Disaggregation of regional GDP series in Poland with a simple smoothing algorithm

Jan Acedański[1]

PRELIMINARY AND INCOMPLETE

This version: 15/02/2021

Abstract

We propose a simple smoothing algorithm to disaggregate regional annual and national quarterly GDP series into regional quarterly GDP data which are usually unavailable for many countries. The method does not rely on auxiliary benchmark variables and can be applied sequentially, when the new annual data become available without the need the revise most of the previously disaggregated values. The method is used to disaggregate the GDP series for Poland. The disaggregated series shed new light on the regional business cycle dating, especially the 2012/13 recession. We also perform a preliminary analysis of the cross-sectional heterogeneity of GDP dynamics.

Keywords: GDP, disaggregation, regions, business cycle

JEL classification: E01, E32, R12

[1]University of Economics in Katowice, 1-go Maja St. 50, 40-287 Katowice, Poland, `jan.acedanski@ue.katowice.pl`.

1 Introduction

...

2 The procedure

2.1 The preliminaries

In the study, we use the following notation. Capital letters represent annual variables whereas small letters denote quarterly data. The procedure takes two series as inputs: the annual data on regional GDP levels $Y_{\tau,j}$, where $\tau = 1, 2, \ldots, T$ is the time index for years and $j = 1, 2, \ldots, M$ is the index for regions, and the quarterly data on country GDP level y_t^c, where $t = 1, 2, \ldots, 4T$ represents time index for quarters. Our aim is to find the quarterly regional GDP levels $y_{t,j}$, $t = 1, 2, \ldots, 4T$, $j = 1, 2, \ldots, M$. Obviously, there is natural correspondence between quarterly and annual data: every consecutive four quarterly values represent four quarters of the appropriate year from the annual series.

The unknown quarterly regional data should satisfy the set of temporal and contemporaneous conditions:

$$\sum_{j=1}^{4} y_{4(\tau-1)+j,i} = Y_{\tau,i}, \quad \tau = 1, 2, \ldots, T, i = 1, 2, \ldots, M \tag{1}$$

$$\sum_{i=1}^{M} y_{t,i} = y_t^c, \quad t = 1, 2, \ldots, 4T \tag{2}$$

The first condition represents aggregation of quarterly data to the observed annual series and the second implies that the regional data aggregate to the observed country series. The conditions (1)-(2) imply that the input series should satisfy the following set of conditions:

$$\sum_{i=1}^{M} Y_{\tau,i} = \sum_{j=1}^{4} y_{4(\tau-1)+j,i}^c, \quad \tau = 1, 2, \ldots, T, \tag{3}$$

so that the sum of annual regional values is equal to the sum of the quarterly annual series for a given year. In practice, condition (3) might not be satisfied due to rounding errors, particularly when the series are created from volume indices. In this case, an initial rebalancing procedure should be applied to establish the series consistency.

2.2 The optimization problem

In this study, we disaggregate the two input series so that the resulting annual growth rates of quarterly regional GDP are as smooth as possible and satisfy the conditions (1)-(2). More formally, let $\Delta y_{t,i}$ denote the annual growth rates for quarterly regional GDP data:

$$\Delta y_{t,i} = \frac{y_{t,i}}{y_{t-4,i}} \tag{4}$$

The set of quarterly regional GDP series $\{y_{t,i}\}$ solves the following optimization problem:

$$\min_{\{y_{t,i}\}} \left[\sum_{i=1}^{M} w_i \left(\sum_{t=6}^{4T} (\Delta y_{t,i} - \Delta y_{t-1,i})^2 \right) \right] \quad \text{s.t. (1)-(2),} \tag{5}$$

where w_i denotes region-specific weights. We use the average region share in country GDP as weights. The weighing scheme plays the important role in the optimization problem, because contributions of different regions to GDP and, in turn, to the contemporaneous constraints (2) differ considerably. Without the weights, the algorithm would prefer smoothing GDP growth rates in small regions over the most important ones.

2.3 Sequential solution to the optimization problem

The optimization problem (5) can be solved either in one step or sequentially. The sequential procedure takes the new input data and disaggregates them without revising most of the already disaggregated series. We find it a very practical feature of the proposed method.

