SVERIGES
RIKSBANK

Sveriges Riksbank Working Paper Series

Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach

AUGUST
2002

No.
138

Tor Jacobson and Sune Karlsson

WORKING PAPERS ARE OBTAINABLE FROM

Sveriges Riksbank • Information Riksbank • SE-103 37 Stockholm
Fax international: +46 8 787 05 26
Telephone international: +46 8 787 01 00
E-mail: info@riksbank.se

The Working Paper series presents reports on matters in the sphere of activities of the Riksbank that are considered to be of interest to a wider public. The papers are to be regarded as reports on ongoing studies and the authors will be pleased to receive comments.

Finding good predictors for inflation: A Bayesian Model Averaging Approach*

Tor Jacobson[†] Sune Karlsson[‡]

Sveriges Riksbank Working Paper Series
No 138
August 2002

Abstract

We consider a Bayesian Model Averaging approach for the purpose of forecasting Swedish consumer price index inflation using a large set of potential indicators, comprising some 80 quarterly time series covering a wide spectrum of Swedish economic activity. The paper demonstrates how to efficiently and systematically evaluate (almost) all possible models that these indicators in combination can give rise to. The results, in terms of out-of-sample-performance, suggest that Bayesian Model Averaging is a useful alternative to other forecasting procedures, in particular recognizing the flexibility by which new information can be incorporated.

Keywords: Variable selection, Markov chain Monte Carlo, Forecast
JEL-codes: C11, C51, C52, C53

*We would like to thank David Cesarini for excellent research assistance. We have benefitted from discussions with Mattias Villani and comments from seminar participants at Sveriges Riksbank; the dept. of Economics, Uppsala University; the dept. of Statistics, Stockholm University and the 2001 Winter Conference at Ammarnäs. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

[†]Sveriges Riksbank, 103 37 Stockholm, tor.jacobson@riksbank.se
[‡]Stockholm School of Economics, Box 6501, 113 83 Stockholm, sune.karlsson@hhs.se

1 Introduction

In 1993 Sweden introduced an explicit inflation target to guide monetary policy and thereby abandoned its long-standing fixed exchange rate regime. Sweden hereby joined a number of small, open economies pursuing inflation targeting; Australia (1993), Canada (1991), Israel (1991), New Zealand (1990) and the United Kingdom (1992). A more recent example is Brazil (1999).

The Swedish target is set to an annual inflation of 2 per cent, with an associated so called tolerance interval of ± 1 percentage unit, which can be interpreted as a limit for acceptable temporary deviations. The practical policy implementation of the Swedish inflation target policy is en route an inflation forecast. That is, the forecast acts as an intermediate target, a procedure known as inflation forecast targeting. The central bank strives towards the target by adjusting its policy instrument - the repurchase interest rate - in accordance to the development of economic conditions. A threatening increase in inflation is counteracted by an increase in the repo rate. Since the policy interventions are assumed to affect the economy with a considerable time lag (a rule of thumb is 6 to 8 quarters for the maximum effect) successful policy hinges on accurate forecasts of relevant variables, especially the target variable.[1] Much of the analysis underlying monetary policy at a central bank is therefore devoted to forecasting exercises. This is probably true in general and in particular so if the bank is an explicit inflation targeter.[2]

The Swedish Riksbank publishes official inflation forecasts in its quarterly Inflation Report. These forecasts are not the outcomes of any one formal forecasting model, but rather an informal blend of forecasts from many models, reflecting a desire to consider a wide selection of inflation indicators. Moreover, the published forecast has been adjusted for judgements. This reflects a need to take account of recent information, which *per se* may or may not be easily evaluated in a formal model, but nevertheless has to be considered despite a binding time constraint. The purpose of this paper is not to suggest ways of improving the judgemental component of the official inflation forecast process. It is rather to the first aspect above we want to contribute by demonstrating a method that will efficiently and systematically evaluate a wide selection of inflation indicators and (almost) all possible models that these indicators in combination can give rise to. Specifically, we want to formally explore the idea of combining forecasts from various indicator models by using Bayesian Model Averaging.

The idea that forecast performance can be improved by combining forecasts from different models dates back at least to the influential work of Bates and Granger (1969). See Clemen (1989) for a review of the literature on forecast combination. Viewing forecast combination as an application of Bayesian Model Averaging has several advantages. Firstly it provides a rigorous statistical foundation for the exercise where the weights assigned to the different forecasts arise naturally as the posterior probabilities of the models. Secondly, the posterior probabilities provide a ranking of the models and, thirdly, we can rank the predictive ability of the variables by their

[1]It is also the case that the central bank cares about output stabilization in the sense that if the economy is hit by shock that threatens price stability, the bank will not at all costs try to immediately eliminate the effect of the shock by a drastic increase of the repo rate, but rather gradually do so over a longer horizon.

[2]For a full account of inflation targeting monetary policy, see Svensson (1999).

posterior probability of being included in a hypothetical true model.

The next section provides the details of Bayesian Model Averaging. Section 3 presents the empirical results and section 4 concludes.

2 Bayesian Model Averaging and Model Selection

Bayesian model averaging (BMA) and model selection has given rise to a large literature, see Hoeting, Madigan, Raftery and Volinsky (1999) for references and an overview and Palm and Zellner (1992) and Min and Zellner (1993) in relation to forecasting.

The Bayesian treatment of model uncertainty is, in principle, straightforward and parallels the treatment of parameter uncertainty. Given a set $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_M\}$ of possible models, prior probabilities of the models, $p(\mathcal{M}_i)$, prior distributions of the parameters in each model, $p(\theta_i|\mathcal{M}_i)$ and likelihoods, $L(\mathbf{y}|\theta_i, \mathcal{M}_i)$ all quantities of interest for model averaging and selection can be obtained by using Bayes rule. The posterior probabilities of the models are given by

$$p(\mathcal{M}_i|\mathbf{y}) = \frac{m(\mathbf{y}|\mathcal{M}_i)\, p(\mathcal{M}_i)}{\sum_{j=1}^{M} m(\mathbf{y}|\mathcal{M}_j)\, p(\mathcal{M}_j)} = \left[\sum_{j=1}^{M} \frac{m(\mathbf{y}|\mathcal{M}_j)\, p(\mathcal{M}_j)}{m(\mathbf{y}|\mathcal{M}_i)\, p(\mathcal{M}_i)}\right]^{-1}$$

where $m(\mathbf{y}|\mathcal{M}_i)$ is the marginal likelihood,

$$m(\mathbf{y}|\mathcal{M}_i) = \int L(\mathbf{y}|\theta_i, \mathcal{M}_i)\, p(\theta_i|\mathcal{M}_i)\, d\theta_i, \tag{1}$$

for model i. The posterior distribution of some quantity of interest, Δ, when taking account of model uncertainty, is then simply the weighted average of the posterior distributions for each model,

$$p(\Delta|\mathbf{y}) = \sum_{j=1}^{M} p(\Delta|\mathbf{y}, \mathcal{M}_j)\, p(\mathcal{M}_j|\mathbf{y}). \tag{2}$$

