Catching a floating treasure: A genuine ex-ante forecasting experiment in real time

Christian Müller[*] and Eva Köberl[†]

12th January 2012

[*]Zurich University of Applied Sciences, Centre for Economic Policy, E-mail: much@zhaw.ch

[†]ETH Zurich, KOF Swiss Economic Institute, E-mail: koeberl@kof.ethz.ch

Abstract

Forecasting real economic activity poses a considerable challenge not only due to hard-to-predict events like the current financial crisis but also due to the fact that targeted variables often undergo significant revisions after their first publication. In this paper we report the results of a truly ex-ante forecasting experiment in real time. It highlights the difficulties of hitting a moving target and proposes some means for coping with the problems involved.

JEL classification: E37, C35

Keywords: forecasting, GDP revisions, survey data, experiment

1 Introduction

When a fisherman spots a treasure box floating in the ocean his luck remains incomplete until he manages to secures the catch. Getting the box on board is not easy at all, however. First, the box is moving with the waves and the wind. Secondly, and maybe more importantly for an inexperienced fisher, the boat itself also budges. The man therefore has to account for both forces before he really turns lucky.

Forecasting real economic activity like the real gross domestic product is akin to the fisherman's challenge. The treasures to be caught are the GDP figures whose ups and downs are caused by economic winds and storms. At the same time, however, the researcher reaches out not from a firm ground but from a basis that is very volatile due to data revisions and updates he or she has to account for.

In fact, a considerable share of the forecasting problem does not arise because of a lack of appropriate statistical tools but is due to a well-known deficiency of the publication process of official GDP data. Literally all countries which provide national account data regularly revise the data published first. Very often revisions occur as late as two or more years after the initial release.

Forecasters thus face multiple problems. They do not only have to find the best suited statistical approach to forecasting, they also have to rise to the challenge of not knowing the most recent past and hence the probably most relevant data. There are at least two prominent ways for addressing this problem. Traditionally, the problem of data revisions has been neglected. However, as Orphanides and van Norden (2005) among others have shown, that ignoring revisions can be very misleading in terms of both, policy conclusions and in terms of scientific empirical work. Therefore, researchers have turned

1

to the use of vintage data in order to put themselves in an ex-ante position. Doing so allows, for example, not only to model the final, revised data, but also to gauge the revision process itself (Graff and Sturm, 2010; Jacobs and van Norden, 2011; Siliverstovs, 2011a). An intended upshot of these rather involved approaches is an improvement of the forecast properties themselves.

As just mentioned, many researchers employ vintage data in order to develop their procedures. Doing so they mimic an ex-ante analysis purposefully turning a blind eye on information they do possess but they do not want to exploit. It is then argued that the method championed by the investigator will be also superior in consecutive real-life applications.

In this paper we want to contribute to the literature by reporting the results of an experiment that takes the ex-ante consideration to an extreme. Instead of simulating the ignorance of available information this experiment is run without actually having this information altogether. The advantage of our approach is that it does exclude the possibility of intentionally or unintentionally being influenced by the actual sequence of past events. For example, a researcher might want to simulate ex-ante forecasting GDP during the last five years or so. As everybody now knows, in 2007/ 2008 the financial crisis kicked in. To be absolutely sure that knowledge of future events does not impact on the forecasting exercise there is only one way: real ex-ante forecasting in real time. It is virtually impossible to truly achieve this goal in a simulated approach.

A further disadvantage of the standard as-if ex-ante posture certainly is the lack of experience when the best performing method is finally implemented. Moreover, due to the fact that the selection of the most potent method very often is a very involved task, repeating this task with every new release pushes up costs even further.

In this paper we thus aim at two things simultaneously. We try to develop a method for forecasting the final release data that is impartial and reliable on the one hand plus efficient on the other.

The remainder of the paper is organised as follows. First, the data release pattern for the model country, that is Switzerland, is described and the historically correct cut-off point for fixing the final release information is determined. Second, the experimental set-up including the benchmark forecasting is reported, and finally, conclusions are drawn on the basis of statistical evidence gathered so far.

2 Experimental design

2.1 Preliminary considerations

While being very appealing from a methodological point of view, the literature on forecasting has little to offer in terms of experimental evidence. In fact, while both areas, experimental economics and economic forecasting are very actively researched fields, the combination of these two is not very popular in the literature. One of the rare exceptions is Ryan Woodard and Fedorovsky (2011) who test their asset pricing models in a very rigorous manner.[1]

It appears pretty natural to assume that many successful forecasting models are partly shaped by the researchers unintentional use of knowledge she or he posses but which would be unavailable in a genuinely ex-ante setting. Other research areas such as medical sciences have long acknowledged these problems and drugs are now regularly required to undergo prospective test studies.

In macroeconometrics, there are at least six frequent confounding factors in as-if ex-ante experiments which are hard to control. These are

- ex-ante model choice,

- the selection of exogenous variables,

- availability of vintage data for the endogenous variable(s),

- availability of vintage data for the exogenous variable(s),

- knowledge of peculiar events during the forecast period,

- the state of the art in econometrics.

---

[1]Gunnar Bårdsen's approach is another:
`http://www.svt.ntnu.no/iso/gunnar.bardsen/nam/evaluation/index.html`.

