Sveriges Riksbank's inflation interval forecasts 1999–2005*

Michael Lundholm

Department of Economics, Stockholm University

May 12, 2008

Abstract

Are Sveriges Riksbank's inflation (CPI and KPIX) interval forecasts calibrated in the sense that the intervals cover realised inflation with the stated *ex ante* coverage probability? In total 150 interval forecast 1999:Q2–2005:Q2 are assessed. The main result is that the forecast uncertainty is understated, but there are substantial differences between individual forecast origins.

JEL: C53, E31 and E37.

Keywords: Inflation, forecast, interval forecast, forecast uncertainty.

*Correspondence to Michael Lundholm, email <mlu@ne.su.se>. Address: Department of Economics, Stockholm University, SE–106 91 Stockholm, Sweden. I am grateful for comments from Jim Albrecht, Mahmood Arai, Martin Flodén, Matthew Lindquist and Peter Thoursie.

1 Introduction

Since 1993 the objective of Sveriges Riksbank (the central bank of Sweden) has been price stability. In 1995 a self–adopted explicit inflation target of a yearly 2 percent increase in consumer prices with a tolerance band of ± 1 was implemented. In order to determine whether a change in main policy instrument of the Riksbank (the repo rate) is necessary the Bank forecasts the inflation rate measured as changes in consumer price index (CPI) and underlaying inflation (KPIX; before November 12, 2007 called UND1X) with 2 year forecast horizon. Since 1999 these point forecasts are accompanied with interval forecasts for 50, 75 and 90 percent coverage with a forecast horizon of 1–25/26 months. Until 2005:Q2 these forecasts where made four times a year and conditional on an unchanged repo rate.[1]

Sveriges Riksbank applies a judgemental inflation interval forecast procedure which is described in Blix and Sellin (1998, 1999) and briefly in Berg (2000). But whereas the corresponding point forecasts in several studies can be shown to be reliable (using standard assumptions and standard criteria such as efficiency (mean forecast error) and accuracy (mean square forecast error)) compared point forecasts from other forecasters, there is hitherto no assessment about the reliability of the interval forecast.[2] The purpose of this study is therefore to throw light on the question whether these interval forecasts are appropriately calibrated. That is, are the interval forecasts for (say) 50 percent coverage actually contain actual inflation in 50 percent of the cases?

The most apparent strategy would be to employ the test procedure suggested by Christoffersen (1998). For each forecast an indicator variable is created, taking the value one when actual inflation is inside the forecasted interval and else zero. Then the following tests are performed: (i) a test for unconditional coverage (the null hypothesis that the sample mean of the indicator series equals the coverage probability is tested against the alternative that it is not), (ii) for independence (the null hypothesis that the indicator variable is serially independent with probability transition matrix given by the sample coverage probability is tested against the alternative that it is a first order Markov process) and (iii) for conditional coverage (the null hypothesis that the indicator variable is serially independent with probability transition matrix given by the *ex ante* coverage probability is tested against

[1] This forecasting targeting type of monetary policy was recently reviewed by Woodward (2007).

[2] See Blix et al. (2002), Konjunkturinstitutet (2002), Jansson and Vredin (2003), Bergvall (2005) and Andersson et al. (2007); only Konjunkturinstitutet (2002) (although the staff at the National Institute of Economic Research responsible for this evaluation had a professional background at Sveriges Riksbank) and Bergvall (2005) do not formally originate from Sveriges Riksbank.

the alternative that it is a first order Markov process).

However, there are several problems of applying this testing procedure to present interval forecasts. The data consists of 25 forecast origins with three interval forecasts at each origin (50, 75 and 90 percent coverage) for two measures of inflation (CPI and KPIX). Each forecast has a horizon of 25 or 26 months. This sums up to 150 individual forecasts each of which has a horizon of 25 or 26 months, or, if each *h*–step–ahead forecast is counted as an individual forecast, about 625 forecasts for each coverage level and inflation measure.

