Analysing the revisions of selected Finnish Macroeconomic Series

Samu Kurri *
Heikki Hella †

1st Draft. Comments are highly welcome.

June 2005

The views expressed are to those of the authors and do not necessarily reflect the views of the Bank of Finland.

* Samu Kurri, Economist / Project Researcher, Monetary Policy and Research Department, Bank of Finland, P.O. Box 101, FIN-00101 Helsinki, Finland. Email: samu.kurri@bof.fi
† Heikki Hella, Economist, Financial Markets and Statistics Department, Bank of Finland, P.O. Box 101, FIN-00101 Helsinki, Finland. Email: heikki.hella@bof.fi

Analysing the revisions of selected Finnish Macroeconomic Series

Samu Kurri and Heikki Hella

In this paper we analyse selected variables from the quarterly Finnish National Accounts (NA) and the current account surplus/deficit series. The data set includes two vintages, the first preliminary and the final figures. The NA data consists of year-on-year growth rates and the Current Account data of deficit/surplus series. The revision is defined as the difference between the final and first preliminary figures. As a first step we estimate the basic distribution statistics of the revision series. The main goal is to study the bias and efficiency of the revisions both by ARMA and structural time series models (Stamp). As we find systematic deviations from a white noise hypothesis, we also forecast biases. The forecasts seem to reduce the gap between preliminary and final GDP growth figures.

Key words: Real-time series, revisions, bias, efficiency, ARMA model, structural time series model, rank transformation

1. Introduction[1]

Official statistics are the main tool for monitoring the economic situation. They are of interest both to policy makers (with regard to monetary and fiscal policy) and to private economic agents. For example, if the GDP growth figures are biased downwards (upwards), a central bank which follows the Taylor rule as a reaction function would accordingly commit itself to too accommodative (restrictive) monetary policy, which could lead to a sub-optimal outcome. The same might be true for private agents: if for example the government deficit is reckoned to be larger than it really is, they might save more and consume less than they would if they had the correct figures. From a research point of view it could be the case that a policy which seems to have been sub-optimal, could turn out to have been optimal in the light of the real-time data.

Making official statistics is not an easy task, and it has become even more complicated as the number of multinational companies has increased and tradable goods and services have become more complex. One specific important issue is change of quality; to measure real growth in GDP (for instance) deflator changes have to be the quality adjusted - and that is not easy. For example, how does one go about measuring how much the quality of a cell phone has changed over the last five years?

The literature on real-time data analysis is quite comprehensive and is growing rapidly. The research (incl. time series and econometric modelling) on the many different data vintages available is focusing on the framework for policy-making and for improving macroeconomic forecasts. The statistical study of data revisions is an essential part of this empirical research. Also some international comparisons have now been published. A remarkable body of activities has been devoted to investigate the efficiency of data production. From the recent studies those of particular note are Croushore and Stark (2001), Egginton et al (2002) and Öller and Hansson (2004).

Savela and Forsman (2004) have studied the revisions of Finnish GDP growth figures with annual data from 1979-2003. Their results indicate, that on average, the preliminary GDP growth rate underestimates the final rate by 0.6 per cent. Öller and Hansson (2004) focused on the revisions of quarterly Swedish National Accounts data in 1980-1998. Their data set was quite detailed, including GDP, private and public consumption, investment, inventories, exports and imports of goods and exports and imports of services. According to their results the preliminary figures also generally underestimate annual growth rates. Neither Savela and Forsman nor Öller and Hansson studied the forecastability in explicit way – they did not try to estimate any time series or other model for the revision of the data.

The outline of this paper is as follows: in Section 2 we present descriptive statistics and graphs of our revision variables. In Section 3 we fit an ARMA and structural time series model (Stamp) for bias based on the original revision series and its rank transformation. In Section 4 a short

[1] We would like to thank Tarja Yrjöla for research assistance and data collection. We would also like to thank Patrick Crowley (Bank of Finland and Texas A&M University - Corpus Christi), Matti Virén (Bank of Finland and University of Turku) and Jouko Vilmunen (Bank of Finland) for discussions on methodology and comments on this initial draft.

forecasting exercise for GDP revision series is shown. A similar exercise for the Current Account revision series is presented in Appendix 4. In Chapter 5 we sum up the results.

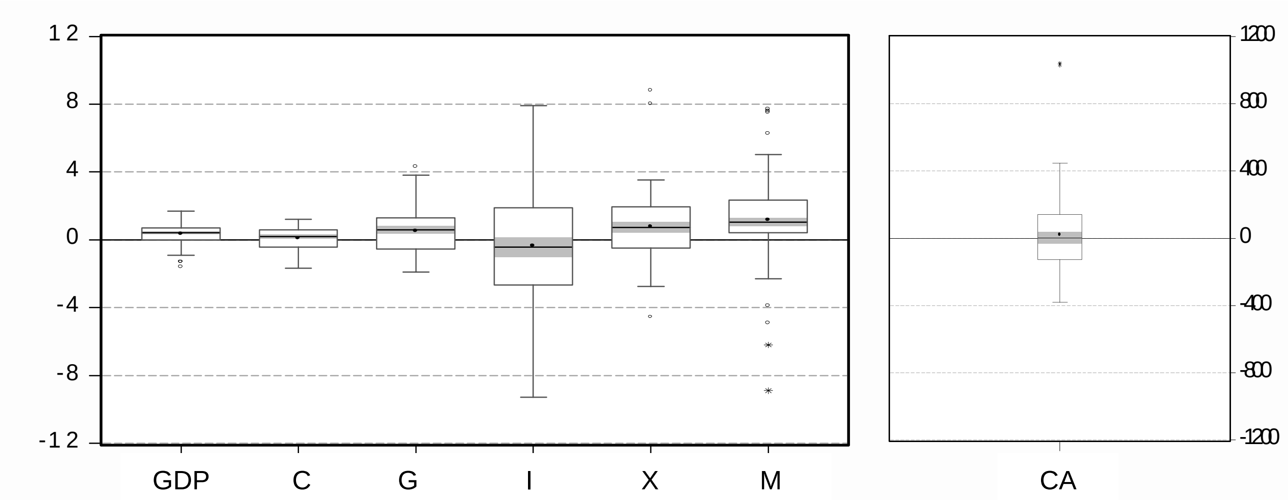
2. Data Description

In this study we analyse how much the main components of the National Accounts (GDP, private consumption, government consumption, investment, exports and imports) and the current account have been revised in Finland between the first preliminary figure and the final one. The definition of the final figure used in this study is the same as Statistics Finland uses - meaning the figure published 2 – 2½ years after the initial one. The revision is defined as the change between the final figure and the preliminary one.