For example, a researcher might pick a model that does best for the whole sample and is then also used in the simulated forecasting exercise. In a realistic setting the future is unknown and so is the benchmark used for model selection.[2] Add to this observation the notorious publication bias and one might suspect that a significant portion of available research results on forecasting will hardly stand the test of practice.

In practical applications, however, forecasters have a strong interest in reliable scientific guidance for accomplishing their tasks. In fact, economic research institutions all over the world perform true ex-ante forecasting while their guiding principles have been derived from theory and predominantly ex-post analyses. It is this gap between the demands of practitioners and the leading literature that is addressed in this paper.

In our genuine ex-ante forecasting experiment the key asset is to make absolutely sure that the unknown really is unknown. The most plausible way to achieve this goal is by a credible commitment to limit oneself to the use of contemporaneous and available information only.

2.2 The commitment

Our experiment started in the third quarter of 2007. At that time we produced a forecast of year-on-year GDP growth for the next quarter and posted the results on the web.[3] In addition, an email was sent to Swiss and foreign forecasters, journalists and lay people with an interest in the matter. The same procedure has been repeated quarterly ever since. While it is in principle possible to manipulate the content of the web publication ex-post it is impossible to alter emails once they have reached their addressees. This set-up is

---

[2]Another example regards the choice of the exogenous variables whose properties with respect to availability and robustness are often known ex-post only.

[3]The address is `http://www.cmueller.ch/8897a3`.

thus strongly binding in terms of making it virtually impossible to use future information for forecasting that is just not available at the time of forecasting.

2.3 The benchmark, data releases and data revisions

Our experiment concerns forecasting quarterly Swiss GDP figures. The (annualised) quarterly growth rates and the year-on-year growth rates are the most widely recognised macroeconomic information which regularly capture the attention of the public. In the following, we will focus on final quarterly year-on-year growth rates of the Swiss real GDP. By looking at the change over the previous year we avoid discussing the choice of the seasonal adjustment.

The Swiss national account data is produced by the Swiss Federal Office of Statistics (SFSO). Every year it publishes annual data on GDP and its components. In the second, sometimes in the third quarter of each year, revisions are announced. The annual data is broken down to quarterly figures by another governmental agency, the Swiss State Secretariat of Economic Affairs (SECO) which also publishes timely quarterly figures for the ongoing year. The latter figures are being made available at the beginning of the third month in each quarter for the past quarter. The SECO (quasi-)final quarterly estimates are going to be used as our benchmark.

Since the official SFSO data is available with a substantial time lag only, all initial SECO publications have to be qualified as forecasts, or rather nowcasts of the true numbers.[4] Moreover, due to data revisions by the SFSO also the SECO updates its quarterly figure many times after the initial publication.

The following graph sketches the publication process. The third month of the first quarter is March. At the end of February the nowcast of the reported

---

[4]We are going to use the term nowcasting for estimating GDP growth in the previous quarter in spite of the more conventional term backcasting. Technically speaking, there is no difference to forecasting, however.

experiment is released. This nowcast is made in about the fourth quarter of the past year and it is denoted CE_1. Within one to two weeks time the SECO publishes its first estimates concerning the same period ($SECO_1$). The same pattern is repeated every three month.

Figure 1: Data release and nowcasting process
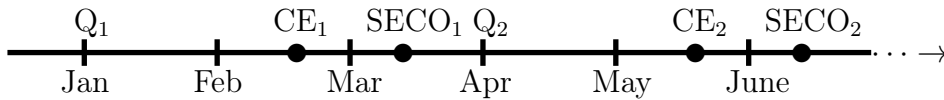
Every time SECO releases its data, the publicised figures of the past may change. The first release according to the national accounting standard ESA95 was in 2006, second quarter for the period 1999q1 through 2006q1. The switch to the new standard represents a structural break in the data series. Therefore, it seems most appropriate to restrict the data sample accordingly.

The figure 2 below depicts the information available at the start of the experiment and contrasts it to the GDP data of the latest release. The difference between these two information sets is obvious and sometimes very impressive. For example, while according to the former release 1999q2 saw a negative y-o-y growth rate, the most recent estimate reports a positive growth.

We further draw a distinction between the data that can currently be regarded the final release data and the data pool that is probably still going to be revised. At the time of writing the 2009q3 SECO release is the one with the longest unchanged data series so far available. Of all data published in 2009q3 the (later) unchanged portion spans 1999q1 through 2005q4. The GDP figures for 2006q1 and later have meanwhile been changed at least one more time. Thus, in order to evaluate the significance of the revisions we distinguish between the history defined by the final release data and all data available.

