Needed: A Real Time Database for the Euro-Area

Brussels, 13-14 June 2005

Hosted by the National Bank of Belgium

HOW RELIABLE ARE GDP FIRST ESTIMATES? IMPLICATIONS FOR SHORT-TERM FORECASTING

Daniel GRENOUILLEAU

Abstract:

Much attention is paid to headline statistical releases such as GDP or HICP estimates. However, the comparison of GDP first (or flash) estimates with final estimates (proxied by the latest estimates available) shows that a sizeable estimation error affects early estimates. The average absolute revision is 0.17 pp over 1999-2003, and moreover the relative revision is -0.08 pp, pointing to a systematic underestimation of growth by first estimates. The first estimates for expenditure components are characterised by a large revision error and (for most of them) a negative bias as well, especially so for investment, exports, imports, the trade balance and of course inventories.

The conclusion is that first estimates should be interpreted with caution, especially in the case of GDP and such components as investment or foreign trade. An important implication for the real-time assessment of "nowcast" or short-term forecast models is that, for out-of-sample experiments, a robustness check with vintage series as input data should always be made. The impact of end-of-sample quarters, for which good proxies for the final estimates of GDP are not yet available, should also be monitored. The impact of GDP revisions on model accuracy is examined for three different models (AR model and dynamic factor model for GDP growth, production-function model for the output gap).

D. Grenouilleau is an economist at the European Commission, Directorate general for Economic and Financial Affairs in the Unit *Econometric models and medium term studies*. Comments are gratefully received at: Daniel.Grenouilleau@cec.eu.int.

1. INTRODUCTION

1.1. The concepts

With the introduction of GDP **flash estimates**, it is now possible to have an estimate of quarterly output as early as 45 days after the end of the quarter considered. **First estimates** of the expenditure components are available slightly later: about 65 days after the end of the quarter, and **second estimates** about 100 days after the end of the quarter. Estimates undergo a revision process that can last up to 2 years. At the end of this process, revisions become negligible and the estimates can be considered as final. Following the US Bureau of Economic Analysis terminology, the benchmark final estimates considered are the **latest estimates** available.

While the first quarterly estimates are those most closely watched by economic policy makers, there is little interest about revisions and estimates available later. Investment bank analysts[1] have occasionally presented anecdotal statistical evidence illustrating a directional bias of Eurostat first estimates. However, by contrast to the US where there is a long tradition of GDP revision studies at the Bureau of Economic Analysis[2], no such study exists for the euro area to the best of our knowledge[3]. Vintage series (i.e. the series as published for the first releases before subsequent revisions) are officially available from Eurostat[4] only for recent quarters.

The focus of this note is limited to a statistical analysis of the revisions to real GDP estimates for the euro area (Greece excluded before 2001). Since quarterly deflator prices are chained, it would not make sense to study real GDP revisions in level. Therefore, we consider revisions to the quarter-on-quarter growth rate of GDP (seasonally and trading-day adjusted) and of the expenditure components.

Several statistics can be used to monitor revisions over time:
- mean revisions (MR) of the aggregates' growth,
- mean absolute revisions (MAR),
- root mean squared revisions (RMSR).

MR statistics might help to identify whether there exists a systematic bias in the estimates or not. Unfortunately, our sample is too short to compute moving averages of mean revisions for a significant time span, and little evidence about a potential change of the bias over time can be gathered with our data. MAR and RMSR might illustrate the magnitude of revisions over time, or in the other words, how wrong a given estimate is on a stand alone basis (compared to the final estimate). Assuming a normal distribution of the absolute revisions, confidence bands around estimates might be derived from the RMSR.

The revision statistics is defined as the difference between the first estimates available (or subsequent estimates) **and the latest estimate available**. If the revision is positive (respectively negative), the first estimate was overestimating (resp. underestimating) growth.

---

[1] Goldman Sachs (2004).

[2] Following Jaszi G. (1965). More recently: Fixler (2004).

[3] Studies comparing revisions across some countries of the European Union are nevertheless available; see for instance Lars-Erik Öller and Karl-Gustav Hansson (2002).

[4] Since Eurostat does not publish internal working papers or documents explaining in details their aggregation method, it is also impossible to replicate their results.

1.2. The data

Given that very little vintage data is available from official sources, we had recourse to historical official data collected at the time of release for 2002 and 2003 releases (2004 releases are excluded from the study since we do not have yet good proxies for the corresponding final estimates). It is noteworthy that a bias exists for 2003 vintages, as the latest estimates available will be significantly further revised. Data for 1999 to 2001 were collected from the European Commission press releases. Unfortunately, in those releases full GDP series are not available in level: only the latest point in level is published, but GDP quarter-on-quarter (qoq) growth for the latest four quarters are also published (they are rounded to the first decimal). We have used these figures, but it should be noted that the rounding is the source of another (slight) bias[5] for revisions from 1999Q1 to 2001Q4.

For the revisions to the expenditure components, we have considered the contributions to the revisions to the change in GDP, i.e. (except for inventories[6]) the revisions to the qoq growth rates of those components weighted by the ratio of the components to GDP.

Finally, the countries' contributions to revision to euro area GDP show that small countries tend to revise more their quarterly national accounts than large countries. Due to aggregation, euro area GDP is less revised than any Member States national accounts.

2. REVISIONS TO EURO AREA GDP, INVESTMENT AND FOREIGN TRADE ARE SIZEABLE BUT ARE LOWER THAN MEMBER STATES' GDP REVISIONS.

2.1. Revisions in GDP estimates are sizeable, even a few quarters after the first estimate

Based on a 20-quarter vintage sample, the **RMSR of euro area GDP is 0.21 pp and the MAR is 0.17 pp**. The magnitude of revisions should be compared to the magnitude of GDP growth itself: the ratio of the MAR to the average absolute growth rate is about 40%. This indicator of absolute imprecision of the first estimate of qoq GDP growth is decreasing quite slowly with subsequent estimates. Three quarters after the first release, the ratio is still about 25%.

