end, the number of characters in the codes was helpful to carry out a first separation: Codes having more than four characters could be safely attributed to the NUTS3 granularity level. NUTS2 have in turn four characters and could be isolated from the lookup table as per this condition. The level of geographical granularity we are interested into is the regional NUTS2, hence we could re-aggregate the NUTS3 information at NUTS2 level by cutting off the last character of the NUTS3 codes.

Payments remitted at the NUTS1 level had a three-character *NUTStitle* in the lookup table. Analogously, those attributed at the member-state level had two characters only.

These two latter categories are those that require to be regionalised, i.e. subdivided at the NUTS2 granularity level. European regions were broken down into four economical *Categories of region*: less developed, transitional and more developed, plus the geographical category *Outermost or Northern Sparsely Populated* areas. These two cross-cutting mappings can be used to entirely regionalise the reported EC payments for all the funds for which these categories have been specified.

The workflow to regionalise all the EC payments is illustrated in Figure 37 below.

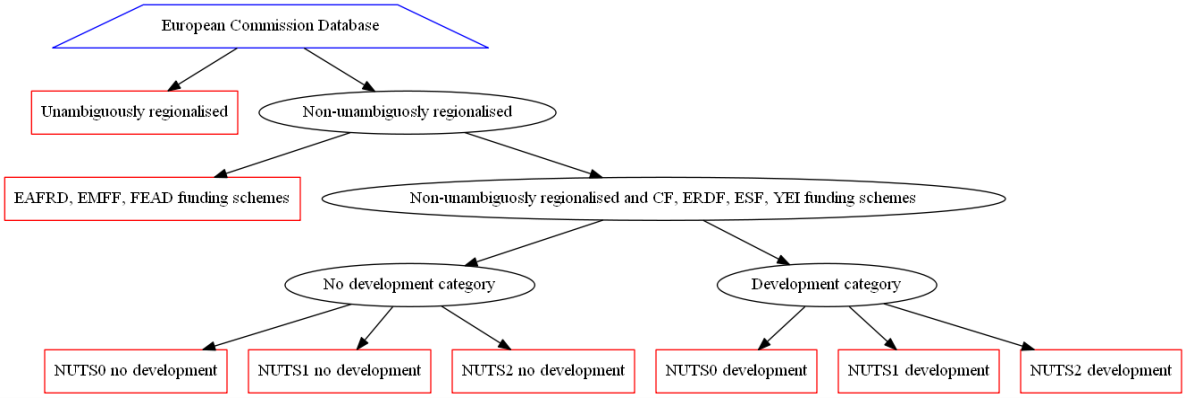

Fig. 37: The payments splitting sequence

First bifurcation - One can firstly separate the payments which have been clearly remitted at NUTS2 level, i.e. those for which it is possible to define an unambiguous one-to-one correspondence between the CCI code of the payment and the NUTS2 area it is remitted to.

Second bifurcation - One can then isolate the payments that are only remitted at the NUTS0 level. This occurs for the funding schemes European Agricultural Fund for Rural Development (EAFRD), European Maritime and Fisheries Fund (EMFF), Fund for European Aid to the most Deprived (FEAD).

Finally, one can disentangle the payments for which an economic/geographical category is specified - *third bifurcation*.

Fourth bifurcation - One can then separate the payments remitted at different NUTS granularity level (NUTS0, NUTS1 or NUTS2) within each of the resulting branches. The payments are finally regionalised on a pro-rata basis as per the population of the NUTS2 in a given year. In case of NUTS1 areas, the payment is not directly attributed at NUTS2 levels. The payment is firstly split on a pro-rata basis between the areas mentioned under the same programme CCI code. The different NUTS1 sub-payment are then split on a pro-rata basis on the NUTS2 composing the given NUTS1.

The only exception are the EMFF and EARDF funds for which metropolitan NUTS2 areas were excluded as out of scope for the typology of these funds.[2]

Let us analyse how the splitting varies dependent upon the specification of the development category.

Let us assume, for instance, that a payment *cash flow P* of 1,000 M euro has been remitted to three NUTS2 areas, named without loss of generality *region A, region B and region C*, whose population amount to 5 M, 3 M and 2M inhabitants, respectively, in the

[2] The following metropolitan regions were excluded : AT13 Wien; BE10 Région de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest; CZ01 Praha; DE30 Berlin; DE50 Bremen; DE60 Hamburg ; EL30 Attiki ; ES30 Comunidad de Madrid ; FR10 Île de France ; SE11 Stockholm ; UKD3 Greater Manchester; UKD7 Merseyside; UKG3 West Midlands ; UKI1 Inner London ; UKI2 Outer London.

year y we are taking into account. Hence, *region A* will receive $1,000M * \frac{5M}{5M+3M+2M} = 500M$. Analogously, *region B* will receive 300 M euro and finally *region C* 200 M euro. Let us now assume the category of development of the first two regions is *transitional*, while the third is *less developed*.