7

Figure 2: Information at the start of the experiment and at the time of writing
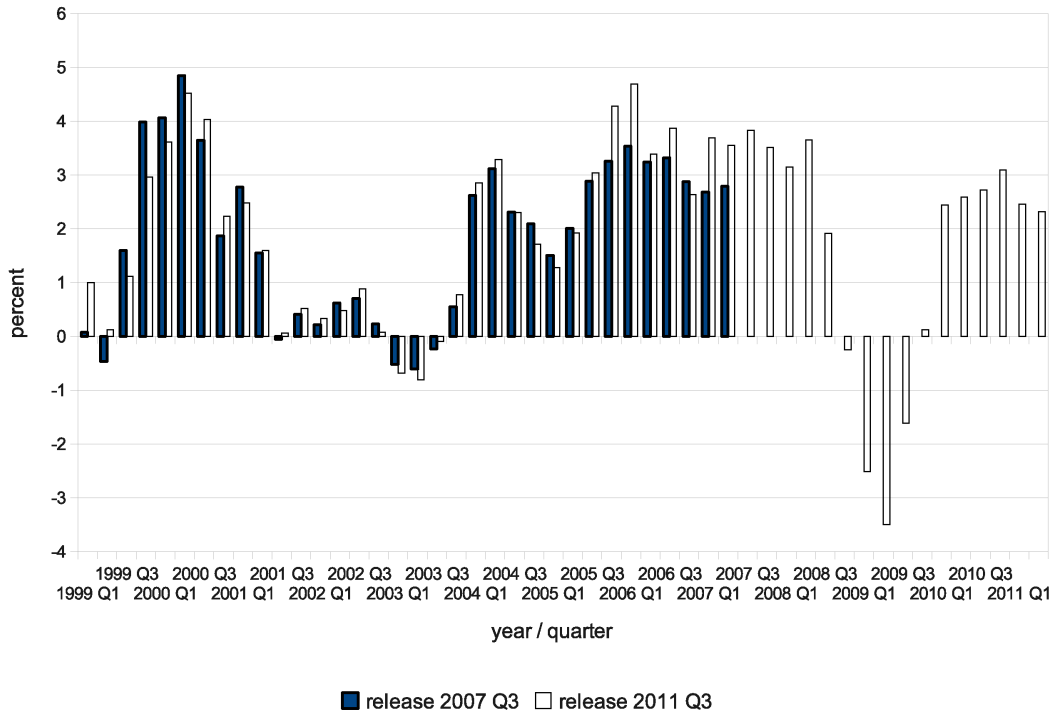

Quarterly year-on-year growth rate of Swiss real GDP available at the beginning of the experiment (solid bars) and at the time of writing 2011q3.
Source: SECO

The figure 3 reports the average size of the revisions of the year-on-year growth rate of real GDP as a function of the time elapsed since the initial publication. It shows that on average the growth rate has been understated by approximately 0.9 percentage points. In the consecutive quarters the error has been reduced monotonically to by and large. The persistence of the initial error can partly be attributed to the choice of the year-on-year growth rate, however. For the first two quarters after an initial release the maximum difference to the final value has always been negative.

Overall, the range within which revisions have been observed appears considerably wide. The standard deviation of revisions is hence pretty large too.

8

It peaks after three quarters at 0.74 and declines down to 0.18 at the last revision. After four and a half years no further revision occurred.

Figure 3: Mean revision of SECO releases (2006q2 – 2011q3)
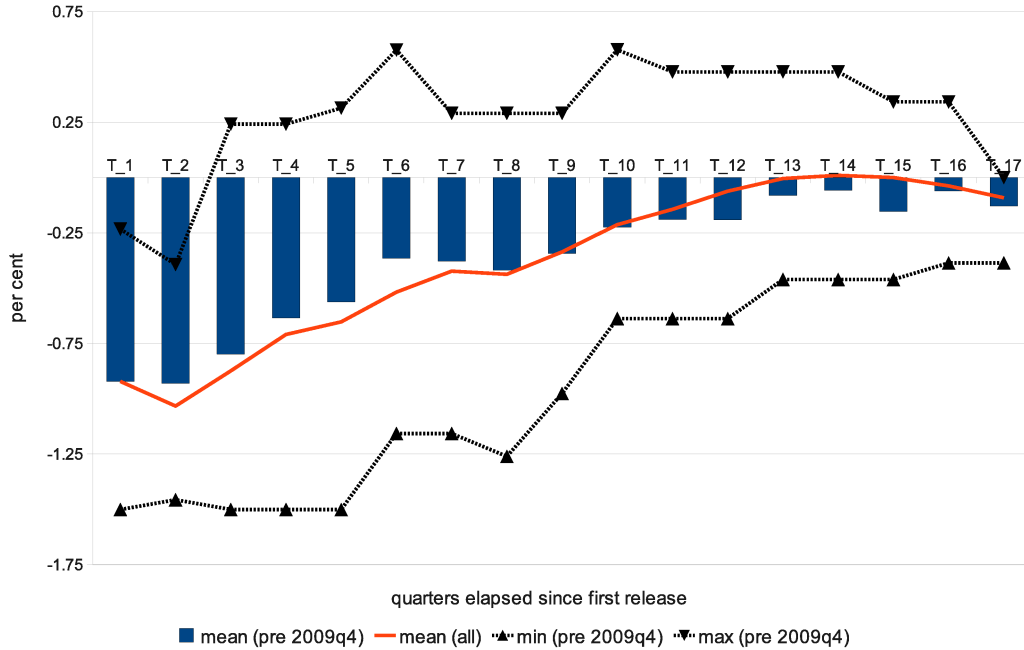
Mean, minimum and maximum size of revisions of year-on-year GDP growth rates by SECO. Comparison of initial releases to finalised data denoted by "pre 2009q4" (2006q2 through 2009q3), comparison to latest release data denoted "all" (2006q2 through 2011q3).
NOTE: The number of observations for the revision estimates is five (pre 2009q4) and varying for the full sample between five and 18.