The **MR is negative (about -0.1 pp),** which means that **growth seems systematically underestimated**. The bias is not significantly reduced in the subsequent releases up to 3 quarters after the first estimate.

Only a limited history is available for the flash estimate. Flash and first estimates are very close and available data suggest for the time being that revisions of the flash estimate in term or MAR or RMSR are only 10% higher than those of the first estimate.

---

[5] The absolute bias on each component is necessarily lower than 0.025 percentage point. The average rounding error for growth rates between -0.1% and 0.1% is 0.025 pp. For growth rates higher than 0.k% in absolute value, the average rounding error is lower than 0.025/k percentage points (in the MR or MAR).

[6] The qoq change in inventories is usually expressed as the change in stocks divided by GDP. For a reason of homogeneity, its contribution becomes the difference between two subsequent changes in stocks divided by GDP in level for the base quarter (the contributions of all components' change should necessarily add up to the change in GDP).

GDP revisions (in % points)	First estimates				One quarter later			
	MR	**MAR**	RMSR	*MAR/ MAG*	MR	**MAR**	RMSR	*MAR/ MAG*
1999-2003	-0.08	**0.17**	0.21	*41%*	-0.08	**0.14**	0.19	*32%*
99Q1-01Q2	-0.12	**0.21**	0.28	*34%*	-0.12	**0.17**	0.24	*27%*
01Q3-03Q4	-0.05	**0.10**	0.12	*49%*	-0.04	**0.08**	0.10	*37%*
	Two quarters later				Three quarters later			
	MR	**MAR**	RMSR	*MAR/ MAG*	MR	**MAR**	RMSR	*MAR/ MAG*
1999-2003	-0.07	**0.13**	0.17	*29%*	-0.07	**0.11**	0.14	*24%*
99Q1-01Q2	-0.10	**0.17**	0.22	*27%*	-0.09	**0.14**	0.18	*22%*
01Q3-03Q4	-0.03	**0.05**	0.08	*23%*	-0.05	**0.05**	0.06	*23%*

Legend:
MR Mean revision
MAR Mean absolute revision
RMSR Root mean squared revision
MAG Mean absolute growth

The MAR over the sub-samples of the 10 first and 10 last quarters are respectively 0.28 pp and 0.12 pp. The discrepancy could be explained by three potential reasons:
1. Revisions could be smaller at the end of the second sub-sample due to the fact that the revision process is not yet completed for those quarters. This end-of-sample revision bias might translate into a further increase in the mean revisions with subsequent releases;
2. Revisions could be scaled to GDP growth, i.e. the lower the magnitude of the actual change in GDP is, the smaller the revisions. The average growth is much higher in the first sub-sample (0.6% in 1999-2001H1) than in the second one (0.2% in 2001H2-2003), and the ratios of the MAR to the average absolute growth rate for both sub-samples are much closer than the MAR. This suggests that a scaling factor could partially explain the lower magnitude of revisions in the second sub-sample;
3. National statistical offices' estimates of QNA might have recently become more reliable. We have not found any evidence supporting this thesis.

In terms of confidence level associated with the estimates, it is noteworthy that under some assumptions[7], **there is about 70% chance that the final growth rate for a given quarter will fall within a range of ±0.2 pp around the first estimate point**. The range width (0.4 pp) should be compared to an average growth rate over the same sample of 0.42%. This reflects a **considerable uncertainty around the first estimate figures**.

2.2. Looking at expenditure components, GDP revisions are mainly due to trade estimates volatility

The table below summarizes mean contributions to GDP revision and mean absolute contributions to GDP revisions over 1999-2003. This means that the revision of each component is implicitly weighed according to the relative size of the component compared to GDP. Only the most striking results will be commented in this note, but these results might deserve further analysis and, possibly, comments from national accounts experts.

The first estimates of the contribution of private consumption are quite accurate, although a systematic downward bias of more than half of the magnitude of GDP bias seems to exist. The accuracy hardly improves with subsequent estimates, which suggests that more reliable data are only available after one or two years (perhaps VAT receipts). The seemingly absence of directional bias in the second sub-sample might be unreliable and might just stem from a general end-of-sample bias problem (the latest estimates available for those quarters will be further revised and are inaccurate proxies of the final estimates).

[7] We have to assume that revisions across vintage series are normally distributed, and that the average growth rate over the time sample 1999-2003 corresponds to the level of trend growth (recall that revisions might be scaled to the magnitude of GDP growth).

