Disaggregation and Nowcasting of Regional GDP Series with a Simple Smoothing Algorithm

Jan Acedański

University of Economics in Katowice, Poland

The 5th Congress of Polish Statistics
Warsaw, 1-3 July 2025

Motivation

- How was the recent slowdown in Poland distributed across the regional level? Did some regions suffer considerably more than others? Do we see some signs of recovery in some regions?
- To answer the questions, we need reliable and up-to-date data on economic activity at regional level
- The official data on GDP for NUTS-2 regions (voivodeships) in Poland published by Statistics Poland are scarce:
  - annual series only
  - substantial publication delays: one year for nominal GDP and two years for real GDP (currently, the real GDP growth data are available for 2022 and nominal series for 2023)
- Other regional series do not solve the problem — narrow scope (industrial production) and poor quality (short time-series, breaks, high volatility and sampling variance)

Spatio-temporal disaggregation

- Spatio-temporal disaggregation methods: given the annual GDP data for regions and the quarterly country-level GDP, calculate the quarterly regional GDP values such that:

Spatio-temporal disaggregation — the idea

Quarter	Country-level	Reg 1	Reg 2	...	Reg M
1q2022	y^c_{1q22}				
2q2022	y^c_{2q22}				
3q2022	y^c_{3q22}				
4q2022	y^c_{4q22}				
Total		$Y_{22,r1}$	$Y_{22,r2}$	...	$Y_{22,rM}$

Spatio-temporal disaggregation

- Spatio-temporal disaggregation methods should be applied: given the annual GDP data for regions and the quarterly country-level GDP, calculate the quarterly regional GDP values such that:
    1. For every region, a sum of disaggregated values for a given year should be equal to the corresponding annual value (intertemporal consistency)

Intertemporal constraint

Quarter	Country-level	Reg 1	Reg 2	...	Reg M
1q2022	y^c_{1q22}	$y_{1q22,r1}$	$y_{1q22,r2}$	...	$y_{1q22,rM}$
2q2022	y^c_{2q22}	$y_{2q22,r1}$	$y_{2q22,r2}$	...	$y_{2q22,rM}$
3q2022	y^c_{3q22}	$y_{3q22,r1}$	$y_{3q22,r2}$	...	$y_{3q22,rM}$
4q2022	y^c_{4q22}	$y_{4q22,r1}$	$y_{4q22,r2}$	...	$y_{4q22,rM}$
Total		$Y_{22,r1}$	$Y_{22,r2}$	...	$Y_{22,rM}$

Spatio-temporal disaggregation

- Spatio-temporal disaggregation methods should be applied: given the annual GDP data for regions and the quarterly country-level GDP, calculate the quarterly regional GDP values such that:
    1. For every region, a sum of disaggregated values for a given year should be equal to the corresponding annual value (intertemporal consistency)
    2. For every quarter, a sum of disaggregated values over regions should be equal to the corresponding quarterly country-level value (contemporaneous consistency)

Contemporaneous constraint

Quarter	Country-level	Reg 1	Reg 2	...	Reg M
1q2022	y^c_{1q22}	$y_{1q22,r1}$	$y_{1q22,r2}$	...	$y_{1q22,rM}$
2q2022	y^c_{2q22}	$y_{2q22,r1}$	$y_{2q22,r2}$	...	$y_{2q22,rM}$
3q2022	y^c_{3q22}	$y_{3q22,r1}$	$y_{3q22,r2}$	...	$y_{3q22,rM}$
4q2022	y^c_{4q22}	$y_{4q22,r1}$	$y_{4q22,r2}$	...	$y_{4q22,rM}$
Total		$Y_{22,r1}$	$Y_{22,r2}$	...	$Y_{22,rM}$