The problem of applying the Christoffersen (1998) procedure is (briefly stated) that there are too few observations at each forecast origin. First, to perform the test for independence between neighbouring observations one need to define four different conditional probabilities, such as the probability that inflation rate is inside the forecasted interval the present period conditional on the event that it was outside the previous period. At no less than 40 forecast origins the outcome is such that at least one of the four events for which conditional probabilities need to be calculated do not occur. Second, even if all four events were observered in a sample of 25, the minimum expected number of observations for each event that is needed for drawing valid inference was in the cases of high coverage levels too low.[3]

The practical solution was to pool all interval forecast for each inflation measure and each coverage level to construct six different data sets with some 625 observations in each and only perform the unconditional coverage test, rather than selecting few forecast origins for which the entire test suite could be performed. However, in addition to testing the interval forecasts of Sveriges Riksbank two additional sets of interval forecasts are tested. These are interval forcasts produced by automatic forecast procedures using ARIMA–models and exponential smoothing models as described by Hyndman and Khandakar (2008). These models were estimated on a real–time inflation data set. The problems noted for the interval forecasts of Sveriges Riksbank was even more severe for the forecasts produced this way.

The main result of the present study using pooled data is that the interval forecats of Sveriges Riksbank have a tendency to understate the forecast uncertainty. Only the forecasted intervals at the 50 and 75 percent coverage level for KPIX passed the unconditional coverage test. The reason in all instances failing the test was that the forecasted intervals were too narrow.

Today there is a small but growing literature assessing different type of uncertainty estimates in the inflation forecasts produced by central banks. Using

[3]For instance, at the 90 percent coverage level the expected number of observations outside the interval is less than three. This should be compared with recent results indicating expected numbers on the level $7--8$ in two–tailed tests; see Andrés and Tejedor (2003).

2–year–ahead *probability forecasts* published by Sveriges Riksbank for a part of the above time period, Dowd (2004) concludes that they understate the forecast uncertainty. Wallis (2003, 2004) and Clements (2004) assess the forecasts of the Bank of England's Monetary Policy Committee (MPC) using the published information about the two–piece normal distribution in the density forecasts. The main result is that the MPC overstates the forecast uncertainty.

The paper is organised as follows. In section 2 the tests the data are described. Section 3 contains the results and ection 4 concludes the paper.

2 Framework for assessment and data

The log likelihood test procedure is described in detail in Christoffersen (1998) and reviewed by Wallis (2003). Compared to Christoffersen (1998), who had a sequence of n one–step–ahead forecasts to be tested, we have n different 1–25/26–step–ahead forecasts pooled from different origins. Necessarily then the pooled data have no particular order and we are restricted to perform only the unconditional coverage test.

Now, an interval forecast is an interval $[L_\tau(p), U_\tau(p)]$, where $L_\tau(p)$ and $U_\tau(p)$ are the lower and upper limits of the point forecast for the coverage probability $p \in \{50, 75, 90\}$ and $\tau = 1, \ldots, n$ is an index identifying (θ, t), where θ is the date of the forecast origin and t is the date for which the forecast applies. That is, since forecasts overlap, forecast made at different origins can apply to the same date.

Let y_t be the observed inflation at date t. An indicator variable $\{I_\tau\}$ is then created such that $I_\tau = 1$ if $y_t \in [L_\tau(p), U_\tau(p)]$ and else $I_\tau = 0$. Also, let $n_1 = \frac{1}{n}\sum_{\tau=1}^{n} I_\tau$ be the number of outcomes in the interval, $n_0 = n - n_1$ the number of observation outside the interval and denote the sample mean of the indicator variable $\pi = n_1/n$. The definition of correct coverage is then that $\{I_\tau\} \overset{iid}{\sim} \text{Bern}(p)\ \forall\tau$; see Christoffersen (1998, Lemma 1 and Definition1).

