Wiadomości Statystyczne. The Polish Statistician, 2021, vol. 66, 2, 7–24 DOI: 10.5604/01.3001.0014.7387
Studia metodologiczne / Methodological studies
Otrzymano/received: 30.03.2020, zaakceptowano/accepted: 13.01.2021

The importance of data revisions for statistical inference

Paulina Ziembińska[a]

Abstract. The aim of the study is a quantitative analysis of revisions conducted by means of a new, real-time macroeconomic dataset for Poland, designed on the basis of the *Statistical bulletin* (*Biuletyn statystyczny*) published by Statistics Poland, covering the period from as early as 1995 until 2017. Polish data have positively verified a number of hypotheses concerning the impact of data revisions on the modelling process. Procedures assessing the properties of time series can yield widely discrepant results, depending on the extent to which the applied data have been revised.

A comparison of the fitted ARIMA models for series of initial and final data demonstrates that the fitted models are similar for the majority of variables. In the cases where the form of the model is identical for both series, the coefficients retain their scale and sign. Most differences between coefficients result from a different structure of the fitted model, which causes differences in the autoregressive structure and can have a considerable impact on the ex ante inference. A prognostic experiment confirmed these observations. For a large number of variables, the total impact of revisions on the forecasting process exceeds 10%. Extreme cases, where the impact goes beyond 100%, or situations where data have a direct impact on the forecast sign, are also relatively frequent. Taking these results into account by forecasters could significantly improve the quality of their predictions. The forecast horizon has a minor impact on these conclusions. The article is a continuation of the author's work from 2017.

Keywords: data revisions, real-time data, time series analysis, forecasting

JEL: C10, C53, C82

Znaczenie rewizji danych dla wnioskowania statystycznego

Streszczenie. Celem pracy jest ilościowa analiza rewizji danych makroekonomicznych w czasie rzeczywistym dla Polski pochodzących z nowego zbioru utworzonego na podstawie „Biuletynu statystycznego” GUS i obejmującego okres od 1995 do 2017 r. Polskie dane pozytywnie weryfikują wiele hipotez dotyczących wpływu rewizji danych na proces modelowania. Procedury oceniające własności szeregów czasowych mogą dawać istotnie różne wyniki w zależności od tego, jak bardzo rewidowane dane zostaną użyte.

Porównanie dopasowanych modeli ARIMA dla szeregów pierwszych i finalnych odczytów wskazuje, że w przypadku większości zmiennych dopasowane modele są podobne. Gdy postać modelu jest taka sama dla obu szeregów, współczynniki zachowują skalę i znak. Większość różnic we współczynnikach wynika z odmiennej struktury dopasowanego modelu, co wpływa na różnice w strukturze autoregresyjnej i może mieć niemały wpływ na wnioskowanie *ex ante*. Potwierdza to eksperyment prognostyczny. Dla dużej części zmiennych całkowity wpływ rewizji na proces prognozowania wynosi powyżej 10%. Nie są też wyjątkiem ekstremalne przypadki, w których ten wpływ przekracza 100%, czy sytuacje, w których dane bezpośrednio wpływają na znak prognozy. Uwzględnienie tych wyników przez prognostów mogłoby znacząco poprawić jakość predykcji. Horyzont prognozy ma niewielki wpływ na te konkluzje. Artykuł jest kontynuacją pracy autorki z 2017 r.

Słowa kluczowe: rewizje danych, dane w czasie rzeczywistym, analiza szeregów czasowych, prognozowanie

---

[a] Uniwersytet Warszawski, Wydział Nauk Ekonomicznych, Katedra Statystyki i Ekonometrii / University of Warsaw, Faculty of Economic Sciences, Division of Statistics and Econometrics.
ORCID: https://orcid.org/0000-0002-8394-4909.

1. Introduction

Statistical agencies struggle with making decisions as to whether publish data quickly or rather postpone their publication in order to improve their accuracy. When they choose the latter option, the ensuing data revisions show a fuller picture of the economy; in other words, they aim at making the data convergent to the 'real' values. The macroeconomic statistical data are clearly very important, for they provide the basis for economic research and are used within a broad range of everyday operations carried out by public and private institutions. Ultimately, they have an impact on all economic entities. Croushore (2011) states that 'Until recently, macroeconomists assumed that data revisions were small and random and thus had no effect on structural modelling, policy analysis, or forecasting. But real-time research has shown that this assumption is false and that data revisions matter in many unexpected ways'. Only when the Federal Reserve Bank in Philadelphia made real-time data for the United States of America public did the development of analyses concerning the revision process cause a change in the perception of this issue, and result in the denunciation of the previously-made assumptions, which, from today's point of view, seem to have overly simplified the matter.

The understanding of the data revision process is of great importance, since it has a twofold impact on economic studies. First of all, data revision allows the determination of a most recent set of information valid at a given time, which is crucial, for example, in evaluating monetary policy or producing forecasts. Secondly, it poses numerous questions concerning the quality of data, which not only induces a more precise determination of the research objectives, but also has an effect on statistical inference. Amir-Ahmadi et al. (2015) show that various results are obtained for the monetary policy models depending on the applied data – either final or real-time. Chang and Li (2018) repeated 23 studies published in leading economic journals and proved that the results could vary to a great extent, depending on which readings have been used in the estimation procedure. A broad review of literature concerning real-time data analyses can be found in Croushore (2010, 2011) and Croushore and Stark (2001, 2002).[1]

In Ziembińska (2017), a new set of real-time data for Poland was presented, designed on the basis of the *Statistical bulletin* of Statistics Poland (*Biuletyn statystyczny – SB*). Since 2016, Statistics Poland has assumed a more formal approach to data revision and communication, which is manifested in a variety of forms, including the publication 'Policy of Revising Statistical Data and Rules of Handling Publication Errors' (GUS, 2016), or the regularly updated 'Statistical Data Revision Calendar'. However, the collected data indicate that before 2016 (i.e. in the years 2003–2015), these updating processes were not regular to the same extent.

[1] Bibliography is published online at https://facultystaff.richmond.edu/~dcrousho/docs/realtime_lit.pdf.

