Applied Economics Quarterly Vol. 55. N° 4 (2009)
Duncker & Humblot GmbH, 12165 Berlin

Do Forecasters Inform or Reassure? Evaluation of the German Real-Time Data

By Konstantin A. Kholodilin and Boriss Siliverstovs

Abstract

The paper evaluates the quality of the German national accounting data (GDP and its use-side components) as measured by the magnitude and dispersion of the forecast / revision errors. It is demonstrated that government consumption series are the least reliable, whereas real GDP and real private consumption data are the most reliable. In addition, early forecasts of GDP, private consumption, and investment growth rates are shown to be systematically upward biased. Finally, early forecasts of all the variables seem to be no more accurate than naïve forecasts based on the historical mean of the final data.

Keywords: Quality of statistical data; real-time data; signal-to-noise ratio; forecasts; revisions

JEL Classifications: C53; C89

> It is a mistake to try to look too far ahead. The chain of destiny can only be grasped one link at a time.
>
> *Winston Churchill*

1. Introduction

The significance of precise knowledge of the current state of the economy as well as of ability to accurately forecast macroeconomic variables in the future is difficult to overestimate. Clearly, no well-defined policy decision could be made without reliable information on the developments in the economy and its future prospects. Consequently, considerable literature has evolved dealing with assessment of macroeconomic forecast accuracy as well as investigating statistical properties of data revisions. In assessing of forecast accuracy, together with testing for forecast rationality and unbiasedness, a considerable attention has been paid to determining the information content of forecasts (e.g., see Parzen, 1982; Öller, 1985; de Gooijer and Klein, 1992; Diebold and Kilian, 2001; Oke and Öller, 1999; Galbraith, 2003; Isiklar and Lahiri, 2007; Öller and Teterukovsky, 2007). For example, Öller (1985), Oke and Öller (1999) and Galbraith (2003) by fitting the ARIMA-type processes attempt to determine the information content of optimal forecasts depending on the length of the forecast horizon. Diebold and Kilian (2001) propose a simple measure of relative predictability, where they compare the expected loss of a short-run forecast to the expected loss of a long-run forecast. de Gooijer and Klein (1992) estimate the optimal forecast horizon for cumulated multi-step predictions. Isiklar and Lahiri (2007) evaluate the flow of new information contained in monthly GDP forecasts provided by Consensus Economics. They report that beyond the 14 months the forecasts have little if any value.

A related strand of research assesses the quality of the current data available in real time by investigating statistical properties of data revisions, i.e., by comparing the first or preliminary data announcements usually made shortly after the end of the forecast period with the ultimate revision figures, often reported years after the initial estimate. If data revisions appear to be rather large and volatile then it is likely that the quality of the first-round estimates is poor (McNees, 1989; Mankiw and Shapiro, 1986; Faust et al., 2005; Swanson and van Dijk, 2006, among others).

In our paper, we evaluate the quality of both forecasts and statistical data revisions using the real-time data set for Germany covering growth rates of the German GDP and its demand-side components. We use both official statistical real-time data and the forecasts provided by the leading German economic institutes during the so-called Gemeinschaftsdiagnose (GD, or Joint Forecast). The choice of the GD forecasts can be justified based on the following considerations. First, GD plays a very important role in providing informational support to the economic policy decisions made by the German government, on behalf (and money) of which these forecasts are made. Second, it is among the few institutions (the Bundesbank, the German Council of Economic Experts (Sachverständigenrat), German federal government) that make macroeconomic forecasts for Germany at such a level of decomposition but it represents not the opinion of a single forecasting institute but rather a consensus attained by several institutes.

In this paper three approaches are used. First, the descriptive analysis of the forecast / revision errors is carried out in order to check for possible biases and degree of informational content of the variables under inspection. Second, the integrated signal-to-noise ratio recently suggested in Öller and Teterukovsky (2007) is applied to measure the overall quality of the data. Third, the entropy measure of Vasicek (1976) is used to examine the reduction in uncertainty at each subsequent vintage.

To the best of our knowledge, such a comprehensive evaluation of the quality of forecasts and statistical data revisions of the GDP and its components for Germany using the real-time data has not been done so far. The previous literature either focuses on forecasts or revisions of the German GDP reported together with forecasts / revisions made for other countries like Faust et al. (2005) and Isiklar and Lahiri (2007), respectively, or it focuses exclusively on data revisions for industrial production like in Jacobs and Sturm (2004) and Knetsch and Reimers (2006).

Our main findings are as follows. The earliest forecasts made for more than one year ahead are seriously flawed. The growth rate forecasts of such variables as GDP, private consumption, total investment, and investment in construction appear to be overly optimistic. That is, the earliest forecasts systematically exceed the final values published by the German statistical office. Moreover, the information content of the forecasts of all variables, except for government consumption, made at the earliest vintages, is virtually zero. Only forecasts made for the current year (so-called nowcasts) become informative about the final value, but still their informativeness remains limited. In this respect, it is worthwhile noting that the data on real GDP and real private consumption can be regarded as more reliable than the rest of the variables. The data on government consumption are the least reliable. Our findings cast serious doubts on the ability of forecasters to accurately reflect the future developments in the economy more than one year ahead.

The remainder of the paper is structured as follows. Section 2 describes the data set used in the analysis. Section 3 introduces the measurement techniques used here to evaluate the quality of statistical data. Section 4 discusses the results of data quality evaluation. Finally, section 5 concludes.

2. Data

Two data sets are used in this study. The first data set contains the forecasts of the annual growth rates made twice a year (Spring, normally in April, and Fall, typically in October) by the leading German macroeconomic forecast institutes during the so-called Gemeinschaftsdiagnose[1]) and covers the period Spring 1995 – Spring 2008. During each GD meeting (except Spring 1995 and Spring 1996) the forecasts for the current and next year are made. The second data set is comprised of the quarterly publications of the quarterly SNA statistics by the Statistisches Bundesamt Deutschlands (StaBu, or German Federal Statistical Office[2]) starting in 1997Q1 and ending in 2008Q2 and covers the real-time data over the period 1995Q2 – 2008Q1. Unfortunately, earlier data are not available, because prior to 1995 StaBu only published West German SNA statistics.

Both data sets include the following eight variables: GDP, private consumption, government consumption, total investment, investment in equipment, investment in construction, exports, and imports. Both nominal and real variables are considered. Based on these data sets a combined data set was constructed, which contains the

---

1 The GD data are taken from its regular forecast publications, which are available for the period Spring 1995 – Spring 2007 on the webpage of the DIW Berlin http://www.diw.de/deutsch/produkte/publikationen/wochenbericht/29864.html and for the period Fall 2007 – Spring 2008 on the webpage of the IWH http://www.iwh-halle.de/asp/publist.asp?Lang=d&Reihe=1.

2 Statistisches Bundesamt, Fachserie 18 Reihe 1.2, Volkswirtschaftliche Gesamtrechnungen, Inlandsproduktsberechnung, Vierteljahresergebnisse.

quarterly vintages of the forecast and revised data on the annual growth rates of the German GDP and its components. The GD Spring and Fall data were assigned to the second and fourth quarters, respectively.

To the best of our knowledge, this is a unique database, which has never been used before to conduct any economic analysis.

3. Revision measures

The purpose of this study is to evaluate the quality of German statistical data. The quality here is measured by the size and dynamics of forecast / revision errors with respect to some "true" value of the variables in question. It is assumed that, although this true value is never observed, it is well approximated by the final revision value. Hence the revision error is defined as:

$$e_t^l = y_t^L - y_t^l \ , \tag{1}$$

where y_t^l is the l-th revision carried out in period t; L is the period, when final revision is made. The final revision is supposed to happen in the period $t+1$, that is, one year after the period, to which it refers. This contradicts the official definition of the final data, according to which the data become final in the period $t+4$ (see Statistisches Bundesamt, 2007, p. 7). However, as our calculations have shown, after one year almost all the revision errors are equal to zero.