The data set for the National Accounts consists of the quarterly year-on-year percentage changes in 1987Q1-2002Q4. There are several reasons to focus to the annual percentage changes: First, by comparing annual changes we can avoid the problems that are related to seasonality: the changes in the seasonal pattern and especially changes in the methods used during the period will not show up in year-on-year data. Second, we can avoid problems related to changes in the base years; in other words changes in the deflators. Third, we should also be able to avoid the problems related to the different definitions in the National Accounts compilation process (eg. the change from SNA68 to ESA95). The data for the current account revisions consists of nominal figures expressed in euros. All the original data series are seasonally unadjusted and are taken directly from the statistical publications.

The null hypothesis in this study is that the revisions are white noise and thus do not include any information that could be used for the prediction of the future revisions. The very first impression of the data is given via Box-Whisker plots (Figure 1). Due the different scales between the national account series and the current account, the plots are shown in two Figures. In the left part of the figure are the national account series (percentage scale) and in the right part the current account (EUR million) series.

Figure 1: Box-Whisker plots of the revision series, %-unit, bill. EUR



In Figure 1 the box plot represents the first and third quartiles if the data (the interquartile range, IQR), middle 50 percent of the revisions). The dot inside the box is the mean, the black line the median and the shadowed areas around it are the confidence intervals for the median. The inner fences ("Staples") represent the observations which are within the first quartile minus 1.5*IQR and the third quartile plus 1.5*IQR. The plots outside the staples are defined as far outliers. As points they are inside ±3.0 * IQR and as star outside that boundary.

None of the National Account series has zero either as a mean or as a median. Only in the case of investments does the confidence interval of the revisions comprise the zero region. Except for investment the preliminary National Account series seems to be downward biased. The revision scale of the National Account components seems to obey the law of large numbers: the largest component has the smallest revision distribution. The very same result was also observed in the Swedish National Account series (Öller and Hansson, 2004, p. 368). The boxes (the interquartile range, IQR), middle 50 percent of the revisions) for GDP and private consumption, as well as the staples, are clearly the smallest ones. Classified by the IQRs, government consumption, exports and imports form the second group of the National Accounts series, while revisions in the investment series have clearly the largest IQR. The same ordering fits roughly also to the inner fences. For government consumption the largest revisions inside the staples are upwards located, while for private consumption and exports the opposite is the case. Also judged by the outer fences, the revisions in investment are clearly the largest, although for imports there seems to be a considerable amount of outliers which have as large values as the investment outer fences.

The revisions made for the current account seems to be surprisingly well behaved. Despite the fact that the mean is slightly upward biased, the median and its confidence interval are located very close to zero. Also the outliers are emphasised only slightly upwards.

The same information is represented in more formal way in the Table 1, where we present the classic distributional statistics: mean, std. deviation, coefficient of variation, skewness, kurtosis, max and min. For the robust location statistic we use median, MAD (median of absolute median deviations) for robust scale estimator and MAD/median as a rough, robust variation measure. The common Jarque-Bera test is carried out to test the normality of the variables.

Table 1: Descriptive Statistics of the National Account changes between first and last vintages

	Mean	Median	Max	Min	Std dev	MAD	Skew-ness	Kurto-sis	Jarque-Bera
GDP	0.332	0.422	1.695	-1.600	0.630	0.478	-0.819	4.134	10.59***
C	0.075	0.200	1.200	-1.672	0.718	0.741	-0.652	2.820	4.626*
G	0.518	0.600	4.300	-1.900	1.333	1.334	0.463	3.167	2.358
I	-0.353	-0.450	7.900	-9.300	3.444	3.399	0.017	2.852	0.062
X	0.780	0.725	8.80	-4.543	2.186	1.816	0.995	6.068	35.664***
M	1.170	1.050	7.700	-8.900	2.754	1.448	-0.618	6.105	29.787***
CA	-5.041	-2.189	377.663	-380.989	179.238	198.791	-0.080	2.446	0.888

The sample is 1987Q1:2002Q4
GDP, C, G, I, X and M are calculated from percentage series, CA from bill EUR.
*10%, ** 5%, ***1%

At first look the most striking result is in the last column: according to the Jarque-Bera normality tests all the series except government consumption (G), investment (I), and current account (CA) fail the test. According to table 1 the revisions in investment are close to normality; its skewness is close to zero (0.017) and kurtosis close to three (3.167). The revision distribution of government consumption is slightly skewed to the right.

With more careful investigation the GDP distribution is negatively skewed, although both the mean and median are greater than zero (0.332 and 0.422). The large mass of the distribution is located on the positive side but at the same time there are several negative changes, which explain the negative skewness. The extreme points of the revisions (the min and max) are located almost symmetrically around zero. The distribution of private consumption (C) reminds one of GDP: the large mass of the revisions is located on the positive area (mean 0.075 and median 0.200) but it has long negative tail (skewness -0.652).

Despite the fact that G passes the Jarque-Bera normality test, its distribution looks bimodal; the largest masses are located around (-1, 0) and (1, 2). It has also some very large positive values.