The introduction of multiple methodological changes in respect to Polish data results mostly from the need to adjust the country's statistical reporting standards to the requirements of the European Union. From the point of view of macroeconomic data users, these changes have had a positive influence on data quality and have allowed the performance of international comparative analyses. On the other hand, frequent methodological revisions cause difficulties in structuring long and consistent data series adequately, which is necessary for econometric analyses. Moreover, the identification of the final reading based on methodologies applied prior to a revision becomes impossible. Data inconsistencies can also affect the latest available data readings, most frequently used in practice.

In Ziembińska (2017), a descriptive analysis of these processes was performed for a broad set of variables and the longest possible data series. Through the study of particular categories of variables, i.e. national accounts, prices, labour market and public finances, basic methodological revisions resulting from the change in the ESA European Methodological Standards or Statistical Classification of Economic Activities were indicated. For the majority of the categories of variables, methodological revisions are statistically significant. However, unpredictable non-methodological revisions can still be frequently of a systematic nature. Tests focusing on the reasons behind revisions do not have much statistical power and often fail to yield un-ambiguous results. However, the applied conservative approach provides several conclusive results. They either indicate the revisions resulting from the extension of an information set or confirm the revisions caused by measurement errors. This is particularly alarming in the light of the statistical properties of the models which use these variables.[2] A broad set of variables analysed in different formats and frequencies allows drawing additional conclusions, indicating that the revision process (e.g. for national accounts) is not trivial – several readings are revised at the same time, which causes a different behaviour of nominal data and growth rates. Seasonal data adjustments also generate significant revisions that are often indefinite, even for variables that have not been subject to revision (e.g. business indicators). Finally, a comparative analysis of various data sets indicates that they are not always methodologically consistent, which can cause substantial problems when using them.

The aim of the study is a quantitative analysis of revisions performed on the basis of a new, real-time macroeconomic dataset for Poland. The article further develops the results previously presented in Ziembińska (2017), where the basic properties and the nature of revision processes were analysed. The next natural step, covered by this article, is the analysis of the impact revision processes have on statistical inference, namely on the properties of time series or fitting ARIMA class models, as well as on forecasting.

[2] It appears that data measurement errors detected through, e.g. revisions, have a significant impact on the asymptotic distribution of test statistics or estimators (Clark & McCracken, 2010).

2. Methodology

The analysis is based on a new set of real-time data for Poland, designed on the basis of the *Statistical bulletin* published between January 2003 and June 2017 (GUS, 2003–2017). The data cover reference periods from as early as 1995. A detailed description and an initial analysis of the dataset is presented in Ziembińska (2017).

As in Ziembińska (2017) and following Croushore and Stark (2001), a three-dimensional information set is defined, representing values of a macroeconomic variable x_{its}[3] – the value of the i-th variable for a given reference period (denoted as t), and available at a given time (denoted as $s \geq t$), as a real-time dataset. The method in Table 1 presents it with respect to a specific economic variable i – each line corresponds to revisions of readings which take place at subsequent points in time s for a given reference period t. Note that $s_1 < s_2 < \cdots < s_d$ and $t_1 < t_2 < \cdots < t_f$. Time intervals between consecutive reference periods (t) and publication periods (s) can be different. In the analysed new dataset s_i represents months, while the frequency of t_i depends on the analysed variable and can be monthly, quarterly or annual. For some more historical reference periods the initial publication might be unknown if it occurred before the first observable publication date (s_1), e.g. for t_1 and t_2 in Table 1. Starting from reference period t_3, the initial publication which happened in s_2 is available and all of its following revisions published in s_3, s_4 up to s_d (the most recent publication date) are also available. The time series $(x_{its})_t$ for a given s is called a vintage.

Table 1. Diagram of a real-time dataset

Reference period	Publication period					
	s_1	s_2	...	s_k	...	s_d
t_1	x_{i,t_1,s_1}	x_{i,t_1,s_2}	...	x_{i,t_1,s_k}	...	x_{i,t_1,s_d}
t_2	x_{i,t_2,s_1}	x_{i,t_2,s_2}	...	x_{i,t_2,s_k}	...	x_{i,t_2,s_d}
t_3	.	x_{i,t_3,s_2}	...	x_{i,t_3,s_k}	...	x_{i,t_3,s_d}
t_4	.	.	...	x_{i,t_4,s_k}	...	x_{i,t_4,s_d}
...	.	.	...	...	...	...
t_n	.	.	...	x_{i,t_n,s_k}	...	x_{i,t_n,s_d}
t_{n+1}	.	.	.	x_{i,t_{n+1},s_k}	...	x_{i,t_{n+1},s_d}
...	.	.	.	.	...	...
t_f	.	.	.	.	.	x_{i,t_f,s_d}

Source: Croushore and Stark (2001).

[3] Variable x_{its} refers to a macroeconomic reading in specific units, e.g. the annual growth rate of the gross domestic product presented in percentages.

The analysis focuses on understanding differences in outcomes of various econometric procedures for different vintages of a particular macroeconomic variable. Specifically, the time series structured according to the three methods below are analysed:[4]

- Method 1: a full sample of the most recent readings available – based on the full columns of Table 1, which – according to the literature – seems to be the most frequently applied approach in the modelling practice;
- Method 2: a full sample of the first readings – based on the data found in the diagonal of Table 1;
- Method 3: a repeated observation method proposed by Croushore and Stark (2002), based on the columns in the top right-hand corner of Table 1. This is the only method where a constant length of the series is maintained.

Certain properties of the analysed data are tested and assumed within the scope of econometric modelling. With regard to univariate time series analyses, the stationarity, autocorrelation and heteroscedasticity should be checked and the normality assumption verified. When data undergo revisions, the results of statistical tests may depend on which readings of a given variable are used. Below is an analysis demonstrating whether the results of basic tests depend on the revision process.[5] Table 2 contains a list of the analysed statistical tests with a defined null hypothesis. As regards each test, the percentage of series generated according to a particular method that provide a consistent conclusion from the conducted test is examined. The analysis also aims to check if various tests produce coherent results.

Furthermore, the study tries to determine whether fitting a simple ARIMA(p,d,q)[6] model to the series yields different results, depending on the applied method of data structuring.