The sequence of GD forecasts and StaBu revisions is illustrated in Figure 1. The first four estimates of the variable of interest, starting from $t-3/2$ and ending in t, are the fore- or nowcasts made by the GD, whereas the last four estimates are the revisions produced by the StaBu.

The forecast / revision errors can be summarized using the measure of the mean squared error (MSE):

$$MSE_l = \frac{1}{T}\sum_{t=1}^{T}(e_t^l)^2 \ . \tag{2}$$

Thus, MSE_l measures the variance of the forecast / revision errors at vintage l.

Based on the MSE, an information measure, also known as signal-to-noise ratio (SNR), can be computed:

$$SNR_l = 1 - \frac{MSE_l}{\sigma_L^2} \ , \tag{3}$$

where σ_L^2 is the variance of the final revision, y_t^L. SNR_l measures information content of each forecast / revision. It can be interpreted as a goodness of fit of each revision. It is expected that as more information comes in, the variance of forecast / revision errors goes to zero and hence SNR_l approaches the value of one.

In order to evaluate the overall quality of data, Öller and Teterukovsky (2007) have introduced the integrated SNR measure, which is defined as:

$$ISNR_L = \frac{1}{2}\sum_{l=0}^{L-1}(SNR_l + SNR_{l+1})\tau(l,\, l+1)\ , \tag{4}$$

where $\tau(l,\, l+1)$ is the interval between the vintages l and $l+1$. Without loss of generality, the time interval between the first forecast and last revision can be re-scaled to the [0,1] interval. In that case, the integrated SNR measure will vary between 0 (complete ignorance about the final value up to the last revision) and 1 (the very first forecast conveys all necessary information about the final value).

Another approach to examining the quality of statistical data is suggested in Patterson and Heravi (1991) who apply the entropy estimator of Vasicek (1976) in order to investigate the reduction in uncertainty as new vintages of data are published. The entropy of a distribution F with a density function f is defined as:

$$H(f) = -\int_{-\infty}^{+\infty} f(x)\log(f(x))dx\ . \tag{5}$$

Vasicek (1976) suggests the following consistent estimator of the entropy:

$$H_{MT} = \frac{1}{T}\sum_{t=1}^{T}\log\left(\frac{T}{2M}e^l_{(t+M)} - e^l_{(t-M)}\right) \tag{6}$$

where $e^l_{(t)}$ is the ordered forecast / revision error such that $e^l_{(1)} \leq e^l_{(2)}, \ldots, \leq e^l_{(T)}$; M is a positive integer smaller than $T/2$. This measure of entropy is robust to the deviations from normality of the forecast / revision errors as well as to non-constant means of the forecast / revision errors observed at different vintages. The latter property is particularly important in our case, where, as shown below, a (declining) bias in forecasts is observed.

4. Results

The presentation of our results unveils as follows. First, we describe the pattern of forecast / revision errors based on the results reported in Table 1 for each variable separately. In doing so, we especially investigate two issues: the unbiasedness and the informational content of forecasts / revisions. The former issue is related to whether forecasts / revisions systematically over- or underpredict final values. The latter issue is investigated based on comparison of the dispersion of forecast / revision errors to that of the final data. The earlier forecasts / revisions are said to be informative if the dispersion of the corresponding forecast / revision errors is (substantially) lower than the dispersion of the final vintage data. We conclude our descriptive analysis summarizing the conclusion reached for each variable. Our descriptive analysis is followed by the assessment of forecast / revision quality using the integrated signal-to-noise measure of Öller and Teterukovsky (2007). Then we present the results based on the entropy measure as suggested in Patterson and Heravi (1991).

4.1 Descriptive analysis

4.1.1 Gross Domestic Product

The descriptive statistics of the forecast / revision errors for the German Gross Domestic Product both in nominal and real terms is displayed in the top panel of Table 1. The reported values of mean and median measure the extent of the bias of the respective forecasts / revisions. The bias, while substantial for the earlier forecasts made at vintages $l = t - 3/2$ and $l = t - 1$, diminishes with every additional vintage. Nevertheless, its magnitude for the two earlier vintages is very remarkable indicating that the growth forecasts made up to one year ahead tend on average to overpredict the realized growth rate by about 1.14 and 0.84 percentage points for nominal GDP and by about 0.83 and 0.57 percentage points for real GDP. Notice that the average annual growth rates of nominal and real GDP were 2.1 and 1.4 percent under the period of investigation, respectively.

The median value recorded for the forecast errors of nominal GDP for the two earliest vintages is even higher: –1.40 and –0.95 percentage points. The rather high values of bias in these forecast errors are also supported by the results of the *t*-test indicating rejection of the null hypothesis of zero mean forecast error at 1% and 10% significance level for nominal and real variables, respectively.

It is also interesting to observe that the bias magnitude drops sharply at vintage $l = t - 1/2$, i.e., when the first forecast for the current year is made. The mean forecast error decreases from –0.84 to –0.18 for nominal variables and from –0.57 to –0.12 for real variables between vintages $l = t - 1$ and $l = t - 1/2$, reflecting a substantial increase in amount of available information to forecasters on the current developments in the economy and on its likely future discourse. This fact is also supported by the *t*-test indicating that we cannot reject the null hypothesis of zero mean forecast / revision error at the usual significance levels for all later vintages starting with $l = t - 1/2$.

The results above suggest that when making the earliest predictions of the GDP growth rates – either in nominal or real terms – the forecasters tend to be overly optimistic. This optimistic nature of forecasts is also reflected in the fact that up to vintages $l = t$ and $l = t - 1/2$ (including) for nominal and real variables, respectively, there is a noticeable asymmetry in the recorded minimum and maximum forecast errors. For example, at the earliest vintage the former is by almost 2.0 and 1.2 percentage points larger in the absolute value than the latter for nominal and real variables, respectively.

Next, consider the informative content of forecasts / revisions which we investigate by comparing the ratio of the mean squared forecast error calculated at every vintage l to the variance of the data at the final vintage, i.e., MSE_l/σ_L^2. In fact, this is the noise-to-signal ratio. This ratio exceeding one indicates complete ignorance about the final values of the predicted variables. In this case, the signal is overwhelmed by the noise. The less noisy the forecasts / revisions (the less ignorant the forecasters) the closer this ratio is to zero. Hence when this ratio is substantially below one, the forecasts / revisions are informative.

The results of such comparison indicate that the earlier forecasts, i.e., those made more than one year ahead, are virtually non-informative. The corresponding ratios calculated for vintages $l = t - 3/2$ and $l = t - 1$ for nominal GDP growth rate forecasts is 4.53 and 3.22, respectively. For real GDP growth rate forecasts the corresponding ratios are 2.23 and 1.28. Only starting with the forecasts made for the current year, the respective ratio falls below one and continuously decreases with every vintage.

In sum, based on the presented results we can tentatively conclude that the earliest GDP forecasts made more than one year ahead are not only too optimistic, but, in addition, they are not informative. The associated uncertainty is so large that a "naïve, forecast based on the historical mean of the GDP growth rates is likely to be more precise than the GD forecasts.

4.1.2 Private consumption

The descriptive statistics of the forecast / revision errors for the German private consumption both in nominal and real terms is displayed in the second panel of Table 1. Private consumption is by far the largest component of GDP accounting in Germany for about 60 % of the GDP. Therefore, it is not surprising that our results obtained for GDP largely apply also for the private consumption. In particular, the private consumption forecasts, especially at the early vintages are upward biased as the mean and median values of the forecast errors reported in Table 1 show. The null of the mean of forecast errors being equal to zero is rejected at 10 % significance level up to and including the vintage $l = t - 1/2$ and $l = t - 1$ for nominal and real values, respectively.

For the private consumption the asymmetry of the forecast errors is even more pronounced when compared to that observed for GDP. For the nominal values at the very first vintage, $l = t - 3/2$, the minimum and maximum are –3.19 and 0.14, whereas for the real values they are –3.12 and 0.48, correspondingly. For the following vintage the corresponding numbers are –2.49 and 0.84, and –2.52 and 1.16. Thus, at the earliest forecast vintage the maximum forecast error is about 23 and 7 times bigger than the minimum forecast error in absolute value for nominal and real data, respectively.