revisions (in % points)	First estimates		One quarter later		Two quarters later		Three quarters later	
	MR	**MAR**	MR	**MAR**	MR	**MAR**	MR	**MAR**
Private consumption								
1999-2003	-0.04	**0.11**	-0.05	**0.10**	-0.05	**0.09**	-0.04	**0.08**
99Q1-01Q2	-0.08	**0.14**	-0.08	**0.12**	-0.09	**0.11**	-0.06	**0.08**
01Q3-03Q4	0.00	**0.06**	-0.01	**0.07**	0.00	**0.05**	-0.02	**0.08**
Government consumption								
1999-2003	-0.03	**0.05**	-0.03	**0.05**	-0.02	**0.05**	-0.02	**0.05**
99Q1-01Q2	-0.02	**0.05**	-0.03	**0.06**	-0.02	**0.05**	-0.02	**0.04**
01Q3-03Q4	-0.04	**0.06**	-0.03	**0.04**	-0.02	**0.04**	-0.03	**0.06**
Investment								
1999-2003	-0.06	**0.13**	-0.02	**0.09**	-0.04	**0.07**	-0.03	**0.06**
99Q1-01Q2	-0.09	**0.14**	-0.02	**0.08**	-0.05	**0.08**	-0.06	**0.07**
01Q3-03Q4	-0.03	**0.11**	-0.02	**0.09**	-0.02	**0.07**	0.01	**0.04**
Inventories								
1999-2003	0.05	**0.20**	-0.01	**0.15**	0.04	**0.14**	-0.04	**0.16**
99Q1-01Q2	0.13	**0.21**	-0.06	**0.16**	0.05	**0.13**	-0.05	**0.15**
01Q3-03Q4	-0.04	**0.20**	0.04	**0.14**	0.03	**0.18**	-0.03	**0.19**
Exports (intra + extra)								
1999-2003	-0.04	**0.24**	-0.04	**0.20**	-0.08	**0.21**	-0.01	**0.20**
99Q1-01Q2	-0.12	**0.27**	-0.04	**0.21**	-0.05	**0.24**	0.05	**0.25**
01Q3-03Q4	0.03	**0.20**	-0.04	**0.17**	-0.11	**0.16**	-0.09	**0.08**
Imports (intra + extra)								
1999-2003	0.05	**0.29**	0.07	**0.20**	0.07	**0.21**	0.08	**0.18**
99Q1-01Q2	0.06	**0.30**	0.10	**0.20**	0.06	**0.19**	0.05	**0.18**
01Q3-03Q4	0.04	**0.28**	0.03	**0.19**	0.09	**0.25**	0.12	**0.19**
Trade balance (extra)								
1999-2003	0.00	**0.17**	0.03	**0.12**	0.00	**0.12**	0.07	**0.15**
99Q1-01Q2	-0.06	**0.16**	0.07	**0.15**	0.01	**0.13**	0.10	**0.15**
01Q3-03Q4	0.07	**0.19**	-0.01	**0.07**	-0.02	**0.09**	0.02	**0.13**

The contribution of government consumption to revisions is very limited, but this is due to the lower magnitude of government spending compared to other expenditure components[8]. As in the case of private consumption or perhaps even more so, it seems to take one to two years to obtain significantly enhanced estimates

Investment first estimates are much less accurate than the previous ones. Although the share of investment in GDP is only 20% (about the same as government spending), **absolute revisions to investment have a sizeable contribution of 0.13 pp** to those to GDP. The negative directional bias (-0.06 pp) is also significant (except, again, in the second sub-sample). However, unlike previous components, about half of the revision is performed after three quarters.

Inventories are strongly revised but this is inherent to the method of construction, according to which stocks are a residual account.

Last but not least, revisions to foreign trade have the largest contribution to GDP revisions. **The trade balance absolute contribution to GDP revision is 0.17 pp**. It is as large as GDP MAR itself, which means that it is somewhat offset by other components' revisions (probably inventories). The directional bias is different over the two sub-samples: **the relative contribution of the trade balance is underestimated over 1999-2001** in a period of depreciation of the euro and **overestimated thereafter (2001-2003)** in a period of appreciation of the euro. Exports and imports first estimates are very poor with respective RMSR of about 0.3 pp and 0.4 pp. It should be stressed

[8] In absolute terms (in terms of revisions and not in terms of contributions to revision), government consumption is in fact less accurately estimated than private consumption.

that, for respective ranges of ±0.3 pp (0.6 pp width) and ±0.4 pp (0.8 pp width) around qoq point estimates, the confidence level is less than 70% under a normal distribution assumption.

The revisions to the foreign trade contribution to GDP is investigated in more details in appendix

2.3. The beauty of aggregation

The same exercise as above is now performed with the second estimate published by Eurostat for all countries. The following table shows that small countries tend to revise their accounts more than large countries. Given the considerable revisions made to Denmark, Finland, Belgium or Portugal national accounts, there might be a gain in accuracy in incorporating some of these countries' data at a later stage than the second estimate.

	1999Q1-2003Q4			1999Q1-2001Q2			2001Q3-2003Q4		
Countries	**MR**	**MAR (MAR/MAG)**	**RMSR**	**MR**	**MAR (MAR/MAG)**	**RMSR**	**MR**	**MAR (MAR/MAG)**	**RMSR**
EA[9]	-0.07%	0.12% (23%)	0.17%	-0.09%	0.16% (21%)	0.22%	-0.05%	0.08% (31%)	0.09%
BE	-0.08%	0.46% (66%)	0.62%	-0.11%	0.70% (76%)	0.84%	-0.04%	0.22% (45%)	0.26%
DK	-0.18%	0.53% (91%)	0.67%	-0.31%	0.54% (77%)	0.68%	-0.05%	0.53% (113%)	0.65%
DE	-0.09%	0.26% (56%)	0.33%	-0.16%	0.34% (49%)	0.42%	-0.02%	0.18% (78%)	0.23%
EL[10]							0.04%	0.19% (16%)	0.41%
ES	-0.15%	0.30% (36%)	0.38%	-0.11%	0.34% (35%)	0.39%	-0.19%	0.27% (38%)	0.37%
FR	-0.10%	0.24% (35%)	0.32%	-0.15%	0.24% (28%)	0.29%	-0.05%	0.23% (44%)	0.34%
IT	-0.02%	0.15% (36%)	0.21%	-0.11%	0.19% (29%)	0.25%	0.07%	0.12% (65%)	0.15%
NL	-0.02%	0.21% (37%)	0.28%	-0.02%	0.23% (29%)	0.32%	-0.03%	0.19% (59%)	0.22%
AT	-0.02%	0.26% (53%)	0.32%	0.09%	0.28% (40%)	0.33%	-0.13%	0.24% (84%)	0.32%
PT[11]	-0.03%	0.46% (73%)	0.59%	0.40%	0.55% (70%)	0.66%	-0.24%	0.41% (90%)	0.54%
FI	-0.15%	0.52% (59%)	0.64%	-0.01%	0.43% (39%)	0.50%	-0.28%	0.60% (90%)	0.75%
SE[12]	-0.10%	0.31% (47%)	0.36%	-0.10%	0.42% (51%)	0.46%	-0.10%	0.22% (43%)	0.25%
UK	-0.10%	0.20% (31%)	0.27%	-0.16%	0.25% (33%)	0.33%	-0.04%	0.16% (29%)	0.20%

Euro area GDP is less revised than countries' GDP. As a first check, one can examine the correlations between countries' revisions. Spanish revisions are indeed negatively correlated to those of most countries. Belgian revisions are also negatively correlated to those of Finland, Greece and Portugal and this feature is likely to minimise the impact of the important revisions of all four countries taken separately.

