

Method of Disaggregating Annual Time Series into Seasons

Richard Hindls¹, Stanislava Hronova^{2,*}

¹Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, Prague, Czech Republic

²Department of Economic Statistics, Faculty of Informatics and Statistics, Prague University of Economics and Business, Prague, Czech Republic

Email address:

hindls@vse.cz (Richard Hindls), hronova@vse.cz (Stanislava Hronova)

*Corresponding author

To cite this article:

Richard Hindls, Stanislava Hronova. Method of Disaggregating Annual Time Series into Seasons. *International Journal of Statistical Distributions and Applications*. Vol. 9, No. 1, 2023, pp. 1-8. doi: 10.11648/j.ijstd.20230901.11

Received: January 15, 2023; **Accepted:** February 1, 2023; **Published:** February 9, 2023

Abstract: A number of indicators (especially economic ones), whose values are monitored with annual periodicity, need to be broken down into the so-called seasons, i.e., time periods shorter than one year. And it is necessary not only to respect the specifics of these seasons, but at the same time to do so without having to specifically survey these values in the seasons. The aim of the paper is to propose a method for distributing the annual values of an indicator into seasons (quarters, months) using a suitably chosen indicator, without subsequent correction. This will preserve the link to the processes evolving within the year, while removing the formalism of splitting the residual generated in the first step. The paper presents a new approach, in which simulated values of the response variables enable us to estimate the series' parameters with the aid of a simple loss function. Such a solution reduces the dependence of formal models on real data, which may or may not be available to the official statisticians who are responsible for statistical surveys. The paper shows the theoretical concept of the disaggregation method, which is the result of research in the given area and was verified on the example of data for the Czech Republic.

Keywords: Disaggregation Methods, Retropolation, Seasonal Variations, National Accounts, Regression Model

1. Introduction

A common practical task we often encounter, especially in economics, is to decompose the values of a certain indicator (let us denote it by the symbol y), which is available as a sequence of its annual values in calendar years $t = 1, 2, \dots, n$, into shorter time periods (into the so-called seasons, of which there are s within each given year). These seasons are naturally repeated from year to year.¹ The purpose of this decomposition is largely pragmatic: in situations like this, it is assumed that the values of y are not observed in each season (most often due to cost-saving or organisational circumstances), but they need to be known for certain reasons (e.g., for economic decision-making at the national-economy level). Despite the pragmatic goals, this procedure is rather non-trivial because of the formal tools that are, from the

mathematical and factual viewpoints regarding the underlying economic indicators, necessary to actually decompose the annual values into the seasons.

Each season can have a completely different character (due to climatic, social, legislative, budgetary, cultural, etc. factors). The aggregate annual value of y cannot therefore be distributed in a trivial way, e.g., in proportion to the corresponding uniform fraction $1/s$, or in any similar simple way. In other words, the qualitatively different frameworks of each of the seasons in question do not admit capturing in such a trivial way.

In order to solve this problem, it is usually necessary to have the values of at least one other indicator (let us denote it by the symbol x), which in its turn is commonly observed within seasons and which is in a significant (regression) relationship to the y indicator. It is undoubtedly linked to the y indicator in a material way. The x indicator thus imprints qualitatively significant traces of the seasons (in the form of seasonal fluctuations) on the seasonally distributed values of the y indicator.

¹ The most frequent instances are those of months, for which $s = 12$, or quarters, for which $s = 4$.

Without going into the details of any particular case, which is not the goal of the present text, we will illustrate the presented idea on a generally relevant example, from which the research in this direction stemmed and which is now cited as a practical implementation of the disaggregation idea. Namely, it is a task typical for economic statistics: the complexity of constructing the annual values of the GDP aggregates, or other aggregates present in the national accounts system, has led to the idea of using the so-called indirect methods for estimating the values of quarterly aggregates; the latter methods make use of existing short-term surveys' results as a basis for disaggregating the values of annual national accounts aggregates into quarterly values. This indirect method of estimating the quarterly values of the national account aggregates, the so-called scheduling method, consists in linking the values of short-term survey (i.e., 'regression-linked') indicators with the values of the annual national account aggregate values. The underlying model for this procedure is based on the relationship between the short-term indicator (x variable) and the annual national account aggregate (y variable). Such a model enables us to distribute the known annual values into individual quarters so that a sum over them is equal to the annual value.