The first step in examining how data revisions can influence the forecasting process is the proper fitting of the model. Cole (1969) proposed a simple method of measuring the direct and indirect impact data have on predictions. It requires the estimation of models for the initial data (corresponding to Method 2) – denoted as Model A, and final data (Method 1) – denoted as Model B. What follows is the comparison of forecasts:

[4] It is assumed here that the data generation process is not subject to changes following a methodological revision. It does not always have to be a correct assumption; it is not, for example, when revisions result from newly emerging information. However, the sample of Polish data is too short to overturn this assumption and analyse the specific sub-samples.

[5] This analysis is complementary to the verification of the size and power of statistical tests for data measurement errors in the form of a revision.

[6] In the ARIMA(p,d,q) model: p is the order (number of time lags) of the autoregressive model (AR(p)), d is the degree of differencing, and q is the order of the moving-average model (MA(q)).

1. based on Model A and the initial data;
2. based on Model A and the revised data;
3. based on Model B and the revised data.

Any difference between predictions (1) and (3) is indicative of an overall impact of revisions on the forecasting process. The direct impact of data, assuming a particular form of the model, is shown by the comparison of forecasts (1) and (2). Any difference between results for (2) and (3) shows the scale of the impact of the data on forecasts, indirectly through the estimation process. A direct impact on model parameters and the functional form is shown in the ARIMA model fit analysis; here an indirect impact of the change of parameters on predictions can be quantified. The described comparisons are limited to comparing forecasts to one another, thus eliminating the question which value is being forecasted and which method minimises the forecast error from the deliberations.

Table 2. List of the analysed statistical tests

Test	H0	Small sample properties
Portmanteau		
Ljung-Box (LB)	i.i.d.	yes (Hope, 1968)
Hosking	i.i.d.	yes (Hope, 1968)
Li-McLeod	i.i.d.	yes (Hope, 1968)
Autocorrelation		
Breusch-Godfrey (BG)	no autocorrelation	no
Durbin-Watson (DW)	no autocorrelation	yes (Farebrother, 1980)
Heteroscedasticity		
Goldfeld-Quandt (GQ)	variance equal in sub-samples	no
Harrison-McCabe (HMC)	variance equal in sub-samples	yes
Normality		
Shapiro-Wilk (SW)	normal distribution	no
Jarque-Bera (JB)	normal distribution	yes (Wüertz & Katzgraber, 2005)
D'Agostino (DA) (skewness/kurtosis /omnibus)	normal distribution	no
Stationarity		
Augmented Dickey-Fuller (ADF)	I(1)	yes*
Kwiatkowski-Phillips-Schmidt-Shin (KPSS)	I(0)	yes*
Phillips-Perron (PP)	I(1)	yes*
Elliott, Rothenberg and Stock (ERS)	I(1)	no

Note. In all tests the alternative hypothesis is two-sided, i.e. H1: ~H0. * means that critical values are interpolated from tables defined in Banerjee et al. (1993). i.i.d. denotes independent and identically distributed. I(0) denotes stationary series and I(1) series integrated of order 1. Small sample properties referred to in the last column specify if any special adjustment of the test statistic distribution was applied due to the fact that the analysed series might not be sufficiently long to rely on the asymptotic distributions.

3. Results

3.1. White noise / autocorrelation tests

The assumption that a series represents white noise is checked by means of the Portmanteau test. Three versions are tested: a standard Ljung-Box test (LB) and its modifications proposed by Hosking, as well as Li and McLeod. A detailed description of the discussed tests can be found in numerous papers, including Mahdi and McLeod (2012), where the authors propose small sample simulated critical values.

The conclusions are as follows: firstly, all three tests provide very consistent results – differences appear only for variables with very short series. Secondly, revisions have an impact on inference for a few variables only, which was found on the basis of the Portmanteau tests. In particular, for a monthly consumer price index (CPI) inflation rate in the food category, the monthly industrial production growth rate and the quarterly data on the current account balance, the inference depends most strongly on the selected data series (in particular for Method 2, i.e. on the initial data).

Table 3 presents the results of the Durbin-Watson (DW) and Breusch-Godfrey (BG) autocorrelation tests and the Ljung-Box test, aggregated for all of the analysed variables,[7] jointly with a number of sample variables referred to in the text. The first three panels present a percentage of the series for a given method in which the null hypothesis has been rejected.[8] Values close to 50% for a given variable[9] indicate inconclusiveness for a given test, depending on the series used (for example, for data on industrial production). Large disparities between values for different methods show a significant impact of revisions on inference. For example, in the case of the monthly industrial production growth rate, for nearly 50% of the series generated with Method 1 and 2 there were no grounds to reject the null hypothesis with regard to the lack of autocorrelation, whereas for all series generated according to Method 3, this hypothesis was rejected. This shows how significant an impact the final value has on the test result. In contrast, data on the current account demonstrate the impact of the first reading on the inference about the autocorrelation of series – the result for Method 2 differs substantially from the ones for the other two methods. It is worth noting that all the three tests yielded identical results which are consistent regardless of the data series used only for Method 3 (apart from the annual data on financial accounts and the average annual gross domestic product (GDP) growth rate). This means that the length of the examined series also affects inference significantly.

[7] The full list of the analysed variables covers 61 macroeconomic variables in different formats and frequencies. The list is available upon request.

[8] With the exception of the KPSS stationarity test.

[9] It needs to be highlighted that the aggregated results for all variables do not indicate consistency of the inference in terms of a specific variable. The aggregation presents a percentage of series (for all variables) for which the null hypothesis is rejected and which does not demonstrate consistency across tests for a given series.