In particular, the minimum mean squared error forecast is given by

$$\widetilde{y}_{T+h} = E(y_{T+h}|\mathbf{y}) = \sum_{j=1}^{M} \widetilde{y}_{T+h.j}\, p(\mathcal{M}_j|\mathbf{y}) \tag{3}$$

for $\widetilde{y}_{T+h.j} = E(y_{T+h}|\mathbf{y}, \mathcal{M}_j)$ the forecast conditional on model j. Madigan and Raftery (1994) note that averaging over all the models in this fashion provides better predictive ability as measured by a logarithmic scoring rule than using any single model in $\mathfrak{M}$.

In a variable selection problem, the posterior probability that variable i is in the "true" model is given by

$$p(x_i|\mathbf{y}) = \sum_{j=1}^{M} I(x_i \in \mathcal{M}_j)\, p(\mathcal{M}_j|\mathbf{y}) \tag{4}$$

where $I(x_i \in \mathcal{M}_j)$ is one if x_i is included in model j and zero otherwise.

Table 1 Interpretation of Bayes Factors

$B_{ij} > 1$	Support for $\mathcal{M}_i$
$10^{-1/2} < B_{ij} < 1$	Very slight evidence against $\mathcal{M}_i$
$10^{-1} < B_{ij} < 10^{-1/2}$	Slight evidence against $\mathcal{M}_i$
$10^{-2} < B_{ij} < 10^{-1}$	Strong to very strong evidence against $\mathcal{M}_i$
$B_{ij} < 10^{-2}$	Decisive evidence against $\mathcal{M}_i$

If a loss-structure satisfying

$$\begin{array}{c} l_{ii} = 0 \\ l_{ij} > 0,\ i \neq j \end{array},$$

where l_{ij} is the loss associated with choosing model i when j is the true model, is available, the posterior expected loss is minimized by choosing model i over model j if the Bayes factor satisfies

$$B_{ij} = \frac{m\left(\mathbf{y}|\mathcal{M}_i\right)}{m\left(\mathbf{y}|\mathcal{M}_j\right)} > \frac{l_{ij}}{l_{ji}} \frac{P\left(\mathcal{M}_i\right)}{P\left(\mathcal{M}_j\right)}.$$

If no formal loss-structure is available, informal decisions can be based on the Bayes factor directly using the rule of thumb suggested by Jeffreys (Table 2). Note that the Bayes factor alternatively can be written as

$$B_{ij} = \frac{P\left(\mathcal{M}_i|\mathbf{y}\right)}{P\left(\mathcal{M}_j|\mathbf{y}\right)} \Big/ \frac{P\left(\mathcal{M}_i\right)}{P\left(\mathcal{M}_j\right)}$$

and is thus a measure of how much our beliefs in model i relative to model j has changed as a result of processing the data.

2.1 Prior specification and the posterior

As the number of models grow large specification of the prior distribution of the parameters in the different models becomes a major difficulty in Bayesian model averaging and model selection. The sheer size of the problem of specifying a coherent set of priors for all models is overwhelming, even if we have quite sharp and well articulated prior beliefs about the effects of some of the variables.

The situation is further complicated by the indeterminacy of the marginal likelihoods (1) when "standard" improper uninformative priors are used. By noting that the marginal likelihoods enter in ratio form in the quantities of interest it is possible to use improper priors on the parameters that are common to all models since the indeterminate normalizations of improper priors cancel for these parameters. In our case of linear regression models we have two parameters that are common to all models, the constant term and the error variance. In order to ensure that the constant term has a consistent interpretation as the unconditional mean of the dependent variable we work with the explanatory variables in deviation form. This also has the advantage of making the constant explanatory variable orthogonal to ordinary explanatory variables and simplifies the posterior calculations. Following Fernández, Ley and Steel (2001) we specify these diffuse priors as

$$p\left(\sigma^2\right) \propto 1/\sigma^2, \tag{5}$$

the usual uninformative prior for the variance, and $p(\alpha) \propto 1$, a uniform prior on the real line. For the ordinary regression parameters Fernández et al. suggest a g-prior (Zellner 1986)

$$p\left(\beta_j|\sigma^2, \mathcal{M}_j\right) \sim N\left(0, c\sigma^2\left(\mathbf{X}_j'\mathbf{X}_j\right)^{-1}\right), \tag{6}$$

that is the prior variance is proportional to the data information and the prior mean is set to zero indicating shrinkage of the posterior towards zero. The constant c remains to be chosen. Based on a Monte Carlo study Fernández et al. recommends choosing c as

$$c = \begin{cases} k^2, & n \le k^2 \\ n, & n > k^2 \end{cases}$$

where k is the number of regressors considered.

This yields a proper posterior on the regression parameters, writing $\theta_j = \left(\alpha, \beta_j'\right)'$ we have the posterior as a t-distribution with $n-1$ degrees of freedom,

$$p\left(\theta_j|\mathbf{y}\right) \sim t_{k_j}\left(\overline{\theta}_j, S_j, \mathbf{M}_j, n-1\right)$$

where

$$\mathbf{M}_j = \begin{pmatrix} n & 0 \\ 0 & \frac{c+1}{c}\mathbf{X}_j'\mathbf{X}_j \end{pmatrix}$$