Let consider a new payment *cash flow Q* of 1,000 M euro, 400 M euro of which are attributed to a transitional area and the remainder 600 M euro to a less developed area. In this case, the 400 M euro will be split across the transitional areas as follows, 400 $M * \frac{5M}{5M+3M} = 250M$ to *region A* and the remaining 150 M euro to *region B*. 600 M euro will be entirely allocated to *region C* as this is the only less-developed region.

The contribution of each of the sub-categories identified in Figure 37 is detailed in Figure 28 below. 24% of the payments are already attributed to a single NUTS2, slightly more at the member-state level. The remainder half is mostly attributed according to development categories (one third of the total, with a prevalence at the member state level). Finally, about one sixth of the payments are attributed without any development criterion, mostly at the member-state level also in this case.

The splitting at different levels of granularity will be analysed more in-depth in task 3 below.

4.3. Task 2.c - Implementation of the existing methodology for estimating the expenditures on the newly added data

Given the reported mismatch between the expenditures reported by MSs authorities and the reported EC reimbursed payments in task 1, it has not been possible to update the algorithm for estimating expenditures. For this reason, we simply applied the previous algorithm to the new data of the programming period 2014-2020.

On the top of this, we briefly highlighted where the mismatch between the MSs and EC data stems from in the case of Italy. The term of comparison for the latter were the figures reported in the document **2007-2013 categorisation cohesion policy FIR.xlsx** provided in July 2019 (based on the categorisation of expenditure reported in the closure reports). The up-to-date EC figures are now exceeding those reported by the MSs authority by around 1 billion euro, 22.5 G euro against 21.5 G euro. The result of breaking down this difference by programme CCI code is illustrated in Figure 38 below. The sign of the difference is variable across programmes and most of it stems from the series 2017IT**161**XXXXX rather than 2017IT**162**XXXXX. For these latter, the differences amount to less than 100 M euro.

These findings are also confirmed in relative terms with figures up to -33% and plenty above 10% in magnitude for 2017IT**161**XXXXX programmes, while 2017IT**162**XXXXX programmes have only one difference that exceeds 10% (Figure 39). The reasons behind these mismatches should be thoroughly investigated to improve the consistency across information sources and update the expenditures estimate model.