Historically, two formal approaches have emerged in this context:

- 1) *distribution with subsequent correction*, i.e., the sum of the seasonal values obtained in the first step from the regression relationship does not equal the annual value, and the observed difference between the known annual value (e.g., the national accounts aggregate) and the sum of the seasonal values must then be re-distributed (according to a chosen formal criterion) into the individual seasons (second step);
- 2) *distribution without subsequent correction*, i.e., the construction of the regression model ensures that, in each year of the chosen time series, the sum of the distributed values (already in the first step) corresponds to the known annual value of the y indicator.

Since our time series always comprises annual values of an aggregate, it is clear that disaggregation into seasons could be performed with the aid of a time series model without using the short-term indicator as an explanatory variable. This approach would also be acceptable from a theoretical point of view (in contrast to our model that uses a short-term indicator), but it would remove responses to the real phenomena occurring in the relevant year.

The aim of the present paper is to propose a method for distributing the annual values of an indicator into seasons (quarters, months) using a suitably chosen indicator, without subsequent correction. This will preserve the link to the processes evolving within the year (given by the choice of an appropriate short-term indicator), while removing the formalism of splitting the residual generated in the first step. This method has resulted from the research in this field and has been validated on an example using data from the Czech Republic. Here we present only the theoretical concept of our method.

2. Theoretical Background

Since the 1960s, many authors have been working on the issue of distributing annual values into seasons. Their interest is implied by pragmatic need; moreover, the non-trivial nature of the task poses a certain challenge not only for users (mainly economists) but also for statisticians and their modelling approaches. Moreover, because of the pragmatic need, the solution to the problem is repeatedly re-considered in attempts to achieve better results than those accomplished in the past. In many of these texts, the term of retropolation is used to describe a procedure that decomposes a particular aggregate indicator into seasons; retropolation thus adequately captures the essence of the problem and can be considered synonymous with disaggregation, which is the term used in the title of the present paper.

In addition to a strong theoretical basis, there are many practical reasons for addressing the problem of disaggregating annual values of a certain indicator into seasons (quarters or months). In particular, the requirements for the rapid publication of short-term (especially quarterly) values of macroeconomic aggregates have led national statistical offices to seek ways to meet the requirement for rapid, reliable quarterly estimates that are consistent with the annual values of the aggregates. Disaggregation methods have gradually emerged, with or without the use of another (quarterly or monthly) indicator as an explanatory variable; in the former case when a short-term indicator is used, both post-correction (i.e., two-step) and non-post-correction (i.e., one-step) distributing methods have been developed.

The first author who came up with the idea of disaggregating annual values into seasons using short-term values of another indicator was undoubtedly [12]. In his work, he formulated a hypothesis of a causal relationship between the short-term indicator and the annual aggregate, and subsequently used a time-series regression model to estimate the quarterly values. The method of disaggregation without using the short-term surveyed indicator, based on the principle of smoothing, was proposed by Boot, J. C. G., *et al.* [2]. They assumed that the unknown short-term trend of the time series could be described by a function of time and that the condition of consistency between the annual value of the aggregate and the sum of its distributed quarterly values was satisfied, as well as the smoothness condition for the transition of the quarterly values from year to year.²

A particular implementation of the two-step disaggregation procedure (i.e., with subsequent correction) in the conditions of government-oriented statistics was then formulated by Nasse [18]. He not only described a general model of distributing the annual values into quarters on the basis of a regression (usually linear) model (the first step), but also proposed a simple method for distributing the differences between the known annual values and the sums of the first-step-distributed values using a simple matrix operator. His model became the basis for the

² This so-called smoothing method later became, among other things, the recommended method for allocating differences between the known annual aggregate values and the sums of the seasonally distributed values ([10] and [9]).

system of quarterly national accounts in France. Subsequently, Bournay and Laroque made important contributions to improving the methods used [3]. Namely, they proposed to estimate a regression model assuming autocorrelation of the residuals, which allowed them to reduce the standard error of the estimation and significantly improve the quality of the disaggregation. In the second step, they replaced the matrix operator by a criterion for minimizing first differences between outliers. Their approach is still used as a basis for not only the French quarterly national accounts, but also as the two-step disaggregation method recommended by Eurostat [10], i.e., the disaggregation-with-subsequent-correction method.

