




Revisions in the Norwegian National Accounts: accuracy, unbiasedness and efficiency in preliminary figures

Magnus Kvåle Helliesen¹ · Håvard Hungnes¹  · Terje Skjerpen¹

Received: 21 April 2020 / Accepted: 29 April 2021 / Published online: 18 May 2021
© The Author(s) 2021, corrected publication 2021

Abstract

This paper investigates the quality of preliminary figures in the Norwegian National Accounts. To address the problem of few observations in such analyses, we use some recently developed system tests for forecast evaluation. We find that preliminary figures for growth rates NA figures (measured in real terms) are accurate, unbiased and efficient. The exception is growth rates for real gross fixed capital formation, which under-predict the final figures. Early published vintages of growth rates for real gross fixed capital formation are often closer to the final vintages than later vintages are.

Keywords Forecasting · Encompassing · Equal predictability · Revision · National accounts data

JEL Classification C12 · C22 · C32 · E01

1 Introduction

In Norway, the final vintage of annual National Accounts figures (based on the normal revision cycle) is published approximately one year and a half after the end of the year for which they apply. Before the final figures are released, preliminary figures are used, among other things, in policy formulation. Therefore, it is important that the preliminary figures are good predictors of the final figures and as accurate as possible.

Data revisions and their implications have been studied for many years. Cole (1969) found that for certain types of forecasts, “The use of preliminary rather than revised data resulted in a *doubling* of the forecast error.” Investigating preliminary National Accounts figures for Germany, Strohsal and Wolf (2020) conclude that the revisions are “biased, large and predictable,” with the noteworthy exception for GDP. Similar results were found for the USA in Aruoba (2008). A recent special issue of *Empirical*

✉ Håvard Hungnes
Havard.Hungnes@ssb.no

¹ Statistics Norway, Oslo, Norway

Economics, see Kunst and Wagner (2020), focuses on forecasting of macroeconomic variables and on the consequence of final vintage of National Accounts figures not being available when forecasts (or nowcasts) are made. For example, Siliverstovs (2020) considers the problem of nowcasting (both point forecasts and density forecasts) when conditioning on variables that are preliminary and finds that a simple univariate model can be better than a sophisticated mixed frequency model to obtain good nowcasts. On the other hand, Claudio et al. (2020) find that a mixed frequency model outperforms forecasts obtained by more traditional single-frequency models when applying data available in real time. Glocker and Wegmueller (2020) consider the problem of dating recessions when taking revisions of GDP into account.

There are only a few previous studies of revisions of the Norwegian National Accounts. Bernhardtsen et al. (2005) consider the problem of estimating the output gap based on preliminary National Accounts figures in Norway. They “find that total revisions of output gap estimates are heavily influenced by uncertainty about the trend at the end of the sample and that data revisions are of less importance.” Jore (2017), studying quarterly Norwegian National Accounts data, finds that first releases of growth in both nominal GDP and its deflator under-predict the final figures. However, these biases cancel out such that “there is no tendency for the first released data [for growth in real GDP] to either over- or under-predict the final data.” Although these two papers analyze revisions in GDP for Norway, they do not study the revisions for all of the main aggregates in the National Accounts.

In this paper, we compare the preliminary published figures for the growth rate of GDP and its main components to the final published figures of these variables.

For the preliminary figures to be characterized as good, they must satisfy certain requirements (see also Aruoba 2008). First, they must be unbiased estimates of the final figures, which we find in tests for most of the figures, both separately and jointly. Second, they must have a small variance (compared to the variance of the final vintage of the figures), and its variance must decrease with new vintages. For most of the National Accounts figures, we find support for both a small and a decreasing variance with newer vintages. Third, they must be efficient; that is, they utilize all available information at the time they are published and, thus, future revisions are unpredictable. In practice, it is impossible to test that there does not exist any available information at the time the preliminary National Accounts figures were made that could have improved them. However, it is possible to test whether one can improve the preliminary figures by using its unconditional mean or earlier vintages of the same variables (also known as weak efficiency). We test weak efficiency both with a test based on Mincer and Zarnowitz (1969) and with an encompassing test (see Chong and Hendry 1986). The efficiency test based on Mincer and Zarnowitz (1969) indicates that the preliminary National Accounts figures cannot be improved by replacing them with a weighted average of the preliminary figures and an unconditional expectation estimate. We show that this test also implies that more “news” is incorporated into the preliminary figures through the revision process. The results from the encompassing tests support this conclusion; new preliminary vintages of the National Accounts figures generally seem to contain all information from earlier vintages, so the latest vintages cannot be improved using earlier vintages.

When testing for efficiency, we do not only apply tests on the variables in the National Accounts separately. We also apply tests of efficiency for the whole vector of variables. The test we use for this is proposed in Hungnes (2018). One of the advantages of testing all of the elements in this vector jointly is that the test can have better power than tests for each variable separately have. Also, it may be easier to draw conclusions from such a joint test if you get divergent results for different variables when testing these separately.

The tests for efficiency might have weak power in small data sets, even if tested jointly. We therefore also apply equal predictability tests, where we examine whether two vintages of the preliminary National Accounts figures are equally good predictors of the final ones. We usually fail to reject this hypothesis when we test the variables separately (with some exceptions). However, when we use the test in Hungnes (2020) to test this hypothesis for all variables jointly, we reject the hypothesis that two vintages of the National Accounts figures are equally good predictors for the final ones. Furthermore, the estimates underlying the tests show that this rejection occurs because the vector of the most recent vintage of preliminary National Accounts figures is significantly better than the vector of an earlier vintage of the same figures.

Although the tests generally show that the preliminary National Accounts figures are unbiased and weakly efficient, there are some exceptions. In particular, this applies to the gross fixed capital formation (a component of gross capital formation). For this variable, it turns out that the preliminary figures are significantly too small for all vintages and that this bias is increasing through the revision process. Thus, the predictions of this variable get worse the closer you get to the time when the final vintage of National Accounts figures is published. The preliminary vintages of the National Accounts figures would have been better if you had kept the figures published in the first vintage of this variable until the final figure is published.

The present paper takes benchmark revisions directly into account. In the main analysis, we do so by attributing the change in the growth rates from the last publication before the benchmark revision to the first publications directly after the benchmark revision to be treated as the effect of the benchmark revision.

Whenever possible, we compare our results with those in Strohsal and Wolf (2020). Strohsal and Wolf (2020) study German quarterly National Accounts data and consider many of the same series as we do here. We find the preliminary figures of growth in Norwegian GDP components to be more accurate than those for Germany. Furthermore, all preliminary National Accounts figures for the Norwegian GDP components are efficient, while this is not the case for either private or public consumption in Germany.

The paper is organized as follows: Sect. 2 describes the revision cycle in the Norwegian National Accounts and gives an overview of the different sources used for the various vintages throughout the revision cycle. The section also presents the National Accounts variables we are considering in this study. Section 3 describes the accuracy of the preliminary vintages of the variables considered, tests for unbiasedness of preliminary vintages, as well as considers two efficiency tests (including an encompassing test) and an equal predictability test. Section 4 concludes.

2 Revision cycles, sources and the National Accounts data

The national statistical office publishes annual National Accounts figures in a relatively fixed cycle: The first to third vintages of the annual figures are preliminary estimates based on the system of the Quarterly National Accounts (QNA). The fourth vintage is the final in the regular revision cycle and is based on the system of the annual National Accounts (NA). The first vintage of the annual figures is published when all the quarters of the year they apply for are available. Since the NA figures of 2006, the first vintage has been published at the beginning of February (about 40 days after the end of the year to which it applies). The second vintage (for the year 2006 and later) has been published in May (about 19 weeks after the end of the year to which it applies), and the third vintage (for the year 2014 and later) has been published in August (about 34 weeks after the end of the year to which it applies). The fourth vintage, which is the final vintage in the regular revision cycle, has (for the year of 2013 and later) been published in August 1 year after the third vintage (about 20 months after the end of the year to which it applies). As a result of benchmark revisions, figures can also be revised after the publication of the fourth vintage.

The times for publishing the different preliminary vintages have changed somewhat over time. For the NA figures for the years 2003–2017, the changes have been exclusively in the direction of higher timeliness. Table 1 shows an overview of the revision cycle for GDP for the mainland Norwegian economy (Mainland GDP). The table also indicates by means of color codes according to which benchmark revision standard each figure was published.

Benchmark revisions usually imply changes in the National Accounts' definitions and guiding principles. If the first vintage of the NA figures for a year was published before a benchmark revision, while the final vintage was published after or as part of a benchmark revision, then revisions from preliminary to final figures could come from definitions and guidance changes, in addition to normal revisions within the regular publishing cycle.

2.1 Sources in the revision cycle

When the first vintage is published, most of the short-term statistics are included in the QNA system. This publication is based on monthly and quarterly figures from the state accounts and KOSTRA (Municipality-State-Reporting), respectively, as the basis for developments in public administration. Several units, including state education and health, have no reporting obligation other than annual figures, so these are estimated in all quarters for the first vintage. For foreign trade, goods data are available, while import and export of services are estimated at a smaller subset for the fourth quarter of the year.

For the second vintage, updated figures for public management are available as annual figures from both the state accounts and KOSTRA are now available. For the state accounts, this implies a full census, while for KOSTRA the reporting deadline for the annual figures is somewhat longer so that several units are missing. For foreign trade, there are some revisions in the goods data, but the main source of revision is

Table 1 Annual growth in GDP Mainland Norway, 1988–2017

year	Vintage 1	Vintage 2	Vintage 3	Vintage 4	Revision
1988	0.4	-0.1	-1.0	-1.7	
1989	-0.9	-1.2	-2.5	-2.2	BR1972
1990	1.2	0.7	1.1	1.1	
1991	0.2	-0.3	-0.6	1.1	
1992	1.3	2.0	2.1	2.2	
1993	2.2	1.9	1.7	2.8	
1994	3.9	4.8	4.3	4.1	
1995	3.3	2.7	3.1	2.9	BR1995
1996	3.2	3.7	4.1	3.8	
1997	3.9	3.8	4.4	4.2	
1998	2.9	3.3	3.3	3.6	
1999	0.8	0.8	1.0	2.7	
2000	1.8	1.8	1.9	2.5	
2001	1.0	1.2	1.7	2.1	BR2002
2002	1.3	1.3	1.7	1.4	
2003	0.7	0.6	0.7	1.4	
2004	3.5	3.5	3.4	4.4	
2005	3.7	3.7	4.1	4.6	BR2006
2006	4.6	4.6	4.3	4.9	
2007	6.0	6.2	6.1	5.6	
2008	2.4	2.6	2.2	1.5	
2009	-1.5	-1.6	-1.3	-1.6	BR2011
2010	2.2	2.1	1.9	1.7	
2011	2.6	2.4	2.5	2.6	
2012	3.5	3.4	3.4	3.8	
2013	2.0	2.0	2.3	2.3	
2014	2.3	2.2	2.2	2.2	BR2014
2015	1.0	1.0	1.1	1.4	
2016	0.8	0.9	1.0	1.1	
2017	1.8	1.9	2.0	2.0	BR2019

For an overview of the different benchmark revisions (BR), see “Appendix A”

that there are complete fourth-quarter figures for service trade. Otherwise, revisions in the short-term statistics can lead to revisions in the NA figures.

The third vintage has full KOSTRA figures, which can give revisions in municipal administration. Also, preliminary structural statistics are available, which will improve the estimates of the market-oriented industries. In connection with the publication of the third vintage, the base year in the QNA is also updated. The shift of base year implies that the short-term indicators in the QNA are weighted together with the NA figures from a more recent year. Changing the base year can provide revisions in itself, even though we have no new information in the short-term statistics.

For the fourth and final vintage, the figures are based on the NA system. The NA system utilizes more sources than the QNA system, and the calculations are done at a more detailed level. In addition, the volume calculations in QNA are mainly done by extrapolating the NA sizes with suitable volume indicators. The NA system consists of accounting sizes at current prices, which are then deflated by suitable price indices to give the figures in real terms.

Table 2 GDP and its main components, 2017

Variable name	Billion NOK	Share of GDP	Growth rate in percent
Gross domestic product (GDP)	3295.4	100.00	2.3
Gross domestic product Mainland Norway (GDPM)	2792.0	84.73	2.0
Final consumption expenditure of household and NPISHs (CP)	1471.7	44.66	2.2
Final consumption expenditure of general government (CO)	791.1	24.01	1.9
Gross capital formation (J)	917.2	27.83	3.1
Gross fixed capital formation (JK)	809.4	24.57	2.6
Total exports (EX)	1197.3	36.33	1.7
Total imports (IMP)	-1081.9	-32.83	1.9

NPISHs non-profit institutions serving households

2.2 National account series considered

In the current paper, we consider revisions in the growth rates (measured in percent growth from the previous year) of gross domestic product (GDP) and its main components, all measured in real terms. The volume and growth rates of the considered series are reported in Table 2 for 2017. Total GDP can be decomposed into *final consumption expenditure of household and non-profit institutions serving households* (CP), which in 2017 makes up about 45% of GDP; *final consumption expenditure of general government* (CO), which makes up about 24% of GDP; *gross capital formation* (J), representing about 28% of GDP; *total exports* (EXP), representing about 36% of GDP; and *total imports*, corresponding to about 33% of GDP.

Gross capital formation includes *changes in stocks and statistical discrepancies*. Since it also includes statistical discrepancies, the figure is derived from the sum of GDP and import minus final consumption expenditure (CP + CO + EXP) and changes in the figure for *gross capital formation* between vintages may not reflect new information on *gross capital formation*. Therefore, we also consider *gross fixed capital formation* (JK), which constitutes the most of *gross capital formation*.

In Norway, GDP Mainland is considered as the most important NA figure. GDP Mainland is defined as the total GDP minus *Petroleum activities and ocean transport*.¹ In 2017, GDP for Mainland Norway constituted about 85% of total GDP.