According to the noise-to-signal ratio, the private consumption forecasts made at the two earliest vintages appear to be uninformative. Only starting from the forecasts of both nominal and real private consumption growth rates made in $l = t - 1/2$, this ratio falls below one, which is similar to what we observed for real GDP.

4.1.3 Government consumption

For the government consumption, as shown in Table 1, we also find an indication of the upward bias in the respective forecasts / revisions. However, the null hypothesis of zero mean forecast error cannot be rejected for earlier vintages, as in case of GDP and private consumption, but for later vintages $l = t + 1/4$ and $l = t + 1/2$. Both mean and median values of the forecast errors are negative for all but the penultimate ($l = t + 1/2$) vintages for nominal data. It is also interesting to observe that bias magnitude only slightly decreases with vintages for nominal data and it fluctuates about the same level of –0.30 for real data. This in sharp contrast to the pattern observed for GDP and private consumption, where bias – rather large for initial vintages – diminishes relatively fast with every additional vintage.

Also here we notice a significant evidence for asymmetry in forecast errors not only for the initial vintages as in case of GDP and private consumption but also for later vintages. For all vintages the negative forecast errors exceed the positive ones, when expressed in absolute value, of course.

It also worth mentioning that already earlier forecasts of government consumption appear to have some informational content regarding the final values as the associated MSE_l/σ_L ratio is below one starting with the first vintage. Nevertheless, this ratio drops at a rather slow rate both for nominal and real values, implying a rather low marginal increase in forecast / revision accuracy. For real values it takes a value of 0.36 even at the penultimate revision $l = t + 3/4$.

4.1.4 Investment

Under this heading we consider the following three variables: total investment and its components (investment in equipment and investment in construction). The common feature of these variables is that they are much more volatile in comparison with GDP as well as private and government consumption, see corresponding values of σ_L in Table 1. This has to be kept in mind when interpreting the results.

The mean forecast error for investment variables is relatively large to what we observe for GDP and (both private and government) consumption at earlier vintages. However, the bias magnitude corrected for the range (mean-to-range ratio) is much smaller for the former than for the latter variables. In addition, although at some vintages the null of mean forecast / revision error equal to zero is rejected at

5 % significance level, no systematic bias pattern can be observed as in the case of the GDP and private consumption.

The noise-to-signal ratio exceeds one at $l = t - 3/2$ and $l = t - 1$ vintages and falls afterwards. Thus, neither of the investment forecasts is informative at the first two vintages.

4.1.5 Exports and imports

Both exports and imports exhibit similar characteristics over different vintages of forecasts / revisions. First, there appear to be no positive bias in these variables (see Table 1). The mean and median are negative for nominal values only at $l = t - 3/2$ and afterwards they stay always positive. Contrary to what we observe for other variables, the hypothesis of the mean forecast / revision error being equal to zero cannot be rejected at the earlier vintages. However, for the exports and imports it can be rejected at the later vintages, from $l = t$ to $l = t + 1/2$. Thus, although the forecasts of exports and imports up to $l = t - 1/2$ appear to be unbiased, their forecasts / revisions at later vintages seem to be systematically underestimated.

Second, the forecasts of both nominal and real exports as well as of nominal imports become informative in $l = t - 1/2$, while the ratio MSE_l/σ_L^2 for the forecasts of real imports falls below one only starting from $l = t$.

4.1.6 Summary of descriptive analysis

Based on the descriptive analysis we can draw two major conclusions. First, there is an evidence that the early forecasts of the growth rates of GDP, private consumption, total investment and investment in construction are excessively optimistic. This observation is also supported by Figures 4 and 5 showing the boxplots of the forecast / revision errors. In overwhelmingly large number of cases up to vintage $l = t$, the interquartile range of these variables is either completely or for the most part below zero. Our finding of upward bias in earlier forecasts generally conforms with the results reported in Batchelor (2007), where bias in real GDP forecasts published by the *Consensus Economics* forecasting service are investigated for G-7 countries. Similarly, Ashiya (2007) finds out that 16-months ahead GDP forecasts of the Japanese government appear also to be too optimistic. More generally, our finding is also supported by Loungani (2001) where upward biases in one-year ahead real GDP forecasts are also documented for 63 countries, both industrialized and developing. As summarized, in Stekler (2008, p. 4) persistent biases in macroeconomic forecasts may be possibly explained by the following reasons: (1) irrationality of forecasters, (2) forecasters face an asymmetric loss functions such that less penalty is expected from making too optimistic rather than too pessimistic forecasts (see discussion in Elliott, Komunjer, and Timmerman, 2005), or (3) strategic behavior of forecasters.

Another interesting observation concerns a seeming lack of informational content in the earliest forecasts made in $l = t - 3/2$ and $l = t - 1$, as the ratio of mean squared error to the variance of observed values of the variables in question in all but few cases exceeds one. This implies that the accuracy of a simple forecast rule based on the observed historical mean often substantially exceeds that made by the GD for a forecast horizon over one year. Our finding conforms with that of Isiklar and Lahiri (2007), who studied the informational content of monthly made GDP forecasts from *Consensus Economics, Inc.* for 18 developed countries. They conclude that the forecasts beyond 18 months are of a little value. In a subsequent paper, where the GDP forecasts from Consensus Economics, Inc. are studied for G-7 countries over the period of 1990-2007, Lahiri and Sheng (2008) also find that "… forecasts for real GDP contain little information beyond 6 quarters". Our findings show that the conclusions reached in these two papers for more-than-one-year ahead forecasts of GDP also extend to similar forecasts of its components. At the same time, we should point out that forecasts of government consumption constitute a noticeable exception. As seen, for that variable the ratio MSE_l/σ_L^2 is always below one and it takes value around 0.8 for three initial vintages and it further decreases in subsequent vintages.

The poor quality of the initial forecast can be best illustrated by comparing with relatively high quality of the first official revision (flash estimate) as shown in Figures 2 and 3. Such a discrepancy between the forecasts and revisions can be attributed, first, to the fact that by the time of the first revision already a lot of relevant information is already accumulated, and, second, that the official statisticians can have access to larger information set.

4.2 Signal-to-noise ratio

Following Öller and Teterukovsky (2007) we illustrate evolution of the information content of the German nominal and real SNA data over vintages using Figures 6 and 7. The horizontal axis shows the vintages from $l = t - 3/2$ till $l = t + 1$, whereas the vertical axis displays the SNR at each vintage. The values of SNR are computed according to equation (3) using the information presented in Table 1. Notice that in constructing of these graphs negative values of SNR are set equal to zero.

Black area in Figures 6 and 7 represents the lack of knowledge about the final value of the variable. When until the very last revision no information is available, the area of the graph should be completely black. In the case of perfect knowledge, the whole information about the true value of variable is known already in the very beginning and hence the area of the graph is completely white. In fact, integrated SNR defined in equation (4) measures the white area of the graph.

The integrated SNR characterizes the overall quality of a variable. The higher the integrated SNR the higher is the quality of the forecasting / revision process. As shown in Table 2 the values of $ISNR_L$ both for nominal and real variables are around 0.5 with the exception of real government consumption which takes a substantially lower value of 0.345. This indicates that for all but one variable the quality of forecast / revision process is roughly equal.

Further insights into the quality of the forecasting / revision process can be gained by addressing the following two questions related to its timeliness: 1) How early does the information exceed a certain level (e.g., $SNR_l = 0.5$)? and 2) When a certain share of all the information about the variable is accumulated (e.g., $ISNR_l = 0.5$)? The earlier the information exceeds a certain level or a certain share of information is accumulated the more timely forecasts / revisions reflect the true value.

