Does it Make Sense to Combine Forecasts from VAR Models? An Empirical Analysis with Inflation Forecasts for Switzerland

Thomas J. Jordan and Marcel R. Savioz, Research,
Swiss National Bank, Zurich

> *If you can look into the seeds of time, and say which grain will grow and which will not, speak then unto me.* *Shakespeare.*[1]

1 Introduction[2]

At the beginning of the year 2000, the Swiss National Bank (SNB) adopted a new monetary policy framework. The new concept is based on an explicit definition of price stability and uses an inflation forecast as the main indicator for guiding monetary policy decisions.[3] Although the forward-looking approach of the SNB has some similarities with both the two-pillar concept of the European Central Bank and with strict inflation targeting – as practised, for example, by the Bank of England – it is in itself an original framework for monetary policy.[4]

What these new monetary policy strategies have in common is the forward-looking orientation of the policy decisions necessitated by the long lag between a monetary impulse and its effect on output and prices. In Switzerland, for instance, the main impact on prices is believed to occur between one and three years after the monetary impulse. Consequently, inflation forecasts over a horizon of one to three years are of special importance for central banks as a basis for decision-making. For this reason, providing long-run inflation forecasts with various methods and analysing their properties and accuracy has recently become an area of intense research. This paper is an empirical contribution to the literature on this topic.

Producing accurate inflation forecasts over a horizon of one to three years is a very difficult task. Uncertainties about the true structure of the economy and the monetary transmission mechanism force central banks to use a variety of approaches for forecasting inflation rather than to rely on a single model.[5] One approach uses large, structural macroeconomic models.[6] These models have the advantage of producing forecasts for many variables and of delivering clear economic intuition behind the dynamics of the forecasts. The problems of large structural models are the restrictive assumptions that have to be made in order to identify the structure of the economy.[7] Vector autoregression models (VAR) constitute a second approach.[8] VARs exploit the information in macroeconomic time series without imposing strong restrictions relating to the structure of the economy. Thus, VARs may suffer less from the problem of data contamination by imposing incorrect restrictions regarding both the economy and the transmission mechanism. A further advantage of VARs comes from the fact that they do not need any assumptions about the course of exogenous variables for the period of the forecasting horizon. All variables in VAR models are endogenous, and the dynamic forecasts are straightforward to compute. When VARs are estimated, however, the problem of a small number of degrees of freedom often arises. The number of parameters to be estimated soon becomes overwhelming when more and more variables are included in a VAR. For instance, given the limited length of the quarterly time series in Switzerland, VARs can often be estimated with only three to five variables. Thus, VARs may fail to use part of the relevant information contained in the macroeconomic data because their size has to be restricted to a small number of variables. One possible way of overcoming this problem is to use a series of small VARs and then to combine their forecasts.[9] The aim of this study is to analyse whether combining forecasts from different small VAR models can improve the accuracy of single forecasts by taking information from more variables into account. The study concentrates on inflation forecasts for Switzerland.[10]

The paper is organised as follows: Section 2 discusses the role of unconditional forecasts in the process of monetary policy decisions. In Section 3, the methods for combining forecasts are explained. Section 4 analyses the time series properties of the variables and the various VAR models used to produce forecasts. Section 5 addresses the question of whether combined forecasts are better than individual forecasts by looking at out-of-sample results. Section 6 contains a summing-up.

1 Quoted from Granger (1989, p. 153).

2 We would like to thank Caesar Lack, Jean-Marc Natal, Samuel Reynard, Enzo Rossi, Martin Schlegel and Peter Stalder for their valuable comments. We appreciated the discussions with the participants in our session at the 2001 Meeting of the Swiss Society for Statistics and Economics in Geneva and with the participants of the 2000 Meeting of the "Arbeitsgruppe Prognoseverfahren der Gesellschaft für Operations Research" at Eichstätt University in Ingoldstadt. Any errors in the paper should be attributed solely to the authors.

3 See Jordan and Peytrignet (2001) for an analysis of the role of inflation forecasts in the new monetary policy framework.

4 See Baltensperger, Fischer and Jordan (2002) for a discussion of the characteristic features of the Swiss monetary policy framework compared to strict inflation targeting.

5 See Kirchgässner and Savioz (1997) for a discussion of various econometric approaches.

6 See Stalder (2001) for the presentation of the large structural model used by the SNB.

7 The classic critique of large structural models was formulated by Sims (1980).

8 See Jordan, Kugler, Lenz and Savioz (2002) for a description of the VAR models used by the SNB in conditional and unconditional forecasting.

9 A different possibility would be to use Bayesian VAR methodology or to restrict some of the coefficients to zero after the appropriate testing.

10 Empirical evidence on GDP forecasts of the Swiss economy may be found in Ruoss and Savioz (2002).

2 Types of Forecasts in Monetary Policy: Conditional and Unconditional

Two types of inflation forecasts are used in monetary policy: On the one hand, conditional forecasts assume a specific path of the future course of monetary policy. Thus, they allow the central bank to evaluate the consequences of alternative policy decisions. On the other hand, unconditional forecasts provide inflation predictions where the future stance of monetary policy over the forecasting horizon is explicitly or implicitly predicted as well.

Unconditional inflation forecasts are produced for three main reasons. First, they provide a benchmark forecast given the observed reaction of the central bank to the macroeconomic situation in the past. Such inflation forecasts are especially important and informative since the assumptions about the course of the policy instrument underlying the conditional forecasts, e.g. a constant interest rate through to the end of the forecasting horizon, are usually unrealistic. Unconditional forecasts are thus important indicators for the general inflation outlook. Second, since conditional inflation forecasts cannot be tested for their accuracy because of their counterfactual nature, the different models have to be evaluated according to their performance in unconditional forecasting. Thus, the central bank needs to produce unconditional forecasts for all its models, including forecasts based on structural models. Third, unconditional inflation forecasts also allow comparisons with forecasts from outside the central bank. This enables the policymaker to judge whether there are differences between the market perception of the inflation outlook and his own analysis.