Table 3. Results of the autocorrelation tests

Variable		LB			DW			BG			LB-BG			LB-DW			BG-DW			Number of series		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
All		**85**	**83**	**0**	**88**	**86**	**0**	**86**	**84**	**0**	**1**	**1**	**0**	**3**	**3**	**0**	**3**	**3**	**0**	**12519**	**13533**	**376**
GDP nsa yy (avg)	y	88	0	0	100	57	100	73	0	0	15	0	0	12	57	100	27	57	100	82	82	48
CPI yy (eop)	y	0	0	0	38	38	0	0	0	0	0	0	0	38	38	0	38	38	0	96	96	48
CPI food mm	m	43	43	0	52	52	0	43	43	0	0	0	0	9	9	0	9	9	0	23	23	23
Industrial production nsa mm	m	65	51	100	59	47	100	64	51	100	1	1	0	7	3	0	6	4	0	199	217	72
Current account in mln EUR	q	80	51	100	84	55	100	80	55	100	0	4	0	4	4	0	4	0	0	153	153	60
	y	100	0	100	100	13	100	100	0	100	0	0	0	0	13	0	0	13	0	69	69	48
Financial account in mln PLN	q	75	78	100	80	84	100	76	82	100	2	4	0	6	6	0	4	2	0	153	153	60
	y	100	100	21	100	100	100	100	100	21	0	0	0	0	0	79	0	0	79	69	69	48

Note. The first 3 panels show the percentage of series in a given method (1), (2) or (3) for which the null hypothesis of a given test (test acronyms are described in Table 2) is rejected. The next 3 panels present the percentage of vintages, for which the conclusions from the two tests are different (at a 10% confidence level). The last panel presents the number of analysed series used to calculate the percentage values. In relation to the definitions of the variables, the following acronyms are used: nsa – not seasonally adjusted, yy – annual growth rate, avg – average, eop – end of period value, mm – monthly growth rate. For each variable its frequency (frequency of the reference periods) is defined as: monthly (m), quarterly (q), annual (y).

Source: author's calculations based on: GUS (2003–2017).

Table 4. Results of the normality tests

Variable		SW			DA			JB			SW-JB			SW-DA			JB-DA			Number of series		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
All		**74**	**77**	**79**	**62**	**64**	**69**	**56**	**60**	**61**	**23**	**20**	**22**	**19**	**17**	**17**	**15**	**6**	**19**	**7832**	**8103**	**3612**
Private consumption in mln PLN	q	83	85	100	73	70	100	59	59	100	24	26	0	10	15	0	14	1	0	139	139	60
Public consumption in mln PLN	q	0	100	0	0	35	0	0	3	0	0	97	0	0	65	0	0	1	0	91	103	60
GDP nsa yy	q	82	65	100	65	54	100	0	0	0	82	65	100	21	36	0	65	1	100	148	169	60
CPI mm	m	75	85	100	0	26	0	0	27	0	75	58	100	75	59	100	0	0	0	189	208	72
Exports in mln EUR	m	59	43	0	75	68	100	49	41	0	10	2	0	17	25	100	26	1	100	146	146	72
Imports in mln EUR	m	62	63	100	71	68	100	53	52	100	10	11	0	10	6	0	18	50	0	146	146	72

Note. For explanation of acronyms and abbreviations see Table 2 and note to Table 3.

Source: author's calculations based on: GUS (2003–2017).

3.2. Normal distribution tests

A key assumption with regard to many econometric procedures is that the data can be approximated by normal distribution. In order to verify whether the results of the normality tests depend on the revision process, the Shapiro-Wilk test (SW), the Jarque-Bera test (JB) (taking account of small-sample critical values, after Wüertz and Katzgraber, 2005) and the D'Agostino test (DA) are used. Table 4 contains results which are analogous to the ones concerning autocorrelation tests. First of all, it needs to be noted that in many cases the three discussed tests lead to dissimilar conclusions. It is an interesting finding, particularly in relation to the differences between the Jarque-Bera test and the D'Agostino test, both of which rely on statistics based on the third and fourth central moments (skewness and kurtosis). Secondly, these results do not substantially depend on the number of observations in series – data on monthly imports could serve as an example, as for them 50% of the tests yielded contradictory results. The impact of revisions on inference is far stronger as well. Only for nine variables are the conclusions consistent in at least 90% for various methods.[10]

3.3. Heteroscedasticity tests

Next, the heteroscedasticity of the series is checked, i.e. whether the variance is equal in the sub-samples. To this end, two tests are used: the Goldfeld-Quandt (GQ) test and the Harrison-McCabe (HMC) test with automatic criteria of selecting a division of a series into two sub-samples. In the latter, simulated critical values are applied, taking account of the appropriate length of a series. Table 5 presents the results. The number of variables for which the results are inconclusive is larger in the heteroscedasticity tests and there are more variables for which the two discussed tests generate different results on the same set of series, which is not necessarily related to the short sample. For example, for the average annual CPI inflation data, the GQ test rejects the null hypothesis more frequently, while the HMC test does not provide grounds to reject it in 100% of cases, regardless of the data generation method. For the majority of variables, it is again Method 3 that generates the most conclusive results for both tests. However, it is noteworthy that they are often opposite to Method 1 and 2. For example, for the quarterly public consumption data we would obtain opposite conclusions concerning heteroscedasticity while using the first series and the final readings.

[10] The above-mentioned variables include: quarterly investments, monthly and quarterly inflation and a monthly harmonised index of consumer prices (HICP) and producer price index (PPI) inflation, monthly data on annual industrial production growth rate, monthly unemployment data, and nominal values of the monetary aggregates.

Table 5. Results of the heteroscedasticity tests

Variable		GQ			HMC			GQ-HMC			Number of series		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
All		**45**	**46**	**40**	**17**	**17**	**17**	**34**	**34**	**28**	**12519**	**13533**	**5699**
Private consumption in mln PLN	q	46	50	100	0	0	0	46	50	100	169	169	60
Public consumption in mln PLN	q	0	43	0	0	0	0	0	43	0	121	133	60
Private consumption nsa yy	q	26	24	0	23	20	98	8	8	98	178	178	60
Public consumption nsa yy	q	37	7	100	36	22	100	2	15	0	121	133	60
CPI yy (avg)	m	83	79	100	0	0	0	83	79	100	199	217	72
CPI yy (avg)	q	69	76	100	0	0	0	69	76	100	177	177	60

Note. For explanation of acronyms and abbreviations see Table 2 and note to Table 3.
Source: author's calculations based on: GUS (2003–2017).