¹ Before 2014, *service activities incidental to oil and gas* were also excluded from the definition of Mainland GDP.

The vintages of Mainland GDP for all the fiscal years 1988–2017 are reported in Table 1.² 1988 was the first year with NA figures for GDP Mainland Norway, and 2017 was the last year with final NA figures (published in August 2019).³

3 Tests of unbiasedness and efficiency

In this section, we consider measures for accuracy and tests for unbiasedness and efficiency. To do so, we must define some variables. Let $y_{t,(j)}^i$ be the j 'th vintage of the growth rate of the NA figure of variable i applying for year t , where $i = 1, 2, \dots, K$ with K as the number of NA variables we are considering. The 4th vintage of this figure is treated as the final value. Thus, the prediction error for the j 'th vintage of variable i for year t is defined as $e_{t,(j)}^i = y_{t,(4)}^i - y_{t,(j)}^i$ ($j = 1, 2, 3$). The variables are measured in percent growth from the previous year with one decimal as in Table 1.

Due to benchmark revisions, for many years, we have that the first vintage of the NA figures is based on one benchmark revision standard, while the final version is based on another benchmark revision standard. We consider here three different approaches to handling this.

In the first approach, we ignore that such revisions have taken place. Although benchmark revisions may change the level of the variable, they will not necessarily change the year-to-year growth of the variable, since the level of the variable for the previous year is also changed. As the first year with NA figures for GDP Mainland Norway was in 1988, and the last year with final NA figures is from 2017, we consider all years from 1988 to 2017—giving us a sample of 30 observations.

In the second approach, we exclude the years where such benchmark revisions have taken place between the first and the last vintage. As there are 15 years where there has been a benchmark revision between the publication of the first and the final NA figures, we only have a sample of 15 years where there have not been such revisions during the publication process.

In the third approach, we adjust for the effect of the benchmark revisions on the figures. If a revision has taken place between vintage j and vintage $j + 1$ of variable i applying for year t (and we expect that when adjusting for this revision the vintage j should be an unbiased predictor of vintage $j + 1$ of the same variable for the same year), the best adjustment for the benchmark revision is the change in the preliminary figure of this variable for year t from vintage j to vintage $j + 1$. Let $R_{t,(j)}$ be an indicator variable, taking the value of 1 if there is a benchmark revision between version j and $j + 1$ for year t , and zero otherwise. Then, the adjusted predictions are given by

² The data considered here are downloadable from https://www.ssb.no/nasjonalregnskap-og-konjunkturer/artikler-og-publikasjoner/_attachment/382739?_ts=169d809c9d0 with the exception for the final vintage for the year of 2017. The final 2017 NA figures are reported in Table 2.

³ After the submission of this paper, NA figures for 2018 have been published. For this year, there have only been minor revisions in the NA figures considered in this paper, and all the results in the paper would have been almost unaltered by extending the sample.

$$y_{t,(j)}^{i,*} = y_{t,(j)}^i + \sum_{k=j}^3 \left(y_{t,(k+1)}^i - y_{t,(k)}^i \right) R_{t,(k)}, \quad (1)$$

such that the adjusted prediction errors are given by⁴

$$e_{t,(j)}^{i,*} = e_{t,(j)}^i + \sum_{k=j}^3 \left(y_{t,(k+1)}^i - y_{t,(k)}^i \right) R_{t,(k)}. \quad (2)$$

For example, consider the Mainland GDP growth rate for 1991 from vintage 3 (see Table 1), which is -0.6 , i.e., $y_{1991,(3)}^{GDP} = -0.6$, i.e., a decline of 0.6% from 1990 to 1991. Before the final figure is published there is a benchmark revision, so we have $R_{1991,(3)} = 1$. Applying (1), we have $y_{1991,(3)}^{GDP*} = y_{1991,(3)}^{GDP} + (y_{1991,(4)}^{GDP} - y_{1991,(3)}^{GDP}) = y_{1991,(4)}^{GDP} = 1.1$, implying that we are revising the GDP growth for vintage 3 up by 1.7 percentage points. The GDP growth for 1991 in vintages 1 and 2 is revised up by the same figure of percentage points, as this is our estimate of the effect of the benchmark revision for GDP in 1991. Furthermore, applying (2), we have $e_{1991,(3)}^{GDP*} = 0$. Thus, if the benchmark revision takes place between the 3rd and the final (4th) vintage, we are essentially comparing the preliminary figures with its 3rd vintage.

The first approach, where we ignore that a benchmark revision has taken place between the first and final vintage (for the normal revision cycle), seems to be the usual approach in this type of analysis (especially when the variables are formulated in percentage growth); see, e.g., Strohsal and Wolf (2020) and Aruoba (2008). Our third approach, where we adjust for benchmark revisions, is more in line with Clements and Galvão (2013), who include “benchmark dummies” to adjust for benchmark revisions. We use this third approach in the main part of the paper. The results obtained by using the other two approaches are reported in “Appendix B.”

3.1 Accuracy

The root-mean-squared error (RMSE) for vintage j of variable i is given by

$$\text{RMSE}_{(j)}^i = \sqrt{\frac{1}{N} \sum_{t \in \mathcal{T}} \left(e_{t,(j)}^i \right)^2}, \quad (3)$$

where the set \mathcal{T} contains the years included in the sample, and N is the number of elements in \mathcal{T} . If we include all years, we have $N = 30$. If we only include years for

⁴ This can also be formulated as

$$\begin{aligned} e_{t,(3)}^{i,*} &= (1 - R_{t,(3)}) e_{t,(3)}^i \\ e_{t,(2)}^{i,*} &= e_{t,(3)}^{i,*} + (1 - R_{t,(2)}) \left(e_{t,(2)}^i - e_{t,(3)}^i \right) \\ e_{t,(1)}^{i,*} &= e_{t,(2)}^{i,*} + (1 - R_{t,(1)}) \left(e_{t,(1)}^i - e_{t,(2)}^i \right) \end{aligned}$$

which there have been no benchmark revision, we have $N = 15$. And if we adjust for benchmark revisions by using (2), we have $N = 30$ for vintages 1 and 2 and $N = 22$ for vintage 3.⁵ When we adjust for benchmark revisions, we replace $e_{t,(j)}^i$ with $e_{t,(j)}^{i,*}$ defined in (3).

The RMSE is a measure of the accuracy of the preliminary figures. It is reported in Table 3 for the adjusted prediction errors (see also Table 8 and Table 13 in “Appendix B” for the results obtained under alternative treatments of benchmark revisions). The average bias in the preliminary figures is $\bar{e}_{(j)}^i = N^{-1} \sum_{t \in \mathcal{T}} e_{t,(j)}^i$. The RMSE can then be decomposed into a prediction variance and a bias component:

$$\text{RMSE}_{(j)}^i = \sqrt{\text{PV}_{(j)}^i + (\bar{e}_{(j)}^i)^2} \quad \text{where } \text{PV}_{(j)}^i = \frac{1}{N} \sum_{t \in \mathcal{T}} (e_{t,(j)}^i - \bar{e}_{(j)}^i)^2.$$

The root of the prediction variance is also reported in Table 3 (whereas the bias component will be considered in the next section). For comparison, Table 3 also reports the root of the variance of the final vintage of the variable, see the final column of the table, where the variance is given as

$$V^i = \frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(4)}^i - \bar{y}_{(4)}^i)^2.$$

This measure can be used as a benchmark for the prediction variance of the preliminary figures. If the preliminary figure for a variable were the same each year (that is, if $y_{1988,(j)}^i = y_{1989,(j)}^i = \dots = y_{2017,(j)}^i$), then the prediction variance of this vintage of the variable would be equal to the variance of the final vintage. Therefore, V^i could be considered as an upper limit for the prediction variance.

From Table 3 (and also Table 8 and Table 13), we draw the following conclusions:

First, the preliminary figures for private consumption expenditure of households and non-profit institutions serving households (CP) are the most accurate of the NA figures we are considering, based on both RMSE and the root of the prediction variance. However, for the third vintage, the preliminary figures for GDP and Mainland GDP (GDPM) are approximately equally accurate as the figures for private consumption expenditure (CP).

Second, the preliminary figures for gross capital formation (J) and its main component gross fixed capital formation (JK) are the least accurate NA figures we are considering here. For the first two vintages, these figures have an RMSE about four times as high as those for private consumption expenditure (CP), the GDP and the Mainland GDP (GDPM). However, this is due to the high volatility in investments over time. The root of the final vintage figures' variance for these two investment types is also about four times as high as those for CP, GDP and GDPM.

⁵ When we adjust for benchmark revisions, we have for vintage 3 essentially $30 - 8 = 22$ observations as the correction conducted in (2) implies that the adjusted prediction error of the preliminary figure of the 3rd vintage of the figure will be zero in all of the 8 years in the sample where there was a benchmark revision between vintage 3 and the final vintage.

Third, for most variables, the accuracy increases for later vintages. Recall that the number of observations is $N = 22$ in vintage 3, implying that vintage 2 and vintage 3 are not directly comparable; for government consumption (CO), the RMSE (and also the root of PV) is equal for vintages 2 and 3. However, this seems to be due to the changed number of observations, as RMSE decreases both when we consider the full sample (see Table 8) and only years for which there are no benchmark revisions (see Table 13). For gross fixed capital formation (JK), the accuracy decreases throughout the revision cycle, as RMSE (and PV) increases with the vintage figure for these variables. For export (EX), the third vintage seems to be less accurate than the second vintage, independent of how benchmark revisions are treated.

Fourth, based on the ratio between prediction variance and the variance of the final vintage (also known as the noise-to-signal ratio, see, e.g., Aruoba 2008), government consumption (CO) has the least accurate figure for all vintages.

Fifth, RMSE for the adjusted prediction errors given by (2) reported in Table 3 is smaller than the corresponding RMSE for the unadjusted prediction errors for the full sample reported in Table 8. This indicates that the correction for benchmark revisions in (2) works well.

Sixth, the small difference between RMSE and the root of prediction variance indicates only small biases. The biggest difference is found for gross fixed capital formation (JK). Below, we will formally test for the absence of bias.

Strohsal and Wolf (2020) consider the accuracy of the first vintage of many of the same NA figures for Germany, though they consider quarterly NA figures. For GDP, Strohsal and Wolf (2020) estimate a noise-to-signal ratio of 0.44, which is in line with the estimates we get for the Norwegian GDP. For private consumption, public consumption and investments, which correspond to CP, CO and J in our analysis, they identify noise-to-signal ratios about twice as large (they report ratios of 0.74, 1.24 and 0.60, respectively) as those we find for Norway. For exports, the estimate for Germany's signal-to-noise ratio in Strohsal and Wolf (2020) is larger than what we identify for Norway (0.43 vs. 0.32).

3.2 Unbiasedness

The test of unbiasedness is based on

$$e_{t,(j)}^i = \mu_{(j)}^i + u_{t,(j)}^i, \quad (4)$$

where $u_{t,(j)}^i$ is a mean-zero error term, and $\hat{\mu}_{(j)}^i = \bar{e}_{(j)}^i$. Let $d_t^i = e_{t,(j)}^i$ and $\bar{d}^i = N^{-1} \sum_{t \in \mathcal{T}} d_t^i = \bar{e}^i$ (where we suppress the subscript j for vintage). A t test statistic for the null hypothesis $\mu_{(j)}^i = 0$ is

$$N^{1/2} \bar{d}^i \left(\hat{q}^i \right)^{-1/2}, \quad (5)$$

Table 3 Accuracy in preliminary figures, 1988–2017 (adjusted preliminary figures)

Variable	Vintage 1 ($N = 30$)		Vintage 2 ($N = 30$)		Vintage 3 ($N = 22$)		Final vintage ($N = 30$)	
	$RMSE_{(1)}^i$	$\sqrt{\frac{PV_{(1)}^i}{Vt}}$	$RMSE_{(2)}^i$	$\sqrt{\frac{PV_{(2)}^i}{Vt}}$	$RMSE_{(3)}^i$	$\sqrt{\frac{PV_{(3)}^i}{Vt}}$	$RMSE_{(3)}^i$	$\sqrt{\frac{PV_{(3)}^i}{Vt}}$
GDP	0.67	0.43	0.57	0.57	0.42	0.40	0.42	0.25
GDPM	0.64	0.35	0.55	0.54	0.44	0.42	0.44	0.23
CP	0.58	0.30	0.53	0.53	0.41	0.41	0.41	0.21
CO	0.93	0.69	0.88	0.88	0.88	0.88	0.88	0.67
J	2.53	2.50	2.10	2.10	1.88	1.86	1.88	0.23
JK	2.29	2.19	2.43	2.25	2.65	2.41	2.65	0.29
EX	1.16	1.11	0.82	0.81	0.97	0.93	0.97	0.27
IMP	1.55	1.49	1.00	1.00	0.79	0.76	0.79	0.17

Root-mean-squared error (RMSE) and root of both the absolute and the relative prediction variance, as well as the root of the variance of the final vintage

where \hat{q}^i is a consistent estimate of the variance of d_t^i . We use

$$\hat{q}^i = \frac{1}{N} \sum_{t \in \mathcal{T}} (d_t^i - \bar{d}^i)^2 + \frac{2}{N} \sum_{l=1}^{\tau} \sum_{t, t+l \in \mathcal{T}} w_l (d_t^i - \bar{d}^i) (d_{t+l}^i - \bar{d}^i), \tag{6}$$

where τ is the order of autocorrelation (where we in the current paper assume $\tau = 1$ due to few observations), and w_l denotes weights (where we follow Newey and West (1987) and use $w_l = 1 - \frac{l}{\tau+1}$). Furthermore, the notation $t, t + l \in \mathcal{T}$ means that we take the sum over all combinations where both t and $t + l$ are elements in \mathcal{T} . The test statistic given by (5) is asymptotically normally distributed. However, in small samples we assume it to be t -distributed with $N - 1$ degrees of freedom.