Simple VARs represent reduced forms, i.e., the parameters of simple VARs have no structural interpretation. Producing *conditional* forecasts from simple VARs is problematic, because a given interest rate is not equivalent to a given monetary policy course. Furthermore, the estimated coefficients are not policy-invariant.[11] However, VARs are an ideal method of producing *unconditional* benchmark forecasts because they rely on only a minimum amount of structural information, i.e. on the choice of the variables and on the lag length. Furthermore, unlike in the case of structural models, exogenous variables do not have to be forecast. In this paper, we look only at unconditional inflation forecasts computed with VAR models and examine whether these forecasts can be improved by combining them.

3 Combining Forecasts

The traditional VAR forecasting procedure is very simple. The forecaster decides on the variables and the number of lags included in the VAR as well as on the assumptions about integration, cointegration, trend, and seasonality of the data. Forecasts are then produced by using the chosen model. The limited length of the time series typically available in macroeconomics highlights three basic problems of the traditional approach. First, only small VARs, i.e., with only a few variables, can be used for forecasting. Thus, potentially useful information may be left out by concentrating on a single model. Second, the small degrees of freedom may lead to estimated parameters with large standard errors, which is likely to influence adversely the out-of-sample performance of the forecasts. The traditional approach also has the disadvantage that the choice of the model does not primarily depend on past out-of-sample forecast performance but on the goodness of the fit.

In order to overcome these shortcomings, we have developed an alternative procedure. It consists of two steps: First, we compute a large number of forecasts using a series of small VARs. The small size of the VARs allows us to preserve a minimal number of degrees of freedom. Second, the forecasts of different models are weighted to produce combined forecasts.[12] We call the forecasts obtained by this procedure "combined VAR forecasts" (CVARFs). While many methods are used for combining forecasts, the present study only covers those most commonly adopted.[13]

11 The production of conditional forecasts with *structural* VAR models is dealt with in Kugler and Jordan (2000) and in Jordan, Kugler, Lenz and Savioz (2002).

12 See Winkler (1989, p. 606), for the basic motivation underlying the proposed procedure: "In most interesting forecasting situations in our uncertain and rapidly changing world, I doubt that such 'true' models are attainable and I think that it is counterproductive to think in terms of 'true' models. The motivation for the combination of forecasts, then, is at its most basic level the simple idea of aggregation of information to achieve a reduction in uncertainty, or an increase in accuracy."

13 For descriptions of methods used to determine the weights, see, for example, Clemen and Winkler (1986), Clemen (1989), and Holden and Peel (1986).

There are three main reasons why combining forecasts can be expected to improve their accuracy: First, it leads to a diversification of the forecast errors and thus diminishes the problem of imprecise estimates of individual models.[14] Second, combined forecasts should be more robust, because they do not depend closely on the specifications applied to an individual model. The damage done by a specification error in a single model may thus be greatly reduced. Third, by taking more variables into account, combined forecasts may be based on a broader information set. This is especially relevant for VAR forecasts which are traditionally based on models with only a few variables. The gains from combining forecasts should arise in particular when the weights depend on the past performance of the individual forecasts.

Further advantages of combining VAR forecasts may exist. First, the weighting of the various forecasts, if determined by their past forecast performance, may provide some information on the relative importance of the different models and variables. Second, a change in the dispersion of the individual forecasts can give an early indication of a deterioration in the forecast accuracy. However, we do not examine these further issues in the present paper. Instead we focus on the question of whether CVARFs are more accurate than VAR forecasts.

The different methods of weighting the forecasts, which are used in the subsequent analysis, are explained with the help of an example where forecasts for inflation in time t from three VAR models are available: $\hat{\pi}_{VAR1,t}$, $\hat{\pi}_{VAR2,t}$ and $\hat{\pi}_{VAR3,t}$.[15] The combined inflation forecast $\hat{\pi}_{CVAR,t}$ is a weighted average of these three individual forecasts

(1) $\hat{\pi}_{CVAR,t} = w_0 + w_1\hat{\pi}_{VAR1,t} + w_2\hat{\pi}_{VAR2,t} + w_3\hat{\pi}_{VAR3,t}$,

where the w_i $i = 0,\ldots 3$ are the weights.

The first very common method of combining forecasts is to take the *simple average* (SA) of the individual forecasts. Accordingly, the weights are equal for all individual forecasts and sum up to one:

(2) $w_0 = 0 \quad w_1 = 1/3 \quad w_2 = 1/3 \quad w_3 = 1/3$.

In the SA method, the weights do not depend on the observed past accuracy of the individual forecasts.

In contrast to the SA method, the other combination methods set weights according to the past performance of the individual forecasts. This is accomplished with the help of a linear regression with the actual inflation rate as the dependent variable and the individual out-of-sample forecasts as explanatory variables. The coefficients may be estimated with restrictions so that they satisfy some or all of the properties of weights ($0 \leq w_i \leq 1, \sum_i w_i = 1$). In our example with three forecasts the regression is:[16]

(3) $\pi_t = \beta_0 + \beta_1\hat{\pi}_{VAR1,t} + \beta_2\hat{\pi}_{VAR2,t} + \beta_3\hat{\pi}_{VAR3,t} + \varepsilon_t$.