Table 6. Results of the stationarity tests

Variable		ADF			PP			ERS			ADF-KPSS			ADF-PP			ADF-ERS			Number of series		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
All		**41**	**42**	**48**	**41**	**38**	**42**	**35**	**34**	**42**	**63**	**62**	**64**	**23**	**22**	**25**	**36**	**36**	**49**	**12693**	**13803**	**5699**
Consumption in mln PLN	q	58	58	100	74	63	100	0	2	0	79	59	100	22	39	0	55	53	100	121	124	60
GDP in mln PLN	q	75	73	100	100	100	100	0	4	0	22	17	0	25	27	0	75	77	100	169	169	60
Consumption nsa yy	q	3	0	0	4	2	0	36	30	0	62	48	100	2	2	0	40	30	0	121	124	60
GDP nsa yy	q	24	51	0	0	24	0	60	93	100	65	62	0	24	27	0	35	42	100	178	179	60
CPI yy (avg)	m	85	0	100	88	5	100	0	5	0	34	91	100	4	5	0	85	5	100	199	217	72
	q	24	100	100	100	80	100	12	10	0	49	58	100	76	20	0	32	88	100	177	177	60

Note. For explanation of acronyms and abbreviations see Table 2 and note to Table 3.
Source: author's calculations based on: GUS (2003–2017).

3.4. Stationarity tests

Beginning with a ground-breaking work by Nelson and Plosser (1982), the impact of the stationarity of series on research results has become a subject of numerous studies, both in terms of methodology and the findings. Surprisingly few results concerning data stationarity have been published in the context of data revisions. Certain examples can be found in a series of articles by K. D. Patterson (compare e.g. Patterson & Heravi, 1991, 2004).

In this section, the impact of the revision process on the unit root tests is discussed, particularly the augmented Dickey-Fuller test (ADF), the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS), the Phillips-Perron test (PP), and the Elliott, Rothenberg and Stock test (ERS) (details concerning these tests can be found in e.g. Phillips and Xiao (1998).

The trend stationarity is tested for nominal variables and for variables expressed as growth rates the level stationarity is tested (all combinations are checked in order to confirm the reliability of the results and the conclusions). An automatic selection of the number of lags is applied, based on the Akaike information criterion (AIC) and the number of observations (following the guidelines from the original articles). The missing data are supplemented according to the procedure proposed by Ryan and Giles (1998). The problem of the low power of these tests in relation to small samples is also addressed. Specifically, in the ADF test critical values proposed by MacKinnon (1996) are used, while in the KPSS test – critical values indicated by Syczewska (2010).[11] Conclusions, which are very similar to the ones drawn in relation to the other discussed tests, are presented in Table 6. The analysis also confirms a certain difficulty in obtaining conclusive results concerning the stationarity of economic variables, which has been widely discussed in the literature (cf. Elliott et al., 1992; Charemza & Syczewska, 1998). This is an effect of the tests' poor properties, but (as confirmed herein) it can also be related to specific data features.

3.5. ARIMA model fit and forecasts

The final aim of the study is to determine whether fitting a simple ARIMA model to the series yields different results, depending on the applied data structure method. Table 7 presents the results for several selected variables.[12] The ARIMA(2,2,2) fully describes the correlation structure of the discussed series – the Breusch-Godfrey autocorrelation test does not provide any basis to reject the null hypothesis on the lack of autocorrelation for errors for the vast majority of the analysed variables. It is not surprising that a more complicated structure of the data generating process can be assigned to data for which longer series are observed, i.e. of a higher frequency – quarterly and monthly.

---

[11] More details concerning small sample properties of stationarity tests can be found in Jönsson (2011).

[12] Due to the volume of the analysed set, only selected results are presented in this study. Full results are available upon request.

Table 7. ARIMA fit

Variable		Drift		Constant		MA(1)		MA(2)		AR(1)		AR(2)		BG test *p*-value	
		first	last	first	last	first	last	first	last	first	last	first	last	first	last
Gross capital formation in mln PLN	y	.	14774.6	.	.	.	.	.	.	.	.	.	.	0.7	1.0
Private consumption in mln PLN	q	2542.0	2551.7	.	.	−1.2	−1.2	0.7	0.6	.	.	.	.	0.8	0.9
	y	38869.5	39063.5	.	.	.	.	.	.	.	.	.	.	0.2	0.2
Public consumption in mln PLN	q	820.8	839.9	.	.	−1.5	−1.5	0.9	0.7	−0.1	−0.1	−0.1	−0.3	0.4	0.5
	y	13384.6	12957.9	.	.	.	.	.	.	0.8	.	0.6	.	0.9	0.7
GDP in mln PLN	q	4809.6	4565.3	.	.	−0.8	−1.3	.	0.8	−0.4	−0.1	−0.3	−0.5	0.2	0.1
	y	71156.9	71178.1	.	.	0.6	.	.	.	.	.	.	.	0.8	0.2
Gross capital formation nsa yy	q	.	.	.	4.7	.	−1.0	.	0.1	0.8	1.8	.	−0.9	0.8	0.6
	y	.	.	5.4	5.8	.	.	.	.	.	.	.	.	0.9	0.9
Private consumption nsa yy	q	.	.	3.1	3.0	0.0	.	0.3	.	0.8	0.8	.	.	0.8	0.8
	y	.	.	3.3	3.2	0.4	.	.	.	.	.	.	.	0.8	0.2
Public consumption nsa yy	q	.	.	1.9	.	.	.	.	.	0.5	.	.	.	0.9	0.5
	y	.	.	2.3	3.3	0.5	.	.	.	.	.	.	.	0.9	0.4
GDP nsa yy	q	.	.	3.5	3.6	−0.3	−0.4	0.0	.	1.6	1.6	−0.7	−0.7	0.7	1.0
	y	.	.	3.9	3.9	0.5	0.6	.	.	.	.	.	.	0.8	0.9
CPI mm	m	.	.	.	.	.	.	.	.	−0.3	−0.3	−0.3	−0.3	0.3	0.4
CPI yy (avg)	m	.	.	.	.	.	.	.	.	0.4	0.4	0.1	.	0.6	0.8
	y	−1.4	−1.4	.	.	.	.	.	.	.	.	.	.	0.6	0.6
HICP yy	m	.	.	.	.	−0.4	.	.	.	0.6	0.3	.	.	0.9	0.8
PPI mm	m	.	.	0.2	0.2	0.3	0.4	.	.	0.2	.	.	.	0.8	0.6
PPI yy	m	.	.	.	.	.	0.6	.	0.1	0.5	.	−0.2	.	0.7	0.9
Industrial production nsa yy	m	.	.	5.3	5.2	.	−0.7	.	0.4	0.4	0.9	0.5	.	0.0	0.9
	y	.	.	5.2	5.2	.	.	.	.	.	.	.	.	0.8	0.9
Unemployment rate (eop)	y	.	.	.	.	0.6	0.6	.	.	.	.	.	.	0.9	0.9
Trade balance in mln EUR	m	.	.	.	.	.	.	.	.	−0.5	−1.4	−0.2	−0.8	0.1	0.9
Current account in mln EUR	q	.	.	−1482.0	−2085.3	−0.7	.	.	.	0.9	0.7	.	.	0.7	0.6
	y	.	.	−7951.9	.	.	.	.	.	.	0.9	.	.	0.2	0.6