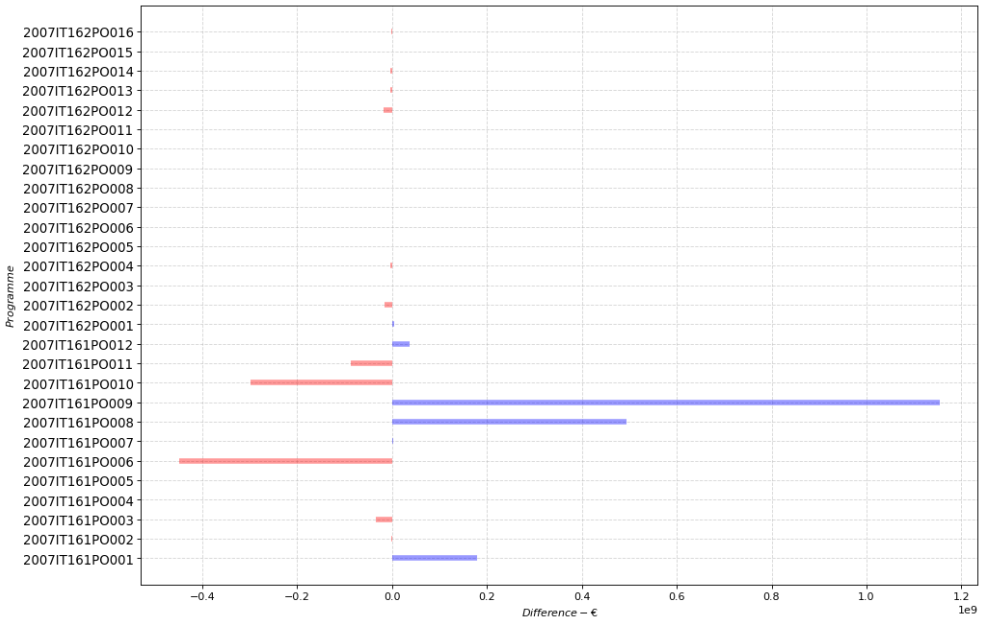

Fig. 38: Absolute diverging bars Programme amounts - Italy, 2007-2013, ERDF

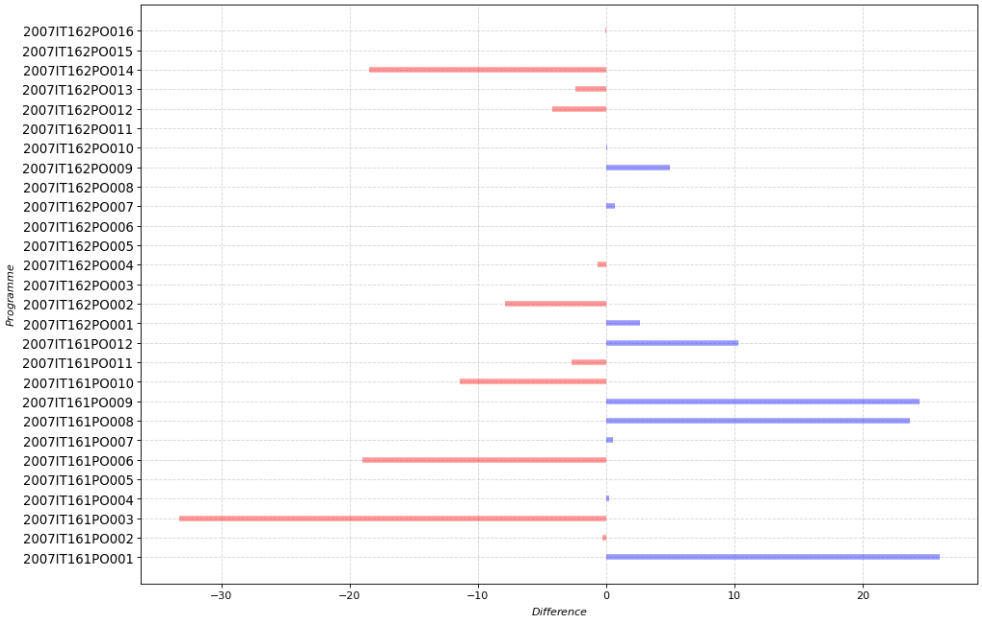

Fig. 39: Relative diverging bars Programme amounts in percentage - Italy, 2007-2013, ERDF

5. TASK 3 - TESTING OF THE 'LOCATION DIMENSION' CATEGORISATION

In this task, we reflect on how the data are produced by the member-state authorities in terms of the granularity level. The figures expressed as share of total allocations and are analysed across two categories: i) the funding scheme; ii) the member states. Results are presented in Table 1 and Table 2 below (Numbers may not exactly add up to 1 due to rounding).

Funding schemes such as EAFRD and FEAD are exclusively remitted at the MSs level. CF follows a similar pattern with 92% of the funds remitted at this level and 8% at NUTS1 granularity. ERDF, European Social Fund (ESF) and Youth Employment Initiative (YEI) present similar trends with around 70% of the payments allocated at NUTS0 (MS) level, this latter fund has also around 15% of the allocations at level NUTS1 and the

remainder NUTS2. The allocations of ERDF and ESF are similar and around 30% at NUTS2 level, 1% or less take place at NUTS3 level, with no share at NUTS1 level.

Briefly, the NUTS3 level comes into play only for these funds, ERDF and ESF, analogously to NUTS2. The payments are also disaggregated for YEI at this level, that also shows payments allocated at NUTS1 level. The same applies to CF.

Table 1

	CF	EAFRD	EMFF	ERDF	ESF	FEAD	YEI
NUTS0	0.92	1.0	1.0	0.68	0.72	1.0	0.67
NUTS1	0.08	NaN	NaN	NaN	NaN	NaN	0.14
NUTS2	NaN	NaN	NaN	0.32	0.28	NaN	0.19
NUTS3	NaN	NaN	NaN	0.01	< 0.01	NaN	NaN