We also consider the joint test of all elements in the vector $(\mu_{(j)}^1, \mu_{(j)}^2, \dots, \mu_{(j)}^K)'$ being zero, by applying the test statistic

$$N \bar{\mathbf{d}}' \hat{\mathbf{Q}}^{-1} \bar{\mathbf{d}}, \tag{7}$$

with $\bar{\mathbf{d}} = (\bar{d}^1, \bar{d}^2, \dots, \bar{d}^K)'$, and $\hat{\mathbf{Q}}$ being a matrix version of \hat{q} defined as

$$\begin{aligned} \hat{\mathbf{Q}} &= \frac{1}{N} \sum_{t \in \mathcal{T}} (\mathbf{d}_t - \bar{\mathbf{d}}) (\mathbf{d}_t - \bar{\mathbf{d}})' \\ &+ \frac{1}{N} \sum_{l=1}^{\tau} \sum_{t, t+l \in \mathcal{T}} w_l (\mathbf{d}_t - \bar{\mathbf{d}}) (\mathbf{d}_{t+l} - \bar{\mathbf{d}})' \\ &+ \frac{1}{N} \sum_{l=1}^{\tau} \sum_{t, t+l \in \mathcal{T}} w_l (\mathbf{d}_{t+l} - \bar{\mathbf{d}}) (\mathbf{d}_t - \bar{\mathbf{d}})', \end{aligned} \tag{8}$$

with $\mathbf{d}_t = (d_t^1, d_t^2, \dots, d_t^K)'$. The test statistic in (7) is asymptotically χ^2 -distributed with K degrees of freedom. In small samples, however, we assume it to be F -distributed with K degrees of freedom in the numerator and $N - 1$ degrees of freedom in the denominator.⁶

Table 4 reports the test results for unbiasedness (see also Tables 9 and 14): Using a 5% significance level, we do not reject the null hypothesis that the preliminary data are unbiased, except for the second vintage of gross fixed capital formation (JK). This is also supported in Table 14, whereas in Table 9 we cannot reject the null hypothesis for JK. Since we are using a 5% significance level for the tests, we will expect 1 out of 20 independent tests to yield rejection even if the null hypothesis is true. Therefore,

⁶ A potential problem with this and the remaining joint test is the approximately linear relationship between the variables measured in percent growth: $\frac{\text{GDP}_t}{\text{GDP}_{t-1}} = s_{\text{CP},t-1} \frac{\text{CP}_t}{\text{CP}_{t-1}} + s_{\text{CO},t-1} \frac{\text{CO}_t}{\text{CO}_{t-1}} + s_{J,t-1} \frac{J_t}{J_{t-1}} + s_{\text{EXP},t-1} \frac{\text{EXP}_t}{\text{EXP}_{t-1}} - s_{\text{IMP},t-1} \frac{\text{IMP}_t}{\text{IMP}_{t-1}}$, where $s_{\text{CP},t} = \frac{\text{CP}_t}{\text{GDP}_t}$ is the private consumption to GDP ratio in year t (and similarly for $s_{\text{CO},t}, s_{J,t}, s_{\text{EXP},t}$, and $s_{\text{IMP},t}$). If these ratios are time-invariant, the covariance matrix in (8) (and also $\hat{\Sigma}$ for the later defined t test) will not be positive definite and, thus, not invertible. If this turns out to be a problem, one can exclude one of the variables.

based on the tests for each combination of variables and vintages, we could argue that the overall conclusion from the test results reported in Table 4 is that they are in line with the hypothesis that the preliminary figures are unbiased. However, the joint hypothesis for unbiasedness, which considers the hypothesis that all preliminary figures are unbiased, is rejected for both vintages 2 and 3. Also, unbiasedness for vintage 1 is close to being rejected at the 5% significance level, which is due to the biased preliminary figures for JK; if we exclude JK in the vector of considered variables, we cannot reject that the vector of preliminary variables is unbiased. Then, our overall conclusion is that preliminary figures for JK are significantly biased and under-predict the final vintage. Our results are also in line with Strohsal and Wolf (2020), who rejected unbiasedness for preliminary German NA figures for vintage 1.

3.3 Weak efficiency

A problem with the test of unbiasedness in the previous section is that we fail to reject the null of absence of bias not only if the estimated bias ($\hat{\mu}_{(j)}^i$) is close to zero but also if the variance in (6) is large. When $\tau = 0$, the variance in (6) is equal to the prediction variance and can be decomposed as

$$\begin{aligned} \frac{1}{N} \sum_{t \in \mathcal{T}} \left(e_{t,(j)}^i - \bar{e}_{(j)}^i \right)^2 &= \frac{1}{N} \sum_{t \in \mathcal{T}} \left(y_{t,(4)}^i - \bar{y}_{(4)}^i \right)^2 + \frac{1}{N} \sum_{t \in \mathcal{T}} \left(y_{t,(j)}^i - \bar{y}_{(j)}^i \right)^2 \\ &\quad - 2 \frac{1}{N} \sum_{t \in \mathcal{T}} \left(y_{t,(4)}^i - \bar{y}_{(4)}^i \right) \left(y_{t,(j)}^i - \bar{y}_{(j)}^i \right), \end{aligned} \tag{9}$$

(where $\bar{y}_{(j)}^i = N^{-1} \sum_{t \in \mathcal{T}} y_{t,(j)}^i$ ($j = 1, 2, 3, 4$), and, then by definition, $\bar{e}_{(j)}^i = \bar{y}_{(4)}^i - \bar{y}_{(j)}^i$) which shows that this variance does not only become high if there is a large observed variance in the variable we want to predict (the first term), but also if there is a large observed variance in the prediction (the second term), and in particular if the preliminary predictions are not highly positively correlated with the variable we want to predict (the third term).

To handle the problem that non-rejection of unbiasedness can be due to a large variance in (6), we also apply the Mincer and Zarnowitz (1969) regression, which usually is used to test the joint hypothesis of unbiasedness and weak efficiency, given as $\beta_0^i = 0$ and $\beta_1^i = 1$ in

$$\begin{aligned} y_{t,(4)}^i &= \beta_0^i + \beta_1^i y_{t,(j)}^i + v_{t,(j)}^i \\ &\Downarrow \\ e_{t,(j)}^i - \bar{e}_{(j)}^i &= \beta_1^{i*} \left(y_{t,(j)}^i - \bar{y}_{(j)}^i \right) + v_{t,(j)}^{*i}, \end{aligned}$$

where $\beta_1^{i*} = \beta_1^i - 1$ and $v_{t,(j)}^{*i} = v_{t,(j)}^i - N^{-1} \sum_{s \in \mathcal{T}} v_{s,(j)}^i$. The OLS estimator for β_1^{i*} is

Table 4 Tests for absence of bias in preliminary figures, 1988–2017 (adjusted preliminary figures)

Variable	Vintage 1 ($N = 30$)			Vintage 2 ($N = 30$)			Vintage 3 ($N = 22$)					
	$\hat{\mu}$	std. err.	test st.	p value	$\hat{\mu}$	std. err.	test st.	p value	$\hat{\mu}$	std. err.	test st.	p value
GDP	0.00	0.15	0.02	0.98	0.02	0.11	0.21	0.84	0.13	0.08	1.60	0.12
GDPM	0.03	0.13	0.25	0.80	0.08	0.10	0.74	0.47	0.11	0.08	1.30	0.21
CP	0.02	0.11	0.15	0.88	0.02	0.10	0.19	0.85	0.03	0.09	0.34	0.74
CO	0.21	0.15	1.38	0.18	0.07	0.18	0.41	0.68	0.03	0.21	0.15	0.88
J	-0.34	0.50	-0.67	0.51	-0.19	0.37	-0.50	0.62	-0.28	0.40	-0.70	0.49
JK	0.69	0.35	1.95	0.06	0.91	0.40	2.29	0.03*	1.10	0.54	2.06	0.05
EX	0.33	0.22	1.50	0.14	0.10	0.15	0.66	0.52	0.27	0.21	1.29	0.21
IMP	0.44	0.29	1.54	0.13	-0.06	0.18	-0.32	0.75	-0.24	0.17	-1.42	0.17
All	2.25	0.05	2.59	0.03*	3.04	0.02*
All exc. JK	0.87	0.54	0.93	0.50	2.09	0.09

Testing the null hypothesis of $\mu_j^i = 0$ in the regression $e_{t,(j)}^{i*} = \mu_j^i + u_{t,(j)}^i$, where $e_{t,(j)}^{i*}$ is the adjusted difference between vintage j and the final vintage of a National Accounts figure. The tests where standard error (std. err.) is reported are t tests with $N - 1$ degrees of freedom when one variable is considered; otherwise, the tests are F tests with K degrees of freedom in the numerator and $N - 1$ degrees of freedom in the denominator

** and * indicate significance at the 1% and 5% level, respectively

$$\begin{aligned} \hat{\beta}_1^{i*} &= \frac{\frac{1}{N} \sum_{t \in \mathcal{T}} (e_{t,(j)}^i - \bar{e}_{(j)}^i) (y_{t,(j)}^i - \bar{y}_{(j)}^i)}{\frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(j)}^i - \bar{y}_{(j)}^i)^2} \\ &= \frac{\frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(4)}^i - \bar{y}_{(4)}^i) (y_{t,(j)}^i - \bar{y}_{(j)}^i)}{\frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(j)}^i - \bar{y}_{(j)}^i)^2} - 1. \end{aligned}$$

Inserting the expression for $\hat{\beta}_1^{i*}$ in (9) yields

$$\begin{aligned} \frac{1}{N} \sum_{t \in \mathcal{T}} (e_{t,(j)}^i - \bar{e}_{(j)}^i)^2 &= \frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(4)}^i - \bar{y}_{(4)}^i)^2 \\ &\quad - (1 + 2\hat{\beta}_1^{i*}) \frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(j)}^i - \bar{y}_{(j)}^i)^2. \end{aligned}$$

Therefore, for the observed prediction variance to exceed the variance in the preliminary version j , we must have $\hat{\beta}_1^{i*} < -0.5$. If $\hat{\beta}_1^{i*} = 0$, we have that

$$\begin{aligned} \frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(j)}^i - \bar{y}_{(j)}^i)^2 &= \frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(4)}^i - \bar{y}_{(4)}^i)^2 - \frac{1}{N} \sum_{t \in \mathcal{T}} (e_{t,(j)}^i - \hat{\mu}_{(j)}^i)^2 \\ &\leq \frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(4)}^i - \bar{y}_{(4)}^i)^2, \end{aligned}$$

i.e., the variance of preliminary figures must be smaller than the variance of the final figures. Thus, the test of $\beta_1^* = 0$ (or the more common joint test of $\beta_0 = 0$ combined with $\beta_1^* = 0$) is also a test of whether the revision from vintage j to the final vintage contains “news,” see Mankiw et al. (1984) and Croushore and Stark (2003).

Furthermore, if $PV_j^i = \frac{1}{N} \sum_{t \in \mathcal{T}} (e_{t,(j)}^i - \bar{e}_{(j)}^i)^2$ decreases with the vintage version (as is the case for most variables in Table 3), it follows that $\frac{1}{N} \sum_{t \in \mathcal{T}} (y_{t,(j)}^i - \bar{y}_{(j)}^i)^2$ increases with the vintage version. Since PV_j^i decreases with the vintage version for most variables (as can be seen from Table 3), the test of $\beta_1^* = 0$ also becomes a test of whether each revision step contains “news.”

The hypothesis of $\beta_1^i = 1 \Leftrightarrow \beta_1^{i*} = 0$ can be tested by defining $d_t^i = (e_{t,(j)}^i - \bar{e}_{(j)}^i) (y_{t,(j)}^i - \bar{y}_{(j)}^i)$ (since \bar{d}^i is the numerator of the estimator for β_1^*) and applying (5). The joint test for all variables in a vintage is conducted by defining $\mathbf{d}_t = (d_t^1, d_t^2, \dots, d_t^K)'$ and applying (7). Table 5 reports the results (see also Table 10 and Table 15).