Spatio-temporal disaggregation

- Spatio-temporal disaggregation methods should be applied: given the annual GDP data for regions and the quarterly country-level GDP, calculate the quarterly regional GDP values such that:
  1. For every region, a sum of disaggregated values for a given year should be equal to the corresponding annual value (intertemporal consistency)
  2. For every quarter, a sum of disaggregated values over regions should be equal to the corresponding quarterly country-level value (contemporaneous consistency)
- The problem has infinitely many possible solutions. Some additional conditions are necessary to make it unique

Spatio-temporal disaggregation methods

1. Benchmarking methods (multivariate Chow-Lin algorithm) — the disaggregated series should follow some high-frequency regional benchmarks as close as possible (Rossi 1982; DiFonzo 1990; Proietti 2011; Cuevas, Quilis, Espasa 2015)

2. Smoothing methods — the disaggregated series (their growth rates) should be as smooth as possible (Acedański 2024)

3. Mixed-frequency time-series methods (Koop, McIntyre, Mitchell, 2022)

Benchmarking and smoothing — pros and cons

1. Benchmarking
   - uses additional, available information (but reliable and long time series are required)
   - convenient for estimating data by the end of the sample where annual regional values are unavailable ("nowcasting")

2. Smoothing
   - no additional series required, simple
   - convenient for estimating historical values as tiny revisions are made when newly-available data are disaggregated
   - deteriorating performance by the end of the sample where annual regional values are unavailable — disaggregation driven solely by the country-level series

"Nowcasted" disaggregated series

Quarter	Country-level	Reg 1	Reg 2	...	Reg M
1q2022	y^c_{1q22}	$y_{1q22,r1}$	$y_{1q22,r2}$	$\cdots$	$y_{1q22,rM}$
2q2022	y^c_{2q22}	$y_{2q22,r1}$	$y_{2q22,r2}$	$\cdots$	$y_{2q22,rM}$
3q2022	y^c_{3q22}	$y_{3q22,r1}$	$y_{3q22,r2}$	$\cdots$	$y_{3q22,rM}$
4q2022	y^c_{4q22}	$y_{4q22,r1}$	$y_{4q22,r2}$	$\cdots$	$y_{4q22,rM}$
Total		$Y_{22,r1}$	$Y_{22,r2}$	...	$Y_{22,rM}$
1q2023	y^c_{1q23}	$y_{1q23,r1}$	$y_{1q23,r2}$	$\cdots$	$y_{1q23,rM}$
2q2023	y^c_{2q23}	$y_{2q23,r1}$	$y_{2q23,r2}$	$\cdots$	$y_{2q23,rM}$
3q2023	y^c_{3q23}	$y_{3q23,r1}$	$y_{3q23,r2}$	$\cdots$	$y_{3q23,rM}$
4q2023	y^c_{4q23}	$y_{4q23,r1}$	$y_{4q23,r2}$	$\cdots$	$y_{4q23,rM}$
1q2024	y^c_{1q24}	$y_{1q24,r1}$	$y_{1q24,r2}$	$\cdots$	$y_{1q24,rM}$
2q2024	y^c_{2q24}	$y_{2q24,r1}$	$y_{2q24,r2}$	$\cdots$	$y_{2q24,rM}$
3q2024	y^c_{3q24}	$y_{3q24,r1}$	$y_{3q24,r2}$	$\cdots$	$y_{3q24,rM}$
4q2024	y^c_{4q24}	$y_{4q24,r1}$	$y_{4q24,r2}$	$\cdots$	$y_{4q24,rM}$

Aim of the study

- To compare the smoothing with the Chow-Lin-type benchmarking algorithm for disaggregation and nowcasting accuracy using Polish regional real GDP data
- The results published as J. Acedański, (2024), *Disaggregation and Nowcasting of Regional GDP Series with a Simple Smoothing Algorithm*, Journal of Official Statistics, Vol. 40(4), 508-529, `https://doi.org/10.1177/0282423X241277716`
- Julia codes available at `https://github.com/JanAcedanski/spatio-temporal-disaggregation-with-smoothing`