The unconditional test has as its null that the sample mean of the indicator variable π equals p against the alternative that it does not, that is $H_0 : \pi = p$ is tested against against $H_1 : \pi \neq p$. The test statistica is then the log–likelihood ratio

$$\text{LR}_u = -2\log\left(\frac{(1-p)^{n_0} p^{n_1}}{(1-\pi)^{n_0}\pi^{n_1}}\right). \qquad (1)$$

If the test statistica is sufficiently large, i.e, $\text{LR}_u > \chi^2_{0.95}(1) \approx 3.84$, then the interval forecast fails the test and we conclude that the width of the interval forecast are not appropriately calibrated. This will happen if π deviates to much from p in either direction.

The data set consists of

1. interval forcasts with 25 forecast origins 1999:Q2–2005:Q2, where Sveriges Riksbank at each origin forecasts the inflation interval for 50, 75 and 90 percent *ex ante* coverage 1–25 or 1–26 months ahead for CPI and KPIX,[4]

2. real time inflation outcome according to CPI and KPIX from a sequence of Inflation reports of Sveriges Riksbank available at their website and used to estimate the ARIMA and exponential smoothing models using automatic forecasting procedures,[5]

3. for each of the forecasts produced by Sveriges Riksbank two reference forecasts are produced using the estimated ARIMA and exponential smoothing models, and

4. revised inflation outcome according to CPI and KPIX (i.e., which contain *ex post* changes in the series due to corrections etc) from the Statistics Sweden home page `http://www.scb.se` against which all forecasts are evaluated. The changes in the methodology to calculate the different inflation series are documented in memoranda and press releases available on this web page.

3 Results

Using the pooled data we can make in total six different tests for unconditional coverage, each based on 640 realisations of the indicator variable. For each inflation measure (CPI and KPIX), each coverage level (50, 75 and 90 percent) and each forecaster (Sveriges Riksbank (SR), ARIMA and exponential smoothing (ETS)), Tables 1–2 reports the test statistica $\mathrm{LR}_{\mathrm{uc}}$ and the sample mean of the indicator variable. The critical value of the test statistica is $\chi^2_{0.95}(1) = 3.84$. All numbers in the table are reported with an accuracy of three digits.

The result is that Sveriges Riksbank's interval forecasts fail the unconditional coverage test in four out of six instances: For CPI in all three instances and for KPIX at the 90 percent coverage level. In all cases the sample mean of the indicator variable is lower than the *ex ante* coverage probability. That is, in these four instances the interval forecasts of Sveriges Riksbank are too narrow and the *a priori* stated forecast uncertainty is too small. The deviation in the sample mean from the *ex ante* coverage is 4–8 percent units, a deviation which is statistically significant. Also, the deviation tend to be larger in absolute terms for the 50 and

[4]The interval forecasts are not found in the Sveriges Riksbank (1996–2006) (except in the form of graphs). They are, however, publicly available for download at `http://www.riksbanken.se`.

[5]These forecasts are produced using the automatic procedures in the package forecast for **R**; see Hyndman (2008). The criteria for model selection etc are described in Hyndman and Khandakar (2008). For more details see the Appendix.

Table 1: Unconditional coverage test for pooled data. CPI. $\chi^2_{0.95}(1) = 3.84$ and $n = 640$.

Coverage level p:		50%	75%	90%
SR	LR_{uc}	8.823	21.398	10.651
	$\frac{1}{n}\sum I(p)$	0.442	0.669	0.859
ARIMA	LR_{uc}	79.485	305.919	134.651
	$\frac{1}{n}\sum I(p)$	0.675	0.989	1
ETS	LR_{uc}	289.571	299.249	134.651
	$\frac{1}{n}\sum I(p)$	0.823	0.988	1

Table 2: Unconditional coverage test for pooled data. KPIX. $\chi^2_{0.95}(1) = 3.84$ and $n = 640$.