If we under the alternative hypothesis impose the restriction that β_1^* is equal across all variables ($\beta_1^{1*} = \beta_1^{2*} = \dots = \beta_1^{K*}$), then the test of the hypothesis that this

Table 5 Test of efficiency in preliminary figures, 1988–2017 (adjusted preliminary figures)

Variable	Vintage 1 ($N = 30$)			Vintage 2 ($N = 30$)			Vintage 3 ($N = 22$)		
	$\hat{\beta}_1^*$	std. err.	p value	$\hat{\beta}_1^*$	std. err.	p value	$\hat{\beta}_1^*$	std. err.	p value
GDP	-0.04	0.06	0.54	-0.04	0.05	0.48	-0.00	0.06	1.00
GDPM	0.05	0.10	0.59	0.01	0.08	0.92	-0.03	0.05	0.51
CP	0.06	0.11	0.59	0.02	0.08	0.84	0.03	0.04	0.41
CO	-0.16	0.18	0.39	-0.16	0.08	0.05*	-0.02	0.11	0.84
J	-0.00	0.05	0.98	-0.02	0.06	0.78	-0.02	0.06	0.68
JK	0.02	0.03	0.68	0.03	0.04	0.48	0.03	0.05	0.52
EX	-0.02	0.06	0.77	0.02	0.03	0.38	-0.04	0.06	0.52
IMP	-0.06	0.05	0.24	-0.05	0.04	0.17	-0.05	0.04	0.28
All (F test)	0.50	0.10	0.51
All (t test)	-0.02	0.03	0.51	-0.00	0.02	0.94	-0.01	0.01	0.40
All exc. JK (F test)	1.07	0.07	0.52
All exc. JK (t test)	-0.02	0.03	0.44	-0.01	0.02	0.77	-0.01	0.01	0.20

Testing the null hypothesis of $\beta_1^* = 0$ in the regression $e_{t,(j)}^{i*} = \beta_0 + \beta_1^* y_{t,(j)}^i + v_{t,(j)}^i$, where $e_{t,(j)}^{i*}$ is the adjusted difference between vintage j and the final vintage of a National Accounts figure. All tests where standard error (std. err.) is reported are t tests with $N - 1$ degrees of freedom when one variable is involved and $NK - 1$ degrees of freedom when there is a joint test involving all K variables; otherwise, the tests are F tests with K degrees of freedom in the numerator and $N - 1$ degrees of freedom in the denominator

** and * indicate significance at the 1% and 5% level, respectively

common parameter is equal to zero can be formulated by defining

$$d_t = (\mathbf{y}_{t,(j)} - \bar{\mathbf{y}}_{(j)})' \tilde{\Omega}^{-1} (\mathbf{e}_{t,(j)} - \bar{\mathbf{e}}_{(j)})$$

where $\mathbf{y}_{t,(j)} = (y_{t,(j)}^1, y_{t,(j)}^2, \dots, y_{t,(j)}^K)'$, $\bar{\mathbf{y}}_{(j)} = (\bar{y}_{(j)}^1, \bar{y}_{(j)}^2, \dots, \bar{y}_{(j)}^K)'$,
 $\mathbf{e}_{t,(j)} = (e_{t,(j)}^1, e_{t,(j)}^2, \dots, e_{t,(j)}^K)'$, $\bar{\mathbf{e}}_{(j)} = (\bar{e}_{(j)}^1, \bar{e}_{(j)}^2, \dots, \bar{e}_{(j)}^K)'$, and
 $\tilde{\Omega} = N^{-1} \sum_{t \in \mathcal{T}} (\mathbf{e}_{t,(j)} - \bar{\mathbf{e}}_{(j)}) (\mathbf{e}_{t,(j)} - \bar{\mathbf{e}}_{(j)})'$,

and using the test statistic given in (5), which is assumed to be t -distributed with $NK - 1$ degrees of freedom. This follows from an analog derivation of the test statistic for the equal predictability test in Hungnes (2020).

The usual approach when applying the Mincer and Zarnowitz (1969) regression is to test the joint hypothesis $\beta_0^i = 0$ and $\beta_1^i = 1$. Here, we consider these tests separately, as we are reporting results of testing $\beta_1^i = 1$ in Table 5, and test results related to the hypothesis $\beta_0^i = 0$ conditioned on $\beta_1^i = 1$ are reported in Table 4.

With only one exception (vintage 2 of Final consumption expenditure of general government, CO), we do not reject the null hypothesis of $\beta_1^{*i} = 0$. Thus—since we concluded that the results in Table 4 indicate that the preliminary figures also are unbiased for all variables excluding gross fixed capital formation (JK)—the results indicate that preliminary figures are both unbiased and efficient for these variables.

We apply two different tests for testing the null hypothesis of joint efficiency. In the F test, we allow the individual β_1^{*i} to differ across variables under the alternative hypothesis. Using the t test, this parameter is restricted to be equal under the alternative hypothesis. The advantage of the latter test is increased power. For both these tests, we do not reject the null hypothesis of $\beta_1^{*i} = 0, \forall i$, which implies that considered jointly the preliminary figures are efficient estimates of the final vintage of the NA figures.

Strohsal and Wolf (2020) consider the joint hypothesis $\beta_0^i = 0$ and $\beta_1^i = 1$ for the first vintage of German NA figures. They find that this hypothesis is not rejected for GDP, gross capital formation and export. However, for private consumption and public consumption (corresponding to CP and CO here), Strohsal and Wolf (2020) need to apply later vintages for the hypothesis not to be rejected.

3.4 Encompassing

Granger and Newbold (1973) and Chong and Hendry (1986) suggest an encompassing test that can be used to test whether one vintage of the data is inferior to a later vintage of the data, i.e., the first of the two vintages contains no additional information. Consider two different vintages of the value for variable i in year t , denoted $y_{t,(j_1)}^i$ and $y_{t,(j_2)}^i$, where $1 \leq j_2 < j_1 \leq 3$ such that j_1 represents the latest vintage. Then, consider the “composite artificial model”

$$y_{t,(4)}^i = (1 - \alpha) y_{t,(j_1)}^i + \alpha y_{t,(j_2)}^i + u_{t,(j_1,j_2)}^i, \tag{10}$$

which is a weighted average of the values of the two preliminary vintages with weight α for preliminary vintage j_2 , $u_{t,(j_1,j_2)}^i$ is an error term.⁷ The encompassing test of the hypothesis $\alpha = 0$ investigates whether vintage j_1 contains all information (i.e., there is no additional information in vintage j_2). This test can be considered as an alternative test for efficiency to the one considered in Table 5; if the latter vintage of the variables is efficient, its prediction cannot be improved by using an earlier vintage of the same variable.

Note that if we subtract $y_{t,(j_1)}^i$ on both sides of (10), we get the formulation

$$e_{t,(j_1)}^i = \alpha \left(e_{t,(j_1)}^i - e_{t,(j_2)}^i \right) + u_{t,(j_1,j_2)}^i. \tag{11}$$

Harvey et al. (1998) show that test of encompassing based on (10) or (11) is related to the equal predictability test put forward by Diebold and Mariano (1995). Define

$$d_{(a),t,(j_1,j_2)}^i = \left(e_{t,(j_1)}^i - e_{t,(j_2)}^i \right) u_{(a),t,(j_1,j_2)}^i, \tag{12}$$

where

$$u_{(a),t,(j_1,j_2)}^i = e_{t,(j_1)}^i - a \left(e_{t,(j_1)}^i - e_{t,(j_2)}^i \right) \tag{13}$$

is the error in (11) when $a = \alpha$. Thus, the encompassing version of the Diebold and Mariano (1995) test is based on $d_{(a),t,(j_1,j_2)}^i$ with $a = 0$, as

$$d_{(0),t,(j_1,j_2)}^i = \left(e_{t,(j_1)}^i - e_{t,(j_2)}^i \right) e_{t,(j_1)}^i. \tag{14}$$

The Diebold and Mariano (1995) test involves testing whether the population equivalence of the mean of $d_{(0),t,(j_1,j_2)}^i$ is equal to zero. The corresponding test statistic is given in (5) with $d_t^i = d_{(0),t,(j_1,j_2)}^i$. The results are reported in Table 6 (see also Table 11 and Table 16). In the last line of the table, results from a vector version of the test is also reported, see Hungnes (2018). This test involves imposing the restriction that the parameter α is equal across all variables and considers the null hypothesis that the value of this common parameter is equal to zero. We define

$$d_t = \left(\mathbf{e}_{t,(j_1)} - \mathbf{e}_{t,(j_2)} \right)' \tilde{\Omega}^{-1} \mathbf{e}_{t,(j_1)}, \text{ where } \tilde{\Omega} = N^{-1} \sum_{t \in \mathcal{T}} \mathbf{e}_{t,(j_1)} \mathbf{e}_{t,(j_1)}'$$

and apply the test statistic given in (5). This test statistic is assumed to be approximately t -distributed with $NK - 1$ degrees of freedom. This follows from an analog derivation of the test statistic for the equal predictability test in Hungnes (2020) that shows that the test statistic is asymptotically normally distributed; see also Hungnes (2018).

⁷ See Ericsson (1993) for a discussion of why this formulation of the encompassing test is preferred to more general formulations in which the weights are not restricted to sum to unity and possibly an intercept is included.

Table 6 Tests for encompassing, 1988–2017 (adjusted preliminary figures)

Variable	Vintage 2 vs. vintage 1 (N = 28)			Vintage 3 vs. vintage 2 (N = 25)			Vintage 3 vs. vintage 1 (N = 30)				
	$\tilde{\alpha}$	std. err.	test st. p value	$\tilde{\alpha}$	std. err.	test st. p value	$\tilde{\alpha}$	std. err.	test st. p value		
GDP	-0.33	0.53	-0.63	0.54	-0.17	0.17	0.34	-0.04	0.11	-0.31	0.76
GDPM	-0.21	0.44	-0.48	0.64	-0.02	0.14	0.89	0.00	0.12	0.03	0.98
CP	-0.08	0.50	-0.16	0.87	-0.14	0.09	1.64	-0.07	0.11	-0.62	0.54
CO	0.40	0.15	2.64	0.01*	-0.65	0.85	0.45	0.15	0.19	0.78	0.44
J	0.22	0.17	1.33	0.20	0.16	0.19	0.86	0.22	0.11	2.03	0.05
JK	0.71	0.31	2.27	0.03*	0.33	0.22	1.55	0.48	0.20	2.38	0.02*
EX	-0.02	0.18	-0.10	0.92	0.54	0.42	1.27	0.11	0.12	0.93	0.36
IMP	-0.14	0.20	-0.72	0.48	0.09	0.11	0.84	0.10	0.06	1.70	0.10
All	0.11	0.05	2.17	0.03*	0.03	0.05	0.60	0.05	0.02	2.22	0.03*
All exc. JK	0.07	0.06	1.29	0.20	0.01	0.05	0.25	0.04	0.02	1.70	0.09

Testing the null hypothesis of $\alpha = 0$ in the regression $e_{t,(j_1)}^{i*} = \alpha(e_{t,(j_1)}^{i*} - e_{t,(j_2)}^{i*}) + u_{t,(j_1,j_2)}^i$, $\{j_1, j_2\} = \{2, 1\}, \{3, 2\}, \{3, 1\}$, where $e_{t,(j)}^{i*}$ is the adjusted difference between vintage j and the final vintage of a National Accounts figure. All reported tests are t tests with $N - 1$ degrees of freedom when one variable is considered and with $NK - 1$ degrees of freedom when there is a joint test involving all K variables

** and * indicate significance at the 1% and 5% level, respectively

First, we test whether vintage 2 encompasses vintage 1. (Here, $N = 28$ as we exclude the years 1994 and 2001, as there is a benchmark revision for these 2 years between vintage 1 and vintage 2.) For government consumption (CO), we get an estimate of 0.40, implying that the best (numerical) estimate of the growth in CO after the first two vintages are published is a weighted average with 0.60 weight on vintage 2 and 0.40 weight on vintage 1. For gross fixed capital formation (JK), the best estimate would be with a weight of as much as 0.71 on vintage 1. However, this may be a result of how we have adjusted for benchmark revisions. When ignoring the effect of benchmark revisions, we do not reject that the second vintage of either CO or JK encompasses the first vintage, see Table 11. And when excluding the years where a benchmark revision has taken place, we only reject that vintage 2 of CO encompasses vintage 1 at a 5% significance level—and with a much smaller estimate of the optimal weight of vintage 1 (see Table 16). Considering all variables jointly, we reject that vintage 2 encompasses vintage 1 (at the 5% significance level).⁸ If we exclude gross capital formation (JK) from the joint test, we do not reject the null hypothesis that vintage 2 encompasses vintage 1.

Second, we test whether vintage 3 encompasses vintage 2. Table 6 shows that we neither reject the null hypothesis for encompassing for the variables individually nor jointly.

Third, we test whether vintage 3 encompasses vintage 1. Again, we reject the null hypothesis for gross fixed capital formation (JK). The rejection of encompassing for JK also leads to rejection of the joint test for encompassing: If we exclude JK from the vector of considered variables, we do not reject the null hypothesis.

The overall conclusion we draw from Table 6, Table 11 and Table 16 is that when excluding gross fixed capital formation (JK), the test results more or less support the null hypothesis that the latest vintage encompasses an earlier vintage, both considered individually and jointly.

3.5 Equal predictability

The equal predictability test implies testing the null hypothesis of $\alpha = 0.5$ in (11). By defining

$$\begin{aligned} d_t^i &= 2d_{(0.5),t,(j_1,j_2)}^i = \left(e_{t,(j_1)}^i - e_{t,(j_2)}^i \right) \left(e_{t,(j_1)}^i + e_{t,(j_2)}^i \right) \\ &= \left(e_{t,(j_1)}^i \right)^2 - \left(e_{t,(j_2)}^i \right)^2, \end{aligned} \quad (15)$$

we can apply the test statistic given in (5). This test statistic is assumed to be approximately t -distributed with $N - 1$ degrees of freedom.

The vector version of the test is derived in Hungnes (2020) and implies restricting α to be equal across all variables and testing the null hypothesis that this common parameter is equal to one-half. To apply this test, we define

⁸ However, this hypothesis is rejected only in the case when we consider the preliminary figures adjusted for benchmark revisions, see Table 11 and Table 16.

$$d_t = (\mathbf{e}_{t,(j_1)} - \mathbf{e}_{t,(j_2)})' \tilde{\Omega}^{-1} (\mathbf{e}_{t,(j_1)} + \mathbf{e}_{t,(j_2)}^i)$$

$$\text{where } \tilde{\Omega} = N^{-1} \sum_{t \in \mathcal{T}} (\mathbf{e}_{t,(j_1)} + \mathbf{e}_{t,(j_2)}) (\mathbf{e}_{t,(j_1)} + \mathbf{e}_{t,(j_2)})'$$

and apply the test statistic given in (5). This test statistic is according to Hungnes (2020) asymptotically normally distributed, and we assume it to be approximately t -distributed with $NK - 1$ degrees of freedom in small samples.

The results of these tests are reported in Table 7. For the individual tests of equal predictability, we only reject the null hypothesis of $\alpha = 0.5$ in a few cases. However, when testing this hypothesis on the vector of all variables except gross fixed capital formation (JK), we reject the null hypothesis of equal predictability. Furthermore, in the joint test, the point estimates of the optimal weight between the two vintages are in all three cases well below 0.5. Hence, the later vintage is significantly better than an earlier vintage of the variables.