Table of contents

Table of contents

The spatio-temporal smoothing algorithm

- A simple multivariate extension of the classic smoothing procedures
- Input: 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.
- Output: the quarterly regional GDP levels $y_{t,j}$.
- The output series must satisfy the following sets of the aggregation constraints:

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

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

Optimization problem

- The output series $y_{t,i}$ solves the following optimization problem:

$$\min_{\{y_{t,i}\}} \left[\sum_{i=1}^{M} w_i \left(\sum_{t=6}^{4T} (\Delta y_{t,i} - \Delta y_{t-1,i})^2 \right) \right] \quad (3)$$
$$\text{s.t. (1)-(2),}$$

where $\Delta y_{t,i} = \frac{y_{t,i}}{y_{t-4,i}}$ represent the annual GDP growth rates and w_i denote region-specific weights (the share of regional GDP in country GDP)

- The optimization problem is solved using the interior point Newton-type algorithm (a very efficient method of solving the benchmarking problems with growth rates; DiFonzo, Marini 2012, 2015; Brown 2010) in Ipopt package in Julia

Balancing input data

- The constraints (1)-(2) imply that the input series must satisfy the following set of conditions:

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

- In practice, (4) is rarely met (due to rounding or chain-linking errors), so a balancing procedure should be applied

Sequential optimization

- The optimization problem can also be solved sequentially
- For example, given some initial disaggregated regional GDP series in a certain year: $y^*_{t,i}$ for $t = t_0 - 3, \ldots, t_0$ and the last regional GDP growth rates in this year $\Delta y^*_{t0,i}$, disaggregate the regional data for the next two years solving the problem (3)
- The sequential solution differs slightly from the global one but has some practical advantages
  - sequential problems are easier to solve numerically
  - new observations can be disaggregated without revising the all previous estimates
- This is the baseline procedure

Multivariate Chow-Lin-type benchmarking procedure

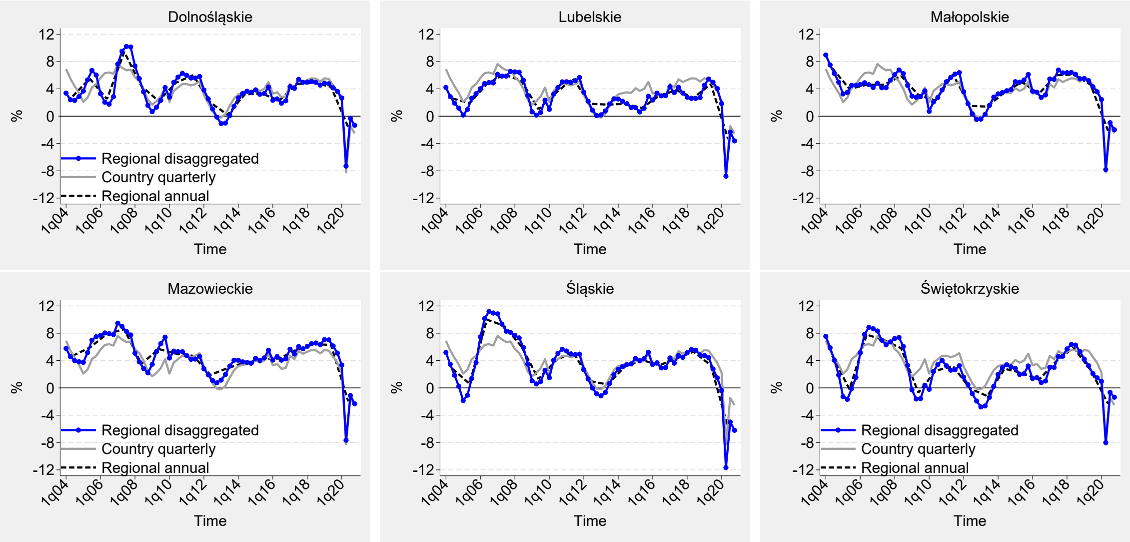
- The disaggregated series should follow a set of regional quarterly benchmarks as close as possible
- A version for the GDP dynamics is applied; it solves the following optimization problem:

$$\min_{\{y_{t,i}\},\{\beta_{j,i}\}} \left\{ \sum_{i=1}^{M} w_i \left[\sum_{t=5}^{4T} \left(\Delta y_{t,i} - (\beta_{0,i} + \sum_{j=1}^{4} \beta_{j,i} x_{j,t,i}) \right)^2 \right] \right\} \quad (5)$$