The key message of figure 3 thus is that SECO's initial Swiss GDP releases have a "pessimist's" bias[5] and adjustments following the first publication are made more than four years afterwards.

––––––––––––––––––––––––––––––

[5]Siliverstovs (2011a) uses a unobserved components approach to correct the systematic bias.

2.4 Model and variable selection

At the time of the start of the experiment the information provided by figure 3 was not yet available. Therefore, a number of choices had to be made with limited information. The first question addressed the use of SECO data. In the absence of superior knowledge, we decided to treat the latest releases of SECO as the most informative source of information about the figures that one day would be regarded the final word on quarterly GDP. Therefore, while we defined our objective to be nowcasting the final data, we used preliminary data as an input for achieving this goal.

Second, in order to mimic traditional pseudo-ex-ante forecasting experiments, we decided to derive a best fitting model which then ought to be used throughout the forecasting period. That model was specified in a general-to-specific approach and it is replicated below.

The next decision had to be made with respect to the use of exogenous data. Using the treasure box analogy, a landing stage greatly simplifies the task of catching the box. Therefore, employing exogenous data which is not subject to revisions helps stabilising the ground from which we reach out. Since we expected SECO to frequently update the GDP data a simple autoregressive process would not serve this goal. Similarly, exogenous variables which are also undergoing repeated revision processes and non-linear seasonal adjustment procedures very likely introduce additional imponderability. Therefore, we opted for explanatory variables which are (almost) not subject to revisions. The variables of choice have therefore been survey data which serves this goal best in principle and applied linear seasonal adjustment.[6]

In particular, we employed the so-called surprise index approach (Müller-Kademann and Köberl, 2008). This approach offers eight potential exogenous

---

[6]Unfortunately, it turned out that this data was not as robust as expected.

variables. This set was further reduced to two by means of a simple correlation analysis using SECO's quarterly GDP data as benchmarks. At the start of the experiment the effective sample size appeared rather small when allowing for up to four lags of the dependent variable. Therefore, both candidate variables were plugged into the general encompassing model separately and the better fitting model was chosen for forecasting (see equation (2)).

After three rounds, that is in the third quarter of 2008, it turned out that the set of exogenous variables had to be revised. On this occasion it was also decided to change the forecasting strategy. Instead of fixing the originally chosen model for the entire forecasting period we thought it more realistic to adapt the optimal model choice to the updated data in each period. In addition, at the same time the general model now included both candidate variables jointly. There has been no further revision of the exogenous data and the strategy has likewise not been changed since. A brief overview of the chronology of the experiment is provided in the appendix.

In order to restrict the computational efforts in every quarter to a reasonable level, we decided to employ standard model selection procedures for picking the best suited model. This model was derived in a general-to-specific approach with the software PcGets (Krolzig and Hendry, 2001; Owen, 2003). This choice implicitly restricted the model class to linear time series models.

Letting y_t denote the logarithm of the Swiss real GDP and $\triangle_i, i = 1, 2, \ldots$ being the difference operator with $\triangle_i y_t \equiv y_t - y_{t-i}$ we can write the general model in the following form.

$$Y_t \;\; = \;\; D + \sum_{i=1}^{4} A_i Y_{t-i} + \sum_{i=0}^{4} B_i X_{t-1} + U_t \tag{1}$$

where

$$Y_t = \begin{bmatrix} \triangle y_t \\ \triangle_4 y_t \end{bmatrix}, X_t = \begin{bmatrix} em_t \\ pm_t \end{bmatrix}, U_t = \begin{bmatrix} u_t \\ 0 \end{bmatrix} \text{ and}$$

$$D = \begin{bmatrix} c & s_1 & s_2 & s_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, A_i = \begin{bmatrix} a_{i,1} & 0 \\ 1 & 0 \end{bmatrix}, B_i = \begin{bmatrix} b_{i,1} & b_{i,2} \\ 0 & 0 \end{bmatrix}.$$

The exogenous variables are coded em_t and pm_t and represent the shares of qualitatively different negative surprises firms have been suffering in the last quarter.

The model has two equations of which only the first ($\triangle y_t$) has unknown coefficients which are to be estimated. The second equation is a simple definition equation we use for handy derivation of year-on-year growth rates from the estimated quarter-on-quarter rates of change. Another simplifying assumption is made with respect to the innovation term u_t: $u_t \sim (0, \sigma^2), \forall t$. Finally, in order to capture seasonal effects we use linear seasonal dummies and denote them $s_{t-i}, i = 0, 1, 2$.

2.5 Estimation results

In the very first run of the procedure the effective sample covered 2000 second quarter to 2006 first quarter. The sample size had been restricted in order to leave space for a pseudo ex-ante forecasting experiment. The resulting estimates of the first equation of the general model have been published in Müller and Köberl (2009) and are repeated here (absolute t-values in parentheses

below the coefficient estimates):

$$\triangle y_t \;=\; -\,1.03\,pm_{t-1} - \underset{(14.1)}{3.66}\,s_{1,t} - \underset{(6.12)}{1.87}\,s_{3,t} + \underset{(11.0)}{2.08} \tag{2}$$
$$\underset{(2.08)}{}$$

$$\hat{\sigma} \;=\; 0.541$$

$$\bar{R}^2 \;=\; 0.89$$

$$T \;=\; 24 \text{ (effective sample size)}$$

The first update was released in 2007, third quarter using all available data points (2000q2 – 2007q2). In each consecutive estimation period the same model was estimated and the corresponding standard error of this regression was published.