$\overline{\theta}_j = \left(\overline{\alpha}, \overline{\beta}_j\right)$ with $\overline{\alpha} = \overline{y}, \overline{\beta}_j = \frac{c}{c+1}\widehat{\beta}_j$, a scaled down version of the least squares estimate, and $S_j = \frac{c}{c+1}\left(\mathbf{y} - \mathbf{Z}_j\widehat{\theta}_j\right)'\left(\mathbf{y} - \mathbf{Z}_j\widehat{\theta}_j\right) + \frac{1}{c+1}\left(\mathbf{y} - \overline{y}\iota_n\right)'\left(\mathbf{y} - \overline{y}\iota_n\right)$. Although the marginal likelihood is indeterminate due to the improper priors on α and σ^2 it is easy to verify that

$$m\left(\mathbf{y}|\mathcal{M}_j\right) \propto (c+1)^{-k_j} S_j^{-(n-1)/2}$$

after having dropped the model invariant factors. The Bayes factors are thus given by

$$B_{ij} = (c+1)^{k_j - k_i}\left(\frac{S_i}{S_j}\right)^{-(n-1)/2}.$$

There are, of course, other possible choices of more or less automatic prior distributions. Smith and Kohn (2000) combine the diffuse prior (5) on σ^2 with a proper data based prior for all the regression parameters conditional on σ^2,

$$p\left[\left(\alpha, \beta_j'\right)'|\sigma^2, \mathcal{M}_j\right] \sim N\left(\left(\mathbf{Z}_j'\mathbf{Z}_j\right)^{-1}\mathbf{Z}_j'\mathbf{y}, n\sigma^2\left(\mathbf{Z}_j'\mathbf{Z}_j\right)^{-1}\right) \tag{7}$$

where $\mathbf{Z}_j = \left(\iota_n, \mathbf{X}_j\right)$. This prior thus takes the scaling factor $c = n$, making the information in the prior comparable to the information in one observation, and centers the prior on the data rather than on zero. Another possibility is to bypass the need to explicitly specify informative priors on the parameters by using the Fractional Bayes factors of O'Hagan (1995).[3] We will, however, not use these approaches in the application since they are problematic from a decision theoretic point of view. The prior (7) is tantamount to using the data twice. The Fractional Bayes factor seems to work well for model choice but it is not clear how one should interpret posterior

[3] Smith and Kohn (2000) attribute the prior (7) to O'Hagan (1995). It should, however, be noted that this prior does not lead to the Fractional Bayes factors suggested by O'Hagan.

model probabilities calculated from Fractional Bayes factors and model averaging can be problematic.

It remains to specify the prior model probabilities, $p(\mathcal{M}_j)$. A useful, informative, specification of the prior probabilities is achieved if we let w_i be the prior probability that variable i is included in the "true" model. The prior model probability under independence is then

$$p(\mathcal{M}_j) \propto \prod_{i=1}^{k} w_i^{\gamma_i} (1 - w_i)^{1-\gamma_i} \tag{8}$$

where γ_i is an indicator variable taking the value 1 if variable i is included in the model and is zero otherwise. In the absence of strong prior beliefs we can set $w_i = 1/2$, making all the models equally likely a priori. The drawback of the latter choice of w_i is that this induces an informative prior on model size, making larger models more likely a priori by virtue of their greater number. If this is deemed undesirable a useful alternative specification of the prior probabilities is

$$p(\mathcal{M}_j) \propto k_j^{-\delta} \tag{9}$$

for $\delta > 1$ and k_j is the number of variables included in model j.

2.2 Traversing the model space

A major difficulty is created by the size of the model space if the number of potential explanatory variables, k, is large. If every possible combination of explanatory variables is considered we have 2^k possible models. In the present we consider 86 possible indicator variables, see Tables A1-A9 in the Appendix. These variables can in combination give rise to roughly 10^{25} possible models. In order to limit the number of possible models and also to avoid ridiculously large models we limit the number of explanatory variables to $k^* = 20$ variables in a single model. This does, however, still yield a very large number of possible models,

$$\mathbb{C}_{\mathfrak{M}} = 1 + \sum_{k=1}^{20} \binom{86}{k} \approx 2.5 \times 10^{19}.$$

Traversing the complete model space, calculating the posterior probabilities and ultimately the sums in (3) and (4) yielding the BMA forecast and the posterior inclusion probabilities of the variables is thus impractical. Madigan and Raftery (1994) suggested reducing the model space by only considering models which receive non-negligible posterior probabilities, i.e. by restricting attention to the subset of $\mathfrak{M}$

$$\mathfrak{M}^* = \left\{ \mathcal{M}_i : \frac{p(\mathcal{M}_k|\mathbf{y})}{\max_j p(\mathcal{M}_j|\mathbf{y})} \geq C \right\} \tag{10}$$

for some predetermined cutoff C. All calculations are then performed conditional on the model set $\mathfrak{M}^*$ rather than the original set $\mathfrak{M}$.

A remaining difficulty is identifying the set $\mathfrak{M}^*$ without, in fact, traversing the full model space. Markov chain Monte Carlo (MC^2) using the reversible jump algorithm of Green (1995) to cope with changing model dimensions turns out to be a computationally convenient method of achieving this. The algorithm can be described as follows. If the current state of the chain is $(\theta_{\mathcal{M}}, \mathcal{M})$

1. Propose a jump to a new model $\mathcal{M}^*$ with probability $j(\mathcal{M}^*|\mathcal{M})$.

2. Generate a vector $\mathbf{u}$ from a continuous distribution $q(\mathbf{u}|\theta_{\mathcal{M}}, \mathcal{M}, \mathcal{M}^*)$.

3. Set $(\theta_{\mathcal{M}^*}, \mathbf{u}^*) = g_{\mathcal{M},\mathcal{M}^*}(\theta_{\mathcal{M}}, \mathbf{u})$ where g is a bijection and $\mathbf{u}$ and $\mathbf{u}^*$ satisfy $\dim(\mathbf{u}) + k_{\mathcal{M}} = \dim(\mathbf{u}^*) + k_{\mathcal{M}^*}$.

4. Accept the move with probability

$$\alpha = \min\left\{1, \frac{L(\mathbf{y}|\theta_{\mathcal{M}^*}, \mathcal{M}^*) p(\theta_{\mathcal{M}^*}|\mathcal{M}^*) p(\mathcal{M}^*) j(\mathcal{M}|\mathcal{M}^*) q(\mathbf{u}^*|\theta_{\mathcal{M}^*}, \mathcal{M}^*, \mathcal{M})}{L(\mathbf{y}|\theta_{\mathcal{M}}, \mathcal{M}) p(\theta_{\mathcal{M}}|\mathcal{M}) p(\mathcal{M}) j(\mathcal{M}^*|\mathcal{M}) q(\mathbf{u}|\theta_{\mathcal{M}}, \mathcal{M}, \mathcal{M}^*)} \times \left|\frac{\partial g_{\mathcal{M},\mathcal{M}^*}(\theta_{\mathcal{M}}, \mathbf{u})}{\partial(\theta_{\mathcal{M}}, \mathbf{u})}\right|\right\}$$

and set $\mathcal{M} = \mathcal{M}^*$ if the move is accepted.