The sequential procedure works as follow. In the first step, the problem (5) is solved for some initial T_0 years. In the second step, we add the data for the next year $T_0 + 1$, take the growth rates of the disaggregated regional GDP series from the last quarter of year $T_0 - 1$ as given and solve the problem (5) for the data covering years T_0 and $T_0 + 1$. In other words, in this step we calculate the disaggregate data for year T_0 and revise the disaggregated series for the previous year. In the subsequent steps, we repeat this last step for the next years and continuing until the last year T.

The sequential procedure is somewhat easier from the numerical point of view than the one-step approach. The former requires solving one optimization problem for $4T \cdot M$ variables whereas the latter consists of one problem with $4T_0 \cdot M$ variables and $T - T0 + 1$

problems with just $8M$ variables.

Of course, the solution obtained with the sequential disaggregation is not identical to the one-step procedure. However, as documented in the next section, the differences between them are small.

The calculations were conducted in Julia. The codes are available upon request.

3 Disaggregating the GDP series for Poland

In this section, we use the method described above to disaggregate the GDP data for Poland. We separately consider GDP in current and constant prices.

3.1 The results for GDP in current prices

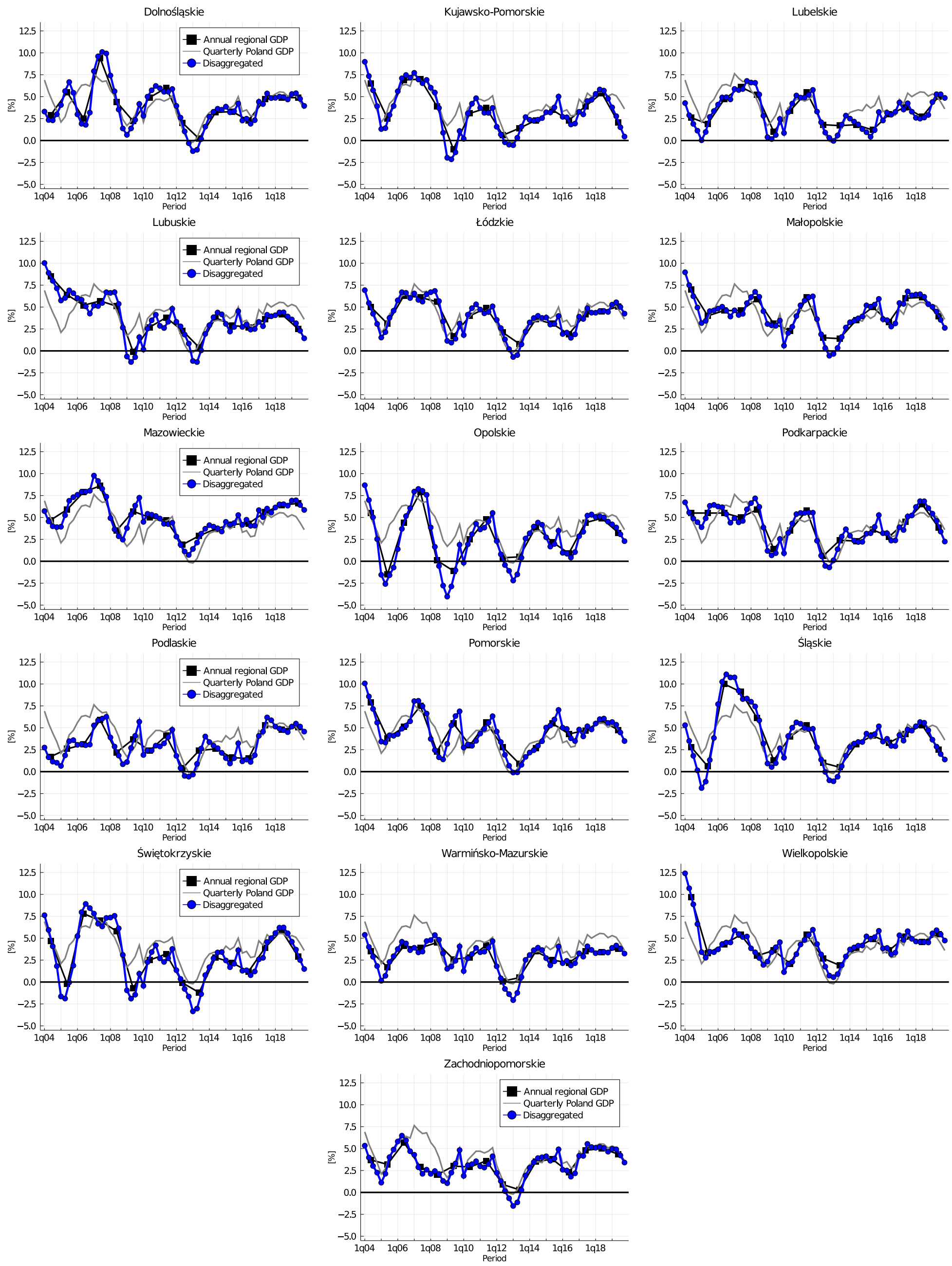
The annual data on GDP for 16 Polish NUTS-2 regions (voivodeships) and the quarterly data on country GDP, both in current prices, are taken as inputs for the procedure. The data covers the period 2002-2019. Because the data do not satisfy condition (3) the regional GDP levels are proportionally adjusted so that the discrepancies no longer exist.