Note. The table presents an ARIMA(p,d,q) model fit for the first and last data series (p is the order of the autoregressive model AR(p), d is the degree of differencing and q is the order of the moving-average model MA(q)). An intercept in the model is represented by ‘constant’ when d = 0 and ‘drift’ when d = 1 (a detailed description of the ARIMA models can be found in e.g. Hyndman and Athanasopoulos, 2018). BG designates the Breusch-Pagan test for the model residuals. For explanation of acronyms and abbreviations see Table 2 and note to Table 3.

Source: author’s calculations based on: GUS (2003–2017).

As regards national accounts, autoregressive indicators are mostly negative for nominal data and positive for annual growth rates. This is fairly non-intuitive and confirms that the revision process for national accounts is not trivial – numerous readings are revised at the same time, causing nominal data and growth rate data to behave differently.

A comparison of coefficients for the first series and the final readings indicates that the fitted models are similar for the majority of variables. In cases where the form of a model is the same for both series, the coefficients retain their scale and sign. Most differences in coefficients result from a different structure of the fitted model, for example for the HICP inflation the series of the first readings has been estimated as ARMA(1,1),[13] while of the last as AR(1), which creates differences between autoregressive coefficients and can have a considerable impact on the ex ante inference, which is analysed below.

The next step is to verify how data revisions can influence the forecasting process. Table 8 presents point forecasts with horizon one and four (corresponding to the data frequency), estimated by means of the fitted ARIMA models, on the basis of the entire data series available and according to the three procedures described in the Methodology section, as well as differences between the results in percentages.[14]

First of all, it must be noted that the forecast horizon has a low impact on the conclusions, i.e. the differences between predictions are similar for the short and long horizon.

Secondly, for a large number of variables, the total impact of revisions on the forecasting process, i.e. the difference between predictions (1) and (3) reaches more than 10%. Extreme cases, in which the impact exceeds 100%, are not an exception either (in the case of variables with low absolute values, e.g. monthly growth rates). For example, a month on month CPI inflation forecast does not depend strongly on the model. However, a direct impact on the sign is observed – a forecast based on initial data yields a positive result, while the one based on final data involves a negative prediction. A trade balance forecast (in million EUR) also changes the sign depending on the model used for estimation. Similarly large differences (also with regard to a sign) can be observed for month on month PPI inflation, where both an indirect impact of the model and a direct impact of the data themselves change the final prediction.

[13] In the ARMA(p,q) model: p is the order (number of time lags) of the autoregressive model (AR(p)) and q is the order of the moving-average model (MA(q)).

[14] For the percentage calculations a median of the three forecasts is adopted as the denominator. Due to the application of a common denominator, the percentage differences can be interpreted as a decomposition of the entire impact of the revisions on the process of forecasting into a direct and an indirect part. Having taken account of proper signs (absolute values have been provided in the table; however, the sign can be extracted from the differences between forecasts in the first columns), the percentage decompositions sums up to 100%.

Finally, it is worth noting that the decomposition of differences between forecasts provides additional information that is complementary to the previous analysis of the data generating process. The consideration of a monthly trade balance (in million EUR) may serve as an example. Table 8 indicates that a short-term forecast is significantly different for procedure (3) – the one that is based on final readings, both with regard to the structure of the model and to the data used for forecasting. The decomposition indicates that differences in predictions (1) and (2) result mainly from an indirect impact of the estimated parameters for quarterly data and from a direct impact of revisions of annual data. However, when comparing this to Table 7, the fitted models prove identical for final and initial data and the coefficients differ only slightly. Therefore, it may be concluded that even with a very similar data generating process revisions alter the level of the process parameters sufficiently enough for different conclusions to be drawn, based, for example, on the predictions.

Table 8. Comparison of the predictions

Variable		Forecast			Differences in %		
		(1)	(2)	(3)	|(1)–(3)|	|(1)–(2)|	|(2)–(3)|
Horizon = 1							
Gross capital formation in mln PLN	y	89657.0	89281.9	89306.2	0.4	0.4	0.0
Private consumption in mln PLN	q	269620.9	271061.3	271087.0	0.5	0.5	0.0
	y	1108054.3	1111159.4	1111353.4	0.3	0.3	0.0
Public consumption in mln PLN	q	80342.3	85946.7	84495.9	4.9	6.6	1.7
	y	343489.9	339977.8	343881.9	0.1	1.0	1.1
GDP in mln PLN	q	480424.1	482354.1	462591.4	3.7	0.4	4.1
	y	1928202.4	1909562.2	1922348.8	0.3	1.0	0.7
Gross capital formation nsa yy	q	4.0	4.0	6.5	62.7	0.0	62.7
	y	5.4	5.4	5.8	7.0	0.0	7.0
Private consumption nsa yy	q	4.5	4.5	4.3	3.0	1.0	4.0
	y	3.5	3.6	3.2	7.3	2.6	9.9
Public consumption nsa yy	q	1.4	1.4	.	.	0.0	.
	y	2.8	2.9	3.3	16.6	0.7	16.0
GDP nsa yy	q	4.5	4.6	4.4	2.2	1.6	3.8
	y	3.4	3.3	3.3	3.1	3.7	0.6
CPI mm	m	0.4	0.0	−0.1	1063.5	1036.2	27.3
CPI yy (avg)	m	2.7	1.3	1.3	101.4	103.0	1.5
	y	−2.0	−2.0	−2.0	0.0	0.0	0.0
HICP yy	m	1.2	1.2	1.2	1.4	0.0	1.4
PPI mm	m	0.1	0.0	0.0	238.1	351.4	113.2
PPI yy	m	1.2	1.8	1.8	32.6	36.3	3.7
Industrial production nsa yy	m	9.9	6.8	5.4	66.2	44.9	21.3
	m	5.2	5.2	5.2	1.2	0.0	1.2
Unemployment rate (eop)	y	7.6	7.6	7.6	0.6	0.5	0.1
Trade balance in mln EUR	m	−20.1	−26.7	388.7	2029.7	32.7	2062.4
Current account in mln EUR	q	−715.6	−679.2	237.1	140.3	5.4	134.9
	y	−7951.9	−7951.9	−757.1	90.5	0.0	90.5