Another principal work that influenced the development of disaggregation methods was undoubtedly that of Chow and Lin [6]. The authors proposed a method of disaggregating annual values using a short-term indicator (or several indicators) as an explanatory variable, with quarterly aggregate values estimated in one step. The existence of the regression model then allowed users ([18], or [3]) to also estimate the quarterly values for the current year (whose annual value is not yet known). The Chow and Lin's method [6] is used by certain national statistical offices (e.g., in Spain and Italy).³

The regression model of Chow and Lin [6] did not contain a dynamic component. Therefore, modified models with different assumptions about the residual component have gradually emerged. To name a few, Fernandez [11] used a random walk model, Litterman [17] and Di Fonzo [8] followed Fernandez [11] and used a Markov model, or Gregoire [13] proposed a simple dynamic model to improve the properties of quarterly estimates in the French national accounts.

The above-mentioned procedures and models today represent not only the basis for quarterly national accounts in many European countries, but also a stepping stone for further considerations on how to formally improve disaggregation models or how to use them outside the scope of national accounts, or even outside the framework of time-wise disaggregation.

A theoretical and formal framework for estimating partial values of time series was presented by Caporin and Sartore [5]. It provided a general approach, which involved a backward decomposition of time series values, but not only in time (which is strongly dominated in the literature) but also in spatial allocation. It also pointed out an important circumstance that linear models might rank among the preferred types of aggregate data allocations. Their methodology was designed with a focus on economic time series, but was also applicable to other statistical domains. It also provided an empirical example of backward decomposition of the EU15 industrial production index and compared it with the Eurostat approach and methodology [10].

An extension of this task is described in the research [4]; the author builds an apparatus based on not only a regression retropolation model but also a regression interpolation one. In constructing the formal methodology, he thus combines both interpolation and retropolation. He not only applies his approach to national accounting aggregates, but also

complements the methodology with a distribution of non-monetary indicators' values, such as employment rate and its decomposition into breakdowns by sectors. The data he uses come from Liechtenstein statistics. He further compares the results of his decomposition method with the official national accounts data and finds a satisfactory agreement between the methods used. Guerrero and Corona mainly present the important role of indicator databases, which can positively affect the quality of the retropolation [15]. They use a seasonally (quarterly) surveyed index of economic activity in Mexico as the reference variable in regression.

Prados de la Escosura discussed comparisons of economic performance over space and time and argued that the quality of the results [20], to a large extent, depends on the links between the actual statistical data from national accounts and their historical estimates. An illustration of this linking is made in his paper with the aid of a specific case – the Spanish economy in the second half of the 20th century. The author is perhaps too sceptical about the possibilities of retropolation, concluding that the usual procedure of linking different indicators by means of retropolation may have its shortcomings, as the method tends to bias the level of GDP upwards and consequently underestimate growth rates, especially for emerging countries undergoing structural changes. He therefore proposes an alternative approach, which is, however, based on classical interpolation methods. Our paper, on the contrary, seeks to use standard regression procedures in disaggregation, so that they are part of, and not an alternative to, retropolation.

Billio, Caporin and Cazzavillan aim to decompose the aggregate time series describing the EU15 business cycle into monthly and quarterly data, using both parametric and non-parametric techniques [1]. The basic idea is to use the monthly industrial production index (IPI) series of the EU15 countries (from 1970 to 2003) as a basis for decomposing the GDP of the EU15 countries into months. In other words, the monthly GDP data is calculated using time disaggregation techniques. Prados de la Escosura follows a similar approach, using a decomposition to calculate the performance of the Spanish economy in the last 170(!) years [21]. In doing so, he relies on estimates of net capital and services as short-term indicators and then describes the different stages of the Spanish economy's development through the prism of investment volumes. However, having in mind the length of the period covered (170 years), the considerations are more or less hypothetical.