The results of the one-step disaggregation procedure are shown in Figure 1. The plots contain the annual growth rates of the disaggregated data as well as the inputs. Despite setting smoothing as the only optimization criterion the resulting series do not seem overly smooth because of the contemporaneous constraint — the quarterly country series is not smooth too. The impact of this constraint can easily be noticed in the first quarter of 2010 and the fourth quarter of 2015. In the former case, there are sudden drops in all the disaggregated series for this quarter. On the other hand, the latter effect is represented by the local peaks.

3.2 The results for GDP in constant prices

The official statistics contain the annual growth rates of GDP in constant average prices from the previous year. The regional data are available for the period 2004-2018. Therefore, we approximate the regional growth rates in constant prices for 2019 by deflating the nominal growth rates for this year by the regional price indices from the previous

Figure 1: Disaggregation of nominal GDP series for Poland



The blue lines with circles represent the disaggregated quarterly regional series; the black ones with squares are annual regional series; the grey lines represent the quarterly series for Poland, common for all plots.

year. Then, we create the chain linked volume GDP series using the growth rates of real GDP described above and the series of GDP in current prices for 2003. Finally, we adjust the regional series to satisfy condition (3), as in the case of GDP in current prices.

The results of the one-step disaggregation procedure are presented in Figure 2 which contains the annual growth rates of the series used by the procedure. Similarly to the disaggregated series of GDP in current prices, one can notice the same effects of the contemporaneous constraints represented by the drops in the first quarters of 2010 and the peaks in the last quarters of 2015.

The disaggregated regional real GDP growth rates can be used for precise regional business cycle dating, which is difficult with the annual series. For example, the quarterly country data suggest that Poland experienced a mild recession at the turn of 2012 and 2013 when the annual GDP growth rates were slightly negative. However, the recession cannot be observed in the official annual data for the regions because for all but Świętokrzyskie regions the growth rates for 2012 and 2013 are positive. The disaggregated quarterly series reveal which regions truly fell into the recession. These are all but Lubelskie, Mazowieckie, and Wielkopolskie. The recession in the remaining regions was short and shallow, as the country data suggests. The disaggregated data also reveals that Śląskie region fall in the short recession at the beginning of 2005.