Coverage level p:		50%	75%	90%
SR	LR_{uc}	0.69	1.44	12.258
	$\frac{1}{n}\sum I(p)$	0.517	0.73	0.856
ARIMA	LR_{uc}	23.051	132.308	109.68
	$\frac{1}{n}\sum I(p)$	0.595	0.925	0.995
ETS	LR_{uc}	31.37	138.048	79.6
	$\frac{1}{n}\sum I(p)$	0.611	0.928	0.986

90 percent coverage levels compared to the 90 percent level. Only the interval forecasts for KPIX with coverage 50 and 75 percent pass the unconditional coverage test with pooled data. In both cases the sample mean of the indicator variable deviates no more than 2 percent units from the *ex ante* coverage level; a deviation which is not statistically significant.

The reference forecast produced with automatic forecast procedures show a completely different pattern. None of these (pooled) forecasts pass the unconditional coverage test but it is for the opposite reason: The interval forecasts are too wide so actual inflation tend to be inside the forcasted intervals too often. One obvious explanation is that, although at each forecast origin the order of the model is reselected and estimated with the additional information gained since the previous forecast origin, these model presume a constant variance.

One should note that the results for the pooled forecasts of Sveriges Riksbank are achieved with all available data, including the 40 forecast origins which contain insufficient variation in the indicator variable. Of these 40 forecasts no less than 18 had no variantion in the indicator variable at all since inflation was always inside the forecast interval. That is, altough the pooled forecast in many instances tend to have too narrow bands, there are several individual forecast origins which have intervals that are too wide:

1. The actual distribution of these 18instances is as follows: 15 occur for the 90 percent interval forecasts (CPI 8 and KPIX 7). The remaining 3 are for the 75 percent interval forecasts divided on CPI and KPIX. This means that for CPI at the 90 percent coverage level this event occurs in no less than 32 percent of the instances (8 out of 25) and for KPIX at the same coverage level in 28 percent of the instances (6 out of 25). For CPI and KPIX at the 90 percent coverage level the average is that 30 percent or 15 out of 50 instances have actual inflation always within the forecasted interval.

2. The *ex ante* probability of such an event (i.e., actual inflation always within the forecasted interval) is about 7 percent.[6] Therefore, at the 90 percent coverage level one should expect only 3–4 such forecasts instead of the observed 15 forecasts.

3. These forecasts do not pass the unconditional coverage test. Their test statistica L_{uc} is equal to 5.27 with 25 observations and 5.48 with 26, which is above the critical level $\chi^2_{0.95}(1) = 3.84$.

For the 90 percent coverage level, then, this means that the problem of too narrow forcast intervals is even more severe.

4 Conclusions

This study has tested whether the inflation interval forecasts of Sveriges Riksbank are calibrated according to the *ex ante* stated coverage probabilities. Due to limitations in data observations from all forecast origins were pooled for each forecasted inflation measure (CPI and and KPIX) and each coverage probability (50, 75 and 90 percent). Hence, six tests for unconditional coverage were performed. They showed that only the forecasted intervals at the 50 and 75 percent coverage level for KPIX passed the unconditional coverage test. The reason in all instances failing the test was that the forecasted intervals were too narrow.

In particular, the interval forecasts at the 90 percent coverage level failed the test for both CPI and KPIX although data included forecast origins (about 30 percent of the total number of origins) where the forecasted intervals were too wide and always included the inflation outcome.

[6] With 25 observations it is $100 \times 0.9^{25} \approx 7.2\%$ and with 26 it is $100 \times 0.9^{26} \approx 6.5\%$.

5 Computational details

All estimations were obtained with R 2.7.0 (see R Development Core Team (2006)) and the packgae **forecast** using a PC running Debian Linux (with a 2.6.15 kernel. The full sources for this document (data, R–code and LaTeX sources) are available at `http://people.su.se/~lundh/SR`.

Appendix

The first class reference forecast is based on an ARIMA model of order

$$(p,d,q)(P,D,Q)_{12}.$$

First a Conova–Hansen test is performed to decide whether there is a seasonal unit root, i.e., $D = 1$ or $D = 0$. Then successive KPSS-tests (Kwiatkowski–Phillips–Schmidt–Shin), were the null hypothesis is no unit root are performed to determine the value on d. Using the AICc criterion a step–wise procedure is used to determine p, q, P and Q. This is then repeated at each of the 20 forcast origins for both CPI and KPIX. The estimation procedure also allows for drift, even if $d + D \geq 1$.