A somewhat less numerous, but certainly not sporadic group of articles is represented by those that present procedures in which the values of certain indicators are decomposed into smaller units rather than retropolation (in the sense of decomposing annual values of an indicator into seasons); the temporal decomposition is, however, not the primary one. For example, Reymann's paper [22] presents a rather original and non-trivial idea of creating a certain future scenario (strategy formation and its implementation) based on the retropolation of future scenarios. He combines the strategy and scenario technique with a method that allows assessing the strategy by

3 Cf. [10].

disaggregating it into sub-scenarios (i.e., subsets) of the strategies. In effect, such an approach to disaggregation creates intermediate steps (i.e., sub-strategies) on the path from the current situation to the desired state. Based on this decomposition, the intermediate goals and objectives are defined that are necessary to implement the overall strategy. In other words, goal-based strategic planning decomposed into individual process strategies is considered there. Guérois works in a completely different application environment, namely, that of the statistical spatial classification and geographical distribution of Local Administrative Units (LAUs) [14]. It is based on the Urban Morphological Zones (UMZ) database. The decomposition of values is traced from the current UMZ 2000 back to 1961. The historical population database of European LAUs allows them to assign the population for the period 1961 through 2011, in the profile and geometry of LAU 2012. The disaggregation model created in this way captures the evolution and changes in the urban districts of the UMZ 2000 back to 1961. Pfeuffer and Scherb, consider the retropolation methods in the context of analysing new business areas [19]. He combines extrapolation and retropolation techniques, but at the same time makes a consistent distinction between them. In doing so, the latter are used to analyse interactions between aggregate strategic requirements on the one hand, and

discrepancies in the fulfilment of partial objectives on the other hand. Overcoming these mismatches or contradictions is embodied in the strategic areas of future-oriented business (predictive strategies).

As can clearly be seen from the literature review, the problem of decomposition of aggregate quantities, either in the temporal or spatial dimension, has been of interest to a number of authors who choose many hands-on applications for illustration. This is evidence of the importance of the topic, which we will also address below. The aim of this paper is to present a method of disaggregation in which, by simulating the values of the explained variable, the parameters can be estimated using a simple loss function.

3. Disaggregation Model

From the formal point of view, we consider the y_{ij} indicator to be disaggregated. Its values also make sense for $j = 1, 2, \dots, s$ (e.g., they are interpreted as quarterly values for $s = 4$, or monthly values for $s = 12$, etc.) in the years $t = 1, 2, \dots, n$. We thus define a variable, y_{ij} , whose values admit disaggregation into the j seasons. There are two possible approaches to estimating them:

- 1) Let us, for the years $t = 1, 2, \dots, n$ and seasons $j = 1, 2, \dots, s$, define an N -component vector of the response variable

$$y = [y_{11}, y_{12}, \dots, y_{1s}, y_{21}, y_{22}, \dots, y_{2s}, \dots, y_{n1}, y_{n2}, \dots, y_{ns}]' \quad (1)$$

where $N = n \cdot s$ and s is the number of seasons in each year. Disaggregation of the values into the seasons can reasonably be interpreted. The information about the values shown in Formula (1) will later be important for assessing the disaggregation quality. Let us further assume that the N -component vector of the response variable is governed by the classical linear model

$$y = X\beta + \varepsilon, \quad E(\varepsilon) = 0_N \text{ a } E(\varepsilon\varepsilon') = \sigma^2 I_N, \quad (2)$$

where X is a deterministic $N \times p$ matrix of known explanatory variables, $1 < p < N$, with a rank of p , β stands for a p -component vector of unknown parameters and ε is an unobservable irregular N -component vector of mutually independent random errors with zero means and constant

variance value σ^2 (fulfilling the homoscedasticity assumptions).

- 2) We will consider a situation in which the y vector defined in Formula (1) is unknown but the annual (that is, seasonally-non-disaggregated) values are known, arranged in an n -component vector

$$Y = [Y_1, Y_2, \dots, Y_n]' , \quad (3)$$

where the following notation

$$Y_t = \sum_{j=1}^s y_{tj}, \quad t = 1, 2, \dots, n, \quad (4)$$

can be used for the annual sums. It means that each annual value equals a sum of its disaggregated values. A linear relation

$$Qy = Y, \quad (5)$$

holds true for vectors y and Y , and

$$Q = \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \end{bmatrix}$$

is a deterministic $n \times N$ matrix.