Starting with the revised surprise index in 2008q3 the optimal model was allowed to vary with each consecutive update.[7] In the table 1 below we report the relative frequency of the variables chosen plus their mean, median and standard deviations (if applicable).

---

[7]The intercept term has always been fixed and maintained.

Table 1: Relative frequencies of variable selection and basic statistics

| | frequency | | coefficient statistics | | |
variable	absolut	relative (%)	mean	median	s.d.
constant	13	100.00	4.2	4.7	1.2
s_t	13	100.00	-3.7	-3.7	0.7
s_{t-1}	6	46.15	-0.6	-0.8	0.3
s_{t-2}	13	100.00	-2.3	-2.1	0.7
y_{t-1}	0	0.00	na	na	na
y_{t-2}	0	0.00	na	na	na
y_{t-3}	8	61.54	-0.3	-0.3	0.1
y_{t-4}	0	0.00	na	na	na
pm_t	0	0.00	na	na	na
pm_{t-1}	8	61.54	21.5	21.5	3.1
pm_{t-2}	4	30.77	21.0	22.2	11.1
pm_{t-3}	6	46.15	-18.2	-18.2	2.5
pm_{t-4}	0	0.00	na	na	na
em_t	13	100.00	10.3	10.5	1.1
em_{t-1}	2	15.38	5.4	5.4	0.1
em_{t-2}	10	76.92	-6.7	-7.1	1.6
em_{t-3}	1	7.69	4.4	4.4	na
em_{t-4}	5	38.46	5.7	5.7	0.2

Frequency of coefficient selection by the automated model
selection procedure. Mean, median and standard deviation
of the selected coefficients.
NOTE: The models considered are those selected between
2008q3 and 2011q3 to ensure data consistency.

The experience hence shows that the most frequently chosen variable next to the deterministic terms is em_t, a variable of the surprise index family. Not surprisingly, changes in the model occurred most likely at the time of new SFSO data releases (not reported).

2.6   Forecast evaluation

2.6.1   The benchmark

The overriding aim of this project is to derive a reliable method for gauging the economy's most recent performance. Therefore, we next turn to the evaluation of the output obtained in the experiment. First of all, the benchmark to which the forecasts are to be compared to has to be defined. It turns out that this task is not easy at all. The main reason is the lack of sufficiently finalised data. As previously mentioned the longest unrevised data series was published in 2009q3 for the period 1999q1 to 2005q4. This implies that as yet we do not even know the final GDP growth rate at the time of the start of the experiment. Therefore, we are going to consider a reasonably finalised data set instead.

Figure 4: Provisional, reasonably final, final GDP data and ex-ante nowcasts
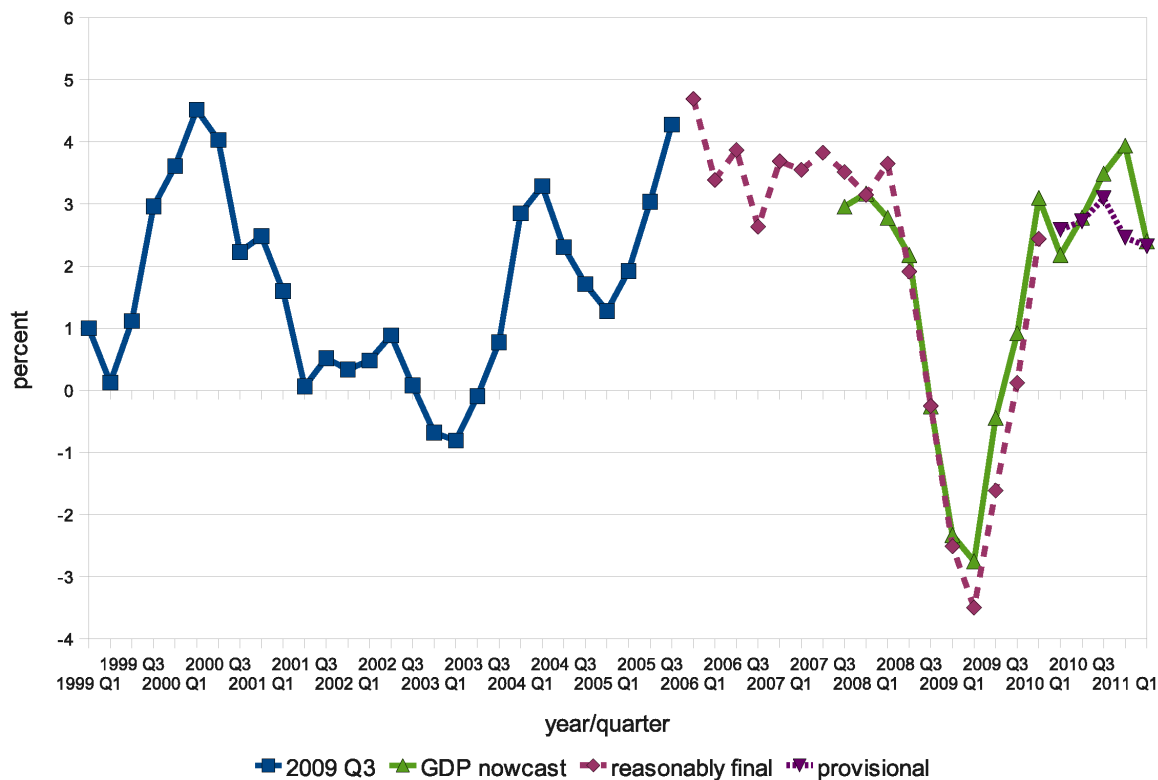
Final data released in 2009q3 for the period until 2005q4 ("2009q3"), observations for which only minor revisions can be expected ("reasonably final"), data points for which significant revisions are expected between 2010q2 – 2011q2 ("provisional"), and the GDP nowcasts of the experiment 2007q3 – 2011q2 ("GDP nowcast").