For nominal values, the earliest vintage at which the SNR_l exceeds 0.5 is $l = t$, except for government consumption for which it occurs one revision later at $l = t + 1/4$. Moreover, the half of all possible information on the growth rates in the reference year is accumulated by the vintage $l = t + 1/2$ for all variables uniformly, i.e., in the second quarter of the succeeding year.

For real variables, our findings are more heterogeneous. For GDP and private consumption forecasts the earliest vintage at which the SNR_l exceeds 0.5 is $l = t - 1/2$, for government consumption – at $l = t + 1/2$, while for the remaining variables – at $l = t$. The half of all possible information on the growth rates in the reference year is accumulated by the vintage $l = t + 1/2$ for all variables except GDP and private consumption for which it accrues earlier at $l = t + 1/4$. Thus, the forecasts of real GDP and real private consumption appear to be most timely, whereas those for government consumption ranks the least in terms of timeliness.

4.3 Measure of entropy

The entropy estimated using equation (6) is reported in Table 3. In addition, based on this we computed the information gain by taking the difference in entropy for two successive vintages. The purpose of this exercise is to assess the process of uncertainty reduction in forecasts / revisions with every additional vintage. Given the fact that the data typically undergo continuous revision process, this could be helpful in determining the vintage after which no further substantial reduction in uncertainty could be realized. If, for example, all the information gain occurs at the early vintages all the subsequent vintages can be ignored.

Based on the results reported in Table 3, we conclude the following. First, the largest information gain typically occurs at vintages starting from $l = t - 1/2$ both for nominal and real variables. Taken together with the fact that at the earlier vintage $l = t - 1$ the information gain is rather small, this supports our conclusions discussed in section 4.1 concerning virtually zero informational content of the forecasts made for more than one year ahead.

Second, the pattern of reduction in uncertainty is not uniform across the variables. For GDP, private consumption, exports, and imports the maximum gain is attained for forecasts in current year and afterwards the uncertainty reduction declines gradually. This implies that for these variables the later vintages cannot be ignored. Hence, for this group of variables, the use of data released at the vintages at which the maximum gain is attained is likely to be associated with high degree of uncertainty and therefore should be exercised with great care.

For the rest of variables the maximum information gain is realized when the first official statistical publication is made (vintage $l = t + 1/4$) except for real government consumption – -at vintage $l = t + 1/2$, and the gains at subsequent statistical revisions are negligible. Hence for this group of variables, all the subsequent revisions are not that important.

5. Conclusions

In this paper, the quality of German SNA real-time data was evaluated. In particular, the forecasting / revision process of the annual growth rates of GDP and its use-side (or expenditure-side) components was investigated.

To this end, three approaches were used. First, the descriptive analysis of the forecast / revision errors was carried out in order to check for possible biases and degree of informational content of the variables under inspection. Second, the integrated signal-to-noise ratio of Öller and Teterukovsky (2007) was used to measure the overall quality of the data. Third, the entropy measure of Vasicek (1976) was used to examine the reduction in uncertainty.

We can draw the following conclusions. The earliest forecasts made for more than one year ahead are seriously flawed. The growth rate forecasts of such variables as GDP, private consumption, total investment, and investment in construction appear to be overly optimistic. That is, the earliest forecasts systematically exceed the final values published by the German statistical office. Moreover, the information content of the forecasts of all variables, except for government consumption, made at the earliest vintages $l = t - 3/2$ and $l = t - 1$, is virtually zero. Our findings cast serious doubts on the ability of forecasters to accurately reflect the future developments in the economy more than one year ahead and raise the question on whether the benefits of such forecasts surpass the corresponding costs.

Only the forecasts made for the current year (so-called nowcasts) become informative about the final value, but still their informativeness remains limited. This is reflected in the fact that the signal-to-noise ratio stays below 0.5 for all vintages up to $l = t$ for most cases and the half of the available informational content as measured by the integrated signal-to-noise ratio typically is attained at the vintage $l = t + 1/2$, i.e., in the second quarter of the succeeding year. In this respect, it is worth noting that the data on real GDP and real private consumption can be regarded as more reliable than the rest of the variables. The data on government consumption are the least reliable reflecting its rather low predictability due to a large discretionary component of fiscal policy.

These findings are also confirmed by the results of the entropy analysis, which suggests that the biggest information gains accrue at vintages $l = t - 1/2$ and $l = t$ for such variables as GDP, private consumption, exports, and imports, and at vintages $l = t + 1/4$ and $l = t + 1/2$ – for the remaining variables. It has to be mentioned, however, that for the first group of the variable substantial information gains further accrue even after the vintage at which one observes the biggest information gain. This implies that for GDP, private consumption, exports, and imports variables subsequent revisions are important and, again, that data at the earlier vintages has to be treated with care.

All in all, our findings suggest that the quality of the German SNA use-side data is rather modest. Only nowcasts appear to have some but still limited informational content. This result also conforms with the results reported in Isiklar and Lahiri (2007) and Lahiri and Sheng (2008) for 18 developed and G-7 countries, respectively. At the same time, we would like to emphasize that the early forecasts are systematically upward biased. However, this finding appear to be a rule rather than an exception, e.g., see Ashiya (2007), Batchelor (2007), and Loungani (2001) for similar evidence documented for other countries. This means that the forecasts, especially the earlier ones, have to be taken with a great reservation as they seem to perform rather reassuring function. They may encourage economic agents to look more optimistically in the nearest future. Unfortunately, these overly rosy forecasts usually are not self fulfilling.

References

Ashiya, M. (2007): Forecast accuracy of the Japanese government: Its year-ahead GDP forecast is too optimistic. Japan and the World Economy 19, 6885.

Batchelor, R. (2007): Bias in macroeconomic forecasts. International Journal of Forecasting 23 (2), 189 – 203.

de Gooijer, J. G. / *Klein,* A. (1992): On the cumulated multi-step-ahead predictions of vector autoregressive moving average processes. International Journal of Forecasting 7 (4), 501 – 513.

Diebold, F. X. / *Kilian,* L. (2001): Measuring predictability: theory and macroeconomic applications. Journal of Applied Econometrics 16 (6), 657 – 669.

Elliott, G. / *Komunjer,* I. / *Timmerman,* A. (2005): Estimation and testing of forecast rationality under flexible loss. Review of Economic Studies 72 (4), 1107 – 1125.

Faust, J. / *Rogers,* J. H. / *Wright,* J. H. (2005): News and noise in G-7 GDP announcements. Journal of Money, Credit and Banking 37 (3), 403 – 419.

Galbraith, J. (2003): Content horizons for univariate time series forecasts. International Journal of Forecasting 19 (1), 43 – 55.

Isiklar, G. / *Lahiri,* K. (2007): How far ahead can we forecast? Evidence from cross-country surveys. International Journal of Forecasting 23 (2), 167 – 187.

Jacobs, J. / *Sturm,* J.-E. (2004): Do Ifo indicators help explain revisions in German industrial production? Technical Report CESifo Working Paper No. 1205, CESifo GmbH.

Knetsch, T. A. / *Reimers,* H.-E. (2006): How to treat benchmark revisions?: The case of German production and orders statistics. Discussion Paper Series 1: Economic Studies 2006, 38, Deutsche Bundesbank, Research Centre.

Lahiri, K. / *Sheng,* X. (2008): Learning and heterogeneity in GDP and inflation forecasts.

Loungani, P. (2001): How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth. International Journal of Forecasting 17 (3), 419 – 432.

Mankiw, N. G. / *Shapiro,* M. D. (1986): News or noise? Analysis of GNP revisions. Survey of Current Business 66, 20 – 25.

McNees, S. K. (1989): Forecasts and actuals: The trade-off between timeliness and accuracy. International Journal of Forecasting 5 (3), 409 – 416.

Oke, T. / *Öller,* L.-E. (1999): Testing for short memory in a VARMA process. International Journal of Forecasting 18, 477 – 487.

Öller, L.-E. (1985): How far can changes in business activity be forecast? International Journal of Forecasting 1, 135 – 141.