4 Conclusions

In this paper, we have investigated the revision process of National Accounts figures for Norway. We have found that the accuracy of most of the preliminary figures increases throughout the revision process, as the root-mean-squared errors (RMSEs) decrease throughout the revision process. For most of the variables considered here, the preliminary figures are unbiased estimates of their final values. The exception is for gross fixed capital formations, which tend to be underestimated in all preliminary versions.

We also conducted two different tests to investigate the efficiency of the preliminary National Accounts figures. The first, a variant of the Mincer and Zarnowitz (1969) test, indicates that the preliminary figures are weakly efficient estimates of the final National Accounts figures.

The second type of test on efficiency in the preliminary figures involves comparing different vintages of the National Accounts figures against each other, both variable by variable but also the full vectors of the variables. We have conducted an equal predictability test between two different vintages of a vector of the National Accounts series that excludes gross fixed capital formation. For each pair of vintages that we have compared, we have rejected the equal predictability hypothesis, and combined with estimated parameters, we find that the most recent vintage of the preliminary National Accounts data is significantly better than an earlier vintage of the data. We have also conducted encompassing tests for the same pairs of the preliminary vintages. The results for the variables tested separately are somewhat mixed. But when tested jointly, we cannot reject the null hypothesis that a later vintage encompasses an earlier one when we exclude gross fixed capital formation. When including gross fixed capital formation in the vector of variables, we reject the null hypothesis for a majority of the compared vintages, indicating that the National Accounts figures for this series are not optimally updated.

Table 7 Tests for equal predictability, 1988–2017 (adjusted preliminary figures)

Variable	Vintage 2 vs. vintage 1 (N = 28)			Vintage 3 vs. vintage 2 (N = 25)			Vintage 3 vs. vintage 1 (N = 30)		
	$\tilde{\alpha}$	std. err.	test st. p value	$\tilde{\alpha}$	std. err.	test st. p value	$\tilde{\alpha}$	std. err.	test st. p value
GDP	-0.33	0.75	-1.11 0.28	-0.17	0.29	-2.27 0.03*	-0.04	0.30	-1.77 0.09
GDPM	-0.21	0.50	-1.42 0.17	-0.02	0.28	-1.85 0.08	0.00	0.29	-1.70 0.10
CP	-0.08	0.62	-0.94 0.36	-0.14	0.18	-3.50 0.00**	-0.07	0.23	-2.42 0.02*
CO	0.40	0.13	-0.78 0.44	-0.65	0.89	-1.29 0.21	0.15	0.22	-1.58 0.13
J	0.22	0.26	-1.05 0.30	0.16	0.15	-2.33 0.03*	0.22	0.14	-1.90 0.07
JK	0.71	0.31	0.66 0.52	0.33	0.26	-0.65 0.52	0.48	0.33	-0.07 0.95
EX	-0.02	0.18	-2.83 0.01**	0.54	0.35	0.11 0.91	0.11	0.16	-2.42 0.02*
IMP	-0.14	0.41	-1.57 0.13	0.09	0.19	-2.13 0.04*	0.10	0.21	-1.96 0.06
All	0.29	0.13	-1.60 0.11	0.14	0.10	-3.45 0.00**	0.20	0.11	-2.69 0.01**
All exc. JK	0.18	0.15	-2.10 0.04*	0.12	0.11	-3.57 0.00**	0.13	0.12	-3.06 0.00**

Testing the null hypothesis of $\alpha = 0.5$ in the regression $e_{t,(j_1)}^{j*} = \alpha(e_{t,(j_1)}^{j*} - e_{t,(j_2)}^{j*}) + u_{t,(j_2)}^{j*}$, $\{j_1, j_2\} = \{2, 1\}, \{3, 2\}, \{3, 1\}$, where $e_{t,(j)}^{j*}$ is the adjusted difference between vintage j and the final vintage of a National Accounts figure. All reported tests are t tests with $N - 1$ degrees of freedom when one variable is considered and with $NK - 1$ degrees of freedom when there is a joint test involving all K variables

** and * indicate significance at the 1% and 5% level, respectively

The figure for total fixed capital formation is given as the sum of fixed capital formation across all industries. For some industries, Statistics Norway has good sources for gross capital formation for preliminary figures. For other industries, few or no sources are available for early vintages of gross fixed capital figures. The lack of sound sources for these industries may be why revisions of preliminary National Accounts figures for gross capital formation are biased and predictable.

Acknowledgements Thanks to Thomas von Brasch, Bjart Holtsmark, and organizers and participants at the 2nd Vienna Workshop on Economic Forecasting, as well as two anonymous referees for valuable comments.

Funding Open access funding provided by Statistics Norway.

Declarations

Conflict of interest Helliesen declares that he has no conflict of interest. Hungnes declares that he has no conflict of interest. Skjerpen declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Benchmark revisions in Norway since 1988

At more or less regular intervals—approximately every 5 years—revisions of the National Accounts series of figures are carried out. These are referred to as benchmark revisions and normally include the incorporation of new definitions and classifications that come with international regulations. Benchmark revisions may also include the incorporation of new source material, new calculation schemes and any error correction of earlier publications without any definition changes being made. Benchmark revisions often lead to level shifts in the time series. In connection with the publication of benchmark revisions, the time series in the National Accounts are updated. This is done to ensure that the time series are consistent and comparable back in time to provide the most accurate picture of developments. For the years 1988–2017, there have been six benchmark revisions (BR): BR1995, BR2002, BR2006, BR2011, BR2014 and BR2019.

BR1995 involved the incorporation of new definitions and guidelines from SNA93, as well as the review and inclusion of new statistical data from the last 10–15 years before the benchmark revision started. Due to the extensive work for preparing the benchmark revision, the final vintage for the year 1991 was delayed and finally published according to BR1995.

BR2002 was a comprehensive revision without new definitions and classifications. The main reason for carrying out the numerical revision was that Statistics Norway compiled new structural statistics for several industries during the 1990s. The results of the revision were published in June 2002, with revised final figures for 1991–1999, as well as new preliminary figures for 2000 and 2001.

BR2006 was published in December 2006. The main reason for the revision was an EU regulation that required the size of indirectly measured banking and financial services to be distributed to end users—such as product intermediates or consumption—rather than being deducted from the gross domestic product in a correction item. This revision resulted in a higher level of GDP.

BR2011 was published in November 2011. The most significant change was the incorporation of a new industry standard, which is consistent with the EU standard NACE Rev. 2. The new industry standard was the reason why the final vintage for the year 2008 was delayed by one year.

BR2014 was published in November 2014. The most significant change that this major revision entailed was that research and development work went from being treated as intermediate consumption to being treated as investments. Therefore, the benchmark revision redistributed costs from intermediate consumption to investments, and the result was a higher level of GDP. The definition of Mainland Norway was also changed and contributed to increased growth in Mainland GDP (see also footnote 1).

BR2019 was published in August 2019, and incorporation of a new data source for salaries and employment (“a-ordningen”) was the most important single cause of the revisions. Transfer of some specific units from market producers to the government sector, as well as a change in how some existing sources are used, has caused other corrections in earlier published figures.

B Estimation result with alternative treatment for benchmark revisions

See Tables 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17.

Table 8 Accuracy in preliminary figures, 1988–2017 (full sample)

Variable	Vintage 1 ($N = 30$)		Vintage 2 ($N = 30$)		Vintage 3 ($N = 30$)		Final vintage ($N = 30$)	
	$RMSE_{(1)}^i$	$\sqrt{\frac{PV_{(1)}^i}{V^i}}$	$RMSE_{(2)}^i$	$\sqrt{\frac{PV_{(2)}^i}{V^i}}$	$RMSE_{(3)}^i$	$\sqrt{\frac{PV_{(3)}^i}{V^i}}$	$\sqrt{\frac{PV_{(3)}^i}{V^i}}$	$\sqrt{V^i}$
GDP	0.86	0.85	0.76	0.76	0.61	0.60	0.38	1.56
GDPM	0.78	0.75	0.75	0.72	0.63	0.59	0.32	1.83
CP	0.75	0.73	0.72	0.69	0.54	0.52	0.27	1.92
CO	1.12	1.11	0.95	0.95	0.88	0.87	0.66	1.32
J	2.92	2.89	2.66	2.64	1.99	1.97	0.25	7.94
JK	4.90	4.87	3.33	3.33	2.82	2.77	0.34	8.24
EX	1.20	1.10	0.90	0.86	0.93	0.91	0.26	3.51
IMP	1.61	1.53	1.08	1.08	0.92	0.89	0.20	4.52

Root-mean-squared error (RMSE) and root of both the absolute and the relative forecast variance, as well as the root of the variance of the final vintage

Table 9 Tests for absence of bias in preliminary figures, 1988–2017 (full sample)

Variable	Vintage 1 ($N = 30$)				Vintage 2 ($N = 30$)				Vintage 3 ($N = 30$)			
	$\hat{\mu}$	std. err.	test st.	p value	$\hat{\mu}$	std. err.	test st.	p value	$\hat{\mu}$	std. err.	test st.	p value
GDP	0.08	0.19	0.42	0.67	0.08	0.16	0.50	0.62	0.13	0.12	1.12	0.27
GDPM	0.21	0.17	1.27	0.21	0.22	0.15	1.46	0.15	0.21	0.12	1.81	0.08
CP	0.21	0.14	1.46	0.15	0.19	0.13	1.48	0.15	0.15	0.09	1.59	0.12
CO	0.10	0.20	0.48	0.64	0.00	0.20	0.00	1.00	-0.01	0.19	-0.05	0.96
J	-0.43	0.61	-0.71	0.49	-0.33	0.53	-0.62	0.54	-0.22	0.38	-0.58	0.57
JK	0.61	0.80	0.76	0.45	0.19	0.71	0.27	0.79	0.52	0.53	0.97	0.34
EX	0.48	0.23	2.11	0.04*	0.26	0.17	1.48	0.15	0.22	0.19	1.17	0.25
IMP	0.50	0.28	1.79	0.08	0.03	0.20	0.17	0.87	-0.25	0.18	-1.40	0.17
All	2.50	0.03*	2.97	0.01*	5.04	0.00**

Testing the null hypothesis of $\mu_{t(j)}^i = 0$ in the regression $e_{t(j)}^i = \mu_{t(j)}^i + u_{t(j)}^i$, where $e_{t(j)}^i$ is the difference between vintage j and the final vintage of a National Accounts figure. The tests where standard error (std. err.) is reported are t tests with $N - 1$ degrees of freedom when one variable is considered; otherwise, the tests are F tests with K degrees of freedom in the numerator and $N - 1$ degrees of freedom in the denominator

** and * indicate significance at the 1% and 5% level, respectively

Table 10 Test of efficiency in preliminary figures, 1988–2017 (full sample)

Variable	Vintage 1 ($N = 30$)			Vintage 2 ($N = 30$)			Vintage 3 ($N = 30$)					
	$\hat{\beta}_1^*$	std. err.	test st.	p value	$\hat{\beta}_1^*$	std. err.	test st.	p value	$\hat{\beta}_1^*$	std. err.	test st.	p value
GDP	-0.02	0.07	-0.35	0.73	-0.05	0.06	-0.87	0.39	0.01	0.06	0.16	0.87
GDPM	0.05	0.11	0.41	0.68	-0.02	0.10	-0.23	0.82	-0.06	0.06	-0.91	0.37
CP	0.04	0.13	0.28	0.78	-0.01	0.10	-0.09	0.93	0.01	0.06	0.13	0.90
CO	-0.23	0.22	-1.02	0.31	-0.15	0.11	-1.37	0.18	-0.06	0.14	-0.44	0.66
J	-0.00	0.05	-0.05	0.96	-0.02	0.06	-0.37	0.71	-0.02	0.04	-0.50	0.62
JK	-0.21	0.17	-1.28	0.21	-0.08	0.08	-0.99	0.33	0.00	0.04	0.09	0.93
EX	-0.06	0.06	-1.09	0.29	-0.03	0.04	-0.94	0.35	-0.04	0.05	-0.90	0.37
IMP	-0.01	0.08	-0.09	0.93	-0.00	0.06	-0.03	0.97	-0.00	0.03	-0.04	0.97
All (F test)	0.97	0.48	1.16	0.35	1.30	0.28
All (t test)	-0.03	0.03	-0.91	0.36	-0.03	0.02	-1.21	0.22	-0.02	0.02	-0.86	0.39

Testing the null hypothesis of $\beta_1^* = 0$ in the regression $e_{t,(j)}^j = \beta_0 + \beta_1^j v_{t,(j)}^j + v_{t,(j)}^j$, where $e_{t,(j)}^j$ is the difference between vintage j and the final vintage of a National Accounts figure. All tests where standard error (std. err.) is reported are t tests with $N - 1$ degrees of freedom when one variable is involved and $NK - 1$ degrees of freedom when there is a joint test involving all K variables; otherwise, the tests are F tests with K degrees of freedom in the numerator and $N - 1$ degrees of freedom in the denominator. ** and * indicate significance at the 1% and 5% level, respectively.