Albeit somewhat arbitrarily, we chose the period 2007q4 to 2010q1 as the evaluation period. Figure 3 lends some support to the cut-off point 2010q1 because it shows that on average more than half of the initial revision error disappears five quarters after the first release. In figure 4 we summarise the key features of the evaluation period. The benchmark of our forecast is thus the last available release of GDP data covering the period 2007q4 through 2010q1 which, however, still is a biased benchmark as figure 3 demonstrates.

Figure 5 depicts the time series of the forecasts made by SECO (first re-

Figure 5: Forecasting benchmarks and competing one-step forecasts for y-o-y Swiss real GDP growth
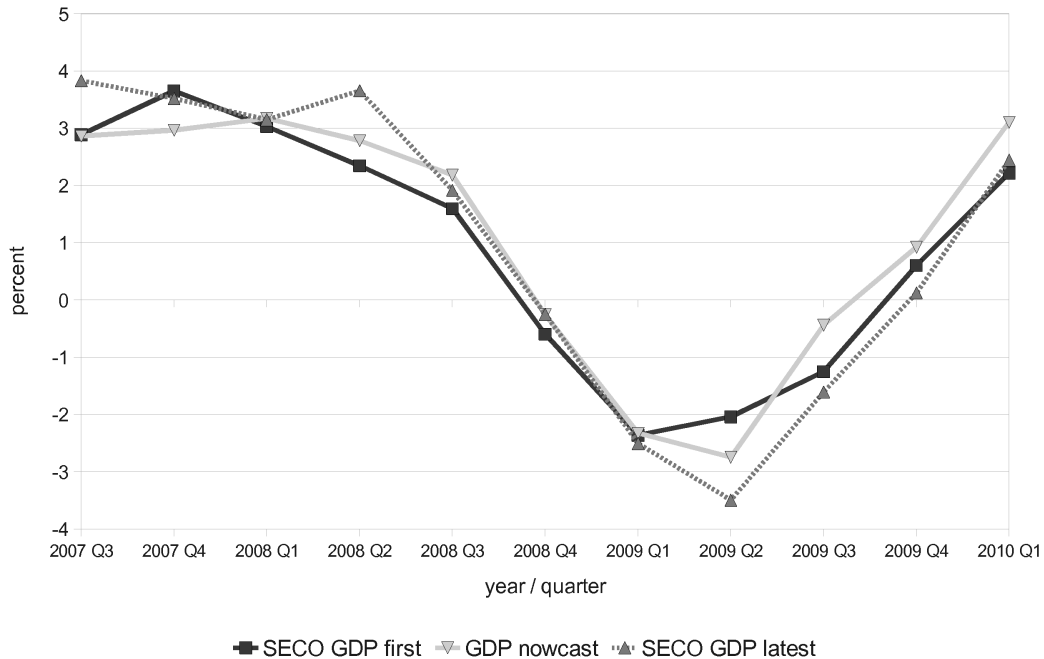
lease) and the modelling approach used in the experiment (labelled GDP nowcast in the figure) for the evaluation period 2007q4 through 2010q1. These two forecast series are going to be compared to the benchmark series given by the latest release.

2.6.2 An almost infeasible horse race

Genuine ex-ante forecast experiments are to be distinguished from pseudo-exante analyses due to several unique features. The following forecast evaluations will illustrate this point. Therefore, next to the forecasting procedure due to SECO (first release forecasts) and this paper's nowcast approach, four more models are used in a horse race.

The first alternative model is a variant of the surprise index model. This alternative simply scales the strongly negative surprises by a linear regression

to fit the historical data best. Doing so provides a new series which directly serves as a quarterly year-on-year GDP estimate. This model is structurally similar to the nowcast approach except that it does not account for the fact that in 2007 the surprise index series had to be adjusted. The differences in the forecasting performance can therefore be attributed to this fact and to the lower sophistication of the whole approach. This alternative will be labelled "*pm* scaled".

Three more alternatives are due to Siliverstovs (2011b) and Siliverstovs (2011a) who has also developed econometric models for forecasting Swiss GDP. These elaborate models can be used for pseudo ex-ante forecasting exercises with almost perfect vintage data.[8]

Siliverstovs (2011b) applies linear autoregressive distributed lag models (ARDL) which are structurally similar to this paper's nowcast model. The key exogenous variable is the so-called KOF baro, a survey data based multi-sectoral economic activity index. Instead of selecting the best suited model in a general-to specific spirit, Siliverstovs (2011b) employs Bayesian model averaging (BMA) techniques. The forecasts of this approach can be considered as ex-ante forecasts in real time. There is no obvious difference to a genuine ex-ante forecasting experiment.