Table 8. Comparison of the predictions (cont.)

Variable		Forecast			Differences in %		
		(1)	(2)	(3)	|(1)–(3)|	|(1)–(2)|	|(2)–(3)|
Horizon = 4							
Gross capital formation in mln PLN	y	90439.2	90085.3	90058.4	0.4	0.4	0.0
Private consumption in mln PLN	q	289415.4	289423.6	289217.8	0.1	0.0	0.1
	y	1224662.7	1227767.8	1228544.1	0.3	0.3	0.1
Public consumption in mln PLN	q	86760.6	87095.6	87568.7	0.9	0.4	0.5
	y	386212.3	385846.1	382755.6	0.9	0.1	0.8
GDP in mln PLN	q	496277.9	497469.6	485160.3	2.2	0.2	2.5
	y	2141673.2	2123032.9	2135883.0	0.3	0.9	0.6
Gross capital formation nsa yy	q	1.8	1.8	9.6	431.0	0.0	431.0
	y	5.4	5.4	5.8	7.0	0.0	7.0
Private consumption nsa yy	q	3.8	3.8	3.7	4.0	0.6	3.4
	y	3.3	3.3	3.2	1.9	0.0	1.9
Public consumption nsa yy	q	1.8	1.8	.	.	0.0	.
	y	2.3	2.3	3.3	45.2	0.0	45.2
GDP nsa yy	q	4.5	4.6	4.5	0.9	2.6	3.5
	y	3.9	3.9	3.9	2.1	0.0	2.1
CPI mm	m	0.5	−0.1	−0.1	742.9	734.0	8.9
CPI yy (avg)	m	2.9	1.1	1.2	135.4	145.0	9.6
	y	−6.0	−6.0	−6.0	0.0	0.0	0.0
HICP yy	m	1.1	1.1	1.2	7.4	0.0	7.4
PPI mm	m	0.2	0.2	0.2	12.4	0.5	11.8
PPI yy	m	1.3	2.1	1.8	30.3	44.5	14.2
Industrial production nsa yy	m	8.0	5.6	5.4	46.1	42.5	3.6
	m	5.2	5.2	5.2	1.2	0.0	1.2
Unemployment rate (eop)	y	7.6	7.6	7.6	0.6	0.5	0.1
Trade balance in mln EUR	m	−45.7	−51.5	450.9	1086.4	12.7	1099.1
Current account in mln EUR	q	−986.0	−962.4	−1297.0	31.5	2.4	33.9
	y	−7951.9	−7951.9	−498.9	93.7	0.0	93.7

Note. (1), (2) and (3) are forecasts corresponding to the processes defined in the Methodology section. For explanation of acronyms and abbreviations see note to Table 3.

Source: author's calculations based on: GUS (2003–2017).

4. Summary

The article provides a quantitative analysis of data revisions on the basis of a new, real-time macroeconomic dataset for Poland. It is an extension of the results introduced in Ziembińska (2017). The studies have confirmed that for numerous variables the revision process is non-trivial and it is often difficult to indicate a systematic reason for revisions. Some revisions are of a systematic nature, while their scale, character and publication time depend on the data format. Polish data also positively verify a number of hypotheses concerning the impact of data revisions on modelling processes. Procedures assessing the properties of time series can yield very discrepant results, depending on the extent to which the data have been revised. The considered autocorrelation tests provide consistent results. However, apart from the length of the sample, the revision process can also have a significant impact on inference. As far as normality, heteroscedasticity and stationarity tests are concerned, the conclusions indeed depend both on the selected test and the analysed series – either taking account of the revision process or not.

Moreover, the impact of revisions on the fit of ARIMA models has been confirmed. A comparison of the coefficients for the first series and final readings shows that the fitted models are similar for the majority of variables. Where the form of a model is identical for both series, coefficients retain their scale and sign. Most differences in coefficients result from a different structure of the fitted model, which causes differences in the autoregressive structure and can have a considerable impact on the ex ante inference, which is analysed through a prognostic experiment. For a large number of variables, the total impact of revisions on the forecasting process exceeds 10%. Extreme cases, where the impact goes beyond 100%, or situations where data have a direct impact on the forecast sign, are not exceptional either. Consideration given to such results by forecasters could vastly improve the quality of their predictions. Examples of variables (e.g. monthly trade balances) have been identified, where the fitted models are identical for both the final and initial data in which case coefficients vary only slightly and both short and long-term forecasts generate significantly different values. Therefore, even with a very similar data generating process, revisions alter the level of process parameters to such an extent that different conclusions based on the predictions are yielded. Additionally, it is worth noting that with regard to simple ARIMA models, the forecast horizon has only a slight impact on the conclusions, i.e. the differences between predictions are similar both for a short and long horizon.

The new set of real-time data allows plenty of interesting analyses, previously unavailable to the Polish economy. These include further research on the significance of revisions to modelling and forecasting. The revisions might have not only a major impact on the predictive models and their accuracy, but also on the forecasts evaluation procedures, where it is unclear which data vintage was actually predicted. The incorporation of revisions might change the interpretations of the actual economic expectations – their formation, embedded information sets, uncertainty and interrelatedness. These research questions shall be answered in the upcoming articles.