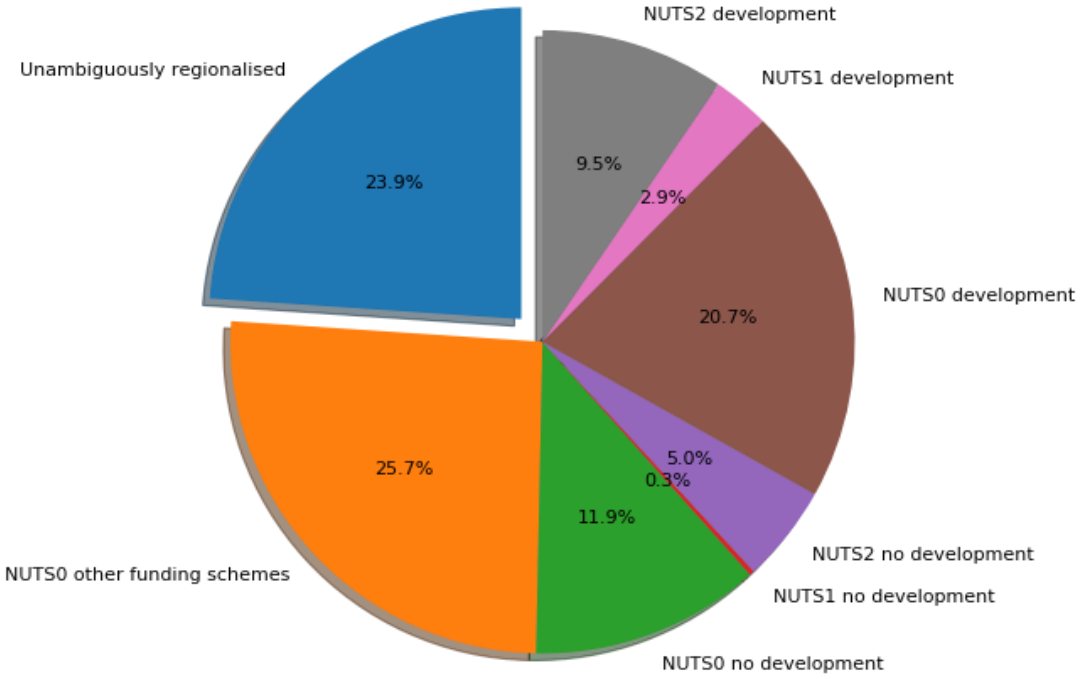

Fig. 40: The share of the various sub-categories of payments

The results for the geographical allocations are shown in Table 2 below (numbers may not exactly add up to 1 due to rounding).

Table 2: Numbers may add up to quantities larger than 1 due to rounding

	AT	BE	BG	CY	CZ	DE	DK	EE	ES	FI	FR	GR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
NUTS0	1.0	0.97	1.0	1.0	0.97	0.84	0.96	1.0	0.78	0.95	0.83	0.98	0.91	0.99	0.90	0.73	0.57	0.69	1.0	0.75	1.0	0.54	0.79	0.99	0.81	0.38	> 0.99	0.82
NUTS1	NaN	0.02	NaN	NaN	0.01	NaN	NaN	NaN	NaN	NaN	< 0.01	NaN	0.09	0.01	0.06	NaN	NaN	NaN	NaN	0.25	NaN	NaN	NaN	NaN	NaN	0.29	NaN	< 0.01
NUTS2	NaN	0.01	NaN	NaN	0.02	0.14	0.04	NaN	0.22	0.05	0.17	0.02	NaN	NaN	0.04	0.27	0.43	0.31	NaN	NaN	NaN	0.46	0.21	0.01	0.18	0.33	< 0.01	0.18
NUTS3	NaN	0.01	NaN	NaN	0.01	0.02	NaN	NaN	< 0.01	NaN	< 0.01	NaN	NaN	NaN	NaN	< 0.01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.01	NaN	< 0.01	NaN

Payments to six member states were exclusively remitted at NUTS0 level, which also played a prominent role for other ten MSs, for which NUTS0 amounted to more than 90%. MSs having the lowest shares of payments remitted at NUTS0 level were Slovenia (38%), Poland (54%) and Lithuania (57%). Allocations higher than 10% at the NUTS1 level were found for Slovenia (29%) and Malta (25%), while the highest at NUTS2 are for Poland (46 %), Lithuania (43%) and Slovenia (33%). Allocations at the NUTS3 level were always below the threshold of 3%, with the highest figure attained by Germany (2%). In general, nine MSs showed remittances at NUTS1 level, nineteen at NUTS2 and

eight at NUTS3 level. Belgium, Czech Republic and France are the only countries who have received payments filed across all the four granularity levels.

The use of data provided in the location dimension does not therefore provide a clear, systematic alternative to the rules of pro rata apportioning of expenditure to the NUTS2 level used to date for programmes that do not coincide with NUTS2 boundaries. The data might, for certain MS, allow a refinement of the pro rata rules on a case by case basis.

6. SUPPORTING MATERIAL

The script underpinning the computations performed is enclosed to the present document. It can also by launched from the project GitHub repository. (https://github.com/Confareneoclassico/DG_REGIO2)

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by email via: https://europa.eu/european-union/contact_en

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

The ESIF Open Data platform contains the consolidated historic EU payments dataset and a data story on the following links:

- Dataset: https://cohesiondata.ec.europa.eu/d/tc55-7ysv
- Data story: https://cohesiondata.ec.europa.eu/stories/s/47md-x4nq

KN-03-20-336-EN-N

Publications Office
of the European Union

ISBN: 978-92-76-19200-8