Let us now define an $n \times p$ deterministic matrix W

$$W = QX \quad (6)$$

and an n -component vector ζ of unobservable irregular random errors such that

$$\zeta = Q\epsilon. \quad (7)$$

Since, within the model for y whose validity we assume, the identity

$$QQ' = s I_n \quad (8)$$

holds true, we get the following formula for the structure of the Y vector:

$$Y = W\beta + \zeta, \quad E(\zeta) = 0_n \text{ and } E(\zeta\zeta') = s\sigma^2 I_n. \quad (9)$$

An original method for disaggregating the annual values into seasons is represented by the following formula:

$$u = s^{-1}Q'Y + (X - s^{-1}Q'W)B, \quad (10)$$

where the N -component vector u can be interpreted as an estimate of the unknown vector y , and B stands for an estimate of the unknown vector β . A specific feature of the estimate is worth mentioning: the formula

$$Qu = Y, \quad (11)$$

holds true for any vector B , which means that this estimate preserves the known annual values.

However, this method of disaggregation brings about another problem, which is usually not referred to in the literature; namely, what should be our choice of the vector B – an estimate of the unknown vector β ?

a) Such a problem would not arise if we knew the vector y in Formula (1) because then the estimate would be derived with the aid of the standard least square method:

$$B_0 = (X'X)^{-1}X'Y \quad (12)$$

for an $N \times p$ matrix X , representing the estimate with the optimum features. Since the following identity

$$E(B_0) = \beta, \quad (13)$$

is valid, it is an unbiased estimate achieving the minimum square of error, namely:

$$R(B_0, f) = E(f'(B_0 - \beta)^2) = \sigma^2 f'(X'X)^{-1}f, \quad (14)$$

where f stands for the given p -component non-zero vector.

In this situation, we know the vector y shown in Formula (1). This is important because the actual values, as mentioned above in connection with Formula (1), will actually be used as criteria for the quality of the estimates. We now proceed to

obtain an estimate of vector β . However, the solution found will only be a starting point for the actual disaggregation comparisons, in particular for a possible specification of the losses that we will have to admit in the parameter estimation.

b) Let us now consider an estimate based on the model for the annual sums, that is,

$$B_1 = (W'W)^{-1}W'Y. \quad (15)$$

With respect to the structure of the vector whose components are the annual values, this estimate can be taken for unbiased. In comparison with the ideal estimate B_0 , this one is a reduced version: instead of the N values contained in the rows of the X matrix, just $n = N/s$ values contained in the W matrix are used, deliberately ignoring the information given by the rows of the X matrix. In other words, the known values of the explanatory variables in individual years' seasons are neglected. This is another reason why we will work on additional estimates below.

Simulating y Vector's Components

Under the circumstances described above, simulation of the elements of the y vector's components appears to be a viable way to improve the statistical properties of the estimates for the β parameter vector. Let us consider several options stemming from two starting points.

a) Recall that knowledge of the response variable's annual values Y_t , $t = 1, 2, \dots, n$, is assumed. The average values of the response variable per season can, in each year, be determined by calculating

$$y_{0,tj} = s^{-1}Y_t \text{ constant for seasons } j = 1, 2, \dots, s \text{ in years } t = 1, 2, \dots, n; \quad (16)$$

in the first approximation, they can be considered the trend components of the response variable time series.

b) We further suppose that we have at our disposal an $N \times 2$ matrix of explanatory variables; after transposition, its first row contains all entries equal to 1 and the second row the values of the sole explanatory variable

$$x_{11}, x_{12}, \dots, x_{1s}, x_{21}, x_{22}, \dots, x_{2s}, \dots, x_{n1}, x_{n2}, \dots, x_{ns}. \quad (17)$$

The β vector of parameters contains two scalar values then: a free parameter β_0 and the slope value β_1 .