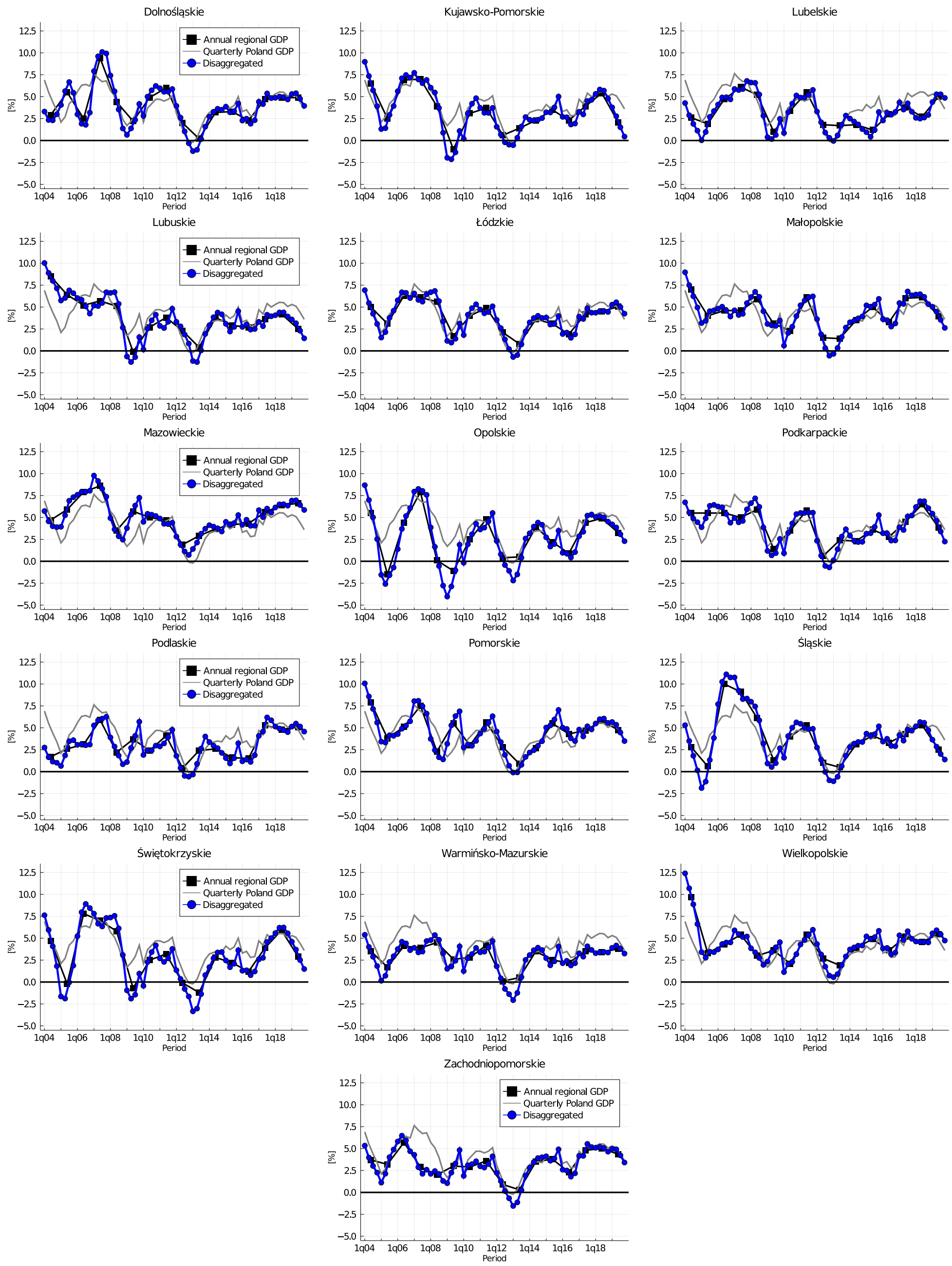
3.3 The analysis of the regional heterogeneity of the GDP growth rates

The disaggregated series can also be used for analysis of the regional heterogeneity of the GDP dynamics. The heterogeneity measured by the standard deviation and the interquartile range is illustrated in Figure (3). The plots show the clear decrease in heterogeneity in 2010, which can be noticed in the input annual data too.

3.4 Comparison of the one-step and the sequential variant of the procedure

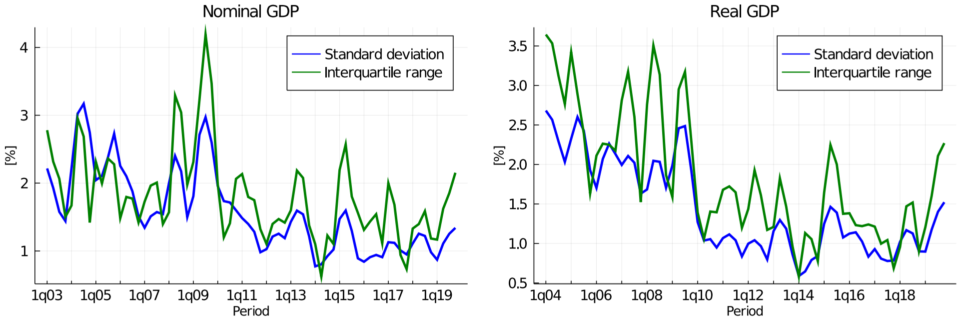
...

Figure 2: Disaggregation of real GDP series for Poland



The blue lines with circles represent the disaggregated quarterly regional series; the black ones with squares are annual regional series; the grey lines represent the quarterly series for Poland, common for all plots.

Figure 3: Regional heterogeneity of the quarterly nominal and real GDP growth rates



4 Conclusion