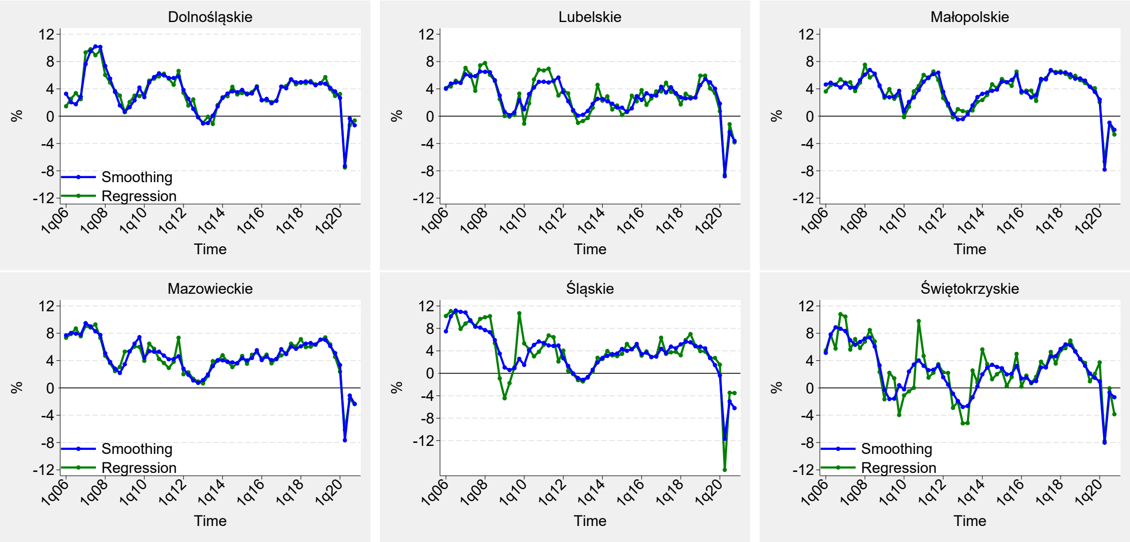
  subject to (1)-(2), where β denote region-specific regression coefficients and x represent values of benchmark indicators
- Four benchmarks: growth rates of industrial production index, investment outlays, and gross salary, as well as CPI

Table of contents

Disaggregation results – sequential smoothing procedure



Comparison with the benchmarking method



The differences between the sequential smoothing procedure and the alternative methods

Method	MD [p.p.] mean/min/max	MAD [p.p.] mean/min/max	RMSD [p.p.] mean/min/max	Pearson mean/min/max
		y-o-y growth rates		
One-step	0.00/0.00/0.00	0.08/0.08/0.08	0.12/0.11/0.13	0.999/0.997/1.000
Regression	0.00/0.00/0.00	0.64/0.53/0.76	0.83/0.71/0.95	0.930/0.871/0.968
		q-o-q growth rates		
One-step	0.00/ 0.00/0.01	0.06/0.03/0.12	0.11/0.04/0.20	0.998/0.994/1.000
Regression	0.01/-0.11/0.03	0.55/0.31/1.10	0.79/0.45/1.66	0.909/0.751/0.978

The comparison of the statistics for the alternative decompositions for the regions