Öller, L.-E. / *Teterukovsky,* A. (2007): Quantifying the quality of macroeconomic variables. International Journal of Forecasting 23 (2), 205 – 217.

Parzen, E. (1982): ARARMA models for time series analysis and forecasting. Journal of Forecasting 1 (1), 67 – 82.

Patterson, K. D. / *Heravi,* S. M. (1991): Direct estimation of entropy and revisions to the national income accounts. The Statistician 40 (1), 35 – 50.

Statistisches Bundesamt (2007): Volkswirtschaftliche Gesamtrechnungen: Qualitätsbericht.

Stekler, H. O. (2008): What do we know about G-7 macro forecasts? Research Program on Forecasting (RPF) Working Paper 2008 – 009, Department of Economics, The George Washington University.

Swanson, N. R. / *van Dijk,* D. (2006): Are statistical reporting agencies getting it right? Data rationality and business cycle asymmetry. Journal of Business & Economic Statistics 24, 24 – 42.

Vasicek, O. (1976): A test for normality based on sample entropy. Journal of the Royal Statistical Society. Series B (Methodological) 38, 54 – 59.

Appendix

Table 1: **Descriptive statistics of forecast / revision errors**

Vintage l	Nominal							Real variables						
	$t-3/2$	$t-1$	$t-1/2$	t	$t+1/4$	$t+1/2$	$t+3/4$	$t-3/2$	$t-1$	$t-1/2$	t	$t+1/4$	$t+1/2$	$t+3/4$
	GDP, $\sigma_L = 0.69$							GDP, $\sigma_L = 0.99$						
Minimum	–2.71	–2.13	–1.17	–0.63	–0.49	–0.16	0.00	–2.52	–2.13	–1.53	–0.55	–0.60	–0.10	0.00
Maximum	0.76	1.46	0.96	0.46	0.50	0.46	0.00	1.36	1.66	1.06	0.56	0.18	0.10	0.00
Range	3.47	3.59	2.13	1.09	1.00	0.61	0.00	3.89	3.79	2.59	1.11	0.78	0.20	0.00
Median	–1.40	–0.95	–0.15	–0.17	0.05	0.05	0.00	–0.60	–0.65	–0.05	–0.14	0.02	0.01	0.00
Mean	–1.14	–0.84	–0.18	–0.15	0.05	0.10	0.00	–0.83	–0.57	–0.12	–0.06	–0.02	0.01	0.00
H_0 : Mean = 0	0.01	0.02	0.26	0.18	0.47	0.11	na	0.09	0.09	0.59	0.54	0.79	0.52	na
MSE_l/σ_L^2	4.53	3.22	0.57	0.26	0.11	0.09	0.00	2.23	1.28	0.47	0.08	0.04	0.00	0.00
	Private consumption, $\sigma_L = 0.98$							Private consumption, $\sigma_L = 1.00$						
Minimum	–3.19	–2.49	–1.79	–0.65	–0.48	–0.48	0.00	–3.12	–2.52	–1.62	–0.70	–0.38	–0.38	0.00
Maximum	0.14	0.84	0.54	0.93	0.70	0.54	0.01	0.48	1.16	0.59	0.89	0.95	0.65	0.00
Range	3.33	3.33	2.33	1.57	1.18	1.02	0.01	3.60	3.68	2.21	1.59	1.33	1.02	0.00
Median	–0.80	–0.77	–0.61	–0.15	0.19	0.15	0.00	–1.12	–0.95	–0.54	–0.15	0.20	0.20	0.00
Mean	–1.20	–0.88	–0.45	–0.07	0.15	0.13	0.00	–0.93	–0.70	–0.36	0.01	0.21	0.17	0.00
H_0 : Mean = 0	0.01	0.01	0.05	0.67	0.13	0.19	0.34	0.04	0.05	0.10	0.95	0.07	0.10	0.22
MSE_l/σ_L^2	2.43	1.51	0.66	0.26	0.12	0.11	0.00	2.04	1.45	0.52	0.30	0.15	0.11	0.00
	Government consumption, $\sigma_L = 0.98$							Government consumption, $\sigma_L = 0.96$						
Minimum	–1.77	–1.74	–1.74	-1.15	–0.84	–0.84	0.00	–1.78	–1.68	–1.68	–1.48	–1.97	–0.90	–0.90
Maximum	0.52	0.60	0.90	0.90	0.27	0.32	0.00	0.78	0.93	1.03	1.42	0.60	0.53	0.16
Range	2.29	2.35	2.65	2.06	1.10	1.16	0.00	2.55	2.61	2.71	2.90	2.57	1.43	1.06
Median	–0.29	–0.28	–0.45	–0.37	–0.22	–0.20	0.00	0.20	–0.17	0.00	–0.43	–0.34	–0.34	0.00
Mean	–0.47	–0.40	–0.38	–0.32	–0.31	–0.22	0.00	–0.30	–0.33	–0.19	–0.28	–0.48	–0.30	–0.10
H_0 : Mean = 0	0.11	0.13	0.15	0.14	0.01	0.07	na	0.35	0.28	0.50	0.279	0.039	0.078	0.294
MSE_l/σ_L^2	0.78	0.76	0.78	0.52	0.21	0.18	0.00	0.88	0.88	0.83	0.73	0.70	0.36	0.09

Continued next page

Table 1 continued

Vintage l	Nominal							Real variables						
	$t-3/2$	$t-1$	$t-1/2$	t	$t+1/4$	$t+1/2$	$t+3/4$	$t-3/2$	$t-1$	$t-1/2$	t	$t+1/4$	$t+1/2$	$t+3/4$
	Total investment, $\sigma_L = 3.65$							Total investment, $\sigma_L = 3.49$						
Minimum	–10.02	–8.82	–6.62	–2.92	–0.63	–0.63	0.00	–9.57	–8.46	–6.66	–2.56	–0.42	–0.62	0.00
Maximum	4.63	5.43	3.33	2.53	0.76	0.73	0.00	4.60	4.70	3.00	2.20	0.97	0.98	0.86
Range	14.65	14.25	9.95	5.45	1.39	1.35	0.00	14.16	13.16	9.66	4.76	1.40	1.60	0.86
Median	–2.04	–1.85	–1.07	–0.17	0.38	0.12	0.00	–1.52	–1.87	–0.87	-0.45	0.07	0.07	0.00
Mean	–3.18	–2.78	–1.50	–0.18	0.21	0.12	0.00	-2.71	–2.28	–1.22	–0.19	0.27	0.22	0.08
H_0 : Mean = 0	0.07	0.04	0.09	0.73	0.19	0.38	na	0.11	0.06	0.21	0.68	0.11	0.24	0.34
MSE_l/σ_L^2	2.19	1.58	0.64	0.19	0.02	0.01	0.00	2.10	1.38	0.69	0.17	0.03	0.03	0.01
	Investment in equipment, $\sigma_L = 5.91$							Investment in equipment, $\sigma_L = 5.78$						
Minimum	–16.17	–12.90	–11.40	–6.10	–1.06	–1.06	0.00	–15.64	–12.97	–11.67	–6.07	–0.88	–0.93	0.00
Maximum	4.00	2.84	3.00	2.65	1.93	1.93	0.00	4.60	3.82	2.82	2.83	2.11	2.11	0.00
Range	20.16	15.73	14.39	8.75	2.99	2.99	0.00	20.23	16.79	14.49	8.91	2.99	3.04	0.00
Median	–2.83	–1.53	0.47	0.20	0.16	0.01	0.00	–0.82	–0.10	1.45	–0.60	0.24	0.00	0.00
Mean	–4.31	–3.11	–1.44	–0.66	0.39	0.27	0.00	–3.08	–2.24	–0.98	–0.64	0.52	0.22	0.00
H_0 : Mean = 0	0.11	0.09	0.33	0.46	0.16	0.30	na	0.25	0.22	0.51	0.46	0.14	0.47	na
MSE_l/σ_L^2	1.88	1.07	0.62	0.23	0.02	0.02	0.00	1.75	1.02	0.64	0.22	0.04	0.03	0.00
	Investment in construction, $\sigma_L = 3.30$							Investment in construction, $\sigma_L = 2.72$						
Minimum	–7.19	–6.39	–3.59	–2.66	–0.69	–0.69	0.00	–6.96	–5.86	–3.74	–2.44	–0.49	–0.49	0.00
Maximum	7.47	7.97	4.87	3.57	0.96	0.50	0.00	5.66	5.46	3.06	2.66	0.45	0.45	0.00
Range	14.67	14.37	8.47	6.24	1.65	1.19	0.00	12.62	11.32	6.80	5.10	0.95	0.95	0.00
Median	–3.40	–3.34	–2.17	–0.10	0.05	–0.06	0.00	–3.36	–2.77	–1.65	–0.03	0.05	–0.06	0.00
Mean	–2.27	–2.45	–1.62	0.03	0.07	–0.01	0.00	–2.37	–2.16	–1.66	–0.07	0.03	–0.03	0.00
H_0 : Mean = 0	0.14	0.09	0.05	0.95	0.63	0.94	na	0.13	0.08	0.02	0.88	0.76	0.68	na
MSE_l/σ_L^2	1.91	1.80	0.70	0.22	0.02	0.01	0.00	2.45	1.80	0.90	0.25	0.01	0.01	0.00