The large mass of the revision distribution for exports (X) is quite symmetric, but it is clearly located in the positive area (mean 0.789, media 0.725). The same can be said for the min and max values (8.80 and -4.53). In the case of exports, also the skewness is clearly positive. While most of the revisions are positive for imports (M), the distribution is negatively skewed. Both mean and median are clearly on the positive area (1.170 and 1.050).

3. Modelling of the Bias

Our goal is to study the systematic and random components of the defined revisions of the selected macroeconomic series. As is well-known these revision series contain a lot of different components like new relevant information, old corrected information and different kinds of error components (see eg. Young, 1995).

We applied both ARMA modelling (Franses, 1998) and state space modelling (structural time series models, Stamp package, see Koopman et al, 2000) to the revision series. Because the revision series are almost all non-normal, robust and resistant estimators are appropriate methods to apply. We selected the "half robust" *rank transformation* approach (RT) (see Conover and Iman, 1981; Kelley and Noel, 1982). This means that we used time series modelling for the series with both the original observations and the rank series (i.e. an observation is replaced by its rank).

As is known in the literature, there is a lot of new theory and applications using *rank-based* analyses and estimators for linear models and ARMA models (eg Hallin et al, 1987; Hallin and Puri, 1988 and Hettmansperger et al, 1997). Because we did not have the computer programs available for these new methods and estimators, we used the RT-approach. Rank transformation and classic nonparametric estimators (like Spearman and Kendall rank correlation) are known to be quite robust for low and medium outlier contaminated series.

The reason for using the structural time series modelling by Stamp is to get the analyses of a changing component model with a final state vector. By using the model we also are able to detect different outliers and their estimates and time points (auxiliary residuals). The other reason is a matter of forecasting: by using a final state model it is more reasonable to predict the near future compared with conventional modelling.

3.1. Cross-correlations

In the next phase we estimated the cross-correlation coefficients between the revision series and the final vintage series. The results are only suggestive and they must be interpreted carefully, because we used the classic Pearson correlation instead of the Spearman rank correlation.

Table 2: Cross Correlations of the data

	GDPF	CF	GF	IF	XF	MF	CAF
GDPFP	**0.459**	0.436	0.189	0.328	0.067	0.090	0.258
CFP	-0.239	**-0.160**	-0.092	-0.189	-0.137	-0.281	-0.457
GFP	0.129	0.227	**0.358**	0.046	-0.068	-0.068	0.0519
IFP	0.441	0.456	0.135	**0.617**	-0.046	0.239	0.227
XFP	0.129	0.027	-0.163	0.057	**0.359**	0.236	0.071
MFP	0.002	-0.060	-0.056	-0.041	0.148	**0.098**	0.163
CAFP	0.359	0.295	-0.120	0.220	0.389	0.216	**0.390**

The sample is 1987Q1:2002Q4

In Table 2 the correlations between the revisions and the final figures are shown on the diagonal. Off diagonal entries show the correlations between the revisions of a variable and its correlation with the final figure of another variable. For example, the second observation from left in the upper row is the correlation between GDP revision and the final private consumption growth.

The cross-correlation results are quite as expected. For instance, the high correlation between GDP, C and I is in line with the consistent revision system (we don't use here significant or not, because we use classic estimator for non-normal variables). The estimates on the main diagonal of the matrix are naturally of quite a high value; there are two exceptional cases: C and M series. The use of rank correlation might give more reasonable result, especially for these two series.

3.2. ARMA-Modelling

Because the data did not support our null hypothesis (white noise series), we try to test if it is possible to find regularities in the revisions. Our starting point for this is to estimate a simple time-series model and test whether we can find parameters that are statistically significantly different from zero. A natural starting point for this is a mixed ARMA(1,1)-model, for two reasons. First, it is a widely used yardstick model and almost as parsimonious as possible. Second, we can find economically meaningful interpretations for the parameters. The AR(1) term can be interpreted to represent the new information accumulating during the data collection process. The MA(1) term catches the non-systematic part of the revisions, such as problems in measuring and data recording.

We estimate an ARMA(1,1) model for two different cases: for the raw data and for the absolute values of the data. The idea is to test, whether we are able to estimate (forecast) the sign of the revision, the size of the revision or both. If for example the AR(1) term is statistically significant for the raw data but not significant for the absolute values, our conclusion is that there is inertia in the sign of the revisions but not in their size.

Due to the possibility of the outliers in the revision series (especially in the current account revisions) invalidating our results, we made experiments by fitting additional ARMA(1,1) models for the rank-transformed data. In this case the true values are replaced with their rank in the data. So for example, the smallest value of the data gets value 1, the second smallest value 2 and so on. The idea behind this transformation is that it cuts the absolute values of the largest outliers. With this transformation we try to find more robust parameter estimates than we could get from the actual data.

In total we have directly fitted four ARMA(1,1) models for each revision series. Before estimating each model, we tested the stationarity of the raw series with Dickey-Fuller tests. On the basis of visual inspection (Appendix 1), other revision series except the private consumption look quite stationary. For some reason the revisions made to private consumption from 1995 on have been negative: the final growth figure has been smaller than the preliminary one. On the basis of Dickey-Fuller tests, however, all the revision series were accepted as I(0). This result also means that we can add a constant in all our ARMA models. In general, if a series had been non-stationary, adding a constant to the model would have introduced a deterministic trend in the model (Harvey 1981, p. 169). However, due the nature of this kind of revision data, the possibility of a trend is not an option: there cannot be a continuously increasing gap between the preliminary and the final data vintages.