Choosing $q(\mathbf{u}|\theta_{\mathcal{M}}, \mathcal{M}, \mathcal{M}^*) = p(\theta_{\mathcal{M}^*}|\mathcal{M})$ and by implication defining the transformation $g_{\mathcal{M},\mathcal{M}^*}$ by $\theta_{\mathcal{M}^*} = \mathbf{u}, \mathbf{u}^* = \theta_{\mathcal{M}}$ simplifies the acceptance probability to

$$\alpha = \min\left\{1, \frac{m(\mathbf{y}|\mathcal{M}^*) p(\mathcal{M}^*) j(\mathcal{M}|\mathcal{M}^*)}{m(\mathbf{y}|\mathcal{M}) p(\mathcal{M}) j(\mathcal{M}^*|\mathcal{M})}\right\}$$

since there is no need to perform steps 2 and 3 of the algorithm.

We will consider two types of model changing moves:

(a) Draw a variable at random and drop it if it is in the model or add it to the model (if $k_{\mathcal{M}} < k^*$). This step is attempted with probability p_A.

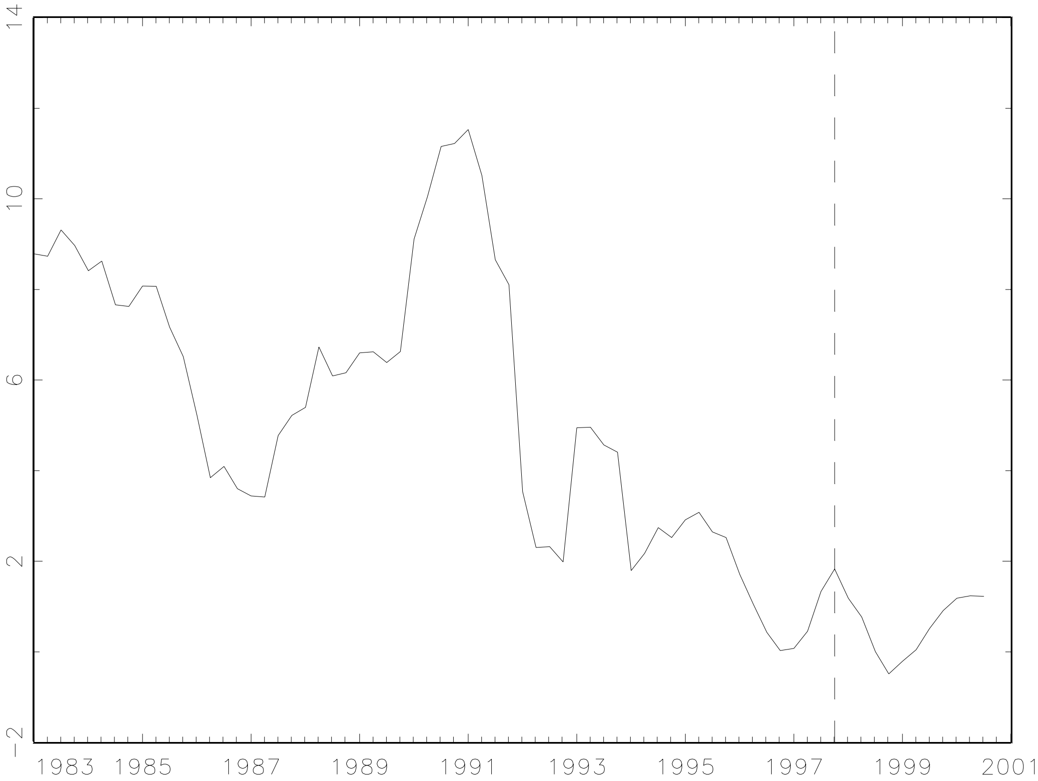
(b) Swap a randomly selected variable in the model for a randomly selected variable outside the model (if $k_{\mathcal{M}} > 0$). This step is attempted with probability $1 - p_A$.

Within each move type the probability of an allowed move is constant, $\frac{1}{k}$ for move (a) and $\frac{1}{k_{\mathcal{M}}} \frac{1}{k-k_{\mathcal{M}}}$ for move (b). Since the move types do not commute the proposal probabilities $j(\cdot|\cdot)$ drop out of the acceptance probability. If, in addition, we use the model prior $p(\mathcal{M}) \propto \prod_{i=1}^{k} w_i^{\gamma_i} (1 - w_i)^{1-\gamma_i}$ with $w_i = 1/2$ the model probabilities are constant and the acceptance probability simplifies further to

$$\alpha = \min\left\{1, \frac{m(\mathbf{y}|\mathcal{M}^*)}{m(\mathbf{y}|\mathcal{M})}\right\}.$$

The Markov Chain outlined above converges to the posterior model probabilities under quite general conditions and provides one way of estimating $p(\mathcal{M})$. We do view this mainly as a way of exploring the model space and identifying the subset $\mathfrak{M}^*$ of "important" models by approximating sampling from the posterior model distribution. As such we are not particularly interested in the convergence of the chain and will run the chain considerably shorter than $\mathbb{C}_{\mathfrak{M}}$ draws. A simple convergence diagnostic is, however available by comparing the estimated model probabilities with the exact model probabilities calculated using a closed form for the marginal likelihood conditional on the set of models visited by the chain.