Table 11 Tests for encompassing, 1988–2017 (full sample)

Variable	Vintage 2 vs. vintage 1 ($N = 30$)			Vintage 3 vs. vintage 2 ($N = 30$)			Vintage 3 vs. vintage 1 ($N = 30$)					
	$\hat{\alpha}$	std. err.	test st.	$\hat{\alpha}$	std. err.	test st.	$\hat{\alpha}$	std. err.	test st.	p value		
GDP	-0.36	0.49	-0.74	0.46	-0.18	0.25	-0.73	0.47	-0.11	0.13	-0.90	0.38
GDPM	0.25	0.46	0.53	0.60	-0.02	0.20	-0.10	0.92	0.12	0.20	0.60	0.55
CP	-0.01	0.63	-0.02	0.98	-0.15	0.14	-1.05	0.30	-0.04	0.15	-0.25	0.80
CO	0.19	0.19	1.03	0.31	-0.09	0.73	-0.12	0.91	0.12	0.24	0.53	0.60
J	0.30	0.17	1.79	0.08	0.12	0.16	0.71	0.48	0.22	0.11	1.97	0.06
JK	-0.03	0.10	-0.26	0.79	0.32	0.17	1.87	0.07	0.17	0.10	1.65	0.11
EX	0.07	0.18	0.37	0.71	0.55	0.33	1.65	0.11	0.27	0.16	1.70	0.10
IMP	-0.13	0.22	-0.61	0.55	0.31	0.17	1.77	0.09	0.18	0.08	2.29	0.03*
All	0.06	0.08	0.82	0.41	0.14	0.07	1.99	0.05*	0.12	0.04	3.25	0.00**

Testing the null hypothesis of $\alpha = 0$ in the regression $e^i_{t,(j_1)} = \alpha(e^i_{t,(j_1)} - e^i_{t,(j_2)}) + u^i_{t,(j_1,j_2)}$, $\{j_1, j_2\} = \{2, 1\}, \{3, 2\}, \{3, 1\}$, where $e^i_{t,(j)}$ is the difference between vintage j and the final vintage of a National Accounts figure. All reported tests are t tests with $N - 1$ degrees of freedom when one variable is considered and with $NK - 1$ degrees of freedom when there is a joint test involving all K variables

** and * indicate significance at the 1% and 5% level, respectively

Table 12 Tests for equal predictability, 1988–2017 (full sample)

Variable	Vintage 2 vs. vintage 1 ($N = 30$)			Vintage 3 vs. vintage 2 ($N = 30$)			Vintage 3 vs. vintage 1 ($N = 30$)					
	$\hat{\alpha}$	std. err.	test st.	p value	$\hat{\alpha}$	std. err.	test st.	p value	$\hat{\alpha}$	std. err.	test st.	p value
GDP	-0.36	0.67	-1.30	0.21	-0.18	0.34	-1.99	0.06	-0.11	0.31	-1.97	0.06
GDPM	0.25	0.41	-0.61	0.55	-0.02	0.30	-1.75	0.09	0.12	0.30	-1.27	0.21
CP	-0.01	0.72	-0.71	0.48	-0.15	0.21	-3.12	0.00**	-0.04	0.23	-2.35	0.03*
CO	0.19	0.22	-1.37	0.18	-0.09	0.73	-0.80	0.43	0.12	0.26	-1.43	0.16
J	0.30	0.27	-0.74	0.47	0.12	0.17	-2.22	0.03*	0.22	0.15	-1.93	0.06
JK	-0.03	0.42	-1.24	0.22	0.32	0.20	-0.94	0.36	0.17	0.26	-1.26	0.22
EX	0.07	0.19	-2.27	0.03*	0.55	0.21	0.23	0.82	0.27	0.13	-1.72	0.10
IMP	-0.13	0.41	-1.53	0.14	0.31	0.16	-1.23	0.23	0.18	0.19	-1.65	0.11
All	0.18	0.13	-2.44	0.01*	0.30	0.08	-2.67	0.01**	0.20	0.09	-3.41	0.00**

Testing the null hypothesis of $\alpha = 0.5$ in the regression $e_{t,(j)}^j = \alpha(e_{t,(j_1)}^j - e_{t,(j_2)}^j) + u_{t,(j_1,j_2)}^j$, $\{j_1, j_2\} = \{2, 1\}, \{3, 2\}, \{3, 1\}$, where $e_{t,(j)}^j$ is the difference between vintage j and the final vintage of a National Accounts figure. All reported tests are t tests with $N - 1$ degrees of freedom when one variable is considered and with $NK - 1$ degrees of freedom when there is a joint test involving all K variables

** and * indicate significance at the 1% and 5% level, respectively

Table 13 Accuracy in preliminary figures, 1988–2017 (excluding main revisions)

Variable	Vintage 1 ($N = 15$)		Vintage 2 ($N = 15$)		Vintage 3 ($N = 15$)		Final vintage ($N = 15$)	
	$RMSE_{(1)}^i$	$\sqrt{\frac{PV_{(1)}^i}{V^i}}$	$RMSE_{(2)}^i$	$\sqrt{\frac{PV_{(2)}^i}{V^i}}$	$RMSE_{(3)}^i$	$\sqrt{\frac{PV_{(3)}^i}{V^i}}$	$RMSE_{(3)}^i$	$\sqrt{\frac{PV_{(3)}^i}{V^i}}$
GDP	0.84	0.82	0.68	0.67	0.38	0.38	0.38	0.16
GDPM	0.74	0.74	0.60	0.60	0.38	0.38	0.38	0.13
CP	0.66	0.65	0.61	0.61	0.40	0.40	0.40	0.12
CO	0.57	0.57	0.57	0.54	0.51	0.47	0.51	0.18
J	3.07	3.03	2.54	2.54	2.15	2.13	2.15	0.27
JK	2.89	2.54	2.81	1.99	2.73	2.25	2.73	0.23
EX	1.29	1.22	0.85	0.84	0.99	0.97	0.99	0.23
IMP	1.83	1.61	1.00	0.99	0.76	0.74	0.76	0.16

Root-mean-squared error (RMSE) and root of both the absolute and the relative forecast variance, as well as the root of the variance of the final vintage

Table 14 Tests for absence of bias in preliminary figures, 1988–2017 (excluding main revisions)

Variable	Vintage 1 ($N = 15$)				Vintage 2 ($N = 15$)				Vintage 3 ($N = 15$)			
	$\hat{\mu}$	std. err.	test st.	p value	$\hat{\mu}$	std. err.	test st.	p value	$\hat{\mu}$	std. err.	test st.	p value
GDP	-0.17	0.26	-0.65	0.53	-0.09	0.19	-0.46	0.65	0.02	0.08	0.25	0.81
GDPM	-0.07	0.22	-0.30	0.77	0.01	0.17	0.08	0.94	0.01	0.08	0.16	0.87
CP	-0.09	0.19	-0.46	0.65	-0.03	0.17	-0.19	0.85	0.02	0.10	0.19	0.85
CO	0.02	0.12	0.17	0.87	-0.19	0.15	-1.28	0.22	-0.20	0.13	-1.48	0.16
J	-0.47	0.90	-0.53	0.61	-0.05	0.68	-0.08	0.94	-0.29	0.57	-0.52	0.61
JK	1.37	0.55	2.51	0.02*	1.98	0.50	3.96	0.00**	1.54	0.61	2.54	0.02*
EX	0.43	0.38	1.14	0.27	0.15	0.25	0.59	0.56	0.21	0.28	0.76	0.46
IMP	0.86	0.42	2.07	0.06	0.15	0.28	0.55	0.59	-0.15	0.22	-0.69	0.50
All	3.62	0.02*	8.76	0.00**	2.96	0.04*

Testing the null hypothesis of $\mu_j^i = 0$ in the regression $e_{t,j}^i = \mu_j^i + u_{t,j}^i$, where $e_{t,j}^i$ is the difference between vintage j and the final vintage of a National Accounts figure. The tests where standard error (std. err.) is reported are t tests with $N - 1$ degrees of freedom when one variable is considered; otherwise, the tests are F tests with K degrees of freedom in the numerator and $N - 1$ degrees of freedom in the denominator

** and * indicate significance at the 1% and 5% level, respectively

Table 15 Test of efficiency in preliminary figures, 1988–2017 (excluding main revisions)

Variable	Vintage 1 ($N = 15$)			Vintage 2 ($N = 15$)			Vintage 3 ($N = 15$)		
	$\hat{\beta}_1^*$	std. err.	test st.	p value	std. err.	test st.	p value	std. err.	test st.
GDP	-0.00	0.11	-0.01	0.99	-0.00	0.08	0.97	-0.02	0.04
GDPM	0.06	0.13	0.49	0.63	0.04	0.10	0.72	-0.02	0.03
CP	0.05	0.09	0.58	0.57	0.02	0.06	0.72	0.01	0.03
CO	-0.01	0.04	-0.30	0.77	-0.04	0.05	0.47	-0.00	0.05
J	0.04	0.12	0.31	0.76	0.04	0.12	0.77	-0.02	0.08
JK	0.03	0.05	0.60	0.56	0.02	0.02	0.36	0.03	0.05
EX	-0.06	0.09	-0.65	0.52	-0.01	0.06	0.94	-0.05	0.09
IMP	-0.09	0.09	-0.99	0.34	-0.06	0.07	0.33	-0.02	0.04
All (F test)	0.78	0.63	0.79
All (t test)	-0.01	0.01	-0.72	0.47	-0.02	0.01	0.17	-0.01	0.01

Testing the null hypothesis of $\beta_1^* = 0$ in the regression $e_{t,(j)}^i = \beta_0 + \beta_1^* v_{t,(j)}^i + v_{t,(j)}^i$, where $e_{t,(j)}^i$ is the difference between vintage j and the final vintage of a National Accounts figure. All tests where standard error (std. err.) is reported are t tests with $N - 1$ degrees of freedom when one variable is involved and $NK - 1$ degrees of freedom when there is a joint test involving all K variables; otherwise, the tests are F tests with K degrees of freedom in the numerator and $N - 1$ degrees of freedom in the denominator. ** and * indicate significance at the 1% and 5% level, respectively.

Table 16 Tests for encompassing, 1988–2017 (excluding main revisions)

Variable	Vintage 2 vs. vintage 1 ($N = 15$)			Vintage 3 vs. vintage 2 ($N = 15$)			Vintage 3 vs. vintage 1 ($N = 15$)					
	$\hat{\alpha}$	std. err.	test st.	p value	$\hat{\alpha}$	std. err.	test st.	p value	$\hat{\alpha}$	std. err.	test st.	p value
GDP	-0.34	0.44	-0.76	0.46	-0.15	0.13	-1.16	0.27	-0.04	0.10	-0.41	0.69
GDPM	-0.21	0.30	-0.71	0.49	0.04	0.13	0.34	0.74	0.00	0.11	0.02	0.98
CP	-0.07	0.37	-0.18	0.86	-0.11	0.05	-1.98	0.07	-0.07	0.06	-1.25	0.23
CO	0.26	0.10	2.50	0.03*	0.04	0.21	0.18	0.86	0.16	0.06	2.45	0.03*
J	0.21	0.16	1.34	0.20	0.12	0.12	1.01	0.33	0.19	0.08	2.30	0.04*
JK	0.02	0.03	0.77	0.46	0.08	0.06	1.48	0.16	0.03	0.02	1.56	0.14
EX	0.03	0.15	0.21	0.84	0.21	0.17	1.19	0.25	0.08	0.09	0.99	0.34
IMP	-0.13	0.20	-0.66	0.52	0.05	0.09	0.59	0.56	0.07	0.05	1.37	0.19
All	0.03	0.02	1.14	0.26	0.04	0.03	1.59	0.11	0.04	0.02	2.66	0.01**

Testing the null hypothesis of $\alpha = 0$ in the regression $e^i_{t,(j_1)} = \alpha(e^i_{t,(j_1)} - e^i_{t,(j_2)} + u^i_{t,(j_1,j_2)}) + u^i_{t,(j_1)}$, where $e^i_{t,(j)}$ is the difference between vintage j and the final vintage of a National Accounts figure. All reported tests are t tests with $N - 1$ degrees of freedom when one variable is considered and with $NK - 1$ degrees of freedom when there is a joint test involving all K variables

** and * indicate significance at the 1% and 5% level, respectively

Table 17 Tests for equal predictability, 1988–2017 (excluding main revisions)

Variable	Vintage 2 vs. vintage 1 ($N = 15$)			Vintage 3 vs. vintage 2 ($N = 15$)			Vintage 3 vs. vintage 1 ($N = 15$)					
	$\hat{\alpha}$	std. err.	test st.	p value	$\hat{\alpha}$	std. err.	test st.	p value	$\hat{\alpha}$	std. err.	test st.	p value
GDP	-0.34	0.62	-1.35	0.20	-0.15	0.25	-2.56	0.02*	-0.04	0.28	-1.90	0.08
GDPM	-0.21	0.34	-2.11	0.05	0.04	0.25	-1.79	0.09	0.00	0.26	-1.92	0.08
CP	-0.07	0.51	-1.12	0.28	-0.11	0.13	-4.73	0.00**	-0.07	0.16	-3.67	0.00**
CO	0.26	0.11	-2.28	0.04*	0.04	0.19	-2.46	0.03*	0.16	0.11	-3.13	0.01**
J	0.21	0.25	-1.17	0.26	0.12	0.13	-2.99	0.01**	0.19	0.12	-2.56	0.02*
JK	0.02	0.15	-3.08	0.01**	0.08	0.15	-2.87	0.01*	0.03	0.15	-3.09	0.01**
EX	0.03	0.15	-3.17	0.01**	0.21	0.19	-1.53	0.15	0.08	0.12	-3.43	0.00**
IMP	-0.13	0.38	-1.64	0.12	0.05	0.16	-2.89	0.01*	0.07	0.18	-2.42	0.03*
All	0.21	0.10	-3.04	0.00**	0.14	0.10	-3.63	0.00**	0.11	0.10	-3.87	0.00**

Testing the null hypothesis of $\alpha = 0.5$ in the regression $e_{t,(j)}^j = \alpha(e_{t,(j)}^j - e_{t,(j_2)}^j) + u_{t,(j_1,j_2)}^j + u_{t,(j_2)}^j$, where $e_{t,(j)}^j$ is the difference between vintage j and the final vintage of a National Accounts figure. All reported tests are t tests with $N - 1$ degrees of freedom when one variable is considered and with $NK - 1$ degrees of freedom when there is a joint test involving all K variables