Method	Mean [p.p.] mean/min/max/std	St. dev. [p.p.] mean/min/max/std	Autocorr. mean/min/max/std
y-o-y growth rates			
Sequential	3.33/2.35/4.64/0.61	2.71/1.83/3.84/0.51	0.74/0.58/0.85/0.07
One-step	3.38/2.34/4.64/0.61	2.73/1.83/3.84/0.50	0.75/0.61/0.85/0.06
Regression	3.24/2.35/4.56/0.55	2.83/2.13/4.61/0.65	0.64/0.32/0.77/0.11
q-o-q growth rates			
Sequential	0.85/0.62/1.15/0.15	1.62/1.46/1.85/0.09	-0.30/-0.41/-0.16/0.08
One-step	0.85/0.61/1.15/0.15	1.58/1.35/1.82/0.13	-0.29/-0.39/-0.10/0.09
Regression	0.81/0.55/1.12/0.14	1.76/1.15/3.26/0.51	-0.31/-0.49/-0.14/0.10

Table of contents

"Nowcasted" disaggregated series

Quarter	Country-level	Reg 1	Reg 2	...	Reg M
1q2022	y^c_{1q22}	$y_{1q22,r1}$	$y_{1q22,r2}$	$\cdots$	$y_{1q22,rM}$
2q2022	y^c_{2q22}	$y_{2q22,r1}$	$y_{2q22,r2}$	$\cdots$	$y_{2q22,rM}$
3q2022	y^c_{3q22}	$y_{3q22,r1}$	$y_{3q22,r2}$	$\cdots$	$y_{3q22,rM}$
4q2022	y^c_{4q22}	$y_{4q22,r1}$	$y_{4q22,r2}$	$\cdots$	$y_{4q22,rM}$
Total		$Y_{22,r1}$	$Y_{22,r2}$	...	$Y_{22,rM}$
1q2023	y^c_{1q23}	$y_{1q23,r1}$	$y_{1q23,r2}$	$\cdots$	$y_{1q23,rM}$
2q2023	y^c_{2q23}	$y_{2q23,r1}$	$y_{2q23,r2}$	$\cdots$	$y_{2q23,rM}$
3q2023	y^c_{3q23}	$y_{3q23,r1}$	$y_{3q23,r2}$	$\cdots$	$y_{3q23,rM}$
4q2023	y^c_{4q23}	$y_{4q23,r1}$	$y_{4q23,r2}$	$\cdots$	$y_{4q23,rM}$
1q2024	y^c_{1q24}	$y_{1q24,r1}$	$y_{1q24,r2}$	$\cdots$	$y_{1q24,rM}$
2q2024	y^c_{2q24}	$y_{2q24,r1}$	$y_{2q24,r2}$	$\cdots$	$y_{2q24,rM}$
3q2024	y^c_{3q24}	$y_{3q24,r1}$	$y_{3q24,r2}$	$\cdots$	$y_{3q24,rM}$
4q2024	y^c_{4q24}	$y_{4q24,r1}$	$y_{4q24,r2}$	$\cdots$	$y_{4q24,rM}$

"Nowcasting" procedure for smoothing algorithm

Simply solve the standard smoothing problem (3) subject to the contemporaneous constraints (2) but without the missing intertemporal constraints (1).

"Nowcasting" procedure for the benchmarking algorithm

1. Estimate the regression coefficients β by solving the benchmarking problem (5) subject to both constraints (1)-(2) for the period when the annual regional values are available
2. Use the estimated coefficients $\hat{\beta}$ from the previous step as well as the available quarterly values of benchmarking regressors to nowcast the missing series by solving the modified benchmarking problem:

$$\min_{\{y_{t,i}\}} \left\{ \sum_{i=1}^{M} w_i \left[\sum_{t=4T+1}^{4T+8} \left(\Delta y_{t,i} - (\hat{\beta}_{0,i} + \sum_{j=1}^{4} \hat{\beta}_{j,i} x_{j,t,i}) \right)^2 \right] \right\} \tag{6}$$

subject to the contemporaneous constraint (2)