Siliverstovs (2011a) differs in that respect because it utilises a mixed frequencies unobserved component model variant due to Mariano and Murasawa (2003) with modifications suggested by Camacho and Perez-Quiros (Camacho and Perez-Quiros, 2010; Camacho and Quiros, 2011) for relating the survey data to the latent factor. Obviously, Siliverstovs uses an econometric technique which was not yet available at the beginning of the experiment.

There are two versions of Siliverstovs's (2011b) forecasts. One is a "direct"

---

[8]We are grateful to the author for providing his forecasts.

Table 2: Forecast errors for $N = 11$ one-step forecasts (2007q4 through 2010q1)

model	MAFE	MAFE ratio	RMSFE	ratio
SECO first	.530	1.000	.700	1.000
GDP nowcast	.568	1.072	.682	.973
pm scaled	.646	1.219	.783	1.117
Baro-ARDL	.647	1.222	.831	1.186
Baro direct	.515	.971	.631	.902
Baro iterative	.518	.978	.700	1.000

NOTE: MAFE and RMSFE abbreviate mean absolute forecast error and root mean squared forecast error respectively, MAFE ratio (RMSFE ratio) provides the absolute forecast errors (RMSFE) in relation to the SECO benchmark. Further: *pm* scaled for the indicator series *pm*'s mean and range being scaled to the year-on-year growth rate using the 2000q2 – 2005q4 data.

forecast which is distinct from its "indirect" counterpart because intermediate results do not enter the final forecast outcome. There are hence three models due to Siliverstovs. For the sake of brevity we will call them "Baro-ARDL" (Siliverstovs, 2011b), "Baro direct", and "Baro iterative" (Siliverstovs, 2011b) respectively. The latter two forecasts do not meet all the criteria of a genuine ex-ante experiment because some ex-post knowledge helped shaping the results (Siliverstovs, 2011b, footnote 2).

All in all there are six forecast approaches prepared for the following horse race. Of these six only two (Siliverstovs's (2011a) latent factor models) could not have been used for real time ex-ante forecasting for obvious reasons. The other two approaches due to Siliverstovs and the so-called indicator scaled method may suffer from hidden hindsight while SECO's estimates and this paper's nowcasts are genuine ex-ante forecasts.

Tables 2 summarises the results. Comparing mean absolute forecast errors shows that the SECO estimates are beaten only by the pseudo ex-ante forecasts due to Siliverstovs (2011b). Siliverstovs's (2011b) models are almost thirty

Table 3: Forecast evaluation for $N = 11$ one-step forecasts

model	test statistics (p-values)		
	sign change	D-M MAFE	D-M RMSFE
SECO first	4 (.753)	n.a.	n.a.
GDP nowcast	5 (1.0)	$-.29$ (.78)	.12 (.90)
pm scaled	2 (.109)	$-.53$ (.60)	$-.35$ (.73)
Baro-ARDL	2 (.109)	$-.88$ (.40)	-1.06 (.31)
Baro direct	4 (.753)	.10 (.91)	.44 (.66)
Baro iterative	3 (.344)	.07 (.94)	-0.0 (.99)

The sign test (binomially distributed) checks whether or not the forecast error changes the sign in a systematic fashion. Under the null hypothesis the sign of the forecast error has a constant 0.5 probability of changing from period to period. D-M MAFE and D-M RMSFE list the test statistics for the Diebold-Mariano test for differences in mean absolute forecast error (MAFE) and root mean squared forecast error (RMSFE) between the actual model and SECO's benchmark (two-sided test with correction for heterogeneity) together with the p-values. Further: *pm* scaled is short for the indicator series *pm*'s mean and range being scaled to the year-on-year growth rate using the 2000q2 – 2005q4 data.

percent more successful than SECO's first releases. This paper's approach delivers forecasts which seven percent off the benchmark when looking at the ratio of the mean absolute forecasts by SECO and the GDP nowcast.

In terms of reliability the picture is a bit different. While the same latent factor model by Siliverstovs (2011b) still does best in relative terms, the root mean squared error of the survey data based approach of this paper does better in comparison to SECO's estimates. This lends support to the choice of survey data in combination with a simple linear model structure.

Overall the differences between the model appear rather small in absolute value. Therefore, formal statistical tests have been used to try to discriminate between the various approaches. Table 3 on page 20 has the details. There are two types of tests. We use the binomial distribution for checking whether or not there is a systematic component in the forecast error, the so-called sign test. If forecast errors were purely random, they should irregularly change their sign with a 50 percent chance of a sign change in each period. Except for the *pm* scaled approach none of the models comes close to rejecting the null. In other words, all approaches do equally well on that account.