For the above-defined sequence of the response variable's annual values, the given variance value comparable to that of the variance corresponding to the sequence $y_{0,tj}$, $t = 1, 2, \dots, n$, $j = 1, 2, \dots, s$, denoted by S^2 , a new sequence of random variables can be simulated

$$y_{1,tj} = y_{0,tj} + \psi_{tj}, \quad t = 1, 2, \dots, n, \quad j = 1, 2, \dots, s, \quad (18)$$

where

$$\psi_{tj}, \quad t = 1, 2, \dots, n, \quad j = 1, 2, \dots, s \quad (19)$$

are the simulated values of the stochastic component generated by a random sample from, say, the normal distribution with zero mean and variance equal to S^2 .

In this context, the model

$$y_{1,tj} = \beta_0 + \beta_1 \cdot {}_a x_{tj} + \eta_{tj}, \quad t = 1, 2, \dots, n, \quad j = 1, 2, \dots, s, \quad (20)$$

can be used for estimating the β vector with two components β_0 and β_1 . In addition to variables already defined for $t = 1, 2, \dots, n$ and $j = 1, 2, \dots, s$, two more sequences are relevant: ${}_a x_{tj}$, which contains seasonally adjusted time series of the explanatory variable x_{tj} , and a sequence of random errors η_{tj} . The estimates in this model can be obtained with the aid of the standard least squares, and the estimate b_1 of the unknown parameter β_1 is calculated as a ratio of covariance and variance, that is,

$$b_1 = \frac{\text{cov}(y_1, {}_a x)}{\text{var}({}_a x)}. \quad (21)$$

Simulations can also be taken as acceptable if they reflect the initial idea, namely, that the simulated explanatory variable is considered rather similar to the explanatory variable, and therefore it is also comparable to it regarding the nature of the seasonal component. In our view, if such a hypothesis is justified, the seasonal component contained in the sequence of the explanatory variable can also be grafted instead of the simulated explanatory variable, thereby further specifying it for our purposes.

This idea can be reformulated as follows: suppose that, in years $t = 1, 2, \dots, n$ and seasons $j = 1, 2, \dots, s$, the explanatory variable x_{tj} can be described with the aid of a multiplicative model, i.e., a system of seasonal indices I_{tj} that systematically change their values in seasons $j = 1, 2, \dots, s$ and are repeated each year, while the following formula holds true:

$$s^{-1} \sum_{j=1}^s I_{tj} = 1; \quad (22)$$

this formula represents an additional assumption that the arithmetic mean of the seasonal indices within one year equals 1 and a sum of those indices equals s . If an assumption is justified that that seasonal effects are equally strong on all variables. the response variable can be simulated in the form

$$y_{3,tj} = y_{2,tj} + \eta_{tj}, \quad t = 1, 2, \dots, n, \quad j = 1, 2, \dots, s, \quad (23)$$

where

$$y_{2,tj} = I_{tj} \cdot y_{0,tj}, \quad t = 1, 2, \dots, n, \quad j = 1, 2, \dots, s \quad (24)$$

is an estimate of the mixed model comprising regular (trend and seasonal) components for the time series in question. We can say that, in this model, the deterministic component of the time series equals a constant value of $I_{tj} \cdot y_{0,tj}$ and the corresponding stochastic component equals the simulated value of ψ_{tj} for all $t = 1, 2, \dots, n$ and $j = 1, 2, \dots, s$ in the sense specified above.

Two comments should be made in this context.

a) Let us consider the sums

$$\sum_{j=1}^s y_{2,tj} = \sum_{j=1}^s I_{tj} \cdot y_{0,tj}. \quad (25)$$

for the t -th year, $t = 1, 2, \dots, n$. From the definition of the variable $y_{0,tj}$ it follows that

$$\sum_{j=1}^s y_{2,tj} = s^{-1} Y_t \sum_{j=1}^s I_{tj}. \quad (26)$$

holds true for all $t = 1, 2, \dots, n$, and $j = 1, 2, \dots, s$. With regard to the identity

$$\sum_{j=1}^s I_{tj} = s, \quad (27)$$

the following formula is obtained

$$\sum_{j=1}^s y_{2,tj} = Y_t, \quad t = 1, 2, \dots, n, \quad (28)$$

which can be interpreted as follows: the variables $y_{2,tj}$ in the given year t , $t = 1, 2, \dots, n$, can be taken for rough estimates of the response variable's unknown values y_{tj} in seasons $j = 1, 2, \dots, s$.