The results for the raw data are shown in Table 3.
Table 3: The Estimated ARMA (1,1) models

	Raw data				Ranked raw data			
	C	AR(1)	MA(1)	Q(15)	C	AR(1)	MA(1)	Q(15)
GDP	0.33**	0.62***	-0.17	6.97	32.32***	0.62***	-0.28	14.87
C	0.07	0.57***	0.05	12.78	32.38***	0.56***	0.01	14.82
G	0.48	0.59***	0.36***	23.68**	31.98***	0.58***	0.30**	19.25
I	-0.4	0.72***	-0.37	5.64	32.10***	0.70***	-0.3	11.48
X	0.75**	-0.26	0.65***	11.88	32.31***	-0.05	0.33	10.97
M	1.13**	0.17	0.34	3.62	32.11***	0.27	0.15	10.13
CA	10.79	0.91***	-0.72***	15.43	31.32***	0.89***	-0.59***	24.80**

*10%, ** 5%. *** 1%

On the basis of the portmanteau (Q) test the residual can be accepted as white noise in all cases except government consumption, and for the rank transformed case for current account. Also both the parameter and standard error estimates for the AR(1) and MA(1) terms remain (in most cases) very close to each other in both the original and rank-transformed estimations; the estimates seem to be rather robust for outliers with the exception of the exports and imports series. Somewhat surprisingly, even the current account model remained rather similar for the both data cases.

For three of the series (GDP, exports, imports) we can estimate a statistically significant constant. The AR(1) term is significant for all the series except exports and imports and its value varies from 0.57 (consumption) to 0.91 (current account). The MA(1) term is statistically significant for government consumption, exports and the current account.

The estimated models for the absolute values of the revision series are shown in Table 4.

Table 4: The Estimated ARMA (1,1) models

	ABS data				Ranked ABS data			
	C	AR(1)	MA(1)	Q(15)	C	AR(1)	MA(1)	Q(15)
GDP	0.578***	0.186	0.062	14.889	32.424***	0.465	-0.309	17.627
C	0.577***	0.0239	0.198	13.668	32.473***	0.026	0.12	7.775
G	1.105***	0.356**	0.253	15.439	32.193	0.262	0.124	18.474
I	2.641***	0.898***	-0.846***	17.605	32.14	0.865***	-0.829***	14.468
X	1.662***	0.321	0.288	11.264	32.49	0.433	-0.06	7.561
M	2.137***	0.603***	-0.164	5.784	32.264***	-0.021	0.656***	7.5
CA	165.095	0.289	-0.165	12.082	32.415***	0.384	-0.179	14.01

*10%, ** 5%. *** 1%

In the case of the absolute valued data, the ARMA(1,1) model residuals pass the portmanteau test in all cases. For some reason the parameter estimates are more sensitive as to whether we use rank data or not, than in the case with absolute value data (Table 3). The only statistically significant and stable estimates with respect to the rank transformation is investment. For GDP, private consumption, exports and current account we were not able to estimate significant AR(1) or MA(1) coefficients in either case. The constant terms are statistically significant in all non-rank transformed cases except the current account.

On the basis of tables 4 and 5, we can say as a conclusion that for the case of GDP the first estimate of the annual growth is on average 0.33 percent too low and there is persistence in the sign of the corrections between previous quarters. However, we are unable to estimate any persistence in the size of the correction. On average, the absolute value of those revisions is close to 0.6 percent. For private consumption, there seems to be as similarly high persistence on the signs of the revisions but neither on the size nor the average value of it. The reason that we were unable to estimate a significant constant term is probably due the fact that it seems to have undergone a structural break in the mid 1990s; before that it was, on average, positive while after the break it has had a negative trend. The mean of the absolute values of the revisions is pretty close to the one estimated for GDP. In the case of government consumption there seems to be persistence both in the signs and the sizes of the revisions. There also seems to be some noise in the raw series because we can estimate a significant MA(1) term. In absolute value, the revisions are on average twice as large as for GDP and private consumption. However, the results for government consumption must be taken with a grain of salt due the failure of the portmanteau test for the residuals.

With investment we can estimate statistically significant AR(1) terms both for the size and the sign. The average absolute value of the revisions is large (2.64), 4 ½ times as large as for GDP and private consumption and almost 2½ times as large as for government consumption. In absolute values, investment revisions are extremely noisy, while we are able to estimate an MA(1) term with a parameter estimate -0.85. Both in raw and absolute values the results seems to be quite robust, while the use of ranked data did not significantly change the parameter estimates.

The mean of export revisions is almost a ¾ percentage point and for import revisions is over one (1.13). The series are rather noisy; we can find a statistically significant MA(1) term for exports in raw form but are unable to estimate significant AR(1) parameters with the exception of the absolute value series of imports. The point estimates are also affected as to whether we use ranked data or not. For the current account we find the signs of the revisions to be highly persistent as the estimate for the AR(1) term is 0.91. The MA(1) term is also really high. However, we are not able to find any persistence in the size of revisions. These results seem to be quite robust as there is not much difference between the estimates made with rank transformed and raw data.

3.3. The Estimated Stamp-models for the data

There seems to be long term movements in the revisions data. For example, during economic downturns the preliminary figures of the GDP have overestimated the final ones, while in other periods the sign has been the opposite (see Appendix 1). Another example is private consumption, where revisions tend to have become negative. It seems quite plausible that for forecasting the current revision, the model which uses all the information available would not necessary lead to a better forecast. For the GDP example, at least the constant term should probably be larger without the recession periods in the data.

The Kalman filter provides a general recursive updating method for estimating a signal plus noise in state-space modelling. As is known, many types of time series models like the ARMA models can be put into a state-space form. So, a proper solution would have been to estimate the ARMA(1,1) models as the Kalman filter model. Unfortunately, that was not possibly at this stage of the study, so we did the estimations with Stamp. Stamp uses Kalman filter to estimate the final state vector of the unobserved varying component model (Koopman et al, 2000) This method should allow us to see whether the constant term is different at the end of the period compared to the whole data.