** and * indicate significance at the 1% and 5% level, respectively

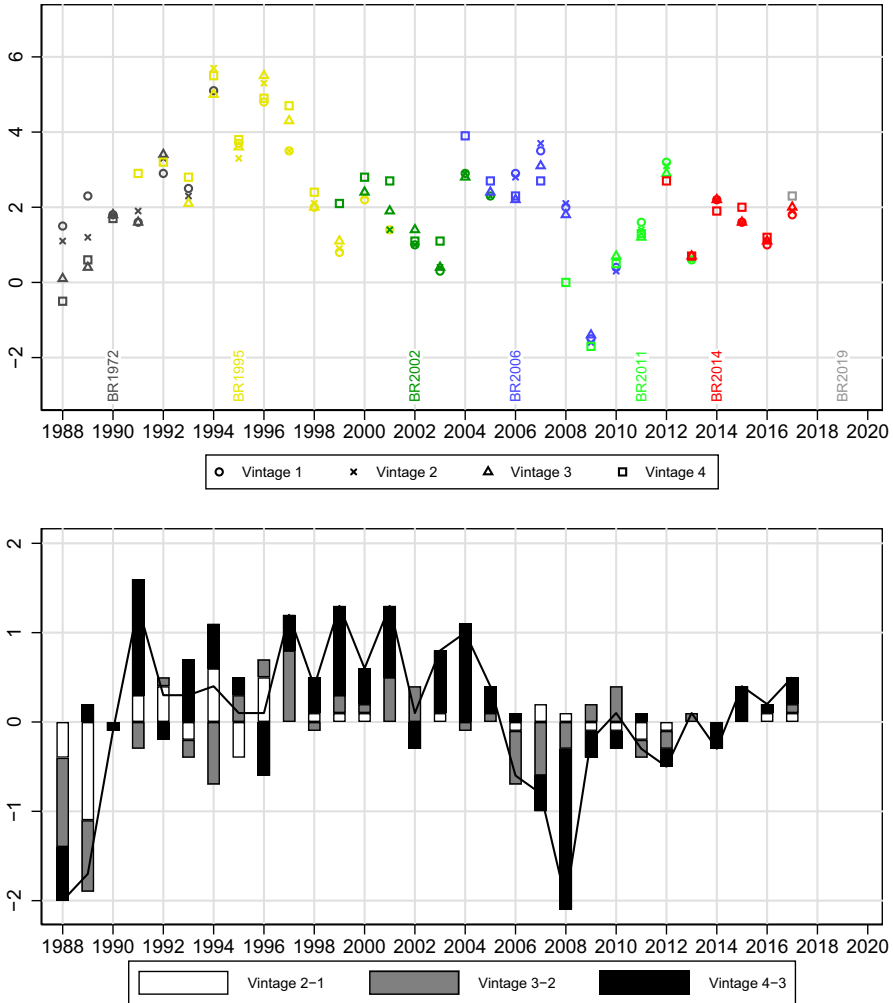


Fig. 1 Growth in gross domestic product (GDP), measured at market prices. Upper part: percentage change from previous year. Lower part: revision in percentage points from on vintage to the next (bars) as well as the total revision from vintage 1 to vintage 4 (line).. *Source:* Statistics Norway

C Graphs

The graphs below show percentage growth rates as published in the first to fourth versions (top), and associated revisions in percentage points (bottom). In the upper graph, the color code indicates which benchmark revision the different versions belong to (Figs. 1, 2, 3, 4, 5, 6, 7 and 8).

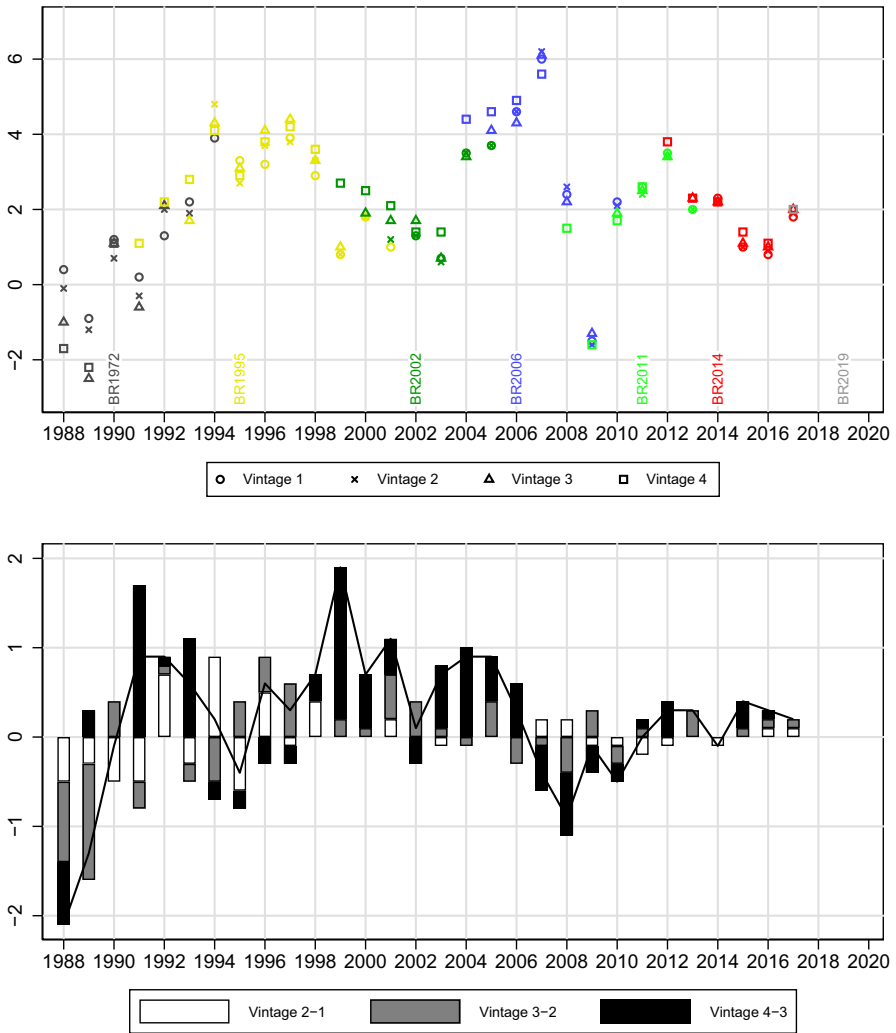


Fig. 2 Growth in gross domestic product Mainland Norway (GDPM), measured at market prices. See Fig. 1