The same holds true for a comparison of the forecast properties of the alternatives to the SECO benchmark. When comparing the SECO results to their competitors, the Diebold-Mariano-statistics (Diebold and Mariano, 1995) clearly indicate that the differences between the SECO forecasts and the forecasts of the other models are the result of chance.[9]

Summarising the test results, those forecasts which have been made on the basis of a pseudo ex-ante forecasting experiment do best in absolute terms. Therefore, the additional skills and possibly the hindsight which have become

---

[9]There is no significant difference between the best performing model and all other models either. The authors would like to thank Boriss Siliverstovs for sharing his Ox code for calculating the Diebold-Mariano statistics.

available after the start of the experiment seem to bear fruit. Hindsight alone, however, does not seem to suffice as the performance of the BMA-ARDL approach suggests. This paper's genuine ex-ante forecasting approach is neither superior nor inferior to the SECO benchmark on all accounts considered. It is doing slightly worse in terms of the mean absolute forecast error but has the edge when the forecast error variances are compared.

Owed to the fact that all differences do not matter statistically, we can conclude that the adopted forecasting strategy of the nowcasting experiment appears to be quite successful both in terms of formal statistical assessment as well in terms of efficiency. Considering the lean production process the forecast performance seems to be accurate and reliable. This reliability can certainly be attributed to the choice of timely available and robust exogenous variables and the simple linear model structure. Taken together, the experiment's outcomes lend significant support to the chosen approach.

3 Summary and conclusions

This paper describes a genuine ex-ante forecasting experiment for the quarterly year-on-year growth rate of Swiss real GDP. Despite its long duration which – in the context of forecasting and controlled economic experiments – is as yet unmatched in the economics literature to the best of our knowledge, the estimation outcome could only be compared to quasi-final data.

A simple horse race assessment has shown that a parsimonious, linear time series modelling approach delivered highly satisfactory results. In particular, the growth rates calculated in a real-time fashion were as good as the official first data releases in terms of average forecast accuracy and on a squared loss function metric. The official data generation requires considerably more resources and is not as timely as the estimates from the time series model.

The key features of this modelling approach are a LSE-type automated general-to-specific model selection procedure (Krolzig and Hendry, 2001) and the use of survey data based surprise index due to Müller and Köberl (2008).

The outcome of this experiment implies that GDP nowcasting with a lean time series model using the surprise index method delivers reliable information which is moreover available more timely than the official data. The final word on this conclusion can only be spoken, however, when the truly final data is going to be available in about four years time.

A   Appendix

A.1   Chronology of the experiment

The chronology of the experiment can be summarised as follows.

date	event	remark
2007q4	start of experiment	variable pre-selection, choice of first generation surprise index, fixing the model to (2), pseudo-ex-ante forecasting.
2007q4	first update (Nov 30)	full sample estimation.
2008q3	revision of exogenous variables	general model expanded to its final version (see eq. (1) on page 11).
2009q3	publication of longest times series without consecutive revisions	the "finalised" data spans 1999q1 – 2005q4
2011q3	intermediate report	last results entering this paper

References

Camacho, M. and Perez-Quiros, G. (2010). Introducing the euro-sting: Short-term indicator of euro area growth, *Journal of Applied Econometrics* **25**(4): 663–694.

Camacho, M. and Quiros, G. P. (2011). Spain-sting: Spain short-term indicator of growth, *Manchester School* **79**(s1): 594–616.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business & Economic Statistics* **13**(3): 253–63.

Graff, M. and Sturm, J.-E. (2010). The information content of capacity utilization rates for output gap estimates, *KOF Working papers 10-269*, KOF Swiss Economic Institute, ETH Zurich.

Jacobs, J. P. and van Norden, S. (2011). Modeling data revisions: Measurement error and dynamics of "true" values, *Journal of Econometrics* **161**(2): 101–109.
http://ideas.repec.org/a/eee/econom/v161y2011i2p101-109.html

Krolzig, H.-M. and Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures, *Journal of Economic Dynamics and Control* **25**(6-7): 831–866.

Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series, *Journal of Applied Econometrics* **18**(4): 427–443.

Müller, C. and Köberl, E. (2008). Business cycle measurement with semantic filtering: A micro data approach, *Working Paper 212*, KOF Swiss Economic Institute, ETH Zürich.

Müller, C. and Köberl, E. (2009). Business cycle measurement: A semantic identification using firm level data, *in* J. R. Oliver Holtemöller and M. H. Stierle (eds), *Characteristics of Business Cycles: Have they Changed?*, Institut für Wirtschaftsforschung Halle, Halle, pp. 0–0.

Müller-Kademann, C. and Köberl, E. M. (2008). Business Cycle Measurement with Semantic Filtering: A Micro Data Approach, *SSRN eLibrary* .

Orphanides, A. and van Norden, S. (2005). The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time, *Journal of Money, Credit and Banking* **37**(3): 583–601.

Owen, D. P. (2003). General-to-specific modelling using pcgets, *Journal of Economic Suryeys* **17**(4): 609 – 628.

Ryan Woodard, D. S. and Fedorovsky, M. (2011). The financial bubble experiment: Advanced diagnostics and forecasts of bubble terminations, volume iii.
http://arxiv.org/abs/1011.2882v2

Siliverstovs, B. (2011a). Are GDP Revisions Predictable? Evidence for Switzerland, *KOF Working papers 11-281*, KOF Swiss Economic Institute, ETH Zürich.

Siliverstovs, B. (2011b). The real-time predictive content of the KOF economic barometer, *Swiss Journal of Economics and Statistics* **147**(III): 353–375.