Acknowledgements

The research was supported by the National Science Centre, Poland (the Preludium 2015/17/N/HS4/00209 grant). The author would like to thank Prof. Ryszard Kokoszczyński for all of his helpful comments.

References

Amir-Ahmadi, P., Matthes, C. & Wang, M. (2015). Measurement errors and monetary policy: Then and now (FRB Richmond Working Paper 15–13). https://fraser.stlouisfed.org/title/working-papers-federal-reserve-bank-richmond-3942/measurement-errors-monetary-policy-531441.

Banerjee, A., Dolado, J. J., Galbraith, J. W. & Hendry, D. (1993). Co-integration, error correction, and the econometric analysis of non-stationary data. Oxford: Oxford University Press. https://doi.org/10.1093/0198288107.001.0001.

Chang, A. C., Li, P. (2018). Measurement Error in Macroeconomic Data and Economics Research: Data Revisions, Gross Domestic Product, and Gross Domestic Income. *Economic Inquiry*, *56*(3), 1846–1869. https://doi.org/10.1111/ecin.12567.

Charemza, W. W., Syczewska, E. M. (1998). Joint application of the Dickey-Fuller and KPSS tests. *Economics Letters*, *61*(1), 17–21. https://doi.org/10.1016/S0165-1765(98)00149-9.

Clark, T. E., McCracken, M. W. (2010). *Testing for Unconditional Predictive Ability* (Federal Reserve Bank of St. Louis Working Paper No. 2010-031A). https://doi.org/10.20955/wp.2010.031.

Cole, R. (1969). Data Errors and Forecasting Accuracy. In J. A. Mincer (Ed.), *Economic forecasts and expectations: analysis of forecasting behavior and performance* (pp. 47–82). Cambridge: NBER. https://www.nber.org/system/files/chapters/c1215/c1215.pdf.

Croushore, D. (2010). An evaluation of inflation forecasts from surveys using real-time data. *The B.E. Journal of Macroeconomics*, *10*(1), 1–32. https://doi.org/10.2202/1935-1690.1677.

Croushore, D. (2011). Frontiers of Real-Time Data Analysis. *Journal of Economic Literature*, *49*(1), 72–100. https://doi.org/10.1257/jel.49.1.72.

Croushore, D., Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, *105*(1), 111–130. https://doi.org/10.1016/S0304-4076(01)00072-0.

Croushore, D., Stark, T. (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics*, *24*(4), 507–531.

Elliott, G., Rothenberg, T. J. & Stock, J. H. (1992). *Efficient tests for an autoregressive unit root* (NBER Working Paper Np. T130). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=995460.

Farebrother, R. W. (1980). Algorithm AS 153: Pan's Procedure for the Tail Probabilities of the Durbin-Watson Statistic. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *29*(2), 224–227. https://doi.org/10.2307/2986316.

GUS. (2003–2017). *Biuletyn statystyczny*.

GUS. (2016). *Policy of Revising Statistical Data and Rules of Handling Publication Errors*. https://bip.stat.gov.pl/dzialalnosc-statystyki-publicznej/polityka-rewizji-danych-statystycznych-oraz-zasady-postepowania-z-bledami-publikacyjnymi/.

Hope, A. C. A. (1968). A Simplified Monte Carlo Significance Test Procedure. *Journal of the Royal Statistical Society. Series B (Methodological), 30*(3), 582–598. https://doi.org/10.1111/j.2517-6161.1968.tb00759.x.

Hyndman, R. J., Athanasopoulos, G. (2018). Forecasting: principles and practice. Melbourne: OTexts.

Jönsson, K. (2011). Testing Stationarity in Small-and Medium-Sized Samples when Disturbances are Serially Correlated. *Oxford Bulletin of Economics and Statistics*, *73*(5), 669–690. https://doi.org/10.1111/j.1468-0084.2010.00620.x.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, *11*(6), 601–618.

Mahdi, E., McLeod, A. I. (2012). Improved multivariate portmanteau test. *Journal of Time Series Analysis*, *33*(2), 211–222. https://doi.org/10.1111/j.1467-9892.2011.00752.x.

Nelson, C. R., Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some Evidence and Implications. *Journal of Monetary Economics*, *10*(2), 139–162. https://doi.org/10.1016/0304-3932(82)90012-5.

Patterson, K. D., Heravi, S. M. (1991). Data revisions and the expenditure components of GDP. *The Economic Journal*, *101*(407), 887–901. https://doi.org/10.2307/2233861.

Patterson, K. D., Heravi, S. M. (2004). Revisions to Official Data on U.S. GNP: A Multivariate Assessment of Different Vintages. *Journal of Official Statistics*, *20*(4), 573–602. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/revisions-to-official-data-on-u.s.-gnp-a-multivariate-assessment-of-different-vintages.pdf.

Phillips, P. C. B., Xiao, Z. (1998). A primer on unit root testing. *Journal of Economic Surveys*, *12*(5), 423–470. https://doi.org/10.1111/1467-6419.00064.

Ryan, K. F., Giles, D. E. A. (1998). Testing for unit roots in economic time-series with missing observations. *Advances in Econometrics*, *13*, 203–242. https://doi.org/10.1108/S0731-9053(1999)0000013010.

Syczewska, E. M. (2010). Empirical power of the Kwiatkowski-Phillips-Schmidt-Shin test (Warsaw School of Economics Department of Applied Econometrics Working Papers, No. 3–10). https://ssl-kolegia.sgh.waw.pl/pl/KAE/struktura/IE/struktura/ZES/Documents/Working_Papers/aewp03-10.pdf.

Wüertz, D., Katzgraber, H. G. (2005). *Precise finite-sample quantiles of the Jarque-Bera adjusted Lagrange multiplier test*. https://arxiv.org/pdf/math/0509423.pdf.

Ziembińska, P. (2017). The importance of data revision in the modeling process. *Wiadomości Statystyczne*, *62*(2), 5–30. https://ws.stat.gov.pl/Article/2017/2/005-030.