	EA	**BE**	**DK**	**DE**	**GR**	**ES**	**FR**	**IT**	**NL**	**AT**	**PT**	**FI**	**SE**	**UK**

[9] These statistics for the euro area are slightly discrepant with first estimates figures. This could be due to a methodological problem for the second estimates releases. For the first two quarters of 1999 and the first of 2000 and 2001, the growth rates are strongly revised compared to the former and subsequent quarter's press releases (0.1 pp, 0.2 pp, 0.2 pp and 0.1 pp respectively).

[10] Data from 1999Q1 to 2001Q3 is missing. Last table includes 9 periods instead of 10, from 2001Q4 to 2003Q4

[11] Data from 1999Q1 to 2000Q1 is missing. Table one includes 15 periods instead of 20, and the second one 5 instead of 10

[12] 2000Q3 value was missing, so the first table includes 19 instead of 20 periods and the second one 9 instead of 10

EA		16%	28%	82%	20%	13%	34%	43%	66%	9%	72%	11%	59%	49%
BE	16%		17%	5%	-25%	-46%	19%	23%	36%	10%	-33%	-16%	33%	-28%
DK	28%	17%		27%	33%	-46%	34%	19%	18%	2%	21%	26%	-2%	43%
DE	82%	5%	27%		-8%	9%	-3%	41%	51%	-14%	59%	12%	36%	50%
GR	20%	-25%	33%	-8%		11%	11%	9%	-22%	12%	24%	25%	-21%	60%
ES	13%	-46%	-46%	9%	11%		-40%	-2%	-14%	17%	32%	-12%	-15%	-2%
FR	34%	19%	34%	-3%	11%	-40%		-12%	6%	5%	-18%	-6%	9%	16%
IT	43%	23%	19%	41%	9%	-2%	-12%		32%	-4%	-12%	-7%	43%	29%
NL	66%	36%	18%	51%	-22%	-14%	6%	32%		2%	58%	1%	50%	13%
AT	9%	10%	2%	-14%	12%	17%	5%	-4%	2%		-15%	50%	9%	-13%
PT	72%	-33%	21%	59%	24%	32%	-18%	-12%	58%	-15%		8%	9%	23%
FI	11%	-16%	26%	12%	25%	-12%	-6%	-7%	1%	50%	8%		10%	38%
SE	59%	33%	-2%	36%	-21%	-15%	9%	43%	50%	9%	9%	10%		30%
UK	49%	-28%	43%	50%	60%	-2%	16%	29%	13%	-13%	23%	38%	30%	

3. SOME IMPLICATIONS FOR MODEL ACCURACY ASSESSMENT

3.1. Should model-based forecasts replicate first estimates?

The relative inaccuracy of first estimates poses a serious challenge for short-term forecasting. Suppose Eurostat model for first estimates and a given forecast model have errors that are not greatly correlated, it is likely that, once in a while, a very large discrepancy will be observed between a flash (or first) estimate and a coincident forecast. Given that, at a 70% confidence level, the range around Eurostat point estimates is ±0.2 pp (hence a total width of 0.4 pp), if a coincident forecast has the same confidence level vis-à-vis the final estimate with a range of ±0.2 pp, the discrepancy between the former estimate and the later forecast could easily amount to 0.3 pp or even 0.4 pp for some quarters.

The solution would be to find a forecast model, of which errors are well correlated with Eurostat estimates. The only way to achieve this would be to make forecasts for the Member States' indicators that are used by Eurostat for the flash estimate, to use the same aggregation scheme (undisclosed by Eurostat) and to apply seasonal and (massive) trading-day adjustments in the same way as some Member States. But it would not make much sense to forecast an estimate which is itself biased and volatile (partly due to noise) with respect to the short-term assessment of the economic situation.

3.2. Are out-of-sample experiments conducted with the latest estimates available reliable?

A model that obtains a good in-sample fit is not necessarily accurate out-of-sample. Putting aside model uncertainty[13] issues, the reason is that the model might be fitted with information that will be

[13] The search for a model involves an unavoidable selection among the economic series that are likely to perform well (data mining). In the OLS framework, confidence bands around the estimates are always conditional to the assumption that the model is the right one. A non-random series selection procedure that targets the best possible fit in-sample is likely to lead to biased confidence intervals due to model uncertainty, i.e. due to the fact that the

effectively available after the date for which the estimation is performed. Out-of-sample experiments solve the problem insofar as the estimation of the model is only based on the observations up to that for which the estimation is performed.

A similar problem might indeed occur where vintage data are not used. Confidence intervals obtained with a model fitted with the latest series available do not take into account **data uncertainty**. In a real time situation, the information available for the model estimation is of a lesser precision than the final series used to perform a "standard" out-of-sample experiment. All models might not have the same sensitivity to the input data quality. In general, the discrepancy between real-time out-of-sample and standard out-of-sample is likely to depend on the weight of the revised variables as explanatory variables[14].

For the same reasons, the impact of the last observations of GDP on out-of-sample results should be carefully monitored. Since GDP can be significantly revised for about two years, there is some uncertainty around the forecast accuracy statistics computed for the latest quarters available. In practice, an out-of-sample experiment limited to the ten last quarters is likely to lead to a wrong assessment since the benchmark (the latest observations of GDP) is itself imperfectly measured.