The second reference forecast is produced using an innovations state space representation of a class of exponential smoothing models ETS, which is acronym for error trend and seasonality. Trend and seasonality could be additive (A), multiplicative (M) or not be present (N); in both cases damped or not. Errors could be additive or multiplicative. The standard exponential smoothing model therefore has order (A,N,N), Holt's linear algorithm has order (A,A,N) etc. Models are selected on the base of minimised AICc, which helps to select between additive and multiplicative models which produce the same point forecasts.

The models were reestimated at each forecast origin using real-time inflation measures. It turns out that the order of the model that minimises AICc for the KPIX is ARIMA$(0,1,0)(0,0,1)_{12}$ with a drift parameter for all forcasting origins. For CPI, however, the model order is not as stable although in 17 out of 20 instances an ARIMA$(1,1,1)(0,0,1)_{12}$ is selected. The model selected for the KPIX and CPI series is of order ETS(A,N,N), with no dampening, in all instances; i.e. a standard exponential smoothing model.

References

Andersson, M. K., Karlsson, G., and Svensson, J. (2007). "En utvärdering av riksbankens prognosförmåga." *Penning– och valutapolitik*, *2007*(3), 59–74.

Andrés, A. M., and Tejedor, I. H. (2003). "On the minimum expected quantity for the validity of the chi–squared test in 2×2 tables." *Journal of Applied Statistics*, *27*, 807–820.

Berg, C. (2000). "Inflation forecast targeting: The Swedish experience." In M. I. Blejer, A. Ize, A. M. Leone, and S. Werlang (Eds.), *Inflation targeting in practice: Strategic and operational issues and application to emerging market economies*, 28–36, International Monetary Fund.

Bergvall, A. (2005). "Uvärdering av Konjunkturinstitutets prognoser.", specialstudie Nr 5, mars 2005, Konjunkturinstitutet.

Blix, M., Friberg, K., and Åkerlind, F. (2002). "An evaluation of forecasts for the Swedish economy." *Penning– och valutapolitik*, *2002*(3), 39–73.

Blix, M., and Sellin, P. (1998). "Uncertainty bands for inflation forecasts.", working paper 65, Sveriges Riksbank.

Blix, M., and Sellin, P. (1999). "Inflationsprognos med osäkerhetsintervall." *Penning– och valutapolitik*, *1999*(2), 12–28.

Christoffersen, P. F. (1998). "Evaluating interval forecasts." *International Economic Review*, 841–862.

Clements, M. P. (2004). "Evaluating the Bank of England density forecasts of inflation." *Economic Journal*, *114*, 844–866.

Dowd, K. (2004). "The Swedish inflation fan charts: An evaluation of the riksbank's inflation density forecasts.", occasional Paper Series 2004–10, Nottingham University Business School.

Hyndman, R. (2008). "The forecasting package."

Hyndman, R. J., and Khandakar, Y. (2008). "Automatic time series forecasting: the forecast package for R." *Journal of Statistical Software*, to appear.

Jansson, P., and Vredin, A. (2003). "Forecast-based monetary policy: The case of sweden." *International Finance*, *6*, 349–380.

Konjunkturinstitutet (2002). "Penningpolitiken 1999–2001.", specialstudie Nr 1, mars 2002, Konjunkturinstitutet.

R Development Core Team (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Sveriges Riksbank (1996–2006). “Inflationsrapport (1996:1–2006:1).”

Wallis, K. E. (2003). “Chi–squared tests of interval and density forecasts, and the Bank of England’s fan charts.” *International Journal of Forecasting, 19*, 165–175.

Wallis, K. E. (2004). “An assessment of Bank of England and National Institute inflation forecast uncertainties.” *National Institute Economic Review, 189*, 64–71.

Woodward, M. (2007). “The case for forecast targeting as a monetary policy strategy.” *Journal of Economic Perspectives*, *21*, 3–24.