Table 5: Stamp results for the revisions data

	Raw data				Ranked data			
	C	Slp	R2	Q(15)	C	slp	R2	Q(15)
GDP	0.657**	0.011	0.15	6.354	43.203***	0.385	0.23	13.861
C	-0.68***	-0.017	0.05	21.907**	13.316*	-0.441	0.06	22.9224**
G	-	-	-	-	-	-	-	-
I	0.348	0.018	0.15	24.778**	23.815***	-0.0957	0.203	13.337
X	-	-	-	-	37.28***	0.184	0.33	15.955
M	1.86	0.029	0.13	12.091	41.727***	0.24	0.24	20.900**
CA	-90.206	-11.992	0.42	15.477	23.391***	-0.094	0.35	24.421**

C is the constant and slp is the slope at the end of the estimation period.
The sample is 1987Q1:2002Q4; 64 observations.

In these estimates we have included in the Stamp model both a stochastic constant and slope but not a seasonal term. For some of the series the Kalman filter approach seems to fit quite well. However, for some of the series Stamp estimation did not reach a convergence.

For GDP and private consumption, the results are as suggested: the values of the constants differ from the ARMA – and from the mean value reported in Table 1 – and the sign of the change was as expected. For GDP, the point estimate of the constant is larger without the depression time period. For private consumption, the insignificant and close to zero point estimate of the constant changes to a negative point estimate which is statistically significant. However, the residuals of the fitted model are rejected by Ljung-Box test.

4. Forecasting the Bias: a small forecasting exercise

The main reason why we are interested in modelling the bias is that it should allow us to make forecasts of future revisions. In the real forecasting situation, we only need to forecast one step ahead: we want to assess how much to correct the latest available (period t) preliminary national account figures and in which direction. The forecast for t+1, t+2, …, T are outside the region of this model; for those time points we need to forecast the whole number (eg. growth rate) instead of the adjustment parameter. However, whether one can forecast time t is not a clear cut issue because by definition the latest available final vintages series is 2 – 2½ years old.

We study the forecasting features of the models with ex-post and ex-ante forecasts. First, we compare the model forecasts for the final revision data by estimating rolling models from 2001Q4-2002Q3 and forecasting one step ahead. As, by definition, we cannot have newer final data than 2002Q4, our ex-ante forecasts exploit the differences between the preliminary figures and the latest available updates. For this case, we have estimated the rolling models from 2002Q4 to 2003Q3 and again calculated the one-step ahead forecasts. At this stage, we restrict our forecasts only to the GDP series.

4.1. Ex post forecast

As a starting point, we forecast the last available final revision series (year 2002Q1-Q4) with the ARMA(1,1) and Stamp models presented in Tables 3 and 5. The results are presented in Table 6:

Table 6: 1 Step Ex Post Forecasts for the GDP Revisions

		ARMA(1,1)			Stamp		
	Actual	Forecast	S.E.	Forecast Error	Forecast	Rmse	Forecast Error
2002Q1	1.2	0.307	0.541	0.893	0.236	0.594	0.964
2002Q2	1.2	0.708	0.549	0.492	0.833	0.602	0.367
2002Q3	0.5	0.799	0.548	-0.299	1.088	0.599	-0.588
2002Q4	0.6	0.477	0.545	0.123	0.742	0.599	-0.142

The forecasts are 1 step forecasts made by rolling estimations. The ARMA(1,1) model includes a constant. The Stamp model has constant, slope and irregular terms.

Despite the simplicity and probable non optimality of our models, they are both able to reduce the bias related to the preliminary figures of the GDP growth. The average forecasting error (actual – forecast) during the inspected period is 0.3 for the ARMA and 0.15 for the Stamp model. The average of the absolute forecasting errors is 0.45 and 0.52 for the ARMA and Stamp models.

Unfortunately all the revisions were positive during the inspected period, so we could not compare the forecasting performance in more complicated situation. At least in principle, the Stamp should have reached better forecasts because it weights the most recent observations more than the ones at the beginning of the estimation period.

4.2. Ex ante forecasts

To be able to simulate the real forecasting situation, we must be able to fill the gap between the latest final data and the latest available data; the gap is around 2 – 2½ years. The most correct way would be to model the convergence between different vintages, instead of just modelling the final adjustment. Due the lack of the intermediate vintages, we must at this stage cheat a bit with the data: From 2002Q4 on we define the revision as the difference between latest available data and the preliminary one. This allows us to estimate the model up to 2004Q3; the latest preliminary figure available is 2004Q4.

The extension of the data did not change the ARMA(1,1) model significantly from the one reported in Table 4. The same is true also for the Stamp model. The results are reported in Table 7:

Table 7: 1 Step Ex Ante Forecasts for the GDP revisions

		ARMA(1,1)			Stamp		
	Actual	Forecast	S.E.	Forecast Error	Forecast	Rmse	Forecast Error
2003Q1	1.6	0.477	0.540	1.123	0.668	0.594	0.932
2003Q2	1.2	0.953	0.554	0.247	1.235	0.600	-0.035
2003Q3	1.1	0.869	0.550	0.231	1.233	0.596	-0.133
2003Q4	0.4	0.817	0.547	-0.417	1.173	0.591	-0.773
2004Q1	0.6	0.458	0.545	0.142	0.735	0.594	-0.135
2004Q2	0.4	0.495	0.541	-0.095	0.664	0.590	-0.264
2004Q3	0.7	0.406	0.537	0.294	0.515	0.587	0.185
2004Q4	-	0.534	0.535	-	0.634	0.583	-

The forecasts are 1 step forecasts estimated with rolling estimations. The ARMA(1,1) model includes a constant. The Stamp model has constant, slope and irregular terms.