3.3. Should short-term economic assessment target GDP growth or a less noisy indicator of the economic situation?

The focus of this note was restricted to the accuracy of GDP first estimates compared to GDP final estimates. A broader issue is that of the meaningfulness of GDP growth estimates for the assessment of the overall economic situation, and more precisely the issue of **the noise surrounding GDP first estimates as well as final estimates**. There is anecdotal evidence of a significant remaining seasonal pattern in seasonally adjusted GDP (due to the aggregation of series adjusted with different methods and unadjusted series in some cases). Moreover, trading-day adjustments seem to have recently introduced even more volatility in GDP growth series. Let us also recall that there is now a substantial discrepancy between quarterly adjusted growth rates and annual growth rates due to the fact that annual data are not trading-day adjusted. Obviously, a precise assessment of the economic situation should discard noise and some of the volatility of the GDP growth measured series. An interesting challenge would be to construct a less noisy indicator of the economic situation.

4. EXAMPLES OF MODEL ACCURACY ASSESSMENT WITH REAL TIME DATA

4.1. An AR benchmark model of GDP growth

An AR model for GDP growth is a tough benchmark especially for several steps ahead forecasts, partially due to the fact that Eurostat GDP contains some seasonality. We examine here whether neglecting the real time availability of the series has an impact on the AR model accuracy assessment.

---

model could just be wrong and another model the right one. The fit could be just due to data mining and the model have little structural meaning or statistical properties.

[14] The revised GDP series is used for filtering the other data ex ante in our factor model. The use of the latest series available for filtering would lead to a much better (but biased) forecast performance.

If the model selection procedure uses all the available information optimally, there is no difference between forecast accuracy statistics based on final or real time data. The table opposite summarizes the results of the following experiment: using the latest GDP series (including a first estimate for 2005Q1), the best AR models for each forecast horizon are selected based on respective out-of-sample RMSE. The RMSE[15] is in a second stage recomputed with series available in real time for each quarter of the out-of-sample exercise. Results are barely unchanged.

AR model for GDP growth (model selected with final data)			
Out of sample RMSE (25 quarters) based on:			
Horizon:	Lags:	final data	real time data
1	1 2 3 5	0.31	0.31
2	2 5	0.34	0.35
3	3	0.38	0.38
4	4	0.35	0.35

The design of the former experiment did not take into account the fact that, in a real time situation, the model selection cannot be based on data available later on. In the following second experiment, the model selection phase is more realistic in terms of real time constraints, insofar as models are chosen according to in-sample statistics computed with data up to the date of beginning of the out-of-sample. Thus, the model selection is not influenced by information available later on. The RMSE is now slightly higher at all forecast horizons: it is noteworthy that the model accuracy was slightly overrated according to the first experiment due to the use (at the stage of model selection) of information that was not available in real time. This caveat also concerns most bridge-equation type models (model selection is based on the full sample available) and explains the deterioration of forecast accuracy of such models in real time. On the other hand, results are virtually unchanged whether final or real time data are used to compute the out-of-sample results. It means that, when only data available in real time is used, data uncertainty does not lead as such to a wrong assessment of the model accuracy, but it might further the choice of a non robust model.

AR model for GDP growth (model selected with data available in real time)			
Out of sample RMSE (25 quarters) based on:			
Horizon:	Lags:	final data	real time data
1	1 3 4 5	0.34	0.34
2	2 3 4 5	0.34	0.35
3	3 4 5	0.39	0.39
4	4 5	0.36	0.36

The conclusion of both experiments is that data uncertainty as such is not a real issue for models that are not very accurate, probably because they do not use the information available optimally. But it can become an issue if the model choice is based on a procedure or decision rule that is not robust enough

4.2. A dynamic factor model of GDP growth

Thanks to the optimal use of a larger amount of information, "approximate" factor models[16] should allow more precise coincident estimates of GDP growth than conventional OLS equations or VAR systems with a limited number of explanatory variables. However, factor models' performances are not necessarily robust to real time data availability constraints: the best series might be available with a substantial delay (especially in the case of the euro area[17]) and the only series coincidently available are limited to survey series (which contain a lot of noise) or financial series. Moreover, for monthly series, only partial information is available about each quarter until the last monthly figure of the quarter is obtained.

The dynamic factor model forecast accuracy assessment introduced by Grenouilleau (2004) is based on a complex procedure that ensures that, for each out-of-sample quarter, only data effectively

---

[15] The reference series is the euro area GDP computed by the OECD as published in May 2005

[16] See Stock and Watson (2004) for a general presentation.

[17] For the euro area, employment data (ECB) are available six months after the end of the relevant quarter, industrial production (Eurostat) three months later, HICP (Eurostat) two months later.

available at a precise point in the course of the quarter are effectively taken on board[18]. The only approximations made concern revised series (except GDP[19]) and seasonal adjustments coefficients.

The following table summarises out-of-sample forecast accuracy (based on the RMSE) at various points in time compared to the date of release of euro area GDP flash estimates[20]. The simulation was performed with a standard calibration[21] and the out-of-sample experiment contains 25 quarters for one-step ahead forecasts, 24 quarters for two-step ahead forecasts and 23 quarters for the three-step ahead forecasts[22]. The reference series is the euro area GDP computed by the OECD as published in May 2005[23].

Dynamic factor model accuracy												
Forecast date	Nine months before	Eight months before	Seven months before	Six months before	Five months before	Four months before	Three months before	Two months before	One month before	Flash estimate release date	One month before	Two months after
	T_f - 9m	T_f - 8m	T_f - 7m	T_f - 6m	T_f - 5m	T_f - 4m	T_f - 3m	T_f - 2m	T_f - 1m	T_f	T_f + 1m	T_f + 2m
Steps ahead [(a)]	3	3	3	2	2	2	1	1	1	0	0	0
RMSE	0.31	0.28	0.27	0.25	0.23	0.22	0.22	0.22	0.21	0.22	0.21	0.21

(a) Compared to Eurostat Flash estimate

Given that data availability constraints are taken into account as precisely as possible, these results suggests that the factor model accuracy is stable (the RMSE improves with greater data availability). Up to four months before the release of the flash estimate, the model forecasts seem quite reliable, but their accuracy deteriorates quickly for predictions performed more than six months ahead of the flash estimate release.