Figure 1 The Swedish inflation rate 1983Q1 - 2000Q3



3 Forecasting inflation

Since our primary goal is forecasting we do not attempt to develop models for the inflation rate with causal interpretations. Instead we focus on simple regression models of the form

$$y_{t+h} = \alpha + \gamma d_{t+h} + \mathbf{x}_t \beta + \varepsilon_t \qquad (11)$$

with the aim to forecast h time periods ahead. The constant term α and a dummy variable, d_t, for the low inflation regime starting in 1992Q1 is always included in the model whereas the members of $\mathbf{x}_t$ are selected from the set of potential predictors. While this might seem an overly simplistic and static model formulation at first there is nothing preventing us from including lags of variables in $\mathbf{x}_t$, the model can thus allow for quite complicated dynamics in the inflation rate. Another, slightly unusual, feature of the model class is the use of the h period lead, y_{t+h}, instead of y_t as the dependent variable. This choice of dependent variable has the great advantage that it does away with the need of forecasting the predictors in $\mathbf{x}_t$ when forecasting y_{t+h}. In essence we view (11) as the reduced form of a joint model for y_t and $\mathbf{x}_t$. The obvious disadvantage of this choice of dependent variable is that it leads to a different model for each forecast horizon. We consider this to be a small price to pay for the considerable reduction in the complexity of the modeling task.

The simplicity of the model class allows us to consider a wide range of explanatory variables and possible forecasting models. For the application at hand we have quarterly data for the period 1983Q1 to 2000Q3 on the 86 predictor variables listed in the Appendix. This set of variables includes a wide range of indicators of real and monetary aspects of the Swedish economy and is close to an exhaustive set of potential predictors for the inflation rate. Note that we include (the current level of) inflation in the set of predictor variables for inflation h periods ahead. Inflation is measured as the 4 quarter percentage change in the consumer price index and the remaining variables are with few exceptions 4 quarter growth rates or 4 quarter log differences.

We evaluate the performance by producing 4 and 8 quarter ahead forecasts for the period 1997Q4 to 2000Q3. The Swedish inflation rate is depicted in Figure 1 with the beginning of the evaluation period indicated by the dashed line.

3.1 Implementation

We use the prior (5, 6) suggested by Fernández et al. (2001) with $c = (k^*)^2 = 400$, corresponding to our upper limit of 20 predictor variables. For the prior model probabilities we use the specification (8) with $w_i = 1/2$, i.e. a uniform prior on the models.

To assess the performance of the Markov Chain we started it at several different models, including the null model and models containing a full set of 20 predictor variables and let it run 500,000 steps.[4] We then calculated the exact posterior model probabilities and variable inclusion probabilities conditional on the models visited by the Markov Chain. In each case this resulted in the same 10 models with highest posterior probability and virtually identical inclusion probabilities for the predictor variables.

3.1.1 Adding lags

To allow for richer dynamics we also include 3 lags of the variables in the set of potential predictors for inflation. Including 3 lags of all the variables would create a prohibitively large set of predictor variables. Instead we use a preliminary run of the Markov chain to select a subset of the variables and use these variables with 3 additional lags as the set of potential regressors in the final run of the Markov chain. To be more precise, we select the 20 variables with the highest inclusion probability in the preliminary run, to this set we add Infla, USD, DEM, R3M, R5Y, R5YR3M, R10Y, Unemp, NAIRU and OutGap if they are not already selected. Let $\mathbf{x}_t^*$ denote this set of variables, the set of potential predictors in the final run is then $\mathbf{x}_t = \left(\mathbf{x}_t^*, \mathbf{x}_{t-1}^*, \mathbf{x}_{t-2}^*, \mathbf{x}_{t-3}^*\right)$. Forecasts, model and variable inclusion probabilities are based on the results from the final run of the chain.

3.2 Models and variables

In the preliminary variable selection run the Markov chain is run 5,000,000 steps visiting between 60,000 and 90,000 models for the different samples. Following George and McCulloch (1997) we use a secondary chain run for 1,000,000 steps and started at a random model to estimate the coverage probability of the primary chain. For the 2000Q3 four step ahead forecast the set $\mathfrak{M}^*$ visited by the primary chain accounts for 92% of the posterior mass. The set of selected (top 20) variables is fairly constant as

[4]The procedure is coded in Fortran. 1,000 steps of the Markov chains takes about 1 CPU second on a 700 MHz Pentium III.

Table 2 Variables selected in first step when forecasting 2000Q3

4 quarters ahead		8 quarters ahead	
Variable	Posterior prob	Variable	Posterior prob
Pp1664	0.99	LPrdF	1.00
InfRel	0.88	AvJob	0.99
Empld	0.47	M0TCW	0.58
PrvEmp	0.46	InfExp	0.45
R5YR3M	0.20	OPrice	0.43
NHouse	0.20	InfImp	0.41
InfFor	0.18	AFGX	0.39
Infla	0.13	RCnsEx	0.21
ExpInf	0.10	DISK	0.18
AFGX	0.10	Pp5064	0.14
BCI	0.09	R10Y	0.12
InfxMr	0.09	R5Y	0.08
M0	0.09	InfHWg	0.04
Pp024	0.05	NUnmp	0.04
OutGap	0.04	Pp6574	0.04
M3	0.04	CnsExp	0.04
NEmpl	0.03	NA4Wrk	0.04
R3M	0.03	WageMM	0.03
USD	0.03	PPP	0.03
InfUnd	0.02	InfMrt	0.02

the sample size increases. Most of the a priori included variables are in the top 20 set and do not have to be added. With the 4 quarter ahead forecast variables are only added for 4 of the 12 forecasts in the evaluation period. With the 8 quarter ahead forecast variables are added for all forecasts, for 4 of the forecasts all of the a priori variables are added to the top 20 variables, for one forecast all but R10Y is added, for one forecast all but R10Y and R5Y are added and for the remaining 6 forecasts Infla, USD, DEM and R3M are added. The variables selected for the final forecasting period, 2000Q3, are displayed in Table 2. Note that the posterior probabilities in all the tables are conditional on the set $\mathfrak{M}^*$ of models visited by the Markov chain.

In the second run the top 20 variables and any of the a priori variables are included in $\mathbf{x}_t^*$ with three additional lags allowing for varying time delays and richer dynamics. The Markov Chain is again run 5,000,000 steps and forecasts calculated based on the 10 models with the highest posterior probabilities for each of the forecast periods and by averaging over the forecasts from all models using the posterior model probabilities.[5] The top 10 models along with the variables with highest posterior probabilities are displayed in Tables 3 and 4 for the 2000Q3 forecast period. The posterior distribution over model sizes in the final step is displayed in Table 5 for the 2000Q3 forecasts. It is quite clear that the restriction to allow no more than 20 variables in the model is not binding. The Markov chain did, in fact, never visit a model with more than 17

[5] In the second run we estimate the probability content of the set $\mathfrak{M}^*$ to 91% for the 2000Q3 four step ahead forecast.