The second combination method we will use is the ordinary least square method (LS). In the LS-combined forecast (1), the estimated coefficients of equation (3) are used as weights to calculate the combined forecast:

(4) $w_0 = \hat{\beta}_0 \quad w_1 = \hat{\beta}_1 \quad w_2 = \hat{\beta}_2 \quad w_3 = \hat{\beta}_3$.

No restrictions are imposed on the estimation of the coefficients in Eq. (3). Note that the coefficient $\hat{\beta}_0$ is equal to zero for unbiased forecasts.

The third combination method is the *constant restricted least square* method (CRLS). The CRLS-combined forecast is assumed to be unbiased and the constant term of the estimated regression is restricted in order to be equal to zero:

(5) $w_0 = \hat{\beta}_0 = 0$.

The fourth combination method is the *equality restricted least square* method (ERLS). In the ERLS-combined forecast a further restriction is that the weights of the forecasts sum up to one. Thus, the regression (3) is estimated with the following restrictions:

(6) $\hat{\beta}_0 = 0$ and $\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = 1$.

The fifth and last combination method we will use is the *non-negativity inequality restricted least square* method (NRLS). In NRLS-combined forecasts the weights are non-negative. The regression (3) is estimated with the following (non-linear) restrictions:

(7) $\hat{\beta}_0 = 0, \hat{\beta}_1 \geq 0, \hat{\beta}_2 \geq 0,$ and $\hat{\beta}_3 \geq 0$.

In the following section, we will make use of these five methods for combined forecasts. Our analysis will seek to establish whether CVAR forecasts achieve a better performance than individual VAR forecasts.

14 See, for example, Granger (1989), Granger and Newbold (1973, 1986). See Jungmittag (1993) for an introductory exposition of the diversification argument.

15 See Aksu and Gunter (1992).

16 The CVARF approach not only provides a forecast but delivers additional information that may be useful for the forecasters and for the conduct of monetary policy. First, CVARFs give a useful indication of the source of forecast performance. This can be read from equation (3). Forecasts that yield no significant coefficient in this regression contain no information not already included in the other forecasts (see Diebold (1989) for a discussion of forecast combination and "forecast encompassing" as well as West (2001) for an appropriate test for "forecast encompassing"). Thus, it is possible to infer which sets of variables are good predictors for inflation at a given forecasting horizon. A non-zero intercept in the regression is an indication that the forecasts are biased (see Holden and Peel, 1989). Second, structural breaks in the inflation process may be identified at an early stage through the analysis of the changes in the estimated weights (see Diebold and Pauly, 1986). The likely source of the structural break may also be inferred from the change in the weights.

4 Data, Time Series Properties, and VAR Models

In order to keep the analysis tractable we restrict ourselves to five variables: The consumer price index P, the money aggregate $M3$, total domestic bank credit C, real GDP Q, and the long-term Swiss franc interest rate R. Money, credit, economic activity, and the long-term interest rate are important determinants of the inflation process according to the main theories of the transmission mechanism of monetary policy.[17]

The sample period covers the first quarter of 1974 to the third quarter of 2000.[18] Table 1 presents the Augmented Dickey-Fuller Unit Root test for the five variables. The logs of all five variables will be assumed to be I(1) in the rest of the paper.[19] We only consider VAR models with stationary variables. Thus, all variables enter the models as first differences. For the purpose of this paper, we do not take cointegration and vector error correction specifications into account.

The VAR models considered include a constant and four lags. No trend or seasonal dummies are allowed for.[20] All VARs have to include at least the inflation rate π_t. Thus, the smallest VAR just includes π_t. The largest VAR includes all 5 variables. Within this setup it is possible to specify 16 different VAR models: 1 model with 1 variable, 4 models with 2 variables, 6 models with 3 variables, 4 models with 4 variables and 1 model with 5 variables.

Combined forecasts are only constructed from forecasts of VAR models with the same number of variables. Given this restriction, 79 different combined forecasts can be specified: 11 combinations of forecasts from models with 2 variables, 57 of forecasts from models with 3 variables and 11 from models with 4 variables (see Table 4).

5 Out-of-Sample Forecasts

In this section, we examine the possibility of achieving better forecasts by combining forecasts. We compare the performance of CVARFs with the individual VAR forecasts (VARFs). Both the average performance of VARFs and CVARFs as well as the performance of the best VARFs and CVARFs will be compared. We concentrate on forecasts of the annual inflation rate

(8) $\pi_t = 100*\log(P_t/P_{t-4})$

and consider the forecasting horizons of one, two, and three years. These are the most relevant horizons for monetary policy.

To assess the accuracy of the forecasts, we use the root mean square error statistics (RMSE)

(9) $RMSE = \sqrt{1/T\sum_{t=1}^{T}(\pi_t - \hat{\pi}_t)^2}$,

where π_t is the actual inflation and $\hat{\pi}_t$ is the predicted inflation for time t. The difference $\pi_t - \hat{\pi}_t$ is the forecast error and T is the number of forecasts. The mean squared forecast error ($MSE = 1/T\sum_{t=1}^{T}(\pi_t - \hat{\pi}_t)^2$) is a measure of the size of the average forecast errors. Due to squaring of the forecast errors, large forecast errors are given a proportionally higher weighting than small ones. The dimension of the root mean squared error (RMSE) corresponds to that of the inflation rate π_t. If the forecasts are perfect, the RMSE is equal to zero.

The analysis considers only out-of-sample forecasts. As put forward by Bernanke (1990) and Thoma and Gray (1994), among others, the ultimate decision about the usefulness of a forecasting model must come from its ability to forecast out of sample. Superior in-sample forecasting ability does not automatically mean superior out-of-sample forecasting ability.