4.3. A production-function model[24] for the output gap

In a recent special box of the Monthly Bulletin of February 2005, the ECB concluded that real-time output gap estimates tend to be of low reliability and that the business cycle analysis should therefore be based on a wider set of indicators. However, the low reliability of output gap estimates is mainly due to the inaccuracy of GDP estimates/forecasts in real time; in other words, potential output is more reliably estimated than GDP itself.

The assessment of the accuracy of output gap[25] real time estimation (or forecast) involves the comparison of two estimates: a real time estimate (or forecast) produced in the past and, as a

---

[18] For example, if the out-of-sample experiment is performed with data up to 15 May 2005, the previous quarter simulated is based on data available up to 15 February 2005, and so forth.

[19] Vintage series for euro area and Member States GDP are used, but not HICP or output vintage series.

[20] The Flash estimate for euro area GDP (Eurostat) is typically available between 40 and 45 days after the end of the relevant quarter. The simulation is performed with data available on the 10th of each month.

[21] The same calibration is used for all forecast horizons: data window of 7 years, 3 factors, no filtering (GDP correlation threshold set to nil), forecasts computed with the EM algorithm, all series are introduced coincidently and with up to three lags.

[22] The results for four-step ahead forecasts close to the performance of the AR model (due to high autocorrelation at order four) are not displayed.

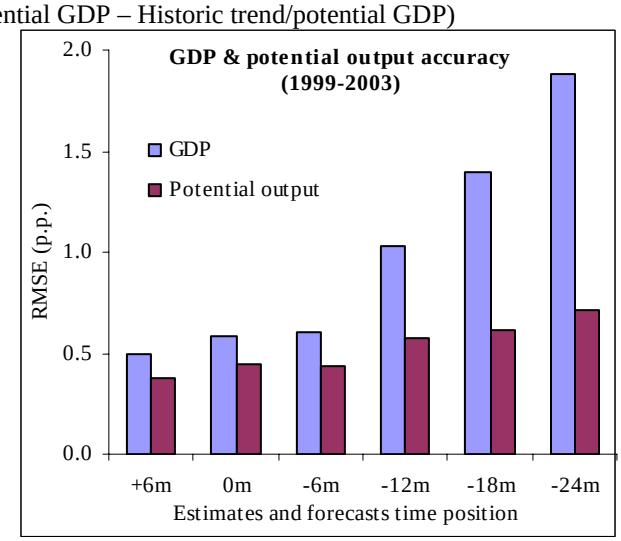
[23] The main difference between the OECD series and that of Eurostat is that the former is adjusted in order to ensure consistency between quarterly and yearly national accounts estimates.

[24] See Denis et al. (2002).

benchmark, a final estimate (the most recently available one) that is supposed to be no longer revised in the future. The following equation immediately shows that part of the output gap error might be completely independent from the issue of model uncertainty (potential output) but simply accounted by GDP revisions:

Output gap error = (Historic GDP - Final GDP) + (Final potential GDP – Historic trend/potential GDP)

An unbiased assessment of the output gap model performance requires disentangling errors due to potential output estimation and those due to GDP estimation. The following graph allows such comparison of both components. The assessment based on these statistics contrasts with the ECB judgement. The potential output accuracy seems rather satisfactory with a RMSE lower than 0.5 percentage point up to 6 months before the first release of national accounts data. Strikingly, its reliability is much better than the reliability of GDP forecasts even though forecasted data are necessarily used for potential output forecasts with a production function.



Note to the graph: The RMSE summarises differences between final estimates and estimates/forecasts produced respectively *x* months (*x*m) after(+)/before(-) the first release of national accounts data. The same sample (1999-2003) is used for potential output and GDP estimates/forecasts. As with previous statistics, estimates published 6 months after the first release of national accounts (NA) are taken from the Autumn forecasts of the subsequent year. Forecasts published 6 months before the first release of NA are taken from the Autumn forecasts of the current year, etc. And forecasts published 24 months before the first release of NA are taken from the Spring forecasts of the year before.

The conclusion with respect to output gap model uncertainty is unambiguously that the model is robust enough to cancel out part of the data inaccuracy. Model uncertainty does not seem to be the main issue. Conversely, the bad quality (see graph below) of GDP estimates for some countries and forecasts (for most countries) is the main source of errors.

Against this background, other indicators than the output gap might provide valuable information for the business cycle analysis only if those indicators are not as much revised as GDP.

---

[25] Note that historic estimates of the output gap according to DG ECFIN refer before 2002 to a concept of trend GDP and not potential GDP (since the Autumn 2002 forecasts). Final (benchmark) estimates are based on potential GDP.

5. CONCLUSION

First estimates should be interpreted with caution, especially in the case of GDP and such components as investment or foreign trade. An important implication for the real-time assessment of "nowcast" or short-term forecast models is that, for out-of-sample experiments, a robustness check with vintage series as input data should always be made. The impact of end-of-sample quarters, for which good proxies for the final estimates of GDP are not yet available, should also be monitored. The impact of GDP revisions on model accuracy examined for three models (AR model and dynamic factor model for GDP growth, production-function model for the output gap) show that data uncertainty tend to alter the assessment of the model accuracy.

6. REFERENCES

Denis, Mc Morrow and Röger (2002). "Production function approach to calculating potential growth and output gaps – estimates for the EU Member States and the US", *Economic Papers* 176, European Commission (Sept)

ECB (2005). “The (un)reliability of output gap estimates in real time”, *Monthly Bulletin*, Feb.

Fixler D. (2004). "Revisions to GDP Estimates in the US". US Bureau of Economic Analysis, Presentation to OECD Workshop on revisions, 7 October 2004.

Goldman Sachs (2004). "Statisticians selling Euroland Short", *Euroland Weekly Analyst*, GS global economic website at https://portal.gs.com.