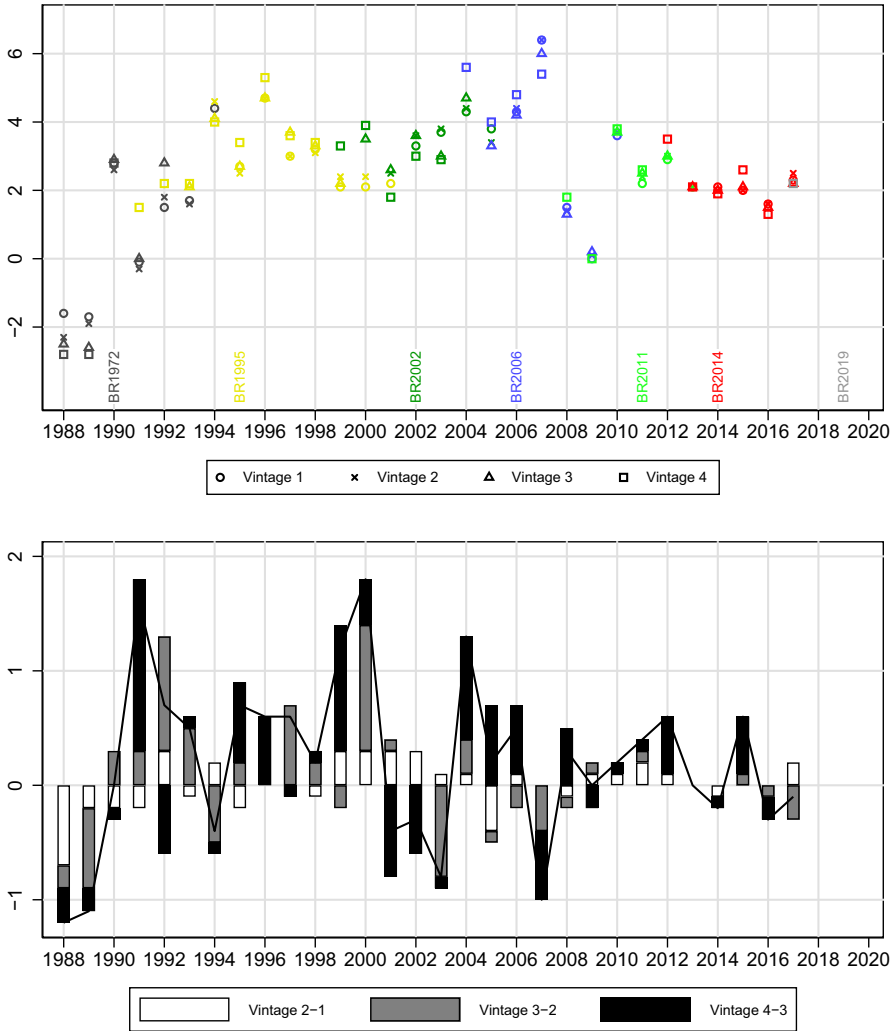


Fig. 3 Growth in private consumption (CP). See Fig. 1

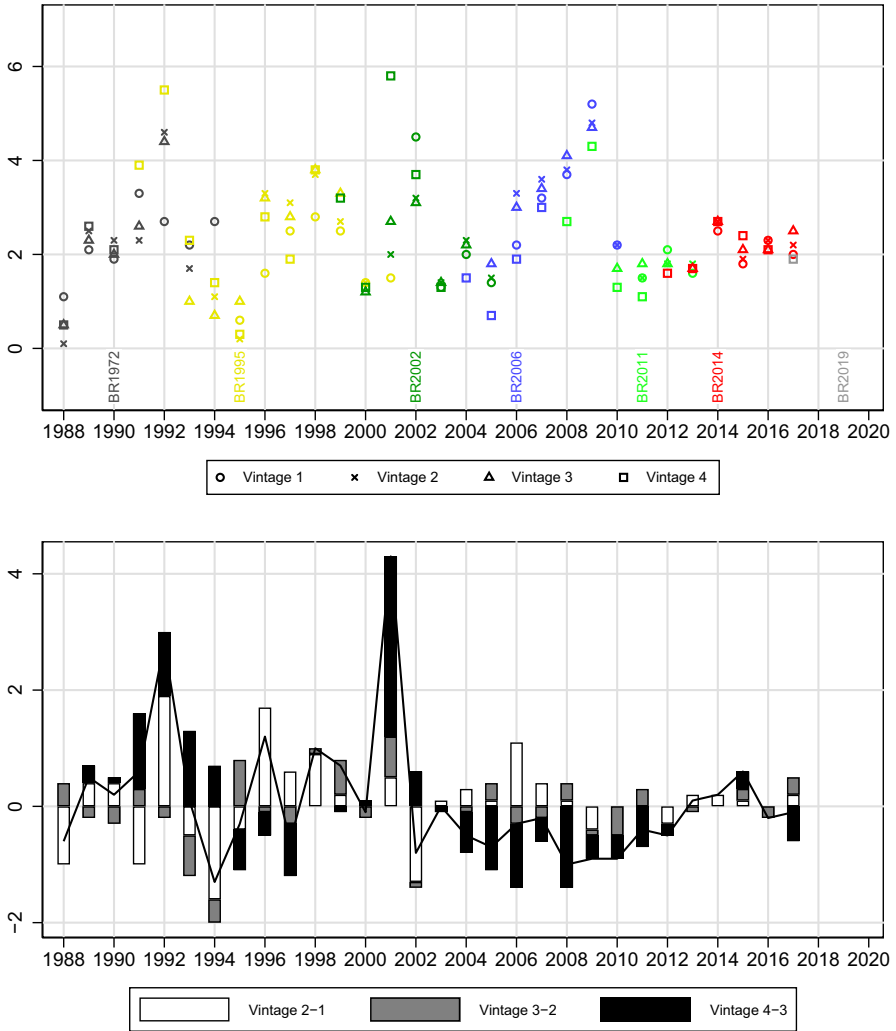


Fig. 4 Growth in general government consumption (CO). See Fig. 1

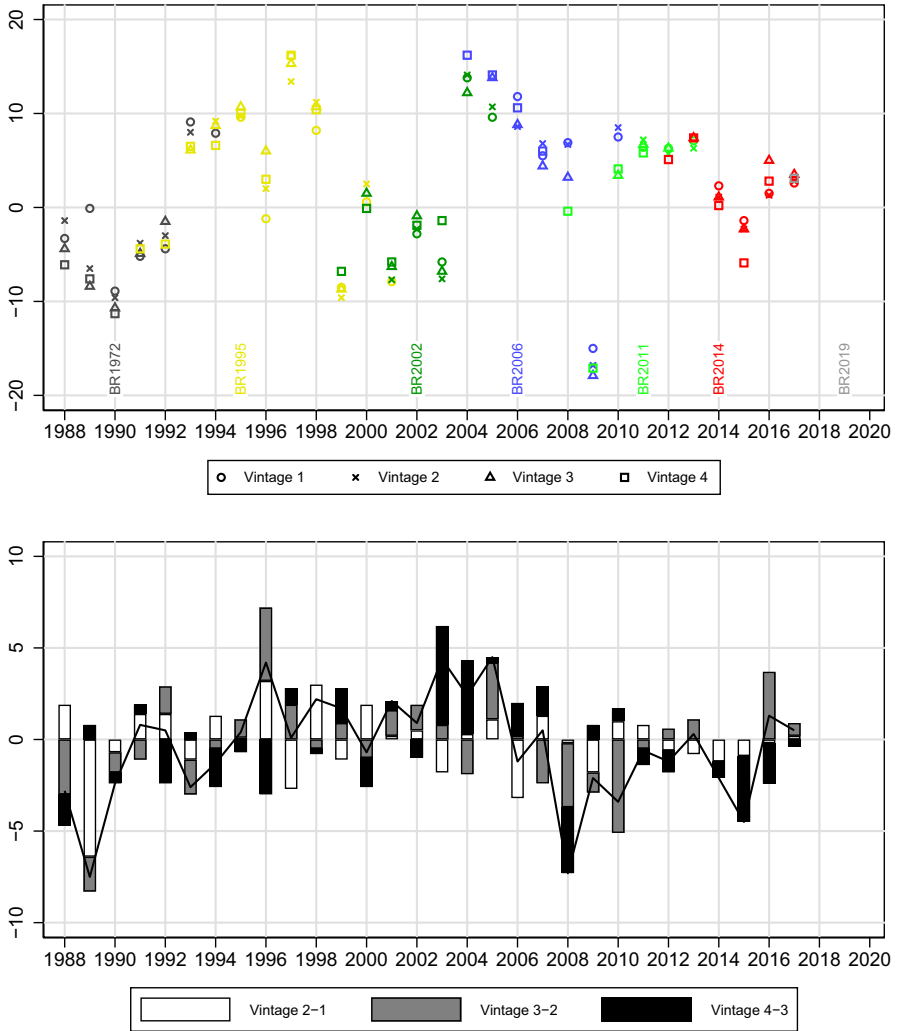


Fig. 5 Growth in gross capital formation (J). See Fig. 1

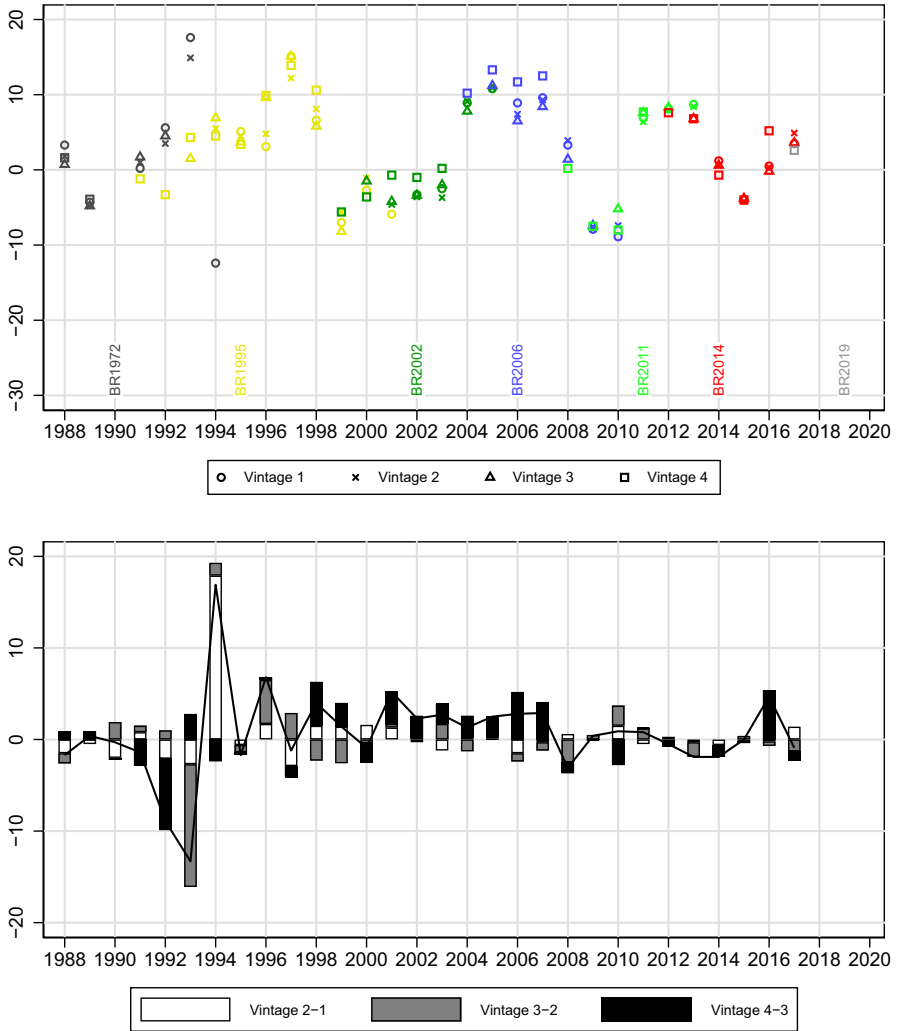


Fig. 6 Growth in gross fixed capital formation (JK). See Fig. 1

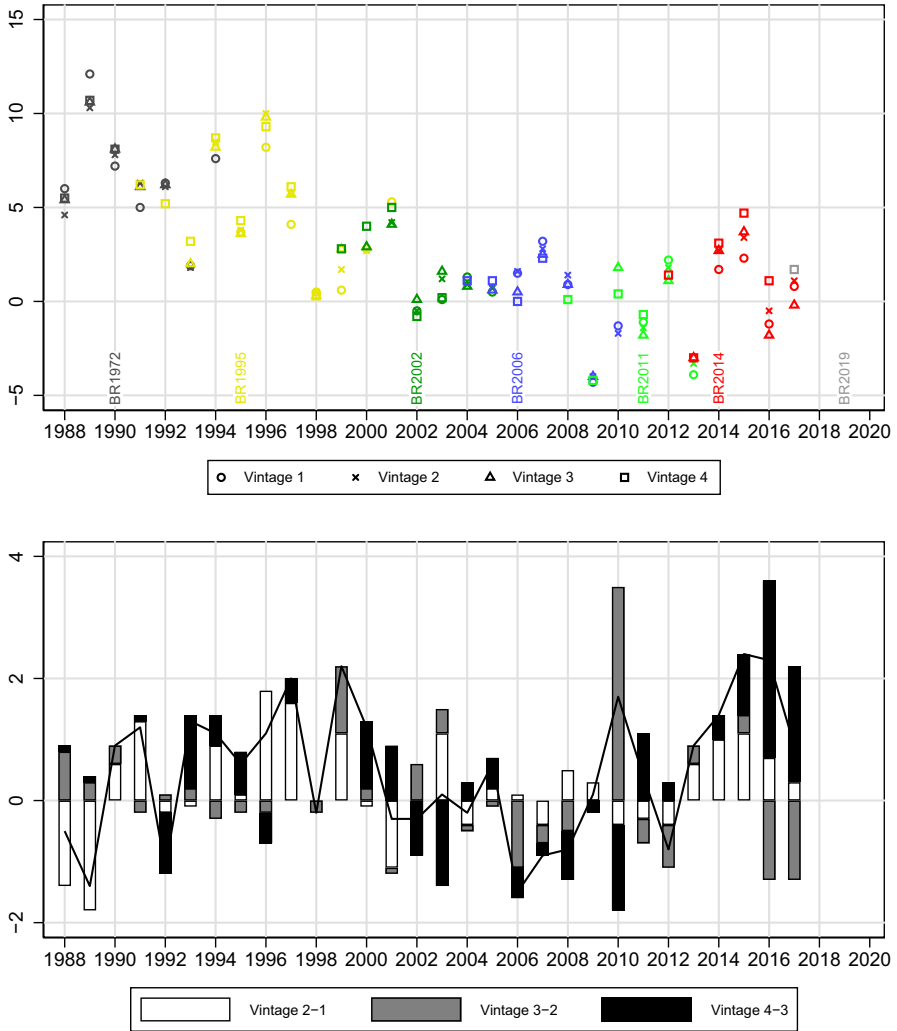


Fig. 7 Growth in total export (EX). See Fig. 1

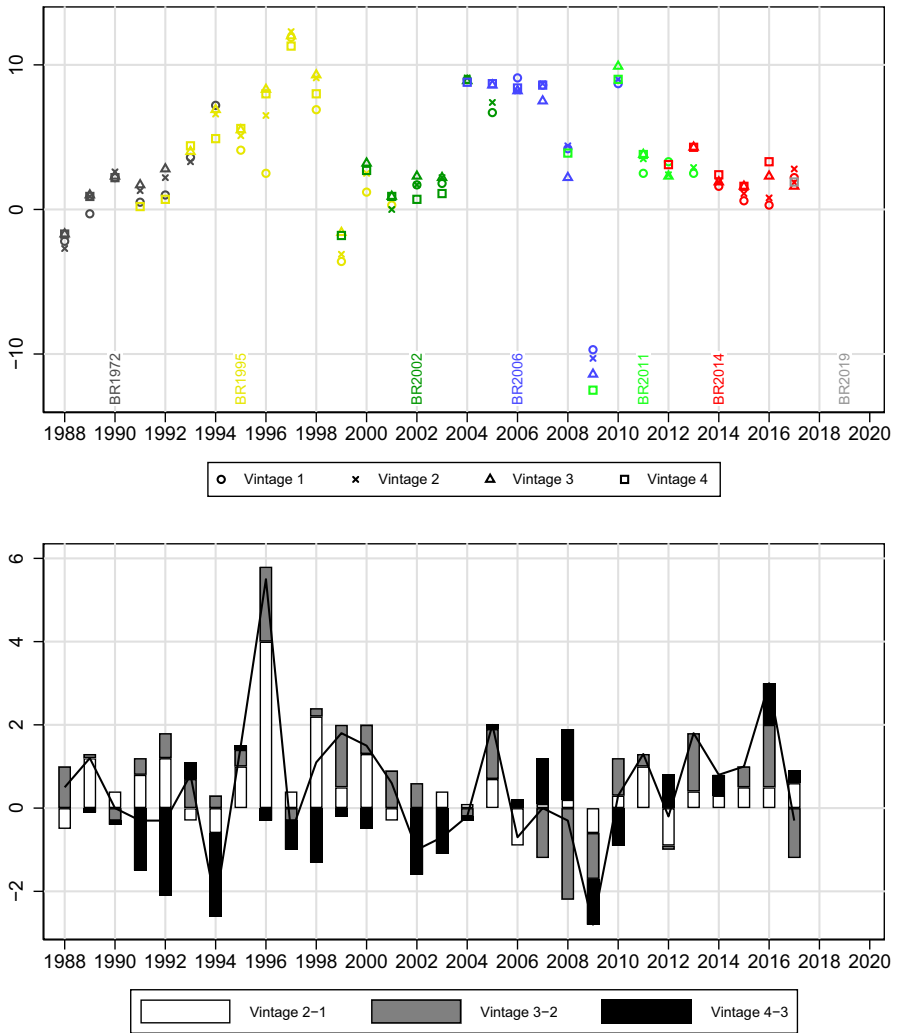


Fig. 8 Growth in total import (IMP). See Fig. 1

References

- Aruoba SB (2008) Data revisions are not well behaved. *J Money Credit Bank* 40(2–3):319–340
- Bernhardsen T, Eitrheim Ø, Jore AS, Røisland Ø (2005) Real-time data for Norway: challenges for monetary policy. *N Am J Econ Finance* 16(3):333–349
- Chong YY, Hendry DF (1986) Econometric evaluation of linear macro-economic models. *Rev Econ Stud* 53(4):671–690
- Claudio JC, Heinisch K, Holtemöller O (2020) Nowcasting East German GDP growth: a MIDAS approach. *Empir Econ* 58(1):29–54
- Clements MP, Galvão AB (2013) Forecasting with vector autoregressive models of data vintages: US output growth and inflation. *Int J Forecast* 29(4):698–714
- Cole R (1969) Data errors and forecasting accuracy. In: Mincer J (ed) *Economic forecasts and expectations: analysis of forecasting behavior and performance*, chapter 2. National Bureau of Economic Research, New York, pp 47–82
- Croushore D, Stark T (2003) A real-time data set for macroeconomists: does the data vintage matter? *Rev Econ Stat* 85(3):605–617
- Diebold FX, Mariano RS (1995) Comparing predictive accuracy. *J Bus Econ Stat* 13(3):253–263
- Ericsson NR (1993) On the limitations of comparing mean square forecast errors: clarifications and extensions. *J Forecast* 12(8):644–651
- Glocker C, Wegmueller P (2020) Business cycle dating and forecasting with real-time Swiss GDP data. *Empir Econ* 58(1):73–105
- Granger CWJ, Newbold P (1973) Some comments on the evaluation of economic forecasts. *Appl Econ* 5(1):35–47
- Harvey DI, Leybourne SJ, Newbold P (1998) Tests for forecast encompassing. *J Bus Econ Stat* 16(2):254–259
- Hungnes H (2018) Encompassing tests for evaluating multi-step system forecasts invariant to linear transformations. Discussion Papers 871, Statistics Norway
- Hungnes H (2020) Equal predictability test for multi-step-ahead system forecasts invariant to linear transformations. Discussion Papers 931, Statistics Norway
- Jore AS (2017) Revisions of national accounts. Staff Memo, Norges Bank, p 6
- Kunst RM, Wagner M (2020) Economic forecasting: editors' introduction. *Empir Econ* 58(1):1–5
- Mankiw N, Runkle DE, Shapiro MD (1984) Are preliminary announcements of the money stock rational forecasts? *J Monet Econ* 14(1):15–27
- Mincer J, Zarnowitz V (1969) The evaluation of economic forecasts. In: Mincer J (ed) *Economic forecasts and expectations: analysis of forecasting behavior and performance*, chapter 1. National Bureau of Economic Research, New York, pp 3–46
- Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3):703–708
- Siliverstovs B (2020) Assessing nowcast accuracy of US GDP growth in real time: the role of booms and busts. *Empir Econ* 58(1):7–27
- Strohsal T, Wolf E (2020) Data revisions to German national accounts: are initial releases good nowcasts? *Int J Forecast* 36(4):1252–1259

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.