	Exports, $\sigma_L = 4.50$							Exports, $\sigma_L = 3.62$						
Minimum	–9.19	–5.84	–6.64	–1.32	–0.42	–0.42	0.00	–6.55	–4.28	–5.08	–0.89	–0.41	–0.51	0.00
Maximum	9.98	7.68	4.50	1.71	1.18	1.09	0.00	8.53	6.33	4.30	2.50	2.71	0.82	0.00
Range	19.17	13.53	11.14	3.02	1.60	1.51	0.00	15.08	10.61	9.37	3.39	3.11	1.33	0.00
Median	–1.52	0.37	2.38	0.58	0.56	0.41	0.00	0.49	0.97	2.71	0.73	0.60	0.34	0.00
Mean	–0.33	0.42	0.88	0.45	0.56	0.35	0.00	0.75	1.32	1.35	0.87	0.59	0.26	0.00
H_0 : Mean = 0	0.87	0.77	0.46	0.18	0.00	0.03	0.34	0.65	0.26	0.19	0.02	0.03	0.08	na
MSE_l/σ_L^2	1.50	0.98	0.69	0.06	0.02	0.01	0.00	1.55	1.05	0.83	0.14	0.07	0.02	0.00
	Imports, $\sigma_L = 5.91$							Imports, $\sigma_L = 3.84$						
Minimum	–11.17	–9.47	–10.47	–0.47	–0.32	–0.59	0.00	–7.96	–7.54	–9.04	–1.34	–1.70	–1.70	0.00
Maximum	11.83	10.93	5.58	2.68	1.42	1.29	0.74	5.50	6.10	6.41	3.51	1.91	1.21	0.00
Range	23.00	20.40	16.05	3.15	1.74	1.88	0.74	13.46	13.64	15.46	4.86	3.61	2.91	0.00
Median	–0.64	1.09	1.85	0.53	0.85	0.15	0.00	1.64	1.90	1.30	1.01	0.88	0.48	0.00
Mean	–0.60	0.09	0.51	0.77	0.58	0.36	0.11	0.17	0.60	0.84	1.15	0.62	0.34	0.00
H_0 : Mean = 0	0.81	0.96	0.75	0.04	0.02	0.09	0.18	0.92	0.66	0.52	0.02	0.06	0.22	na
MSE_l/σ_L^2	1.33	1.00	0.70	0.05	0.02	0.01	0.00	1.41	1.20	1.11	0.21	0.09	0.05	0.00

Note: σ_L^2 is the variance of the final revision, y_t^L.

Table 2

Integrated SNR measure

	Nominal variables			Real variables		
	$ISNR_L$	Vintage when $SNR_l \geq 0.5$	Vintage when $ISNR_l \geq 0.5$	$ISNR_L$	Vintage when $SNR_l \geq 0.5$	Vintage when $ISNR_l \geq 0.5$
GDP	0.526	t	$t + 1/2$	0.589	$t - 1/2$	$t + 1/4$
Private consumption	0.508	t	$t + 1/2$	0.524	$t - 1/2$	$t + 14$
Government consumption	0.498	$t + 1/4$	$t + 1/2$	0.345	$t + 1/2$	$t + 1/2$
Total investment	0.542	t	$t + 1/2$	0.532	t	$t + 1/2$
Investment in equipment	0.538	t	$t + 1/2$	0.532	t	$t + 1/2$
Investment in construction	0.525	t	$t + 1/2$	0.481	t	$t + 1/2$
Exports	0.553	t	$t + 1/2$	0.504	t	$t + 1/2$
Imports	0.549	t	$t + 1/2$	0.455	t	$t + 1/2$

Table 3: **Conditional distributions of forecast / revision errors: Entropy and gain**

Vintage	GDP	Private consumption	Government consumption	Total investment	Investment n equipment	Investmen in construction	Exports	Imports
				Nominal variables				
				Entropy				
$t-3/2$	1.05	0.93	0.68	2.49	2.81	2.35	2.81	3.02
$t-1$	0.95	0.86	0.74	2.29	2.58	2.24	2.52	2.88
$t-1/2$	0.38	0.63	0.85	1.95	2.31	1.56	2.07	2.30
t	–0.02	0.15	0.55	1.55	1.95	1.54	0.97	0.92
$t+1/4$	–0.40	–0.02	–0.06	0.13	0.98	0.31	0.17	0.32
$t+1/2$	–0.67	–0.16	0.04	0.23	0.97	0.00	0.28	0.52
$t+3/4$	na	na	na	na	na	na	na	na
				Gain[a)]				
$t-1$	0.10	0.07	–0.06	0.20	0.23	0.11	0.29	0.13
$t-1/2$	**0.56**[b)]	0.23	–0.11	0.33	0.27	0.68	0.45	0.58
t	0.40	**0.48**	0.30	0.41	0.36	0.02	**1.11**	**1.38**
$t+1/4$	0.38	0.17	**0.61**	**1.42**	**0.97**	**1.23**	0.80	0.60
$t+1/2$	0.27	0.14	–0.10	–0.11	0.01	0.32	–0.11	–0.19
$t+3/4$	na	na	na	na	na	na	na	na

Continued next page

Table 3 continued

Vintage	Real variables							
	GDP	Private consumption	Government consumption	Total investment	Investment in equipment	Investmen in construction	Exports	Imports
				Entropy				
$t-3/2$	1.21	1.02	0.71	2.48	2.79	2.28	2.60	2.44
$t-1$	1.10	1.07	0.85	2.34	2.60	2.09	2.25	2.45
$t-1/2$	0.77	0.62	0.86	2.10	2.06	1.58	1.79	2.46
t	–0.19	0.31	0.83	1.44	2.00	1.48	1.12	1.44
$t+1/4$	–1.05	0.06	0.72	0.18	0.92	–0.16	0.67	1.01
$t+1/2$	–1.74	–0.12	0.24	0.19	0.92	–0.22	0.19	0.73
$t+3/4$	na	na	na	na	na	na	na	na
				Gain				
$t-1$	0.11	–0.05	–0.14	0.14	0.19	0.20	0.34	–0.01
$t-1/2$	0.33	**0.45**	–0.01	0.24	0.53	0.51	0.47	–0.01
t	**0.96**	0.31	0.03	0.66	0.07	0.10	**0.67**	**1.02**
$t+1/4$	0.86	0.25	0.11	**1.27**	**1.08**	**1.63**	0.45	0.43
$t+1/2$	0.69	0.17	**0.48**	–0.01	0.00	0.07	0.48	0.28
$t+3/4$	na	na	na	na	na	na	na	na

[a] Gain stands for difference in entropy between two successive rows.