Grenouilleau D. (2004). "A sorted leading indicators dynamic factor model for short-run euro-area GDP forecasting", *European Economy - Economic Paper* n°219, European Commission (Dec.).

Jaszi G. (1965). "The quarterly National Income and Product Accounts of the Unites States: 1942-62". In *Short-Term National Accounts and Long-Term Economic Growth*, ed. S. Goldberg et al., 100-187. London: Bowes & Bowes.

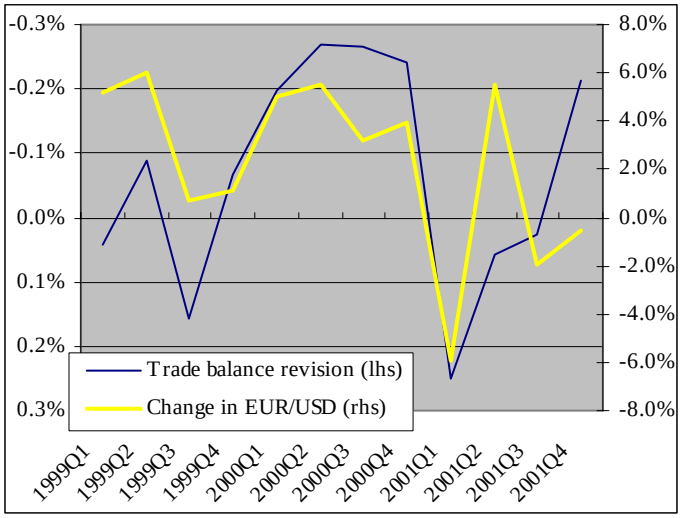
Öller Lars-Erik and Karl-Gustav Hansson (2002). *Revisions Of Swedish National Accounts 1980-1998 and an International Comparison*, Statistics Sweden.

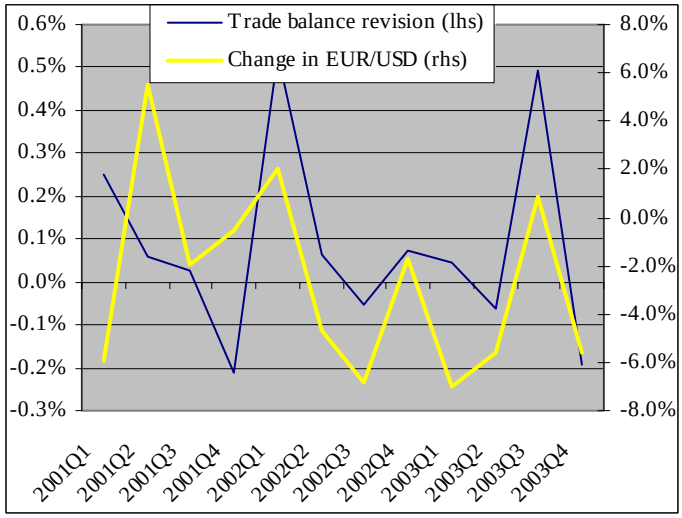
Stock J. H. and M. W. Watson (2004). "Forecasting with many predictors", survey prepared for the *Handbook of Economic Forecasting*.

7. APPENDIX:
MORE INSIGHT ABOUT THE REVISIONS TO THE FOREIGN TRADE CONTRIBUTION TO GDP

The distinct profile of relative revisions over our two sub-samples suggests that there could be a link between currency movements and trade revisions through deflators' estimation.

The two graphs below show the patterns of exports and imports contributions to growth. The coincident change in the EUR/USD seems somewhat correlated to the trade balance revision, except in the transitory period around the turning point in the EUR/USD trend. In a regime of euro depreciation (1999-2001), changes to the EUR/USD seem negatively correlated with an underestimation of the trade balance contribution, while an appreciating euro seems positively correlated to an overestimation of its contribution.





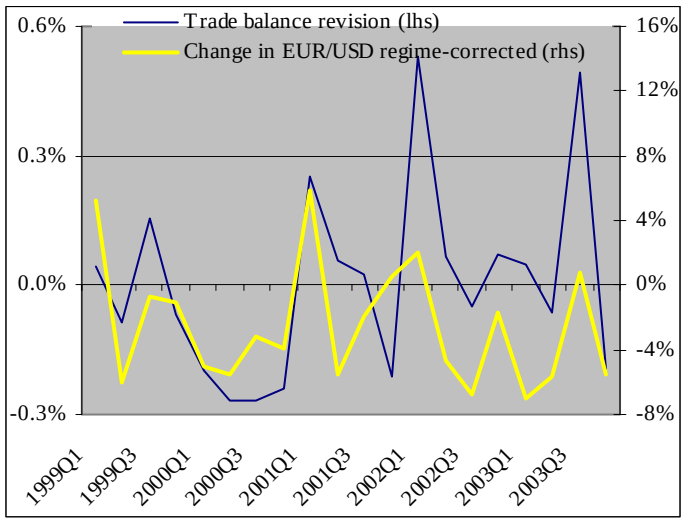
A transformation of the change in EUR/USD could thus be tested in order to incorporate potential non linearity and lags in the regime shift. An analysis of correlation between trade components and such a transformation is presented in appendix.

We examine in more details correlations between the change in EUR/USD (transformed) and real national accounts trade aggregates' revisions. A transformation of the EUR/USD is here tested in order to incorporate potential non linearity and lags in the regime shift. Revisions to real aggregates incorporate revisions to deflators and revisions to nominal trade. The following comments are based on the assumption that correlations between real aggregates and currency change are dominated by the link between the estimation of deflators and currency changes[26]. A more rigorous approach would obviously require disentangling revisions to nominal aggregates from revisions to deflators.