Table 3 Posterior Model and Variable Probabilities, 4 quarter ahead forecast for 2000Q3

Variables		Models									
	Prob	1	2	3	4	5	6	7	8	9	10
$Pp1664_t$	1.00	x	x	x	x	x	x	x	x	x	x
$R5YR3M_{t-1}$	0.98	x	x	x	x	x	x	x	x	x	x
$InfRel_t$	0.96	x	x	x	x	x	x	x	x	x	x
$PrvEmp_t$	0.46		x	x	x	x			x	x	x
$PrvEmp_{t-1}$	0.43	x					x	x			
$M3_{t-1}$	0.33	x				x			x	x	
$M0_t$	0.26			x			x				
$AFGX_{t-1}$	0.18		x							x	x
$AFGX_t$	0.16				x				x		
$Empld_t$	0.10										
$M3_{t-2}$	0.06										x
Posterior Prob×100		11	7	4	4	4	3	3	2	2	1

variables.

3.3 Forecast performance

The forecast performance is summarized in Table 6 and compared with the results for a random walk (no change) forecast. The four quarter and eight quarter ahead forecasts are displayed in Figures 2 and 3. The performance of the four quarter ahead BMA forecast is quite good with a Root Mean Square Error (RMSE) of 0.63 and clearly outperforms the random walk forecasts with a RMSE of 1.30. The BMA forecast also compares favorably with the track record of about 10 professional forecasters. Blix, Wadefjord, Wienecke and Ådahl (2001) survey the forecast performance of professional forecasters, for the time period and forecast horizon closest to our four quarter ahead forecasts we find an RMSE of 0.99 for the professional forecasters.[6]

The performance of the eight quarter ahead forecasts is, in contrast, quite disappointing. From Figure 3 we see that the failure occurs in the beginning of the evaluation period when the forecasts extrapolate the downwards trend without catching the trend break. This is presumably due to the simplicity of the forecasting model; the dynamics is not rich enough, or that the built-in delay of eight quarters between the dependent and the explanatory variables is simply too large.

4 Conclusions

This paper has demonstrated how a Bayesian treatment of model uncertainty - in an efficient and systematic way - can help us specify indicator models when the number of potential indicators becomes very large. We apply the method to the problem of

[6]Table 2 of Blix et al. (2001) gives the professional forecasters RMSE for January forecasts of 1998, 1999 and 2000 current year inflation as 1.66, 0.25 and 0.33, yielding an overall RMSE of 0.99 for the three years.

Table 4 Posterior Model and Variable Probabilities, 8 quarter ahead forecast for 2000Q3

Variables	Prob	1	2	3	4	5	6	7	8	9	10
$AvJob_t$	0.97	x	x	x	x	x	x	x	x	x	x
$LPrdF_t$	0.72	x	x					x		x	x
PPP_{t-1}	0.38	x	x					x		x	x
$InfImp_t$	0.36			x	x	x	x		x		
$OPrice_t$	0.33			x	x	x	x		x		
$M0TCW_t$	0.33						x				x
$WageMM_{t-2}$	0.31			x		x	x		x		
$LPrdF_{t-2}$	0.30			x	x	x	x		x		
$Pp5064_t$	0.25	x					x	x			x
$AFGX_t$	0.16										
$DISK_t$	0.15			x	x						
$Pp5064_{t-1}$	0.12		x								
$InfHWg_t$	0.11							x			
$Pp6574_t$	0.09						x				
$RCnsEx_t$	0.08										
$InfImp_{t-2}$	0.07										
$InfExp_t$	0.07										
PPP_{t-2}	0.06										
$NA4Wrk_t$	0.06										
$WageMM_{t-3}$	0.04										
$R3M_t$	0.04					x					
$InfImp_{t-1}$	0.04										
$CnsExp_{t-1}$	0.04										
$AFGX_{t-3}$	0.03										
$M0TCW_{t-1}$	0.03										
$RCnsEx_{t-1}$	0.02										
$Pp5064_{t-2}$	0.02									x	
Posterior Prob×100		9	6	2	2	2	2	2	2	1	1

Table 5 Posterior Distribution of Model Sizes, 2000Q3

Model Size	3	4	5	6	7	8	9	10	11
Posterior Prob, 4 step	0.00	0.04	0.44	0.39	0.12	0.01	0.00	0.00	0.00
Posterior Prob, 8 step	0.00	0.01	0.22	0.25	0.25	0.18	0.07	0.02	0.00

Table 6 Root mean square error of forecasts of inflation 1997Q4 to 2000Q3

		Top 10 models										Rand.
	BMA	1	2	3	4	5	6	7	8	9	10	walk
4 step	0.63	0.76	0.74	0.54	0.77	0.71	0.79	1.06	0.82	0.91	0.68	1.30
8 step	2.95	2.72	3.91	3.21	4.39	2.78	3.80	4.18	3.48	2.90	3.32	0.63