In general, the ex ante forecast are able to reduce the gap between the preliminary and later vintages. The average forecasting error (actual – forecast) during the forecasting period is +0.22 for the ARMA model and -0.03 for the Stamp model. On absolute values of the forecasting error, both models perform similarly, while the averages are 0.36 and 0.35 respectively. As was the case with the ex post forecasts, none of the actual revisions were negative.

On the basis of ex post and ex ante forecasting exercises, our simple ARMA and Stamp models seem to be able to forecast the bias in the GDP growth. Due to the simplicity of the models – we did not estimate the ARMA structure in the traditional way but fitted the ARMA(1,1) model to the data – the results are far from perfect, but they seem to add a valuable set of information for the real forecasting situation: whether to take the preliminary GDP growth figures as given or to do some adjustment. The ability to forecast the bias even – although not perfectly – supports our earlier findings of the non-randomness of the revision data.

5. Conclusions

In this paper we have analysed the revisions of selected Finnish macroeconomic series. The revision was defined as the difference between the preliminary (first) and final vintages. Our null hypothesis was that the revisions are white noise and thus are not forecastable. This hypothesis was rejected quite clearly on the basis of distribution statistics: in most cases, neither the mean nor the median are zero, and the normality tests for GDP, private consumption, exports and imports revisions were all rejected. In all cases the mean of the revisions were found to be positive. The width of the revision distribution was also very different between the series.

As a second step we tried to model the bias, as we fitted ARMA(1,1) models to the revision series. We estimated the ARMA models both for the raw data and its absolute value transformations, so we were able to investigate whether we can find persistency in the sign of the revisions, in the size of them, or both. Due the heterogeneity and detected outliers in the data, we also estimated the same models for the rank transformed data. This allowed us to study the stability (for outliers) of the parameter estimates. Due the possible structural changes in the revision series we also applied Stamp (structural time series models) to the data. In general, we were able to fit statistically significant models to the data.

As a third and last step, we tested whether we could exploit the observed persistence and forecast the revisions. We tested the forecasting performance for the GDP revisions both with ex post and ex ante experiments. Although the forecasts were far from perfect, in both cases we were able to improve the accuracy of the preliminary published information.

The results from this paper suggest that there is systematic bias in most of the Finnish National Account series. Furthermore, they also suggest that we are able and should try to model and forecast the bias while interpreting the latest information and producing forecasts for eg. monetary policy purposes. Even a simple model – and probably far from the optimal one – turned out to be useful for forecasting the GDP revisions. The analysis of the revisions allows us to also improve our future forecasts.

For future research one should pay more attention to the convergence process from the preliminary vintages to the final ones. This is especially important for forecasters, because it allows them to interpret more carefully the latest figures and their expected future revisions and thus improve the future forecasts. Also a more careful modelling strategy should improve the results further. In addition robust and resistant methods and estimators could be used to adjust estimates in case of outliers in revision series.

References:

Conover, W.J. and Iman, R.L. (1981) "Rank transformations as a bridge between parametric and nonparametric statistics", The American Statistician, 35, 124-129.

Croushore, D. and Stark, T. (2001) "A real-time data set for macroeconomists", Journal of Econometrics, 105, 111-130.

Egginton, D.M., Pick, A. and Vahey, S.P. (2002) "Keep it real!: a real-time UK macro data set", Economics Letters, 77, 15-20.

Franses, P.H. (1998) Time Series Models for Business and Economic Forecasting. Cambridge University Press.

Hallin, M. Ingenbleek, J.-FR. and Puri, M.L. (1987) "Linear and quadratic rank tests for randomness against serial dependence", Journal of Time Series Analysis, 8, 409-424.

Hallin, M. and Puri, M.L. (1988) "Optimal rank-based procedures for time series analysis: testing an ARMA model against other ARMA models", The Annals of Statistics, 16, 402-432.

Harvey, A.C. (1981) Time Series Models, Philip Allan.

Hettmansperger, T.P., McKean, J.W. and Sheather, S.J. (1997) "Rank-based analyses of linear models" in Handbook of Statistics 15: Robust Inference (eds. G.S. Maddala and C.R. Rao) 145-173, Elsevier, Amsterdam.

Kelley, G.D. and Noel, D.E. (1982) "The rank transformation as a method for dealing with time series outliers", in Applied Time Series Analysis, O.D. Anderson & M.R. Perryman (eds.), North-Holland Publishing Company.

Koopman, S.J., Harvey, A.C., Doornik, J.A. and Shephard, N. (2000) Stamp, Structural Time Series Analyser, Modeller and Predictor, Timberlake Consultants Ltd.