Pseudo-real time "nowcasting" experiment

- Investigated variable: annual growth rates of GDP in constant (previous-period) prices for 16 Polish voivodeships
- Period of analysis: 2005-2020
- "Nowcasting" horizon: 8 quarters
- 8 extending samples:
  - The first one — disaggregation period: 2005-11 (7y/28q); nowcasting/verification period: 20012-13 (8q)
  - ...
  - The last one — disaggregation period: 2005-18 (14y/56q); nowcasting/verification period: 2019-20 (8q)
- Official data revisions are not accounted for — the most recent dataset is used in the study

Calculating the "true" values for verification periods

- True disaggregated quarterly values are unknown
- They are estimated using the two disaggregation procedures applied for the whole sample 2005-20
- The mean values of results of the two procedures are taken as "true" series

Table of contents

Nowcasting accuracy – all samples

Horizon	ME		MAE		RMSE	
(quarters)	Smooth	Regress	Smooth	Regress	Smooth	Regress
1	-0.01	0.26	0.92	2.62	1.09	4.32
2	0.05	-0.19	1.06	2.16	1.27	3.30
3	0.00	-0.16	1.14	2.01	1.37	3.10
4	-0.03	-0.19	1.21	2.55	1.43	4.05
5	0.05	-0.72	1.39	2.75	1.70	4.40
6	-0.11	-0.38	1.54	2.25	1.95	3.55
7	-0.17	-0.29	1.58	2.32	1.97	3.37
8	-0.20	-0.41	1.59	2.67	1.94	4.10

Nowcast accuracy for different sample lengths

Mean sample length (q/y)	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$	$h = 8$
	RMSE for smoothing relative to RMSE for regression							
30/7.5	**0.19**	**0.28**	**0.33**	**0.26**	**0.25**	**0.28**	**0.31**	**0.26**
34/8.5	**0.30**	**0.32**	**0.53**	**0.36**	**0.38**	**0.54**	**0.55**	**0.46**
38/9.5	**0.55**	**0.49**	**0.67**	**0.69**	**0.83**	**0.94**	**0.78**	**0.68**
42/10.5	**0.74**	**0.87**	**0.92**	**0.96**	1.13	1.19	1.18	1.19
46/11.5	**0.68**	**0.83**	**0.85**	**0.91**	1.45	1.55	1.52	1.35
50/12.5	**0.99**	1.04	1.12	1.01	1.29	1.55	1.70	1.41

All the values are based on four consecutive samples of different lengths. The relative values of RMSE for smoothing lower than one indicate that smoothing nowcasts were more accurate than nowcasts generated by the regression-based method. Bolded are the relative errors that are lower than 1.

Size of the revisions of the disaggregated series

Year	Mean standard deviation [p.p.]				Maximum revision [p.p.]			
	All samples		Last 5 samples		All samples		Last 5 samples	
	Smooth	Regress	Smooth	Regress	Smooth	Regress	Smooth	Regress
2006	0	0.95	0	0.13	0	18.1	0	2.00
2008	0	1.18	0	0.14	0	34.4	0	1.99
2010	0.05	1.31	0	0.11	0.64	50.8	0	1.29
2012	0.05	0.34	0	0.12	0.56	6.80	0	2.08
2014	0.08	0.12	0.08	0.12	0.62	2.20	0.62	2.20
2016	0.06	0.06	0.06	0.06	0.38	0.44	0.38	0.44

Table of contents

Summary

- Smoothing algorithm – simple, computationally efficient, does not require revisions of already disaggregated series, but does not account for additional information, can be poor at nowcasting
- Statistical properties of the series disaggregated using smoothing and the Chow-Lin-type benchmarking method are similar
- For short horizons ($h \leq 4$), smoothing performed not worse in nowcasting than benchmarking
- For longer horizons ($h \geq 5$), smoothing was better only in short samples
- Benchmarking suffers from substantial revisions, particularly in short samples

Thank you for the attention!