[b] Bold font indicates maximum gain for a given variable.

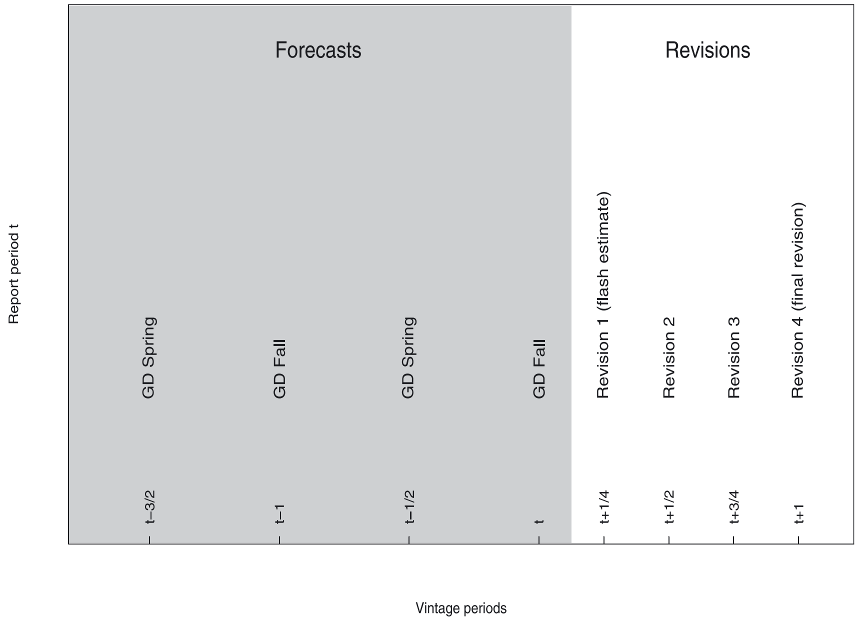
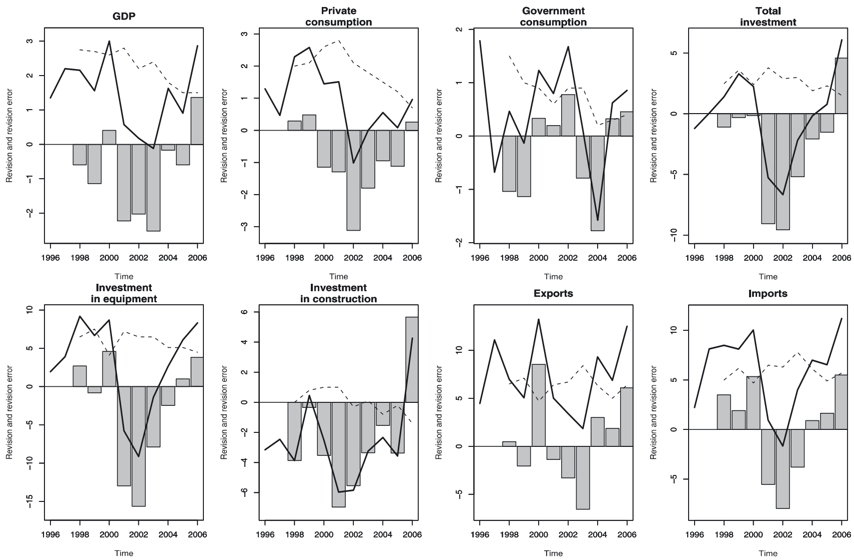


Figure 1: Sequence of forecasts / revisions for the reporting period *t*

(a) Nominal variables



(b) Real variables

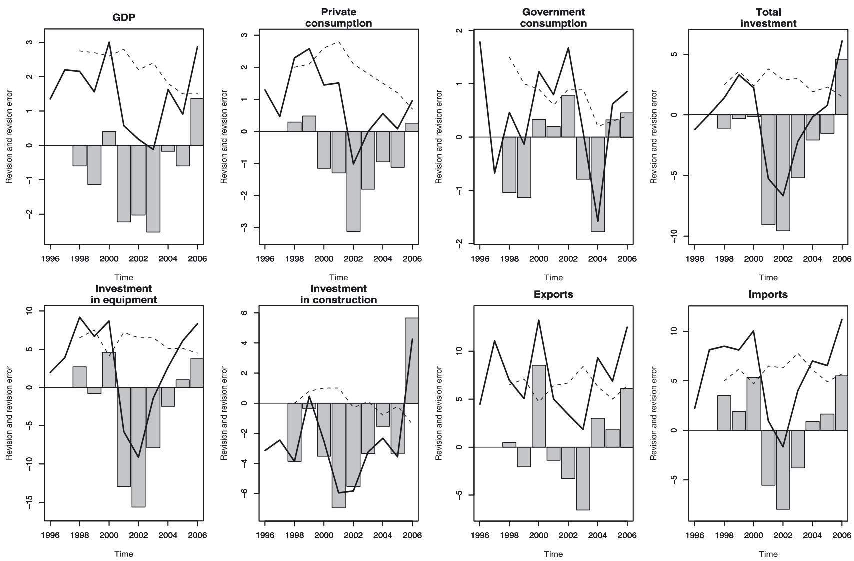
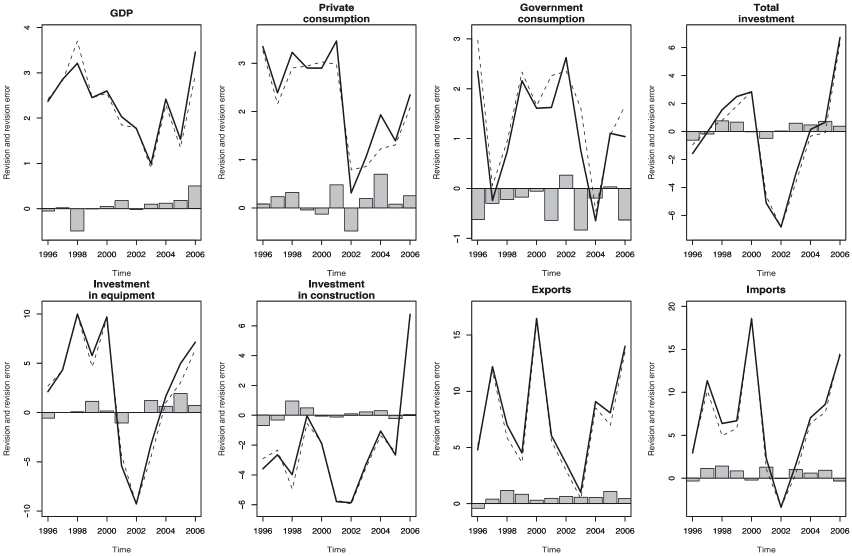


Figure 2: First forecasts ($l = t - 3/2$, dashed line) vs. final data ($l = t + 1$, bold continuous line), 1996 – 2006

(a) Nominal variables



(b) Real variables

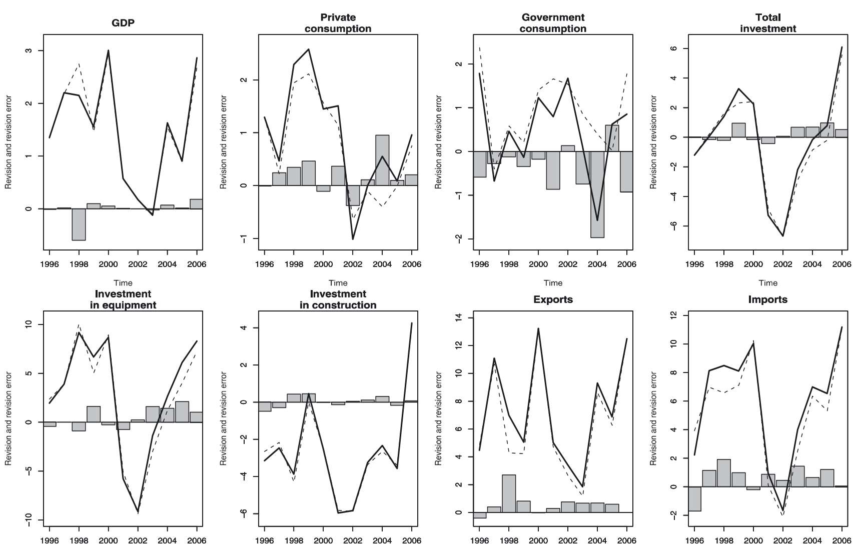


Figure 3: First statistical publication ($l = t + 1/4$, dashed line) vs. final data ($l = t + 1$, bold continuous line), 1996 – 2006