The forecasts are computed with rolling regression methods: A rolling estimation of the VAR models yields a series of out-of-sample individual VAR forecasts for different forecasting horizons $k = 4, 8, 12$. The forecasts for the horizon of k quarters are computed as follows: First the VAR is estimated with observations running from time s to time $s - 29$, where s is the period after which the first forecast starts.[21] The estimated coefficients are then used to compute the forecast for time $s + k$. For the forecast for time $s + k$ only information available in time s is used. Then, the sample is enlarged by one period and the equation is re-estimated with data running from

17 The study by Jordan (1999a) has shown that credit aggregates are good predictors of inflation. For the importance of monetary aggregates for inflation, see Baltensperger, Jordan and Savioz (2001) as well as Kirchgässner und Savioz (2001).

18 Note that the variability of inflation has recently been very small around a low level of inflation. Therefore the latest data are not very appropriate for testing the performance of models to forecast (high) inflation.

19 See Miller, Clemen, and Winkler (1992).

20 The variables are seasonally adjusted.

21 Thus no VAR is estimated with less than thirty observations.

$s+1$ to $s-29$. The re-estimated coefficients are now used to compute the forecast for time $s+1+k$. This procedure is continued until the end of the available data, but the estimation sample is held constant when it reaches 50 observations. With this technique, 73 one-year-ahead individual forecasts from 1982:3 to 2000:3, 69 two-year-ahead individual forecasts from 1983:3 to 2000:3 and 65 three-year-ahead individual forecasts from 1984:3 to 2000:3 can be computed.

Combined forecasts are obtained by using weights computed by the five methods discussed above. For the calculation of these weights, rolling regression techniques are applied. The regression of Eq. (3) is run with the first 30 individual forecasts to produce the first combined forecast for time $s+2k+29$, where $s+k$ is the quarter for which the first individual forecast with horizon k is available. Then the regression is run with the first 31 individual forecasts to produce the combined forecast for quarter $s+2k+30$. This procedure is continued until the end of the available data, but the number of individual forecasts used in the regression is again held constant after reaching 50. This procedure allows us to produce 39 one-year-ahead combined forecasts from 1991:3 to 2000:3, 31 two-year-ahead combined forecasts from 1993:3 to 2000:3 and 23 three-year-ahead combined forecasts from 1995:3 to 2000:3.[22]

During the period from 1991:1 to 2000:3 actual inflation averaged 1.9%, with a maximum of 6.1% and a root mean square ($RMS = 1/T\sum_{t=1}^{T}\sqrt{\pi_t^2/T}$) of 2,62%. For the period 1993:1 to 2000:3 inflation averaged 1.2%, with a maximum of 3.4% and an RMS of 1.54%. During the period from 1995:1 to 2000:3 inflation averaged only 0.9% and was consistently below 2%. The RMS was 1.09%. Since it is preferable to evaluate the forecasting performance over a period when inflation shows some variation, we do not present the results of the different forecasting horizons for a common sample.

The change in both the volatility and the level of inflation makes it more difficult to assess the forecast performance of the models between different forecasting periods and horizons. One possibility to compare the forecast accuracy between different periods is to look at Theil's U[23]

$$(10)\quad U = \frac{RMSE}{RMS} = \frac{\sqrt{1/T\sum_{t=1}^{T}(\pi_t - \hat{\pi}_t)^2}}{1/T\sum_{t=1}^{T}\sqrt{\pi_t^2/T}}.$$

The U statistic relates the RMSE of the inflation forecasts to the RMS of the actual inflation. This scales the forecast errors relative to the level of inflation because absolute forecast errors tend to be smaller in periods of low inflation than in periods of high inflation. Furthermore, Theil's U allows the performance of a model to be judged relative to a simple forecast of no change in the price level (i.e. zero inflation). The inequality coefficient U is equal to one if the model has the same predictive power as the simple forecast. If U is smaller (bigger) than one, the model yields more (less) precise forecasts than the simple forecast of no change. One should, however, be aware that a simple forecast of no change in the price level may in some circumstances be a very good forecast. A U statistic higher than one would then not indicate a poor forecast performance per se. This is especially valid for the period from 1995:1 to 2000:3, when a forecast of a constant price level would have been quite acceptable. Over this period of almost six years, the increase in the price level was only 4.3%, or 0.75% a year. Consequently, for forecasts within the period from 1995:1 to 2000:3 with a horizon of three years, a U statistic that is bigger than one may not necessarily reflect a bad performance.

In Table 2, we show the results of the out-of-sample performance of individual VAR forecasts.[24] We report average results for different groups of VAR models. In general, as the RMSE and the U statistics show, the performance deteriorates as the length of the forecasting horizon increases. The row with one variable ($n=1$) corresponds to an AR(4) model of the change in the price level. This model serves as a benchmark. On average, the other VARFs outperform this benchmark. The forecasts with VARs with four variables ($n=4$) perform best on average for the one-year forecasting horizon. For the two-year-forecasting horizon VARs with three variables ($n=3$) perform best on average, but the difference to the VARs with four variables ($n=4$) is only very small. For the three-year-ahead forecasts, VARs with three variables ($n=3$) achieve on average the best results, followed by bivariate VARs ($n=2$) and the VARs with four variables ($n=4$). The performance of the single VAR with five variables ($n=5$) is not particularly good. This may be attributable to a small number of degrees of freedom in the estimation of this VAR. The results show clearly, at least for one-year-ahead and two-year-ahead forecasts, that including additional variables improves the performance.

22 When the forecast horizon increases by one year, the sample for the evaluation of the combined forecasts decreases by two years (see Table 2 and following). The first year is lost because fewer forecasts can be computed with a given data set when the forecast horizon becomes larger. Similarly, the second year is lost because fewer combined forecasts can be computed with a given set of forecasts when the forecast horizon becomes larger.

23 The Theil U statistic is defined here as in Greene (2000, p. 310).

24 See Jordan (1999b) for another piece of evidence.

5.1 Comparison with the average performance of the VARFs

How do CVARFs perform compared to VARFs? In the following we will compare the performance of the CVARFs to the *average performance* of the VARFs as reported in the first row of table 2. Table 3 reports the average RMSE and U of the 79 CVARFs for the different combination methods. The table also indicates the improvement from combining forecasts relative to the average of all individual VAR forecasts. The simple average method (SA) performs quite well for each forecasting horizon. For the one-year forecasting horizon, SA is even the best method for weighting the individual forecasts. An advantage of this method is that the weights do not have to be estimated. The least square method (LS) performs very poorly at any forecasting horizon. Restricting the constant to zero (CRLS) improves the performance substantially for the two-year and the three-year forecasting horizon.[25] For the three-year forecasting horizon, CRLS is the best method. Imposing the equality restriction, which requires the weights to sum up to one (ERLS), improves the performance only for the one-year forecasting horizon.[26] The method of restricting the weights of the forecasts to be non-negative (NRLS) achieves good results for the two- and three-year-ahead forecasts. It is even the best method for the two-year forecast horizon.

Imposing constant weights (SA) works well especially for short forecasting horizons. However, for longer forecasting horizons, it seems important to let the weights change over time by re-estimating them in each period. But, as the poor performance of LS shows, restrictions should be imposed in the estimation of the weights. Whereas ERLS seem not to impose the right restriction to estimate the weights, CRLS and NRLS perform well. If the individual VAR forecasts are very similar but not identical, CRLS is still numerically feasible, whereas the NRLS may not be.[27]

For the subsequent analysis we concentrate on the best methods for each forecasting horizon. For the one-year horizon, we use SA. For the two-year horizon, we may choose between CRLS and NRLS. The results are very similar. In order to avoid possible numerical problems, we opt for CRLS rather than NRLS. CRLS is also the best method for the three-year horizon.

Table 5 reports the results for different subgroups of the CVARFs. As a benchmark, the results of the average of the VARFs and the CVARFs are reported in Table 5 as well. We form two different subgroups of 79 combined forecasts. First, we form subgroups for all combined forecasts that are computed from the same number of VARFs. For instance, the group $m = 2$ consists of all 27 combined forecasts that are formed with two individual forecasts. These individual forecasts may stem from VARs with 2, 3, or 4 variables. Second, we form subgroups of combined forecasts that are computed from VARFs with the same number of variables. For instance, the subgroup ($n = 2$) consists of all 11 combined forecasts that are computed from forecasts of individual VARs with 2 variables. The combined forecasts may include 2, 3, or 4 individual forecasts. The subgroups are illustrated in Table 4.

The results show very clearly that the more forecasts are combined, the better the forecast performance becomes. The improvement is monotone with an increasing number of combined forecasts. If 6 forecasts ($m = 6$) are combined, the RMSE decreases by more than 20% compared to the average of the VARFs. This improvement may be due to either the diversification effect (more forecasts) or the information effect (more variables). The findings also show that combinations of forecasts of VARs with 3 or 4 variables ($n = 3$, $n = 4$) achieve on average a bigger improvement of the RMSE than combinations of forecasts from bivariate VARs ($n = 2$). Somewhat odd are the results for the two-year horizon, where the combinations of forecasts from VARs with 4 variables ($n = 4$) do not perform very well.

25 On the CRLS-method see also Granger and Ramanathan (1984).
26 On the ERLS-method see also Clemen (1986).
27 The weights of the previous regression are used if the new regression yields no numerical results.

5.2 Comparison with the best VARFs

The results of the analysis so far show that combining forecasts from different models can, in general, improve the precision of the forecasts, provided that an adequate combination method is chosen. This finding was based on a comparison of the average performance of CVARFs with the average performance of VARFs. Two questions arise: First, do any individual VARFs substantially outperform the averages of the combined forecasts? Second, is it possible to identify which combinations perform better and whether it might be advantageous to concentrate on some specific combinations? We will therefore look more closely at the performance of the *best performing* VARFs and CVARFs.

Table 6 presents the results for the three best VARFs for each forecasting horizon. A comparison with Table 3 shows that for the one-year forecasting horizon the best VARFs outperform the average CVARFs (all weighting methods). For the two-year-ahead forecasts only the two best VARFs outperform the average CVARF computed with the CRLS method. The best VARF, however, is not better than the average CVARF computed with the NRLS method. For the three-year forecasting horizon, the best VARF performs worse than the average CVARF computed either by the CRLS or the NRLS method. Note also that the best VARs do not include all variables, especially at long forecasting horizons. The results set out in Table 6 indicate that combining forecasts is especially important for long-term forecasts.

Table 7 compares the result of the best VARF with the three best CVARFs for each forecasting horizon. Using the simple average (SA) method, the best CVARF outperforms the best VAR forecasts by 9% for the one-year forecasting horizon. However, for longer forecasting horizons, the best CVARF (CRLS) improves the RMSE by more than 30%. This is a substantial improvement in the forecast accuracy. For the two (three)-year forecasting horizon 59 (49) out of the 79 possible CVARFs achieve lower RMSE and U than the best VARF.

Can the best combination be identified? It is interesting to note that the three best models for any forecasting horizon include, without exception, all five variables considered in this analysis. Note also the following interesting result: For the one-year forecast horizon, only three out of fifteen possible combinations (number of two forecasts of trivariate VARs) use information from all variables. The best CVARF is among these three combinations. This may point not only to a "diversification advantage" but also to an "information advantage" of combined forecasts over individual forecasts. One piece of practical advice may be to combine VARFs in such a way that a large number of variables is taken into account in the CVARF.

6 Conclusions

Unconditional inflation forecasts are important for the conduct of monetary policy, and VAR models are well-suited to producing such unconditional forecasts. Given the limited amount of data typically available in macroeconomic studies, however, VARs have the disadvantage of being restricted to a small number of variables. We have developed a modelling approach to compute unconditional forecasts that overcomes the problem of a limited number of variables. The procedure exploits the properties of combined forecasts and proceeds in two steps: First, a large number of VAR forecasts, stemming from models specified in various ways and including different variables, are computed. Second, the weights for combining the forecasts are determined according to the past performance of the forecasts. Then the unconditional combined forecasts are computed using these weights. With this approach, "real-time" information on the forecast accuracy and possible structural breaks may be extracted from the change of weights. Furthermore, the weights may show which group of variables incorporate information about future inflation for a specific forecasting horizon.

The results of this paper show that, on average, the combined inflation forecasts computed with the developed approach outperform the best VAR forecast. This is especially true for long-run forecasts. The superiority of combining forecasts can be attributed to three features: First, forecast errors are diversified. Second, combined forecasts do not rely heavily on the specification of a single VAR model and may thus be less sensitive to specification errors. And third, combined VAR forecasts are usually based on more information than single VAR forecasts.

The present study contains some shortcomings. To minimise the computational and programming workload, we restricted ourselves in two respects. First, we only used five variables. The literature on the monetary transmission mechanism underlines the importance of further variables like exchange rates, import prices, different interest rates, etc. Second, we did not examine whether the results hold if different time series properties of the data are assumed. This would drastically increase the number of VARFs and CVARFs to be taken into account because VARs in levels and Error Correction Models would have to be included. Because an inefficient (and unbiased) forecast can always be improved by combining it with another forecast, we would expect some improvement in forecast accuracy if these two restrictions were relaxed. The interesting question, however, is whether this improvement would be as large as the one shown in this study. Finally, a fundamental limitation of the method presented here must be mentioned: Combined VAR forecasts are only suitable when unconditional forecasts are needed. For structural simulations (impulse respons, variance decomposition, conditional forecast) the single VAR approach has to be used.

To sum up, this paper demonstrates that, in order to produce unconditional inflation forecasts, it may be more sensible to work with many "small" VARs than to use a single VAR model. The empirical results reveal that this may be especially true for long-run forecasts of inflation. Reverting to the introductory quote by Shakespeare, we may draw the following conclusion: If you cannot say which grain will grow, it is wise not to choose a single seed but to plant them all. This paper and the wealth of literature on combined forecasts show that if the purpose is to forecast rather than to analyse the economy, it is more appropriate to use more than just a single model. The SNB's inflation forecasts are in fact based on several models, and the results of this paper help to underpin this pluralist approach.

7 References

Aksu C. und Gunter, S.I. 1992. An empirical analysis of the accuracy of SA, OLS, ERLS, and NRLS combination forecasts. *International Journal of Forecasting* 8: 27–43.

Baltensperger, E., Fischer A.M. and Jordan T.J. 2002. Abstaining from inflation targets, Understanding SNB rhetoric in the inflation targeting debate. Mimeo, Swiss National Bank.

Baltensperger, E., Jordan T.J. and Savioz M.R. 2001. The demand for M_3 and inflation forecasts: An empirical analysis for Switzerland. *Weltwirtschaftliches Archiv/Review of World Economics* 137(2): 244–272.

Bernanke, B.S. 1990. On the predictive power of interest rates and interest rate spreads. *New England Economic Review* November–December: 51–61.

Clemen, R.T. 1986. Linear constraints and the efficiency of combined forecasts. *Journal of Forecasting* 5: 31–38.

Clemen, R.T. 1989. Combining forecasts: A review and annotated bibliography, *International Journal of Forecasting* 8: 559–83.

Clemen, R.T. and Winkler, R.L. 1986. Combined economic forecasts. *Journal of Business and Economic Statistics* 4: 39–46.

Diebold, F.-X. 1989. Forecast combination and encompassing: Reconciling two divergent literatures. Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series.

Diebold, F.-X. and Pauly, P. 1986. Structural change and the combination of forecasts. Board of Governors of the Federal Reserve System, Special Studies Section, Discussion Paper 201.

Granger, C.W.J. 1986. *Forecasting Economic Time Series*. London: Academic Press.

Granger, C.W.J. 1989. *Forecasting in Business and Economics*. London: Academic Press.

Granger, C.W.J. and Newbold, P. 1973. Some comments on the evaluation of economic forecasts. *Applied Economics* 5: 35–47.

Granger, C.W.J. and Newbold, P. 1986. *Forecasting Economic Time Series*. Second Edition, London: Academic Press.

Granger, C.W.J. and Ramanathan, R. 1984. Improved methods of combining forecasts. *Journal of Forecasting* 14(3): 367–79.

Greene, W.H. 2000. *Econometric analysis*. Fourth Edition, London: Prentice-Hall.

Holden, K. and Peel, D.A. 1986. An empirical investigation of combinations of economic forecasts. *Journal of Forecasting* 5: 229–42.

Holden, K. and Peel, D.A. 1989. Unbiasedness, efficiency and the combination of economic forecasts. *Journal of Forecasting* 8: 175–882.

Jordan, T.J. 1999a. The information content of bank credit to forecast output and inflation: The case of Switzerland. Mimeo, Swiss National Bank.

Jordan, T.J. 1999b. Inflationsprognosen mit VAR-Systemen. Mimeo, Swiss National Bank.

Jordan, T.J. and Peytrignet, M. 2001. Die Inflationsprognosen der Schweizerischen Nationalbank, La prévision d'inflation de la Banque nationale suisse. *Quarterly Bulletin of the Swiss National Bank* 2: 54–61.

Jordan, T.J., Kugler, P., Lenz, C. and Savioz, M.R. 2002. Inflationsprognosen mit vektorautoregressiven Modellen, Prévisions d'inflation par des modèles vectoriels autorégressifs. *Quarterly Bulletin of the Swiss National Bank* 1: 40–66.

Jungmittag, A. 1993. Die Kombination von Prognosen: Ein Überblick mit Anwendungen, *Jahrbücher für Nationalökonomie und Statistik* 212 (1–2): 58–72.

Kirchgässner, K. and Savioz, M.R. 1997. Empirische Forschung in den Wirtschaftswissenschaften: Ein Überblick. *Homo oeconomicus* 16: 507–538.

Kirchgässner, K., and Savioz, M.R. 2001. Monetary policy and forecasts for real GDP growth: An empirical investigation for the Federal Republic of Germany. *German Economic Review* 2: 339–365.

Kugler, P. and Jordan, T.J. 2000. Structural vector autoregressions and the analysis of monetary policy interventions: The Swiss case. Mimeo, University of Basel.

Miller, C.M., Clemen, R.T. and Winkler, R.L. 1992. The effect of nonstationary on combined forecasts. *International Journal of Forecasting* 7(4): 515–529.

Ruoss, E., and Savioz, M.R. 2002. How accurate are GDP forecasts – a survey for Switzerland. *Quarterly Bulletin of the Swiss National Bank* 3: 43–63.

Sims, C. 1980. Macroeconomics and reality. *Econometrica* 48: 1–49.

Stalder, P. 2001. Ein ökonometrisches Makromodell für die Schweiz, Un modèle macroéconomique pour la Suisse. *Quarterly Bulletin of the Swiss National Bank* 2: 63–89.

Thoma, M.A. and Gray, J.A. 1994. On leading indicators: Getting it straight. 1994 Federal Reserve Bank of Dallas *Texas conference on monetary economics Paper No. 4*.

West, K.D. 2001. Tests for forecast encompassing when forecasts depend on estimated regression parameters. *Journal of Business & Economic Statistics* 19(1): 29–33.

Winkler, R.L. 1989. Combining forecasts: A philosophical basis and some current issues. *International Journal of Forecasting* 5: 605–609.

Appendix:

Augmented Dickey-Fuller Unit Root Test 1974:1 – 2000:3 Table 1

Variable	*k*	*r*	*t*
ΔP	5	0.699	−2.699(*)
ΔQ	10	0.058	−3.17*
ΔM	4	0.570	−2.845(*)
ΔC	0	0.833	−2.905*
ΔR	0	0.412	−6.736**

* indicates that the null hypothesis of a unit root is rejected at the 5% significance level. ** and (*) indicate a rejection at the 1% and 10% significance level. *k* is the number of lagged (endogenous) variables entering the Augmented Dickey-Fuller test equation. *k* is the lag between 0 and 10 with the smallest value of the AIC-criterion. *r* is the estimated unit root and t is the test statistic. The critical values of MacKinnon are used.

Individual Out-of-sample VAR Forecasts
Average results of the 16 VAR forecasts Table 2

Number of variables included in the VAR:	One-year-ahead forecasts 1991:1 – 2000:3 39 forecasts		Two-year-ahead forecasts 1993:1 – 2000:3 31 forecasts		Three-year-ahead forecasts 1995:1 – 2000:3 23 forecasts	
	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U
All VARS (16)[1]	*0.892*	*0.340*	*1.259*	*0.817*	*1.418*	*1.302*
n = 1 (1)	1.198	0.457	1.365	0.886	1.464	1.345
n = 2 (4)	0.966	0.369	1.252	0.812	1.417	1.301
n = 3 (6)	0.860	0.328	**1.242**	**0.806**	**1.402**	**1.287**
n = 4 (4)	**0.801**	**0.306**	1.245	0.808	1.425	1.309
n = 5 (1)	0.842	0.321	1.340	0.869	1.445	1.328

Note:
For each forecasting horizon, the best statistic is reported in bold font. n is the number of variables entering a VAR. All VARs are of order 4. For example, for n = 1 the forecasts are computed with an AR(4) model. For n = 2 (n = 3) the forecasts are computed with a bivariate (trivariate) VAR(4), and so on. Starting with 30 observations the sample used for the estimation of the VARs is augmented until the sample size of 50 is reached.

1 The number of VARs with n variables is given in brackets. For example, "n = 3 (6)" means that six forecasts were computed with six trivariate VAR(4). The result reported in the table is the average of the six forecasts.

Combined Out-of-sample VAR Forecasts

Average result for the 79 combined forecasts

Table 3

Method	One-year forecast horizon 1991:1 – 2000:3 39 forecasts RMSE	*U*	%	Two-year forecast horizon 1993:1 – 2000:3 31 forecasts RMSE	*U*	%	Three-year forecast horizon 1995:1 – 2000:3 23 forecasts RMSE	*U*	%
VARF	*0.892*	*0.340*	*100%*	*1.259*	*0.817*	*100%*	*1.418*	*1.302*	*100%*
SA	**0.749**	**0.286**	−15.9%	1.167	0.757	−7.3%	1.374	1.262	−3.1%
LS	1.069	0.408	20%	1.714	1.112	36.1%	2.528	2.322	78.3%
CRLS	0.942	0.360	5.9%	1.142	0.741	−9.3%	**1.261**	**1.158**	−11.1%
ERLS	0.868	0.331	−2.6%	1.302	0.844	3.3%	1.552	1.426	9.5%
NRLS	0.904	0.345	1.5%	**1.088**	**0.706**	−13.6%	1.268	1.165	−10.5%

Note:
VAR(4) and combinations are estimated with 50 observations.

Subgroups of Combined VAR-Forecasts

Table 4

Number of variables in the VAR (n):	Number of combined forecasts (m): *m* = 2	*m* = 3	*m* = 4	*m* = 5	*m* = 6
n = 2	6	4	1	–	–
n = 3	15	20	15	6	1
n = 4	6	4	1	–	–

Note:
n is the number of variables entering the VARs. *m* is the number of forecasts combined.

Combined Out-of-sample VAR Forecasts (Subgroups)
Average result for different subgroups of the 79 combined forecasts

Table 5

	One-year-ahead forecasts 1991:1 – 2000:3 (39 forecasts) SA			Two-year-ahead forecasts 1993:1 – 2000:3 (31 forecasts) CRLS			Three-year-ahead forecasts 1995:1 – 2000:3 (23 forecasts) CRLS		
	RMSE	U	GAIN	RMSE	U	GAIN	RMSE	U	GAIN
VARF	*0.892*	*0.340*	*100%*	*1.259*	*0.817*	*100%*	*1.418*	*1.302*	*100%*
CVARF	*0.749*	*0.286*	*−15.9%*	*1.142*	*0.741*	*−9.3%*	*1.261*	*1.158*	*−11.1%*
Subgroups formed by the number of VAR-forecasts combined									
m = 2 (27)	0.782	0.298	−12.4%	1.208	0.784	−4.0%	1.327	1.219	−6.4%
m = 3 (28)	0.745	0.284	−16.5%	1.179	0.765	−6.4%	1.264	1.161	−10.8%
m = 4 (17)	0.721	0.275	−19.1%	1.080	0.701	−14.2%	1.201	1.104	−15.2%
m = 5 (6)	0.705	0.269	−20.9%	0.914	0.593	−27.4%	1.147	1.054	−19.0%
m = 6 (1)	**0.698**	**0.266**	−21.8%	**0.783**	**0.508**	−37.9%	**1.100**	**1.011**	−22.4%
Subgroups formed by the number of variables entering the VARs									
n = 2 (11)	0.869	0.332	−2.4%	1.389	0.901	−10.3%	1.285	1.181	−9.3%
n = 3 (57)	0.734	0.280	−17.6%	**1.073**	**0.696**	−14.9%	1.258	1.156	−11.2%
n = 4 (11)	**0.706**	**0.270**	−20.6%	1.256	0.815	−0.2%	**1.251**	**1.149**	−11.8%

Note:
m is the number of forecasts combined.

Note:
n is the number of variables entering the VARs.

The Best Individual VAR Forecasts

Table 6

Rank	VAR	RMSE	Theil's U
	One-year-ahead forecasts: 1991:1 – 2000:3 (39 forecasts)		
1	*P, M, C, R*	0.716	0.273
2	*P, M, C, Q*	0.732	0.280
3	*P, C, R*	0.742	0.283
	Two-year-ahead forecasts: 1993:1 – 2000:3 (31 forecasts)		
1	*P, M, C, Q*	1.088	0.706
2	*P, M, C*	1.128	0.732
3	*P, M, Q*	1.152	0.748
	Three-year-ahead forecasts: 1995:1 – 2000:3 (23 forecasts)		
1	*P, C, R*	1.318	1.210
2	*P, C, Q, R*	1.351	1.241
3	*P, C*	1.358	1.247

Note:
VAR(4) and combinations are estimated with 50 observations.

The Best Combined VAR Forecasts

Table 7

Rank	CVARF	RMSE	Theil's U	Gain
	One-year-ahead forecasts (SA)			
	Best VARF	*0.716*	*0.273*	*100%*
1	*P,M,C + P,Q,R*	0.650	0.248	−9.2%
2	*P,M,C + P,C,R + P,Q,R*	0.658	0.251	−8.1%
3	*P,M,C,Q + P,M,C,R + P,M,Q,R*	0.662	0.252	−7.7%
	Two-year-ahead forecasts (CRLS)			
	Best VARF	*1.088*	*0.817*	*100%*
1	*P,M,C + P,M,Q + P,M,R + P,C,Q + P,C,R + P,Q,R*	0.783	0.508	−38.8%
2	*P,M,C + P,M,R + P,C,Q + P,C,R + P,Q,R*	0.787	0.511	−37.5%
3	*P,M,C + P,M,R +P,C,R + P,Q,R*	0.817	0.530	−35.1%
	Three-year-ahead forecasts (CRLS)			
	Best VARF	*1.318*	*1.210*	*100%*
1	*P,M,C + P,C,R + P,Q,R*	0.921	0.846	−30.1%
2	*P,M,C + P,C,R + P,Q,R + P,M,R*	0.951	0.874	−27.8%
3	*P,M,C + P,C,R + P,Q,R + P,M,Q*	0.952	0.874	−27.8%

Note:
VAR(4) and combinations are estimated with 50 observations.