b) If the hypotheses stated in this paragraph about the simulated values of the response variable are justified, the linear model

$$y_{3,tj} = \beta_0 + \beta_1 x_{tj} + \omega_{tj}, \quad t = 1, 2, \dots, n, \quad j = 1, 2, \dots, s \quad (29)$$

can be used and the estimate b_2 of the parameter β_1 can be calculated by the least squares method, i.e., using the known formula

$$b_2 = \frac{\text{cov}(y_1, x)}{\text{var}(x)}. \quad (30)$$

It turns out that, by simulating the values of the response variable, we can achieve a relatively simple approach to modelling, namely, estimating the necessary parameters using a simple loss function provided by the least square method. In fact, any simplification of this – not entirely trivial – task alleviates the dependence of formal models (in this situation, of course, parametric) on real data, which may (or may not) be available to the official statisticians conducting the surveys.

The context of simulations is also important for another reason. The deterministic component of the response variable's time series is equal to the constant value of $I_{tj} \cdot y_{0,tj}$, and the corresponding stochastic component is obtained by simulation. This is important from the point of view of the future work, as discussed at the end of the present paper.

4. Conclusion

From a mathematical point of view, our approach is formally classified as parametric. However, from a pragmatic

point of view and from our past experience, it is difficult to determine which of the two approaches – parametric vs. non-parametric – leads to better results. Classification and evaluation of such methods is difficult, among others due to the fact that the quality of the disaggregation achieved depends not only on the chosen approach, but also on the specific data actually available from statistical surveys. Put another way, the specific data given in the non-stochastic matrix of known explanatory variables in Formula (2) is important. Disaggregation levels of the y vector according to Formula (1) are also important. Certain variables have a high degree of aggregation (e.g., national account aggregates), but it may be difficult to find a suitable explanatory reference variable for them. And even the reverse is true. Hence in practice, disaggregation methods are not directly used to estimate quarterly GDP or GVA of the national economy, but to estimate quarterly values of, for example, output and intermediate consumption values by sector, from which GVA by sector and then GVA of the national economy are subsequently estimated.

To a certain extent, a separate problem of the disaggregation problem is the formulation of the seasonality model's concept: how the character of seasonal variations should be translated into the disaggregation. These variations, of course, evolve in time, changing their qualitative manifestations as well as their relationships to the trend. In the concept of this paper (cf. Formula (22)), we assume that the arithmetic mean of the seasonal indices within a given year is equal to one and their sum is exactly equal to s . Our simulation concept is based on these assumptions. However, there is no doubt that other concepts of seasonality could be incorporated into the model. Therefore, we would like to address to this issue in the future, or switch to a different assumption about the type of the seasonal components.

Last but not least, our considerations also largely depend on the length of the time series we have available. The literature review in the Theoretical Background Section shows that some authors try allocation procedures on time series of various indicators that are many decades long. This is undoubtedly tempting from a purely statistical point of view; on the other hand, the relationships between the x and y variables y and x are not constant in time and are subject to significant variations (leaving aside the changing methodological and legislative practices under which the statistical indicator values are collected). Therefore, the choice of disaggregation methodology should not be based on unimaginative stereotypes.

In future our work, we would therefore like to arrive at a certain pragmatic taxonomy based on our empirical experience: how the quality of the variables that enter the parametric model should be determined and what variables represent the relevant outcome of statistical surveys.

Acknowledgements

This work was supported by Faculty of Informatics and Statistics, Prague University of Economics and Business.