DOI https://doi.org/10.3790/aeq.55.4.269

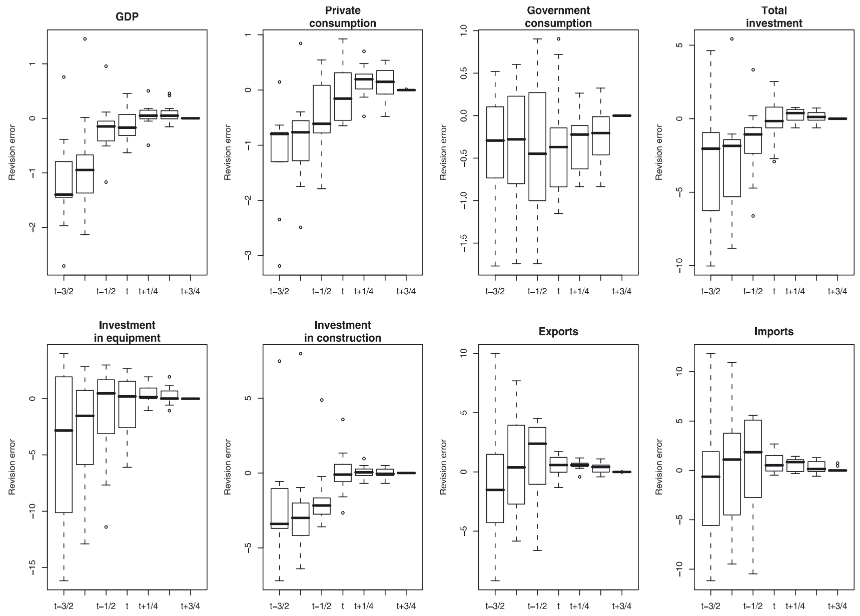


Figure 4: Nominal variables: Distribution of the forecast / revision errors

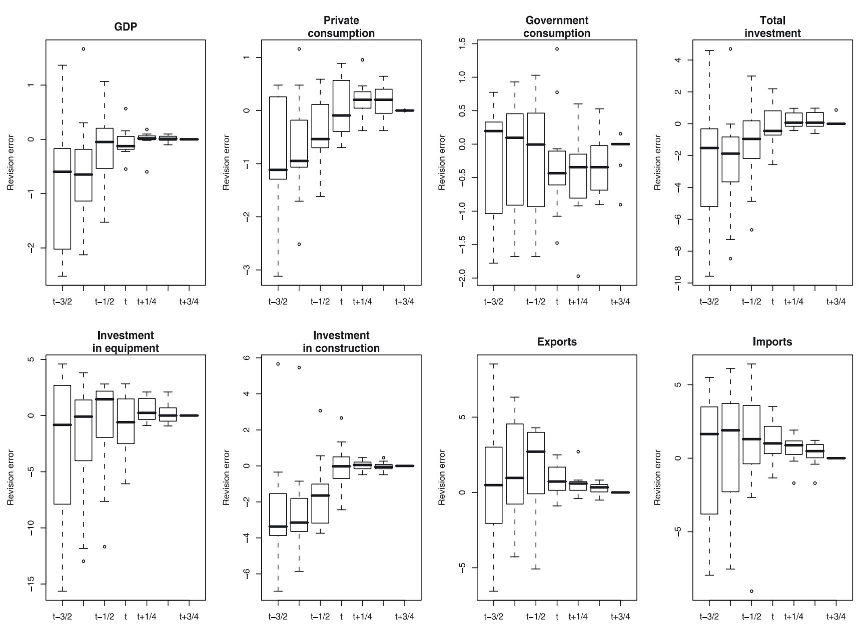


Figure 5: Real variables: Distribution of the forecast / revision errors

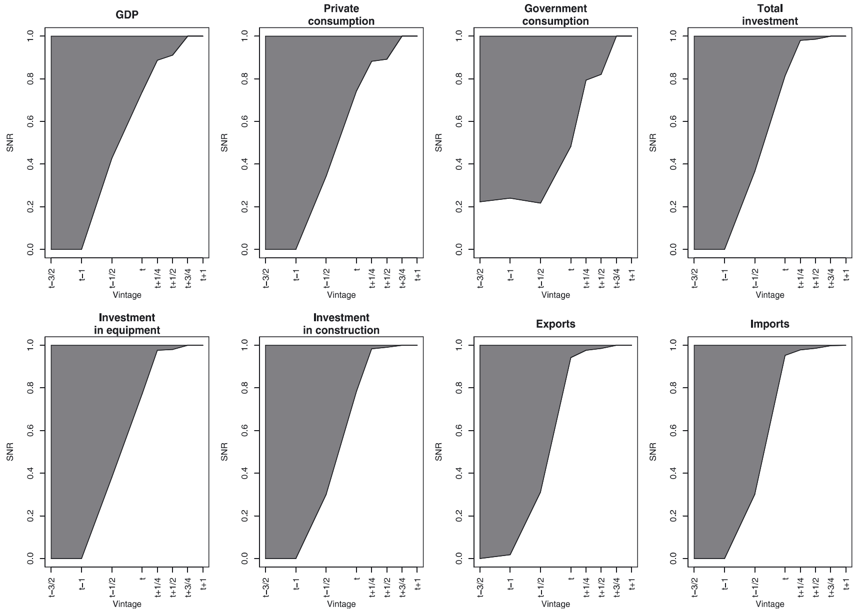


Figure 6: Nominal variables: Signal-to-noise measure

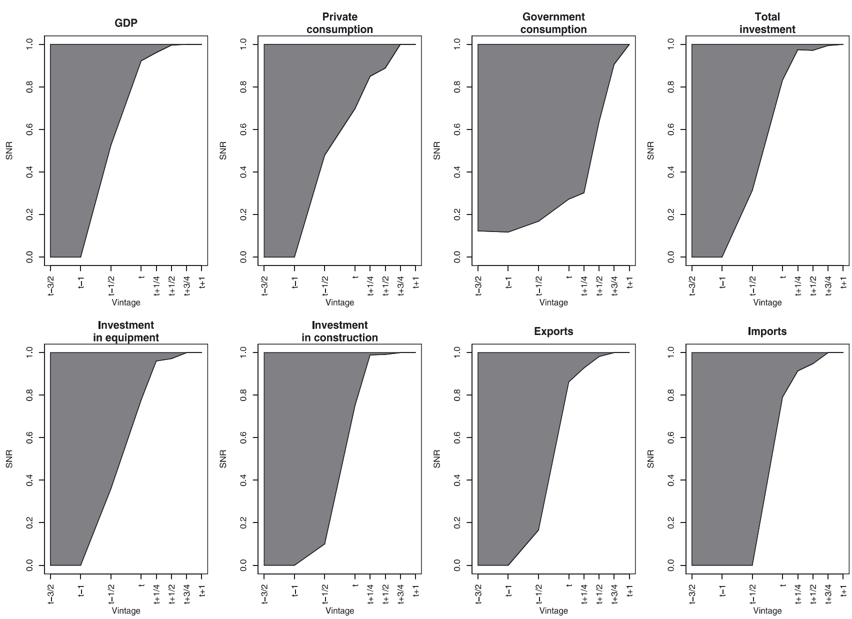


Figure 7: Real variables: Signal-to-noise measure