Figure 2 Four quarter ahead forecasts

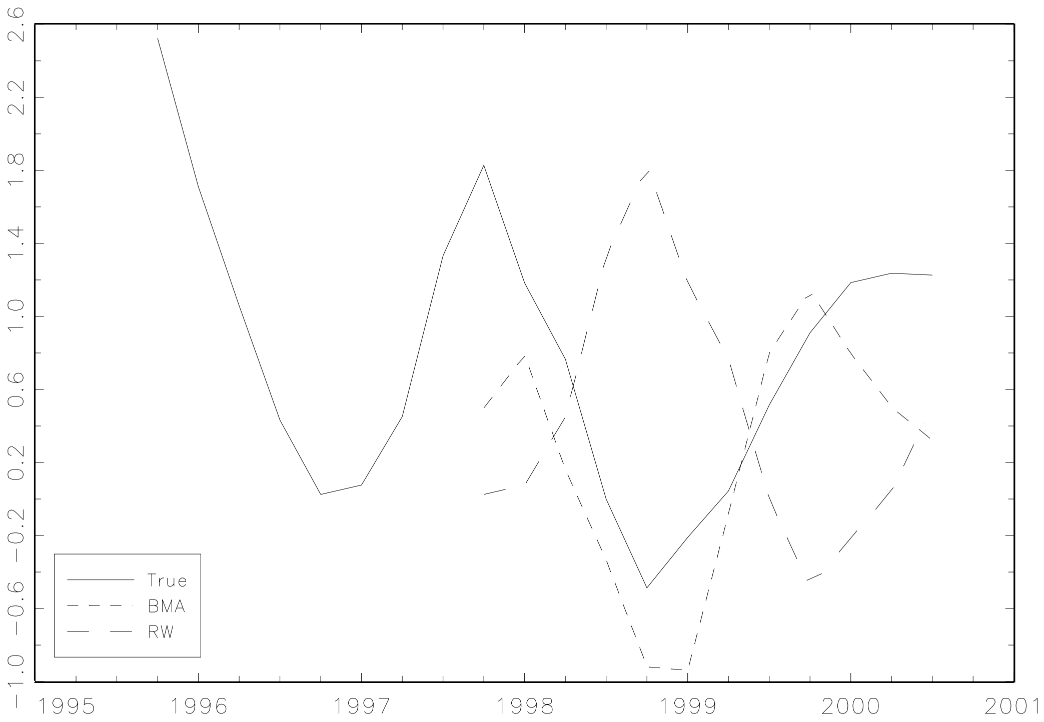
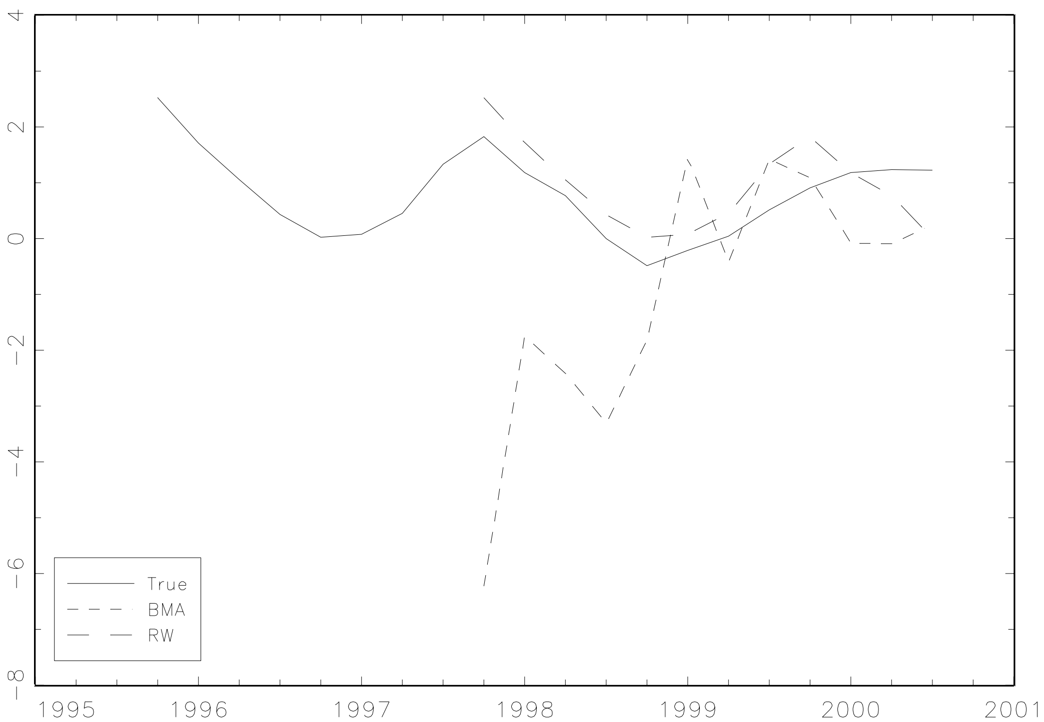


Figure 3 Eight quarter ahead forecasts



specifying forecasting models for Swedish consumer price index inflation using some 80 quarterly indicators. Exploring the idea of Bayesian model averaging, we find that combining forecasts from the 10 highest ranked indicator models (in terms of posterior probabilities) yields robust forecasts with, in general, smaller root mean squared errors than the included individual models display.

The central bank focuses on inflation forecasts 4 to 8 quarters ahead in its management of monetary policy. According to our results there is a dramatic difference in forecasting performance between these two horizons. The RMSE for the 4-quarters ahead Bayesian average forecasts is 0.63, which compares favorably with the track record of professional forecasters and is less than half of what a naive, random walk model gives. In other words, on average predictions by the combined indicator model easily outperforms the predictions that inflation a year from now will be what inflation is today. Given the notorious high persistence in inflation, this is not an obvious outcome. However, when predicting 8 quarters ahead we find that the indicator forecasts are considerably worse, the RMSE is almost 5 times larger than for the 4 quarter ahead predictions and for a random walk model.

It is interesting to note that the 10 top models for the 4 quarter ahead forecasts for 2000Q3 all involve the same three indicators; the share of total population in the ages 16 to 64, the yield curve lagged one quarter, and the ratio of domestic to foreign inflation. Other important indicators are; employment, in total and in the private sector, broad and narrow money, and the stock market index. All these variables, and the first three in particular, share an attractive feature with respect to forecasting. They are not likely to be subject to measurement errors, hence the real-time observation will in general not be revised subsequently. This means that in practise a four-quarter forecasting horizon will indeed be four quarters and not shorter. It is also re-assuring that the included variables pertain to a reasonable blend of the nominal and real sectors; with a wide coverage of changes in economic activity. This should yield a certain robustness to the forecasts in comparison with a case where one particular sector dominates, e.g., the financial markets.

References

Bates, J. M. and Granger, C. W. J. (1969), 'The combination of forecasts', *Operations Research Quarterly* **20**, 451–468.

Blix, M., Wadefjord, J., Wienecke, U. and Ådahl, M. (2001), 'How good is the forecasting performance of major institutions?', *Sveriges Riksbank Economic Review* (2001:3), 38–68.

Clemen, R. T. (1989), 'Combining forecasts: A review and an annotated bibliography', *International Journal of Forecasting* **5**, 559–583.

Fernández, C., Ley, E. and Steel, M. F. J. (2001), 'Benchmark priors for Bayesian model averaging', *Journal of Econometrics* **100**, 381–427.

George, E. I. and McCulloch, R. E. (1997), 'Approaches for Bayesian variable selection', *Statistica Sinica* **7**, 339–373.

Green, P. J. (1995), 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika* **82**, 711–732.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999), 'Bayesian model averaging: A tutorial (with discussion)', *Statistical Science* **14**, 382–417. Corrected version available at http://www.stat.washington.edu/www/research/online/hoeting1999.pdf.