References

- [1] Billio, M., Caporin, M. & Cazzavillan, G. (2008). Dating EU15 Monthly Business Cycle Jointly Using GDP and IPI. *Journal of Business Cycle Measurement and Analysis* 2007/3: 333-366. DOI: <https://doi.org/10.1787/jbcma-v2007-art16-en>.
- [2] Boot, J. C. G., Feibes, W. & Lisman, J. H. C. (1967). Further methods of derivation of quarterly figures from annual data. *Applied Statistics* 16: 65-75. DOI: <https://doi.org/10.2307/2985238>.
- [3] Bournay, J. & Laroque, G. (1979). Réflexions sur la méthode d'élaboration des comptes trimestriels. *Annales de l'INSEE* 36: 3-30. DOI: <https://doi.org/10.2307/20075332>.
- [4] Brunhart, A. (2012). Identification of Liechtenstein's historic economic growth and business cycles by econometric extensions of data series. KOFL Working Papers 14. Universität Liechtenstein, Konjunkturforschungsstelle (KOFL), Vaduz. Available at: <http://hdl.handle.net/10419/97330>.
- [5] Caporin, M. & Sartore, D. (2006). Methodological Aspects of Time Series Back-Calculation. University Ca' Foscari of Venice, Dept. of Economics. Research Paper Series 56/06. DOI: <http://dx.doi.org/10.2139/ssrn.950923>.
- [6] Chow, G. & Lin, A. L. (1971). Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *Review of Economics and Statistics* 43: 372-375. DOI: <https://doi.org/10.2307/1928739>.
- [7] Denton, F. T. (1971). Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization. *Journal of the American Statistical Association* 66: 99-102. DOI: <https://doi.org/10.2307/2284856>.
- [8] Di Fonzo, T. (1987). La stima indiretta di serie economiche trimestrali. Padova: Cleup.
- [9] Dureau, G. (1991). Les comptes nationaux trimestriels. INSEE – Méthodes 13. Paris: INSEE.
- [10] Eurostat. (2013). Handbook on quarterly national accounts. Luxembourg: Eurostat.
- [11] Fernández, R. B. (1981). A Methodological Note on the Estimation of Time Series. *The Review of Economics and Statistics* 63: 471-476. DOI: <https://doi.org/10.2307/1924371>.
- [12] Friedman, M. (1962). The interpolation of time series by related series. *Journal of the American Statistical Association* 57: 729-757. DOI: <https://doi.org/10.2307/2281805>.
- [13] Gregoir, S. (2003). Propositions pour une désagrégation temporelle basée sur des modèles dynamiques simples. In Workshop on Quarterly National Accounts. Paris-Bercy 05/12/1994-06/12/1994: 151-166. Luxembourg: Office for Official Publications of the European Communities. Available at: <https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/ks-ao-02-001>.
- [14] Guérois, M. et al. (2019). A harmonized database to follow the demographic trajectories of European cities, the TRADEVE database. *Cybergeo: European Journal of Geography*. Document 892. DOI: <https://doi.org/10.4000/cybergeo.32077>.

- [15] Guerrero, V. M. & Corona, F. (2018). Retropolating some relevant series of Mexico's System of National Accounts at constant prices: The case of Mexico City's GDP. *Statistica Neerlandica* 72: 495-519. DOI: <https://doi.org/10.1111/stan.12162>.
- [16] Kozák, J., Hindls, R. & Hronová, S. (2000). Some remarks to the methodology of the allocation of yearly observations into seasons. *Statistics in Transition* 4: 815–826.
- [17] Litterman, R. B. (1983). A Random Walk, Markov Model for the Distribution of Time Series. *Journal of Business & Economic Statistics* 1: 169–173. DOI: <https://doi.org/10.2307/1391858>.
- [18] Nasse, P. (1973). Le systeme des comptes nationaux trimestriels. *Annales de l'INSEE* 14: 119-165. DOI: <https://doi.org/10.2307/20075174>.
- [19] Pfeuffer, V. & Scherb, B. (2016). Identification of New Business Fields and Development Directions Using Contradictions. *Procedia CIRP* 39: 197-202. DOI: <https://doi.org/10.1016/j.procir.2016.01.188>.
- [20] Prados de la Escosura, L. (2016). Mismeasuring Long Run Growth. The Bias from Spliced National Accounts. *Cliometrica* 10: 251-275. DOI: <https://doi.org/10.1007/s11698-015-0131-4>.
- [21] Prados de la Escosura, L. (2022). Capital in Spain 1850–2019. *Cliometrica* 16: 1–28. DOI: 10.1007/s11698-020-00221-2.
- [22] Reymann, F. (2013). Verfahren zur Strategieentwicklung und-umsetzung auf Basis einer Retropolation von Zukunftsszenarien. Paderborn: Heinz Nixdorf Institut.