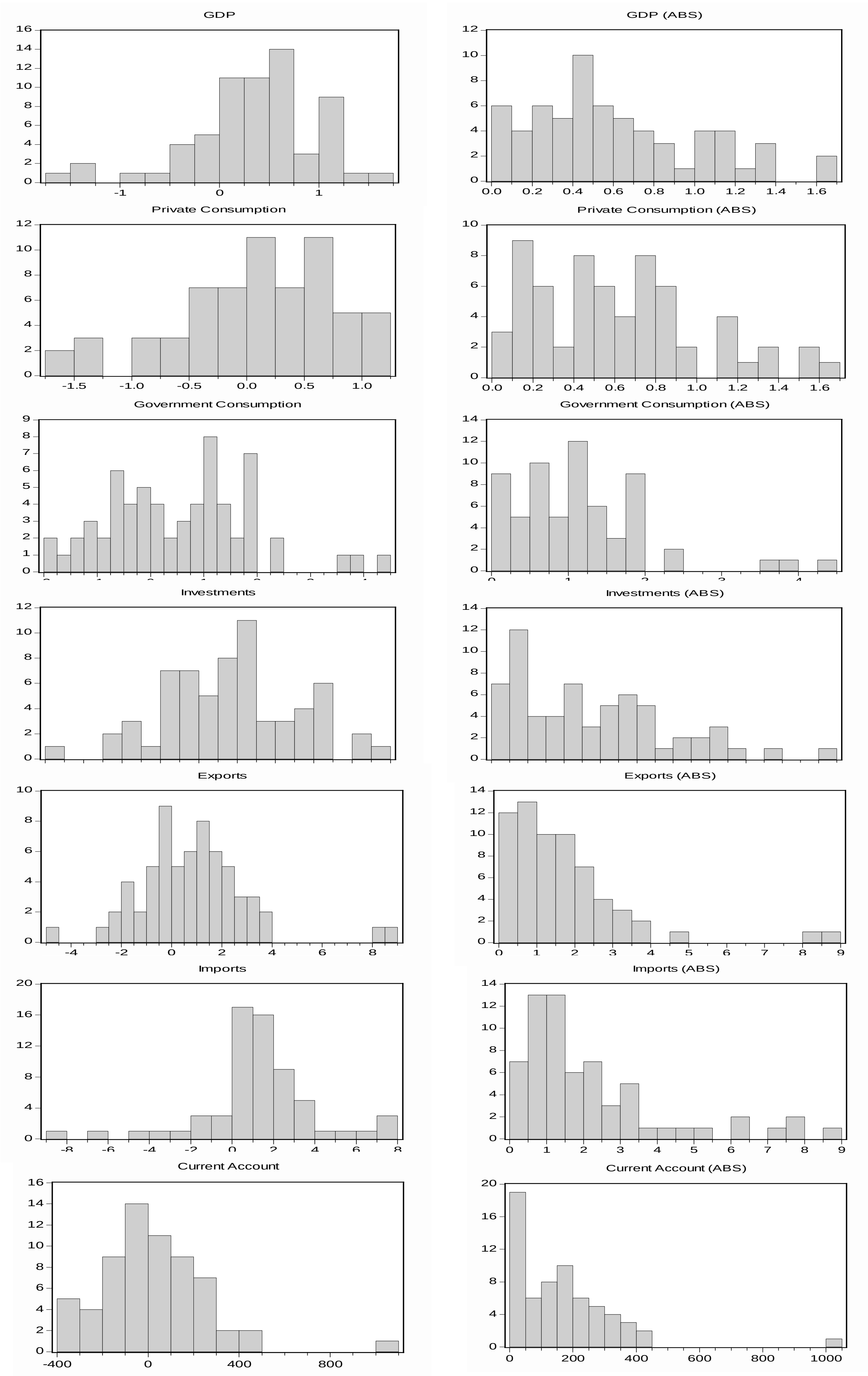
Savela, O. and Forsman, P. (2004): “The Accuracy of Preliminary National Accounts Data”. Bank of Finland Bulletin, 2004:3, 93-99.

Young, A.H. (1995) "Reliability and accuracy of quarterly GDP estimates: a review", in The New System of National Accounts, ed. J.W. Kendrick, Kluwer Academic Publishers, 423-455.

Öller, L.-E. and Hansson, K.-G., (2004) "Revision of National Accounts: Swedish expenditure accounts and GDP", Journal of Business Cycle Measurement and Analysis, 1, 363-385.

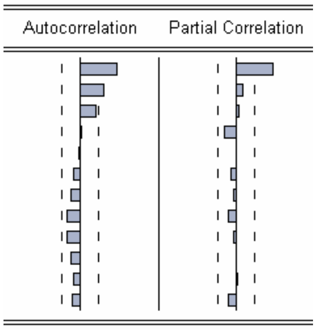
Appendix 1: The Revision Series

Appendix 2: The histograms of the revision series

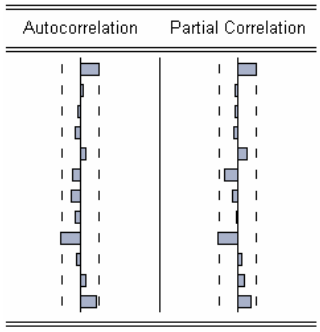


Appendix 3: Autocorrelation and partial autocorrelation functions (lag 12)

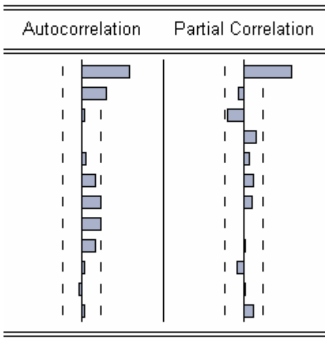
GDP



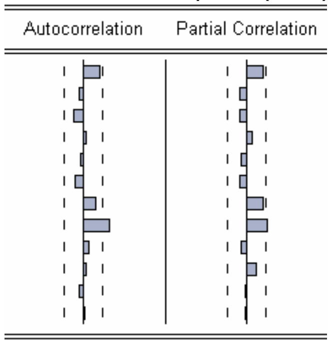
GDP (ABS)



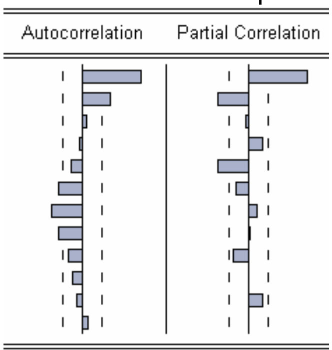
Private Consumption



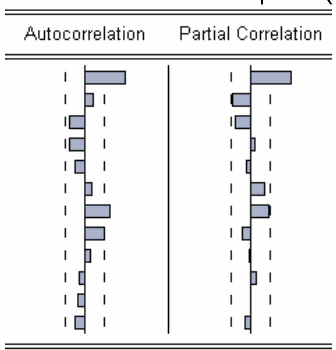
Private Consumption (ABS)