Madigan, D. and Raftery, A. E. (1994), 'Model selection and accounting for uncertainty in graphical models using Occam's window', *Journal of the American Statistical Association* **89**, 1535–1546.

Min, C.-K. and Zellner, A. (1993), 'Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting and international growth rates', *Journal of Econometrics* **56**, 89–118.

O'Hagan, A. (1995), 'Fractional Bayes factors for model comparison (with discussion)', *Journal of the Royal Statistical Society, Ser. B* **57**, 99–138.

Palm, F. C. and Zellner, A. (1992), 'To combine or not to combine? issues of combining forecasts', *Journal of Forecasting* **11**, 687–701.

Smith, M. and Kohn, R. (2000), 'Nonparametric seemingly unrelated regression', *Journal of Econometrics* **98**, 257–281.

Svensson, L. E. O. (1999), 'Inflation targeting as a monetary policy rule', *Journal of Monetary Economics* **43**, 607–654.

Zellner, A. (1986), On assessing prior distributions and Bayesian regression analysis using g-prior distribtutions, *in* P. K. Goel and A. Zellner, eds, 'Bayesian Inference and Decision Techniques - Essays in Hounour of Bruno de Finetti', North-Holland, chapter 15, pp. 233–243.

A Appendix

Table A.1 Financial variables

Variable	Description	Transformation Level	$\ln y_t - \ln y_{t-4}$
GovDebt	Government debt		X
Bank2Pub	Bank lending to public		X
CI2Pub	Credit inst. lending to public		X
CI2Priv	Credit inst. lending to priv. sector		X
AFGX	Affärsvärlden stock index		X
REPO	Repo rate	X	
DISK	Discount rate	X	
R3M	3 month money market rate	X	
R5Y	5 year government bond rate	X	
R5YR3M	Yield curve	X	
R10Y	10 year government bond rate	X	

Table A.2 Exchange rates

Variable	Description	Transformation Level	$\ln y_t - \ln y_{t-4}$
NFX	Effective exchange rate (TCW)		X
RFX	Effective real exchange rate (TCW)		X
USD	SEK/USD exchange rate		X
DEM	SEK/DEM exchange rate		X

Table A.3 Money supply

Variable	Description	Transformation Level	$\ln y_t - \ln y_{t-4}$
M0	Narrow money		X
M3	Broad money		X
M0TCW	TCW-weighted M0		X
M3TCW	TCW-weighted M3		X
M3EU	Broad money (EU-harmonized)		X

Table A.4 Population

Variable	Description	Transformation $\ln y_t - \ln y_{t-4}$
PpTot	Total population	X
Pp1664	Share in ages 16-64	X
Pp014	Share in ages 0-14	X
Pp1529	Share in ages 15-29	X
Pp3049	Share in ages 30-49	X
Pp5064	Share in ages 50-64	X
Pp6574	Share in ages 65-74	X
Pp75+	Share 75 and older	X

Table A.5 Labor costs

Variable	Description	Transformation $\ln y_t - \ln y_{t-4}$
WCSS	Wages incl. social security	X
WgCst	Wages excl. social security	X
ULC	Unit labor cost	X
WageMM	Hourly wages, mining and manufacturing	X
LabCHr	Hourly labor cost	X

Table A.6 Labor market variables

Variable	Description	Transformation		
		Level	$\ln y_t - \ln y_{t-4}$	$y_t - y_{t-4}$
AvJob	# of available jobs		X	
LabFrc	# in labor force		X	
NLFrc	# not in labor force		X	
RelLF	LabFrc/PpTot			X
Empld	# employed		X	
PrvEmp	# privatly employed		X	
PubEmp	# publicly employed		X	
Av4Wrk	# available for work		X	
NA4Wrk	# not available for work		X	
NUnemp	# unemployed		X	
Unemp	Unemployment	X		
U02W	# unemployed $<$ 2 weeks			X
U314W	# unemployed 3 - 14 weeks			X
U1552W	# unemployed 15 - 52 weeks			X
U52W+	# unemployed more than 52 weeks			X
NewJob	New jobs			X
NEmpl	1 - Empld/PpTot	X		

Table A.7 Real activity and Expectations

Variable	Description	Transformation		
		Level	$\ln y_t - \ln y_{t-4}$	$\frac{y_t - y_{t-4}}{y_{t-4}}$
IndProd	Industrial production			X
NewCar	New cars	X		
NewHouse	New single family houses	X		
HourWork	Hours worked		X	
GDP	GDP		X	
RGDP	Real GDP		X	
NAIRU	NAIRU	X		
OutGap	Output gap	X		
LackPrdF	Lack of production factors	X		
BCI	Business confidence indicator	X		
HExpsSWE	Household exp. Swedish economy	X		
HExpOwn	Household exp. own economy	X		

Table A.8 Prices

Variable	Description	Transformation	
		Level	$(y_t - y_{t-4})/y_{t-4}$
InfFor	Foreign CPI (TCW)		X
InfRel	Relative CPI		X
PPP	Real exchange rate		X
Infla	Swedish CPI		X
InfNet	Swedish NPI		X
InfHse	House price index		X
InfMrt	Mortgage interest component of CPI		X
InfxMr	CPI excluding mortgage interest	X	
MrtWgh	Weight of mortgage interest in CPI		X
InfUnd	Underlying inflation		X
InfFd	Food component of CPI		X
InfFl	Housing fuel and electricity comp. of CPI		X
InfHWg	Factor price index, housing incl. wages		X
InfCns	Construction cost index		X
InfImp	Import price index		X
InfExp	Export price index		X
GDPInf	GDP deflator		X
ExpInf	Households exp. of inflation 1 year from now	X	
OPrice	Oil price, SEK		X

Table A.9 Consumption and income

Variable	Description	Transformation $\ln y_t - \ln y_{t-4}$
CnsExp	Consumption expenditure	X
RCnsExp	Real CnsExp	X
DspInc	Disposable income	X
RDspIn	Real disposable income	X
RetSls	Retail sales	X

Earlier Working Papers:

SVERIGES
RIKSBANK

Sveriges Riksbank • Visiting address: Brunkebergs torg 11 • Mail address: se-103 37 Stockholm
Website: www.riksbank.se • E-mail: registratorn@riksbank.se • Telephone: +46 8 787 00 00 • Fax: +46 8 21 05 31