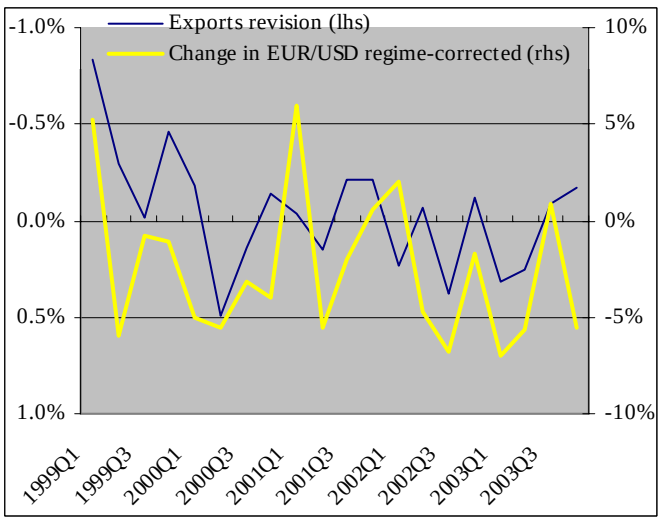
We use the following transformation of the EUR/USD. Where the three-quarter moving average of changes in EUR/USD is positive (regime of depreciation), then the coincident change in EUR/USD is replaced by its opposite value. Otherwise (regime of appreciation), the coincident change is kept as is.

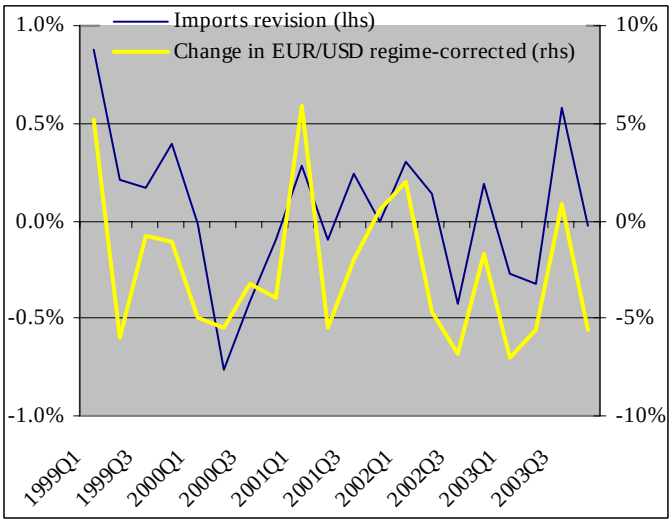
[26] In other words, revisions to nominal data are supposed to be uncorrelated with currency changes in the short run.

For the trade balance, as well as for exports and imports[27], incorporating a regime shift results in a better fit in terms of high frequency fluctuations of the vintage quarters' revisions. However, an additional level shift seems to be unveiled by this transformation. This could be the result of an asymmetry in the elasticity of deflators to currency appreciation or depreciation or, conversely, an asymmetry in models that are used to estimate deflators.



Consider for instance the revisions to the trade balance: in the appreciation[28] regime, a large appreciation seems linked to a low revision whereas a small appreciation is correlated to a large underestimation of the trade balance. In a regime of appreciation, the equations that are used for exports might underestimate export[29] deflators and overestimate exports when the appreciation is strong. The overestimation of imports[30] is symmetric, so there is no impact on the trade balance. On the other hand, equations used for imports underestimate[31] imports when the appreciation stops (probably due to the use of autoregressive terms), which is not the case with exports. As there is in this case no symmetry with exports, the underestimation of imports translates into an overestimation of the trade surplus when the appreciation stops.





---

[27] The contribution of imports to growth is affected by a minus sign, meaning that, where the change in imports is underestimated (negative revision), the contribution of this revision to GDP revision is positive. An underestimation of exports (negative revision) has a negative contribution to the GDP revision. Note that on the graphs above, the scale is reversed for exports.

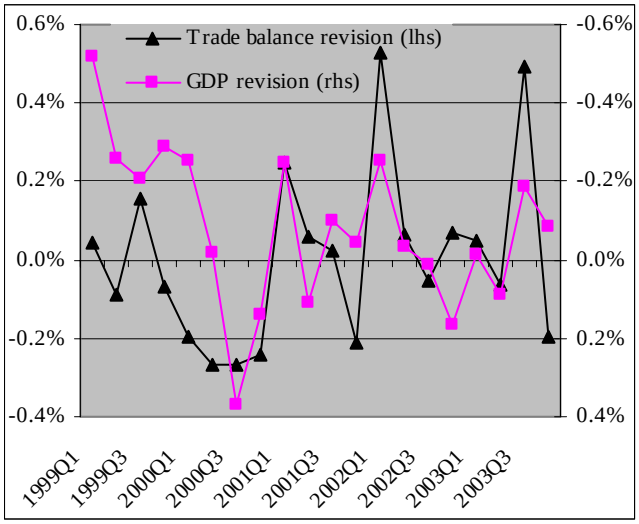
[28] In the depreciation regime, a large depreciation is mirrored by a large trade balance revision (recall that the EUR/USD transformation involves taking the opposite sign of the parity change in a regime of depreciation). The reason is that where imports are overestimated, exports are less overestimated and the resulting trade balance is underestimated given that imports have a negative contribution.

[29] Here, we implicitly assume that the main source of revision would come from deflators, not from exports values. See additional graphs in next page.

[30] As imports have a negative contribution to growth, a positive revision (overestimation) to imports leads to a negative revision in the contribution to growth revision.

[31] A negative revision of imports volume growth translates into a positive contribution of imports to the GDP revision, see previous footnotes.

A further link between trade balance and GDP revisions could be investigated. Strangely enough (see graph opposite), revisions to the trade balance are poorly (negatively) correlated with those to GDP. The only explanation is that there is no direct link between the estimations of the former and the latter. The inventories account adjusts for the difference.



If this analysis also applied to German QNA, the estimated peak in German exports for 2005Q1 might be unbiased against the background of a halt in the euro appreciation in Q1. On the other hand, import volumes might have been underestimated and the trade balance surplus overestimated with little expectable impact on GDP revision. On the other hand, export volumes for the previous quarter (strong appreciation in 2004Q4) might have been overestimated.