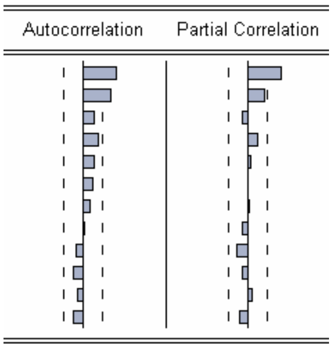
Government Consumption



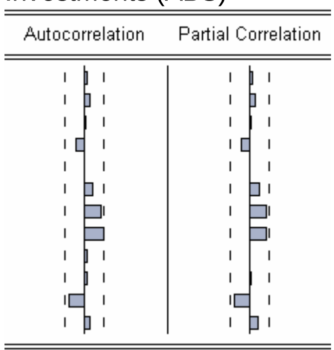
Government Consumption (ABS)



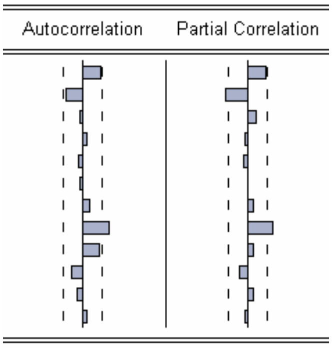
Investments



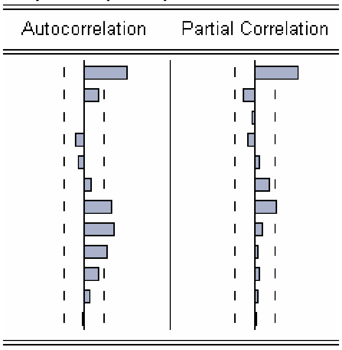
Investments (ABS)



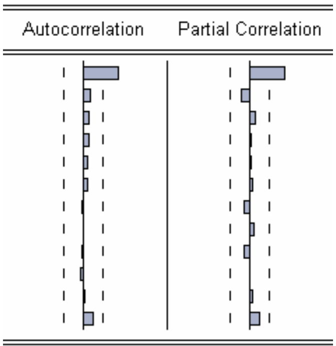
Exports



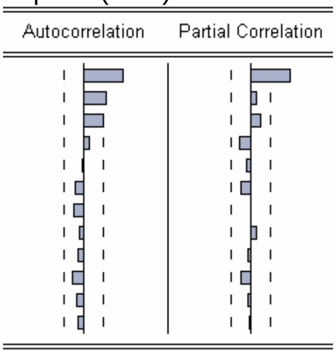
Exports (ABS)



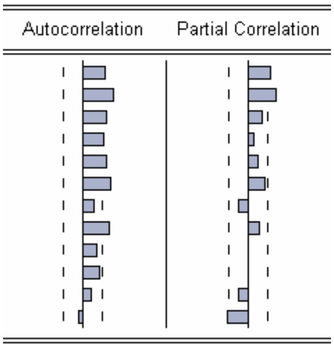
Imports



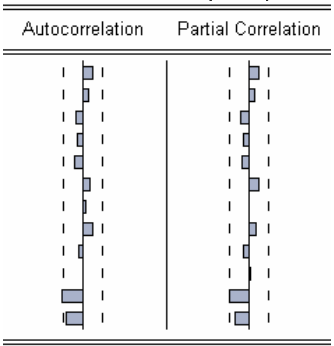
Imports (ABS)



Current Account



Current Account (ABS)



Appendix 4: The Case of CA surplus/deficit revision (CAFP)

We carried out the following special experiment for our revision series CAFP:

a) first we calculated the robust filter (4253 H: Statistica package) from this series
b) in Figure A4.1 is presented the original CAFP and its filtered series CAFP4253H
c) based on the Figure X and numerical listing we found that in time points 95/Q4, 98/Q2, 01/Q4 there are isolated, additive (AO) outliers[1]
d) these three outliers were replaced by the value of the corresponding filtered value; so we had now original CAFP and its outlier corrected series, CAFPoutl
e) for ARMA modelling we formed the rank transformation version of the original revision series and got series CAFPrank.

Now we had three series CAFP, CAFPoutl and CAFPrank for the modelling stage.

The goal was:
To evaluate the two "robust" ARMA estimation based 1) on CAFPoutl and 2) on CAFPrank.

Of course we compared these results with an ARMA estimation based on the original series, CAFP. We estimated all the ARMA(1,1) models with a constant term. The series CAFP and CAFP4253H are displayed in Figure A4.1.

Comments on the estimation results are:

1. In case of original series both AR- and MA-parameter estimates received statistically extra significant values, the constant term had no significant t-values; for the residual series its autocorrelation and partial autocorrelation estimates were not significant in any of 1-15 lags

2. For series CAFPoutl the results were similar, but the numerical values of AR- and MA-parameters received a little lower value.

3. In case of CAFPrank series the ARMA(1,1) parameter estimates received clearly higher values compared with the case of CAFPoutl series. For residual series autocorrelation and partial autocorrelation estimates were higher, but they were also in lags 1-5 inside 95% confidence limits; res. Q-value was larger than in case of CAFPoutl.

The assessment:

Rank transformation seems to be superior to obsolete way of adjusting detected outliers before modelling phase. Nowadays in robust time series modelling outliers are, in general, not adjusted before modelling phase.

---

[1] Detection of outliers may be a difficult task; here we can use besides a graphical analysis Stamp model estimation and checking its auxiliary residual series with different threshold value (see Koopman et al, 2000).

ARMA(1,1) estimation results:

Variable	C	t-value	p-value	AR1-phi	t-value	p-value	MA1-theta	t-value	p-value	res. Q(15)	p-value
CAFP	6.18	0.08	0.94	0.921	13.29	0.00	-0.721	7.04	0.00	15.87	0.39
CAFPoutl	-14.47	-0.24	0.81	0.847	8.13	0.00	-0.404	2.26	0.03	20.25	0.16
CAFPrank	30.78	4.48	0.00	0.907	12.04	0.00	-0.599	4.66	0.00	25.66	0.04

Figure A4.1: