Forecasting the Swiss Economy Using VECX* Models: An Exercise in Forecast Combination Across Models and Observation Windows

Katrin Assenmacher-Wesche and M. Hashem Pesaran

Forecasting the Swiss Economy Using VECX* Models: An Exercise in Forecast Combination Across Models and Observation Windows

Katrin Assenmacher-Wesche[†]
Swiss National Bank

M. Hashem Pesaran[†]
Cambridge University, CIMF, and USC

February 13, 2008

Abstract

This paper uses vector error correction models of Switzerland for forecasting output, inflation and the short-term interest rate. It considers three different ways of dealing with forecast uncertainties. First, it investigates the effect on forecasting performance of averaging over forecasts from different models. Second, it considers averaging forecasts from different estimation windows. It is found that averaging over estimation windows is at least as effective as averaging over different models and both complement each other. Third, it examines whether using weighting schemes from the machine learning literature improves the average forecast. Compared to equal weights the effect of alternative weighting schemes on forecast accuracy is small in the present application.

JEL classification: C53, C32

Key words: Bayesian model averaging, choice of observation window, long-run structural vector autoregression

1 Introduction

Forecasting macroeconomic variables is of importance for market participants and policy makers alike. Although great care is generally taken in designing a specific forecasting model, the true forecast uncertainty is often underestimated because various sources of forecasting errors, like parameter and model uncertainties are not taken into account properly .[1]

This paper considers the problem of forecast uncertainty in the context of a long-run structural vector error correcting model of the Swiss economy. The model includes the effective nominal exchange rate of the Swiss franc, real gross domestic product (GDP), the real money stock, measured by M2, the three-month interest rate, inflation and the ratio of domestic to foreign prices as endogenous variables, and foreign output, the foreign interest rate and the oil price as exogenous variables. We first present an overidentified long-run vector error correction model with exogenous variables (VECX* model) and use it for forecasting. The model contains five long-run relations identified as the purchasing power parity, money demand, output convergence, uncovered interest parity, and the Fisher equation.

We then allow for forecast uncertainty along three different dimensions. First, we deal with model uncertainty. When deciding on a specific model, one always has to make choices like, e.g., the number of lags to include, the number of cointegrating relations to assume, the long-run restrictions to impose, and the data-generating processes to adopt for the exogenous variables. In this paper, we confine ourselves to a class of models that differ only with respect to these characteristics instead of considering entirely different model types. To allow for model uncertainty we apply Bayesian model averaging and combine forecasts from several plausible specifications of the model.

Second, economic relations can be subject to structural breaks. Pesaran and Timmerman (2007) proposed to take this into account by estimating a model over different observation windows and then pooling the forecasts. While estimation is more efficient if all available data are used when the models are stable, the occurrence of structural breaks, which are often difficult to identify and measure accurately with statistical methods, might bias the forecasts. One pragmatic way to deal with this is to average forecasts from models estimated over different estimation windows. Since economic theory is more informative regarding the nature of the long-run relations, in this exercise we do not allow for parameter uncertainty of the

---

[1]For a recent review of the literature on forecasting see Elliott and Timmermann (2007).

long-run coefficients, but consider alternative estimates of the short-run coefficients computed over different observation windows starting in the fourth quarter of each year between 1965 and 1976.

Third, we assess the usefulness of different weighting schemes in model averaging, such as equal weights, Akaike (AIC) weights and weighting schemes advanced in the machine-learning literature (Yang 2004; Sancetta 2006).

We find that averaging forecasts from different models reduces the forecast error considerably. In addition, averaging the forecasts over estimation windows is at least as effective as model averaging in improving forecast precision. Moreover, averaging across the two dimensions complements each other, and leads to further reductions in forecast errors. By contrast, in our application choice of the weights when combining forecasts does not seem to be that important.

The paper is set out as follows. Section 2 discusses the econometric methodology and presents the estimates for the baseline version of our forecasting model. Section 3 evaluates the forecasts from the baseline version of the model. Section 4 explores the effect of averaging forecasts across different models and estimation windows. We find that the forecast average across all models and estimation windows outperforms our long-run restricted VECX* model as well as a univariate AR(1) benchmark model. In addition, we try different weighting schemes and assess their influence on the forecasting performance of the model. Though one would expect that excluding models performing poorly from the average forecast should improve results, we find that schemes weighting models approximately equally perform better. Finally, Section 5 offers some conclusions.

2   The VECX* model

The model used for forecasting is a structural cointegrated vector error-correction model that relates the core macroeconomic variables of the Swiss economy (denoted by the vector $\mathbf{x}_t$) to current and lagged values of a number of key foreign variables (denoted by the vector $\mathbf{x}_t^*$), which we call the Swiss VECX* model. The model is developed along the same lines as the model for the UK in Garratt, Lee, Pesaran and Shin (2003a, 2006). A detailed documentation of the model can be found in Assenmacher-Wesche and Pesaran (2008).

We use quarterly data starting in 1965, so that after differencing and accounting for the necessary lags the model is estimated on data starting in 1965Q4. We stop the estimation in 1999Q4 and use data from 2000Q1 to 2006Q3 to evaluate the

recursive out-of-sample forecasting performance.

The choice of the variables is influenced by the purpose of the model, namely forecasting the rate of inflation and modelling the monetary transmission process. Therefore, the model will incorporate those key relations from economic theory that can be expected to have an impact on the inflation rate. One of these relations is money demand, which postulates a long-run relation between the real money stock, denoted by m_t, the logarithm of real gross domestic product (GDP), y_t, and the nominal interest rate, r_t, which we take to be the three-month LIBOR rate.[2] Another is the Fisher interest-rate parity which establishes a long-run relation between the interest rate and inflation, π_t. For Switzerland as a small, open economy the exchange rate, e_t, has an important influence on economic activity. Therefore, purchasing power parity, which links the nominal exchange rate to the ratio of the domestic to the foreign price level, $p_t - p_t^*$, is included. In addition, we consider the price of oil, p_t^{oil}, as the most important commodity price, which is expected to have direct and indirect impacts on domestic as well as on world inflation. Finally, international business cycles and interest-rate cycles are allowed to have an influence on the domestic economy by considering long-run relations between domestic and foreign real GDP and interest rates.[3] The latter two variables, together with the oil price, are regarded as weakly exogenous variables.

Foreign output, y_t^*, and the foreign CPI, p_t^*, are computed as weighted averages, using three-year moving averages of the trade shares with Switzerland. For example, the foreign output is computed as

$$y_t^* = \sum_{j=1}^{N} \bar{w}_{jt} y_{jt},$$

where y_{jt} is the log real output of country j, and $\bar{w}_{jt}$ are the average trade weights.

The trade weights are based on Switzerland's 15 largest trade partners and are computed as averages of Switzerland's imports from and exports to the country in question divided by the total trade of all the 15 countries.[4] Trade to these 15 countries on average accounts for 82 percent of total Swiss foreign trade. For the

---

[2] We measure real money by the logarithm of M2, deflated with the consumer price index (CPI).

[3] Interest rates are expressed as $0.25 \ln(1 + R/100)$ where R is the interest rate in percent per annum to make units of measurement compatible with the rate of inflation, which is computed as the first difference of the logarithm of the quarterly price level.

[4] Data on imports and exports are from the 'Eidgenössische Zollverwaltung'. We use trade data up to period t since these data arrive in a timely fashion. In forecasting we therefore to not make use of information that are not available at the time the forecasts are made.

construction of foreign financial variables we use weighted averages of the US and the euro area variables, with the weights based on the three-year moving averages of the trade shares of these two regions with Switzerland. Specifically, the foreign interest rate is computed as the weighted average of the three-month interest rates of the euro area and the US, and the Swiss exchange rate is computed as the weighted average of the log exchange rate of the Swiss franc in terms of the US dollar and the euro.[5] This seems justified considering the dominant role played by these economies in the evolution of the financial market interconnections of the Swiss economy and the rest of the world.[6]

Before turning to the estimation of the VECX* model we perform unit root tests, which indicate that the variables can be regarded as $I(1)$. Initially all estimations and tests were carried out over the period 1965Q4 to 1999Q4. We reserve the rest of the available data to investigate the forecasting performance of the model. When computing recursive out-of-sample forecasts, we only use information that would have been available to a forecaster at that point in time. Nevertheless, we use final vintage data so that our exercise is not an analysis of how well forecast averaging performs in real time.[7]

We start from a conditional VECX* model for the endogenous variables in error-correction form with a restricted trend,

$$\Delta\mathbf{x}_t = -\mathbf{\Pi}_x\left[\mathbf{z}_{t-1} - \boldsymbol{\gamma}(t-1)\right] + \mathbf{\Lambda}\Delta\mathbf{x}_t^* + \sum_{i=1}^{p-1}\mathbf{\Psi}_i\Delta\mathbf{z}_{t-i} + \mathbf{c}_0 + \boldsymbol{v}_t, \tag{1}$$

and a marginal model for the exogenous variables,

$$\Delta\mathbf{x}_t^* = \sum_{i=1}^{p-1}\mathbf{\Gamma}_{x^*i}\Delta\mathbf{z}_{t-i} + \mathbf{a}_{x^*0} + \mathbf{u}_{x^*t}. \tag{2}$$

The 9×1 vector of variables $\mathbf{z}_t = (\mathbf{x}_t', \mathbf{x}_t^{*\prime})'$ in the model contains six endogenous variables, $\mathbf{x}_t = \{e_t, m_t, y_t, r_t, \pi_t, p_t - p_t^*\}$ and three weakly exogenous variables, $\mathbf{x}_t^* = \{y_t^*, r_t^*, p_t^{oil}\}$.

Regarding the lag order of the underlying VAR(p) the Akaike criterion indicates two lags whereas the Schwarz and the Hannan Quinn criteria prefer a single lag.

---

[5]The interest rate and the exchange rate for the euro area are linked to German data before 1999.

[6]The appendix in Assenmacher-Wesche and Pesaran (2008) contains detailed information on how the data were constructed.

[7]Clark and McCracken (2006) investigate forecast averaging as a method to deal with data revisions in real time.

Though we start with the same number of lags on the endogenous and exogenous variables in equation (1), we later will set certain coefficients in Ψ_i to zero and distinguish between the number of lags on the endogenous variables, p_x, and the exogenous variables, p_{x^*}. In addition, the lag length in the marginal model can differ from the number of lags considered in the conditional model. In the forecasting exercise below we shall also consider the effects on the forecasts of the endogenous variables of choosing different marginal models for the exogenous variables, $\Delta \mathbf{x}_t^*$.

We next test for the number of cointegrating relations. Using simulated critical values the trace and the maximum eigenvalue (λ-max) statistics suggests that $r = 3$ at the 10% level of significance, though the trace test only marginally rejects the hypothesis that $r = 4$. However, Assenmacher-Wesche and Pesaran (2008) using data over a more recent sample (1976 to 2006) find five cointegrating relations, which is more in line with the long-run theory. In what follows we also assume five long run relations, as economic theory suggests, but investigate the effect of dropping cointegrating relations later when dealing with model uncertainty.

We proceed to imposing economically meaningful over-identifying restrictions on $\boldsymbol{\beta}$ that are in accordance with theoretical priors, namely the purchasing power parity (PPP), money demand (MD), output convergence based on the gap between domestic and foreign output (GAP), interest rate parity between the domestic and foreign interest rate (UIP), and a Fisher equation linking the domestic interest rate with inflation (denoted by FIP). The estimates of these relations, computed over the sample period 1965Q4-1999Q4, and their 95 percent confidence bounds, based on a non-parametric bootstrap with 1000 replications, are as follows:

$$\text{PPP: } e_t - (p_t - p_t^*) = b_{10} \underset{(0.004, 0.015)}{+0.009t} + \xi_{1t},$$

$$\text{MD: } m_t - y_t = b_{20} \underset{(-32.22, -18.00)}{-24.89r_t} + \xi_{2t},$$

$$\text{GAP: } y_t - y_t^* = b_{30} \underset{(-0.0042, -0.0029)}{-0.0036t} + \xi_{3t},$$

$$\text{UIP: } r_t - r_t^* = b_{40} + \xi_{4t},$$

$$\text{FIP: } r_t - \pi_t = b_{50} + \xi_{5t}.$$

$$(3)$$

We impose a unitary income elasticity of money demand since the estimated coefficient was close to one. We do not report estimates for the constant term because it will be re-estimated in the recursive out-of-sample forecasting exercise.

A likelihood ratio (LR) test of the 22 over-identifying restrictions gives a test statistic of 106.21, compared to a bootstrapped critical value of 61.66 for the five percent level of significance and of 71.04 for the one percent significance level. The test therefore rejects the restrictions at conventional significance levels (the p-value is 0.1 percent).

Since the purpose of this paper is to assess the effect of model uncertainty on forecast performance we impose all theoretically motivated constraints on the long-run relations in the long-run restricted VECX*(2,2) model and investigate the effects of relaxing some of these restrictions later. Moreover, model uncertainty of this type can be taken into account using Bayesian model averaging techniques, which gives a theoretical framework for considering forecasts from various specifications (see Geweke and Whiteman 2006). We therefore not only explore the forecast results for our long-run theory-consistent VECX*(2,2) specification, but also consider the effects of changes in the number of cointegrating relations, the identification restrictions and the lag order on the forecasting performance of the model.

3   Forecasting with the VECX*(2,2) model

Macroeconometric forecasting is subject to different types of uncertainties that may impact on the accuracy of a model's forecasts. These include future uncertainty, parameter uncertainty (for a given model), and model uncertainty.[8] Future uncertainty refers to the uncertainty that surrounds the realization of future shocks (innovations) to the model under consideration. Parameter uncertainty refers to the robustness of forecasts with respect to a given set of parameter values (for a specific model).

The standard approach to future and parameter uncertainty is to report confidence intervals instead of point forecasts. Nevertheless, confidence intervals are of limited usefulness if forecasts from multiple models are considered. Model uncertainty arises because there is no consensus about the true model. Though tests can be applied to search for an appropriate model specification, results are often inconclusive and depend on the order the tests performed, so that different plausible specifications can be maintained at the end of the search process. In addition, macroeconometric models are likely to be subject to structural breaks due to policy changes and shifts in tastes and technology. As Clements and Hendry (1998, 1999,

---

[8] See, e.g., Garratt, Lee, Pesaran and Shin (2003b).

2006) emphasize, structural breaks are often the main source of forecast failure and represent the most serious form of model uncertainty.

In this paper we follow Pesaran and Timmermann (2007) and attempt to deal with model uncertainty and structural breaks by pooling forecasts from the same model but estimated over different sample periods, as well as by pooling forecasts estimated over the same sample period but obtained from different models. The latter type of pooling has been the subject of an extensive literature on classical methods of forecast combination and Bayesian model averaging, whilst the former is new and to our knowledge has not been applied before.[9] The pooling of forecasts from models estimated over different estimation windows is viewed as a relatively robust and simple procedure to dealing with possible structural breaks that are difficult to detect and to exploit in forecasting in a timely manner. On this also see Pesaran and Pick (2007).

In the following, we shall first examine the forecasting performance of the VECX* (2,2) model discussed in Section 2 that imposes the 22 over-identifying restrictions derived from economic theory. We refer to these as 'long-run restricted VECX*(2,2)' forecasts. We shall then proceed to investigate how forecasts change with different specifications of the conditional and the marginal model, and whether forecasts improve when they are averaged over different model specifications. When pooling forecasts from different estimation windows we will consider windows starting between 1965Q4 to 1976Q4 and assess whether averaging of forecasts from different estimation windows helps improve the forecasting performance.[10] Since we will average forecasts over different model specifications and over different estimation windows, we need a terminology to distinguish between these two types of averaging. We will refer to the average forecast over different models for a specific estimation window as the *AveM* forecast, whereas the average forecast over estimation windows for a specific model will be denoted by *AveW*. We also consider pooling of forecasts from different models, estimated over different estimation windows. We shall refer to these as *AveAve* forecasts to highlight the two distinct dimensions over which

---

[9]Timmermann (2006) surveys the literature on forecast combinations, while Geweke and Whiteman (2006) discuss forecast combinations in a Bayesian setting. Clark and McCracken (2004) combine rolling and recursive forecasts but do not average over forecasts derived from different models estimated over different observation windows.

[10]As discussed in Pesaran and Timmermann (2007), it is also possible to combine forecasts from different estimation windows using time-varying weights based on the past performance of different forecasts using a cross-validation approach. However, such a procedure is data intensive and does not seem suitable for quarterly macroeconometric forecasting.

the forecast averaging has been carried out. Finally, we will assess the effect on forecasting performance of using different weighting schemes to construct the *AveM* forecast.

To construct the forecasts we need both the conditional and the marginal models as set out in equations (1) and (2). Combing them we have

$$\mathbf{z}_t = \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{z}_{t-i} + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t,$$

where $\mathbf{z}_t = (\mathbf{x}_t', \mathbf{x}_t^{*\prime})'$, $\mathbf{\Phi}_1 = \mathbf{I}_m - \mathbf{\Pi}_x + \mathbf{\Gamma}_1$, $\mathbf{\Phi}_i = \mathbf{\Gamma}_i - \mathbf{\Gamma}_{i-1}$, $i = 2, ..., p-1$, $\mathbf{\Phi}_p = -\mathbf{\Gamma}_{p-1}$. The coefficient matrices $\mathbf{\Gamma}_i, \mathbf{a}_0$ and $\mathbf{a}_1$ include the parameters from both the marginal and conditional models and are defined as $\mathbf{\Gamma}_i = \begin{pmatrix} \mathbf{\Psi}_i + \mathbf{\Lambda}\mathbf{\Gamma}_{x*i} \\ \mathbf{\Gamma}_{x*i} \end{pmatrix}$, $\mathbf{a}_0 = \begin{pmatrix} \mathbf{a}_{x0} + \mathbf{\Lambda}\mathbf{a}_{x*0} \\ \mathbf{a}_{x*0} \end{pmatrix}$ and $\mathbf{a}_1 = \begin{pmatrix} \mathbf{\Pi}_x \gamma \\ \mathbf{0} \end{pmatrix}$. In order to avoid deterministic trends in interest rates, $\mathbf{a}_{x*0}$ is set to zero in the foreign interest-rate equation.

We consider forecast horizons of up to eight quarters ahead since this is the relevant time horizon for central banks when setting interest rates. Our strategy for forecast evaluation is as follows: The model is estimated to the end of 1999Q4 and one- to eight-quarter-ahead forecasts are then produced for 2000Q1 to 2001Q4. The sample period is extended by one observation, the short-run parameters are re-estimated to the end of 2000Q1 and another set of forecasts is generated, this time for 2000Q2 to 2002Q1. Since the long-run coefficients of the model presumably change only slowly, we do not re-estimate them. This procedure is repeated until the end of the available sample, 2006Q3, is reached. At the end of the sample, however, we are not able to evaluate the forecasts for longer time horizons. For the model estimated up to 2006Q2, for instance, we can only compare the one-quarter-ahead forecast with the actual data for 2006Q3. For that reason, the forecast statistics rely on a different number of forecasts for each horizon, ranging from 27 forecast errors for the one-quarter-ahead forecasts to 20 for the eight-quarter-ahead forecasts.

The forecasting performance clearly depends on the evaluation period chosen. In this respect, the period from 2000Q1 to 2006Q3 provides a number of challenges for the various forecasts that we consider. Over the whole of the forecast period, inflation was low and the quarterly changes of the price level fluctuated in a narrow band between -1.0 and 2.3 percent per annum. Similarly, interest rates were low compared to their historical values whereas real money growth was strong during 2002 and 2003 and peaked at 28 percent per annum in 2003Q2. Since the evaluation

period is somewhat atypical, it will be particularly interesting to see if the *AveAve* pooling of forecasts can lead to forecast improvements as compared to forecasts from the best (in-sample) model.

We evaluate the forecasts in terms of their root mean squared forecast error (RMSFE), which constitutes a specific, although widely used loss function.[11] Let z_{t+h} be the level of the variable that we wish to forecast, i.e., the level of output, inflation, or the interest rate. Denote the forecast of this variable formed at time t by $\hat{z}(t+h, t)$, and define the h-step ahead forecasted changes as $\hat{x}_t(h) = [\hat{z}(t+h, t) - z_t]/h$ and the associated h-step ahead realized changes as $x_t(h) = (z_{t+h} - z_t)/h$. The h-step ahead forecast error is then computed as

$$e_t(h) = x_t(h) - \hat{x}_t(h) = [z_{t+h} - \hat{z}(t+h, t)]/h.$$

For a forecast evaluation period from $T+1$ to $T+n$, the RMSFE is defined as

$$RMSFE = 100\sqrt{(n-h+1)^{-1}\sum_{t=T}^{T+n-h} e_t^2(h)}.$$

For convenience, we report the RMSFE in percent.

Table 1 shows the RMSFE per quarter in percent for the forecasts based on the VECX*(2,2) model for the longest estimation window, using all available data from 1965Q4 onward. The forecasts for the exogenous variables are from a marginal model that regresses the change in the exogenous variables, $\Delta\mathbf{x}_t^*$, on the change in the endogenous and exogenous variables, $\Delta\mathbf{z}_{t-1}$. We denote this marginal model by $\mathfrak{M}_a^*$, which is also estimated sequentially over the same sample period as the conditional model.[12] The average RMSFE per quarter decreases with a longer forecast horizon. The reason is that we focus on the average change per quarter in the variable over h quarters. Though the change per quarter at longer forecast horizons is small, this generally accumulates to a substantial deviation of the forecast level from the actual level of the variable at long horizons. The RMSFE for output growth is between 0.57 and 0.33 percent per quarter, whereas the RMSFE for inflation is 0.27 percent for the one-quarter horizon but decreases to 0.07 percent at the eight-quarter horizon. The RMSFE for the interest rate is lowest, lying around 0.07 percent per quarter.

---

[11]Other possible loss functions are the bias, measuring how far the mean of the forecast is from the mean of the actual series or the proportion of correctly predicted directions of change in a variable. Pesaran and Skouras (2002) discuss other decision-based methods for forecast evaluation.

[12]We shall discuss the effects of using different marginal models and estimation windows on the forecast performance later on.

Summing up, the long-run restricted VECX*(2,2) model performs reasonably well and we will take it as one of our reference models when investigating if forecasts can be improved by double averaging (i.e., by following the *AveAve* procedure discussed above).

4 Pooling of forecasts

There is now a sizable literature showing that averaging over different forecasts can lead to forecast improvements. The problem of interest can be described as estimating the forecast probability density function, $\Pr(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{w,T})$, of a vector of variables $\mathbf{Z}_{T+1,h} = (\mathbf{z}_{T+1}, ..., \mathbf{z}_{T+h})$ conditional on the available observations at the end of period T, $\mathbf{Z}_{w,T} = (\mathbf{z}_{T-w+1}, \mathbf{z}_{T-w+2}, ..., \mathbf{z}_T)$, where h denotes the forecast horizon, and w is the size of the observation window. For a given model, $\mathfrak{M}_m$, and a given estimation window, w, the forecast probability density function $\Pr(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{w,T})$ can be estimated by $\widehat{\Pr}(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{w,T}, \mathfrak{M}_m)$, which involves estimating model $\mathfrak{M}_m$ over the estimation window of size w from the end of estimation sample at T. In the face of model uncertainty, assuming that there are M models under consideration and using Bayes formula, we have the familiar Bayesian Model Averaging expression given by

$$\widehat{\Pr}(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{w,T}) = \sum_{m=1}^{M} \widehat{\Pr}(\mathfrak{M}_m \mid \mathbf{Z}_{w,T})\widehat{\Pr}(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{w,T}, \mathfrak{M}_m), \qquad (4)$$

where $\widehat{\Pr}(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{w,T}, \mathfrak{M}_m)$ is the predictive density of $\mathbf{Z}_{T+1,h}$ conditional on model $\mathfrak{M}_m$ and the observation window w, and $\widehat{\Pr}(\mathfrak{M}_m \mid \mathbf{Z}_{w,T})$ is the posterior probability of model $\mathfrak{M}_m$, also estimated over the observation window w.

If a particular model, $\mathfrak{M}_m$, is stable over time, the best estimator of $\Pr(\mathbf{Z}_{T+1,h} \mid \mathfrak{M}_m)$ would be based on all available information, i.e, the longest estimation window possible. Standard applications of Bayesian Model Averaging *implicitly* assume that all models under considerations are stable. But in reality some or all the models under consideration could be subject to structural breaks and different choices of estimation samples might be warranted. The optimal choice of the estimation window depends on the nature of the breaks (their frequency and intensity) and is in general rather difficult to ascertain. In the presence of unknown structural breaks averaging over different estimation windows is recommended (Pesaran, Pettenuzzo and Timmermann 2006, Pesaran and Timmermann 2007). While leaving out obser-

vations at the beginning of the sample will lead to less precise coefficient estimates, one probably discards observations that stem from a different regime and thus deteriorate forecasts. If the structural breaks are unknown, there is a trade-off between both effects.

A pragmatic solution to the model instability problem would be to consider a number of alternative windows, starting from a minimum window size to the largest permitted by the available data set, and then average the forecasts across the windows, what we have termed as the *AveW* forecasts.[13] The minimum window size can be determined as a multiple of the number of parameters being estimated, or could be based on information regarding a known structural break nearest to the forecast date, T. The maximum window size can be set, subject to data availability, to be sufficiently large so that a satisfactory approximation to the asymptotic theory that underlies the estimation of model $\mathfrak{M}_m$ can be achieved. In most macroeconomic applications, including the one in this paper, the maximum window size coincides with the longest observation window that is available. This might not, however, be the case when forecasting high frequency financial data.

Allowing for both model and estimation window uncertainty yields the following *AveAve* formula

$$\widehat{\Pr}(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{T,T}) = \sum_{m=1}^{M} \sum_{w=T}^{T-W+1} \widehat{\Pr}(\mathfrak{M}_m \mid \mathbf{Z}_{w,T})\widehat{\Pr}(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{w,T}, \mathfrak{M}_m),$$

where $\mathbf{Z}_{T,T} = (\mathbf{z}_1, ..., \mathbf{z}_T)$ denotes all the available observations, $\widehat{\Pr}(\mathfrak{M}_m \mid \mathbf{Z}_{w,T})$ is the weight attached to model $\mathfrak{M}_m$, $m = 1, 2, ..., M$, estimated over the estimation window $w = T, T-1, ..., T-W+1$, at the end of period T; the windows are arranged from the longest window of size T, to the shortest window of size $T - W + 1$.

Bayesian model averaging requires the specification of the prior probability of model $\mathfrak{M}_m$ and of the prior probability of the model's coefficients, $\boldsymbol{\theta}_m$, conditional on $\mathfrak{M}_m$, for $m = 1, 2, ..., M$. In our applications we focus on equal weights. This approach is justified if the data-generation process is subject to structural breaks and uncertainty over which model is the right one is diffused. It entails the risk, however, that one considers bad models in the average that should better have been left out. We first present forecast averages that weight all forecasts equally, before we investigate other weighting schemes that have been proposed in the literature.[14]

---

[13]See Pesaran and Pick (2007) for some theoretical results on the *AveW* procedure.

[14]The weighting schemes are discussed in the appendix.

4.1 Average over different model specifications (*AveM*)

When averaging forecasts from different model specifications, we first need to define the class of models to be considered. To improve forecast performance by pooling forecasts from several models, it is important that the models considered are statistically viable and economically meaningful. This is especially relevant when equal weights are used since they do not take account of past model performance. With this in mind we make the following choices. We base our choice of alternative models on the long-run restricted VECX*(2,2) model developed in Section 2. First, we consider uncertainty regarding the number of cointegrating relations. Second, we will vary the order of the lags on the endogenous and the exogenous variables, p_x and p_{x^*}, in the VECX*(p_x,p_{x^*}) specification in equation (1). Third, we shall consider different specifications for the model we use to forecast the exogenous variables.[15]

In general, one would expect that imposing long-run equilibrium relations should improve the forecasting performance of a model, at least over the medium to long term horizons. Testing the restrictions implied by economic theory in Section 2, however, gave ambiguous results as to whether these restrictions are consistent with the data. Therefore, the first set of models we shall consider differ with respect to the long-run restrictions that are imposed. While economic theory suggested five long-run relations, the statistical tests pointed to the existence of only three or possibly four cointegrating vectors. One way to deal with this uncertainty is to estimate several models with different restrictions and to average forecasts across these models. Since we are uncertain about the true cointegration rank, r, of $\mathbf{\Pi}_x$ we consider all possible ranks between $r = 1$ and $r = 5$. When having less than five cointegrating vectors, we do not know which of the over-identified economic relations, i.e., PPP, money demand, output gap, uncovered interest parity or the Fisher relation, to impose. We therefore compute forecasts with all possible combinations of over-identifying restrictions. Specifically, we have five possible combinations of long-run restrictions when $r = 1$, ten possible combinations when $r = 2$, and so on. In total, this gives 31 different model specifications.[16] In addition, we consider models with one to five exactly identified cointegrating vectors. This gives a total of 36 different model specifications.

---

[15]Of course, it would be possible to consider other alternatives, such as VECX* models in inflation and output growth but with fewer or more variables than considered in this paper. However, this particular strategy for generating alternative forecasting models will not be pursued here.

[16]Precisely, there are $2^5 - 1$ combinations since we exclude the model without any long-run restrictions.

Averaging over forecasts from different specifications of the long-run restrictions generally improves on forecasts based on the VECX*(2,2) model. Table 2 shows the RMSFE for output growth, inflation and the interest rate for the average over the 36 different model specifications, applying equal weight to each model when computing the average.[17] At the one-year horizon, we find a reduction in the RMSFE of between 10 and 20 percent for output and the interest rate and of even more than 50 percent for inflation.[18]

Next, we will consider different lag lengths for the endogenous and exogenous variables in the conditional model. Using the estimation sample ending in 1999Q4, the Akaike criterion pointed to the inclusion of two lags whereas the Schwarz and the Hannan-Quinn criteria favoured one lag. We therefore consider all possible combinations of one and two lags for the endogenous and exogenous variables, i.e., in addition to our long-run restricted VECX*(2,2) model we compute forecasts from a VECX*(2,1), a VECX*(1,2) and a VECX*(1,1) model. Testing for cointegration in these additional three models (again for the estimation sample ending in 1999Q4), we find a cointegration rank of either $r = 3$ or $r = 4$. We therefore compute averages over the same 36 model specifications discussed above also for the VECX*(2,1), VECX*(1,2) and VECX*(1,1) models.

Averaging forecasts from all models is likely to improve forecast performance further. In the following, we present the RMSFEs for the average forecasts from the 36 different specification of the long-run relations up to four quarters ahead. The first column in Table 3 shows that the average forecast based on the VECX*(2,2) model performs best for inflation, whilst those based on the VECX*(2,1) model produce best forecasts for output and interest rates.

Next we investigate the effect on the forecasting performance of using different marginal models for the exogenous variables. We will consider two different specifications. First, we regress the change in the exogenous variables, $\Delta\mathbf{x}_t^*$, on $\Delta\mathbf{z}_{t-1}$ (i.e., the first lagged change in the endogenous and exogenous variables), see equation (2). We call this the $\mathfrak{M}_a^*$ model. Second, we include only the lagged changes in the exogenous variables, $\Delta\mathbf{x}_{t-1}^*$ as regressors in the marginal model for $\Delta\mathbf{x}_t^*$. This latter choice can be motivated by Switzerland being a small economy that has no

---

[17] We still consider a VECX* model with two lags of the endogenous and the exogenous variables and the $\mathfrak{M}_a^*$ marginal model for the exogenous variables. Both models are estimated over the longest estimation window, starting in 1965Q1.

[18] The advantages of averaging forecasts from different models when forecasting Swiss inflation are documented in Jordan and Savioz (2003).

influence on foreign variables. We refer to this marginal model as $\mathfrak{M}_b^*$ model. For forecasting both marginal models, $\mathfrak{M}_a^*$ and $\mathfrak{M}_b^*$ are estimated sequentially over the same sample period as the conditional model. While we include a constant in the equations (in first differences) for foreign output and the oil price, the equation for the foreign interest rate is estimated without a constant in order not to generate a trend in the level of the interest rate.

To assess the improvement coming from an explicit marginal model for the exogenous variables, we also compute forecasts with the exogenous variable set to their unconditional sample mean ($\mathfrak{M}_c^*$). In effect, this corresponds to regressing each of the exogenous variables on a constant only. Note that also in this case the mean is computed sequentially over the same period as the conditional model (i.e., up to and including period T, $T + 1$, etc.) so that no post-sample information is used in computing the forecasts of $\mathbf{x}^*$. Finally, we set the exogenous variables to their realized values, which we call the $\mathfrak{M}_d^*$ model.[19] As at the time of forecasting the realized values of $\mathbf{x}^*$ are unknown, these forecasts are not feasible and are provided as a benchmark against which the other feasible marginal models can be assessed.

Averaging the forecasts from different lag specifications and marginal models is also likely to result in forecast improvements. Table 3 shows that the $\mathfrak{M}_b^*$ marginal model produces a lower RMSFE for output and the interest rate, while the $\mathfrak{M}_a^*$ model generates better forecasts of inflation. Perhaps not surprisingly, the RMSFE is smallest if the realized values for the exogenous variables are used. But setting the exogenous variables to their sample means also produces a low RMSFE that is comparable to those of the other marginal models. A possible reason is that changes in the exogenous variables, in particular the oil price, are close to a random walk and thus difficult to forecast. Finally, the *AveM* results based on forecasts across the different marginal models are shown in the third column of Table 3. We compute averages over the $\mathfrak{M}_a^*$ and $\mathfrak{M}_b^*$ models only since $\mathfrak{M}_c^*$ and $\mathfrak{M}_c^*$ do not constitute proper models for the exogenous variables.

The last row in each panel of Table 3 shows the RMSFE for forecasts that are averaged across different conditional models. Of particular interest is the average over both the different conditional and the marginal models, which is in the third column of the last row in each panel of Table 3. Averaging over all model dimensions produces an RMSFE that is close to the lowest of all individual RMSFEs in the table. This leads us to expect a further improvement in forecast performance if different

---

[19]The $\mathfrak{M}_d^*$ model corresponds to what is done in so-called 'scenario forecasts' where the exogenous variables are assumed to be known.

estimation windows are taken into account; an issue that we will explore next.

4.2   Averages over estimation windows ($AveW$)

We investigate the effect of changing the estimation window by estimating each model on a sample starting in 1965Q4 and then reducing the estimation sample successively by leaving out one year at a time at the beginning of the sample. Our shortest estimation window starts in 1976Q4, which is just after the break-down of the Bretton-Woods-System that has changed the behaviour of many macroeconomic variables considerably.[20] This gives a total of twelve different estimation windows. For the over-identified models, the long-run slope coefficients are kept constant at their 1965Q4-to-1999Q4 values and are not re-estimated over the shorter sample periods.[21] Since the long-run relations are based on economic theory we can expect them to be more stable across time than the short-run adjustment coefficients, which are estimated from the data without any restrictions. Moreover, there is little agreement in economic theory on the forces that drive the short-run adjustment of macroeconomic variables to their long-run equilibrium values. Note that the just-identified β vectors are re-estimated since we cannot attach an economic interpretation to them.

Figures 1 to 3 indicate that averaging forecasts from models estimated over different estimation windows improves the forecasts. The figures display the distribution of quarterly RMSFEs for forecasts of inflation, output growth and the short-term interest rate over the next year for each model, estimated over twelve different estimation windows, starting between 1965Q4 and 1976Q4. The estimation windows are shown on the horizontal axis and the RMSFE on the vertical axis. Since we have 36 different specifications for β, four different lag lengths and two marginal models, this gives a total of 288 models for each estimation window. The whiskers of the error bars indicate the 15th percentile and the 85th percentile of the RMSFEs, while the lower end of the box marks the 25th percentile and the upper end the 75th percentile. The line inside the box represents the median. RMSFEs falling outside the 15th and the 85th percentile are marked by dots. The RMSFE from our long-run restricted VECX*(2,2) model is identified by an asterisk. We see that for the longer estimation windows the VECX*(2,2) does not perform particularly well, whereas

---

[20]Since the model contains a fairly large number of estimated coefficients, a further reduction in sample size does not seem appropriate.

[21]The constants in equation (3) are re-estimated together with the short-run coefficients.

its RMSFE for output growth and inflation is in the lower quartile range for the estimation windows starting after 1974. This suggests the presence of a structural break, but this information is, of course, not available *ex ante.*

4.3   Averaging over models and windows (*AveAve*)

In the following, we will discuss how forecasting performance improves when forecasts are averaged both across models and estimation windows. From Figures 1 to 3 it is apparent that considerable variability in RMSFEs is present, both across the model and the window dimensions. In particular, windows starting in 1973Q4 and 1974Q4, i.e., at the time of the first oil-price shock, display comparatively large RMSFEs. One can also see, however, that not all models are affected in the same way by the choice of estimation window. The straight line in Figures 1 to 3 represents the RMSFE for forecasts that are averaged both across models and estimation windows, denoted as the '*AveAve*' forecast. In all cases the *AveAve* forecast lies in the lower part of the distribution of RMSFEs.

Figures 4 to 6 show the RMSFE across different forecast horizons. For each forecast horizon the *AveAve* RMSFE is marked by an asterisk and the *AveM* RMSFE for the longest estimation window by a circle. Since we consider forecasts from all models estimated over all estimation windows, we have 3456 forecasts at each forecast horizon. Again, the AveAve forecast performs well compared to the RMSFE of individual models while for inflation the *AveM* forecast for the longest window performs almost as well as the *AveAve* inflation forecast. For output growth and the interest rate averaging forecasts from models estimated over different estimation windows results in a further improvement of forecasts, especially at longer forecast horizons. Note, however, that the *AveM* RMSFE for inflation is in the lowest quartile at all forecast horizons already so that the scope for further improvement is small.

Averaging forecasts across different dimensions is an attractive strategy to improve forecast performance. Though some models beat the *AveAve* forecast, these models are not the same for the different variables and also change with the estimation window. It is thus apparent that the *ex ante* information needed to pick the best model is not available in practice. By considering the average over different windows the forecaster is able to hedge against a bad forecasting performance from a particular window. Since *a priori* one does not know how the choice of the sample period will affect the forecasting performance, averaging forecasts from models

estimated over different windows seems a useful practical way of dealing with this uncertainty.

4.4   Evaluating the *AveAve* forecast

While it is apparent that the *AveM* and the *AveAve* forecasts perform well, it is interesting to know how much one would have gained if one had picked the best model instead of using average forecasts. Two useful measures are the percentage of models that have a lower RMSFE than the *AveM* forecast and the difference between the average RMSFE of the models with a lower RMSFE and the *AveM* RMSFE. Table 4 provides these summary statistics for the performance of the *AveM* forecast relative to the individual forecasts. Regarding the results for the different estimation windows, only ess than 20 percent of the models manage to beat the AveM forecast of inflation whereas less than 32 percent of the models outperform the *AveM* forecast for output growth. For most estimation windows, however, these figures are considerably lower. For the interest rate, the *AveM* forecast performs slightly worse but still beats at least 50 percent of the individual model forecasts.

When it comes to the *AveAve* forecast, results are even more supportive of the averaging strategy. For inflation and output only 11 percent of the individual RMSFEs are lower than the RMSFE for the *AveAve* forecast, whereas for the interest rate this figure rises to 32 percent of the individual models. The average gain of using the better performing models in terms of the percentage reduction in RMSFE is small and amounts to about 15 percent for output and the interest rate, and 25 percent for inflation. One needs to keep in mind, however, that the information needed to pick the best performing model/window is not known *ex ante*.

We now turn to a comparison of the predictive accuracy of the *AveAve* forecasts relative to the forecasts from the long-run restricted VECX*(2,2) model, and an alternative simple benchmark model, namely a univariate AR(1) model.[22]   To assess whether the improvement in forecasting accuracy is significant, we apply the test of predictive accuracy proposed by Diebold and Mariano (1995). The test is based on a comparison of forecast errors from two different models, i and j, according to some loss function, $\mathcal{L}_t$, of the forecast errors, and tests whether the loss differential of two different forecasts is significantly different from zero. We consider the squared loss $\mathcal{L}_{ij,t}^s = (x_t - \widehat{x}_{t,h,i})^2 - (x_t - \widehat{x}_{t,h,j})^2$ and the absolute loss,

---

[22]In the literature univariate AR(1) models are often chosen as benchmark for forecast evaluation since they are hard to outperform despite their simplicity.

$\mathcal{L}_{ij,t}^{a} = |x_t - \widehat{x}_{t,h,i}| - |x_t - \widehat{x}_{t,h,j}|$, where i is the *AveAve* forecast and j the forecast from either the long-run restricted VECX*(2,2) model or a univariate AR(1) model. When considering forecasts more than one-step ahead, the loss differentials will be serially correlated. To estimate the variance of the loss differential we therefore use a heteroscedasticity and autocorrelation consistent estimate of the variance and correct for serial correlation of order $h - 1$, where h is the forecast horizon. We follow Harvey, Leybourne and Newbold (1997) who suggest to apply a correction factor to the Diebold-Mariano test statistic and to evaluate significance relative to the critical value from the Student's t distribution. We consider only forecasts up to four steps ahead since for longer horizons the number of independent observations becomes too small to expect significant results.

The upper panel of Table 5 shows that the *AveAve* forecast outperforms the forecast of the long-run restricted VECX*(2,2) model, which is indicated by a negative test statistic. In particular, the *AveAve* forecast is significantly better than the long-run restricted VECX*(2,2) model when considering the squared forecast errors, except for output growth three and four quarters ahead, and inflation one, three and four quarters ahead. Dechamps (2007) notes that even for $h = 1$ forecast errors need not be serially uncorrelated if the parameter values of the true model are unknown, and hence a semiparametric estimate of the variance may be necessary also in this case. Indeed, if a correction for first-order autocorrelation is applied, the test statistic becomes -1.807 and thus significant. Regarding the absolute loss, the *AveAve* forecast is significantly better than the VECX*(2,2) forecast for the interest rate and output growth up to two quarters ahead, but not for inflation.

The lower panel Table 5 shows that, compared to the forecast from a univariate AR(1) model, the *AveAve* forecast improves significantly for inflation and the interest rate but not for output.[23] Again, the one-step-ahead test statistic for the squared loss for inflation becomes -3.869 and thus significant if serial correlation is allowed for. The fact that the *AveAve* forecast does not lead to a better prediction of output growth indicates that the additional information coming from the other variables in the model does not help to improve forecasts over the information embodied in past output growth. This, however, might be a consequence of the particular forecast period chosen, which includes a high degree of uncertainty in the financial markets during 2001/2002 that subsequently led to a recession, and a steep rise in the oil price in 2004 that coincided with an economic recovery.

---

[23]These results remain unchanged when the average over estimation windows (*AveW*) for the AR(1) model is considered as benchmark instead.

Summing up, averaging forecasts from different windows and models seems to perform well and is worthy of further consideration.

4.5 Results for different weighting schemes

While up to now we have used equal weights, we next turn to the question of how best to combine the forecasts from different models, i.e., the effect of different weighting schemes on the average forecasts. In addition to equal weights we consider weighting by the AIC criterion (see Pesaran, Schleicher, and Zaffaroni, 2007), the weighting scheme proposed by Yang (2004) and the online weights discussed in Sancetta (2006). A description of the weighting schemes can be found in the appendix. First, we discuss the evolution of weights during the forecast horizon before we look at the influence on the RMSFE for the inflation forecast for up to four quarters ahead. The alternative weighting schemes are compared with respect to the conditional models only, and the uncertainty associated with the choice of the marginal models is dealt with by simple averaging.

Different weighting schemes imply markedly different weights with which the forecasts from a particular model enter the average. Figure 7 shows the evolution of the weights for the longest estimation window over the forecast period. Since it is impossible to depict the weight for each individual model, we show the sum of weights for the VECX*(2,2), the VECX*(2,1), the VECX*(1,2) and the VECX*(1,1) models. The online weights stay close to the equal weights, whereas the AIC weights tend to place most of the weight on the VECX*(1,2) model with only the long-run output gap relation imposed. The weighting scheme by Yang (2004) starts out with equally weighted models for the first period but re-adjusts weighting quickly, favouring a single model type at the time.

In choosing the weights, the forecaster faces a trade-off. On the one hand, the worst (historically) performing models should be excluded from the combined forecast. On the other hand, if model averaging is to provide a hedge against the failure of a particular model, convergence of the weights to a single model is not attractive. Since the AIC weights use the exponential difference between model m's AIC and the maximum AIC over all models, small differences in the log-likelihood will result in a large change in the weight. There is no guarantee, however, that the historically best model according to the AIC will always produce good forecasts. Therefore, weighting schemes that retain a broader portfolio of models, even if their performance was not among the best ones, may work better in practice.

Table 6 shows the RMSFE for the inflation forecast up to four quarters ahead with different weighting schemes. Apparently, equal weights perform quite well when compared to more sophisticated weighting schemes.[24] The online-weighting scheme is able to reduce the RMSFE slightly as compared to equal weights for some of the estimation windows but not for the *AveAve* forecast. By contrast, the AIC weights and the weighting schemes by Yang (2004) are unable to outperform equal weighting. This may be due to the fact that we consider quite similar models so that the advantages of keeping a large portfolio of models outweigh the benefit of excluding the worst performing ones.

5 Conclusions

In this paper, we developed a long-run structural model for Switzerland and tested for long-run relationships derived from economic theory. We found five cointegrating relations that we identified as PPP, money demand, international output growth, uncovered interest parity and the Fisher interest parity. We then investigated forecasting performance of different versions of this model, maintaining different assumptions with respect to the long-run relations, the lag length and the specification of the marginal model. Furthermore, we considered forecasts constructed from models that were estimated over different estimation windows.

We found that forecast averaging tends to improve forecasting performance and provides a hedge against poor forecast outcomes. While averaging across different models lowers the RMSFE of forecasts, averaging over estimation windows leads to an additional reduction in the forecast error and is thus at least as important as model averaging. Finally, we found that equal weights perform reasonably well when aggregating forecasts. The rationale behind this finding is that convergence of weights towards a single model is not attractive in practice if the researcher does not know whether the true model is among the set of models under consideration. In such cases, average forecasts based on portfolio of models estimated over different windows are likely to perform better than forecasts based on a single model considered to be the 'best' based on in-sample information.

---

[24]This result is often found in the forecast combination literature though not completely understood yet (see Timmermann 2006). Smith and Wallis (2007) explain this 'forecast combination puzzle' by the finite-sample error in estimating the combining weights.

A Appendix: Weighting schemes

Let f_{mth} be the mth model's h-step ahead forecast of a scalar random variable, z, formed at date t for date $t + h$, with $m = 1, 2, ..., M$, $t = 1, 2, ...$ Let $\omega_{mth} > 0$, $\sum_{m=1}^{M} \omega_{mth} = 1$, be the weight to be attached to this forecast at time t in arriving at the pooled forecast defined by

$$f_{t,h}(\omega) = \sum_{m=1}^{M} \omega_{mth} f_{mth}.$$

Many different weighting schemes can be considered.

One possibility is to use equal weighted combinations defined as

$$f_{t,h}(1/M) = \frac{1}{M} \sum_{m=1}^{M} f_{mth}.$$

Another one is to approximate $\Pr(\mathfrak{M}_m \mid \mathbf{Z}_{T,T})$ by Akaike weights or Schwartz weights. The latter give a Bayesian approximation when the estimation sample is sufficiently large (see Pesaran, Schleicher and Zaffaroni, 2007).

Here, we will consider AIC weights that are computed as follows:

$$\overline{\omega}_{m,t-1} = \frac{\exp(\Delta_{m,t-1})}{\sum_{j=1}^{M} \exp(\Delta_{j,t-1})},$$

where $\Delta_{m,t-1} = AIC_{m,t-1} - \max_j(AIC_{j,t-1})$ and $AIC_{m,t-1} = LL_{m,t-1} - \theta_m$ and $LL_{m,t-1}$ indicates the maximized logarithm of the likelihood function of model m with θ_m parameters.[25]

Yang (2004) proposes the following weights for $h = 1$ (see his equation (4) on page 186)

$$\mathring{\omega}_{m,t} = \frac{\frac{\pi_m}{\prod_{\tau=1}^{t-1} s_{m\tau}} \exp\left\{ -\frac{1}{2} \sum_{\tau=1}^{t-1} (z_\tau - f_{m\tau})^2 / s_{m\tau}^2 \right\}}{\sum_{j=1}^{M} \frac{\pi_j}{\prod_{\tau=1}^{t-1} s_{j\tau}} \exp\left\{ -\frac{1}{2} \sum_{\tau=1}^{t-1} (z_\tau - f_{j\tau})^2 / s_{j\tau}^2 \right\}},$$

where f_{mt} is the one-step ahead forecast of z_t formed at time t, and the model priors, π_m, can be set to $1/M$. This formula uses an expanding window for the construction

---

[25] For the exactly identified models, θ_m is given by $\theta_m = kk_x(p-1) + (k+k_x+1)r - r^2 + (k_{x^*}+1)k_x$.

of weights and can be modified to use a rolling window of size $\mathfrak{h}$,

$$\breve{\omega}_{m,t} = \frac{\frac{\pi_m}{\prod\limits_{\tau=t-\mathfrak{h}}^{t-1} s_{m\tau}} \exp\left\{-\frac{1}{2}\sum\limits_{\tau=t-\mathfrak{h}}^{t-1}(z_\tau - f_{m\tau})^2/s_{m\tau}^2\right\}}{\sum\limits_{j=1}^{M} \frac{\pi_j}{\prod\limits_{\tau=t-\mathfrak{h}}^{t-1} s_{j\tau}} \exp\left\{-\frac{1}{2}\sum\limits_{\tau=t-\mathfrak{h}}^{t-1}(z_\tau - f_{j\tau})^2/s_{j\tau}^2\right\}},$$

An h-step ahead version can be written as

$$\breve{\omega}_{mth} = \frac{\frac{\pi_m}{\prod\limits_{\tau=t-\mathfrak{h}-h+1}^{t-h} s_{m\tau}} \exp\left\{-\frac{1}{2}\sum\limits_{\tau=t-\mathfrak{h}-h+1}^{t-h}(z_\tau - f_{m\tau h})^2/s_{m\tau h}^2\right\}}{\sum\limits_{j=1}^{M} \frac{\pi_j}{\prod\limits_{\tau=t-\mathfrak{h}-h+1}^{t-h} s_{j\tau}} \exp\left\{-\frac{1}{2}\sum\limits_{\tau=t-\mathfrak{h}-h+1}^{t-h}(z_\tau - f_{j\tau h})^2/s_{j\tau h}^2\right\}},$$

where $s_{m\tau h}^2$ are computed from an expanding window (or a rolling window of size $\mathfrak{h}' > \mathfrak{h}$)

$$s_{m\tau h}^2 = \frac{\sum\limits_{i=\tau-\mathfrak{h}'-h+1}^{\tau-h}(z_i - f_{mih})^2}{\mathfrak{h}'},$$

where $\mathfrak{h}' = \tau - h$ in the case of an expanding window.

Alternatively, weighting schemes from machine learning literature can be used. One such scheme uses the following algorithm (Sancetta 2006): Let $t = \tau$ be the initial forecast date and set $\omega_{m\tau h} = 1/M$. For date $t = \tau + h, \tau + h + 1, ...$, use the following formula to update the weights

$$\begin{aligned}
\omega_{mth} &= \lambda_{t,t-h,h}\widetilde{\omega}_{m,t-h,h}, \text{ if } \widetilde{\omega}_{m,t-h,h} \geq \frac{\gamma}{M}, \text{ for } t = \tau, \tau+1, ..., \\
&= \frac{\gamma}{M} \text{ if } \widetilde{\omega}_{m,t-h,h} < \frac{\gamma}{M}
\end{aligned}$$

where

$$\widetilde{\omega}_{m,t-h,h} = \frac{\omega_{m,t-h,h} \exp\left\{\eta_{t-h} f_{m,t-h,h}(z_{t-h} - f_{t-h,h})\right\}}{\sum\limits_{m=1}^{M} \omega_{m,t-h,h} \exp\left\{\eta_{t-h} f_{m,t-h,h}(z_{t-h} - f_{t-h,h})\right\}},$$

z_{t-h} is the realized value of z at the end of date $t - h$,

$$\lambda_{t,t-h,h} = \frac{1 - \left(\frac{\gamma}{M}\right)\sum\limits_{m=1}^{M} I(\frac{\gamma}{M} - \widetilde{\omega}_{m,t-h,h})}{\sum\limits_{m=1}^{M} \widetilde{\omega}_{m,t-h,h} I(\widetilde{\omega}_{m,t-h,h} - \frac{\gamma}{M})},$$

$$\eta_t = At^{-\alpha}.$$

Note that by construction the new weights satisfy $\widetilde{\omega}_{m,t-h,h} > 0$, and $\sum_{m=1}^{M} \widetilde{\omega}_{m,t-h,h} = 1$. In the empirical application we set $A = 10^5$, $\alpha = 0.5$ and $\gamma = 0.05$.[26]

---

[26]We chose a value of A much higher than recommended by Sancetta (2006) since otherwise the online weights were indistinguishable from equal weights. It is interesting to note that results remain basically unaffected if we change the weights ex-post by choosing $\alpha = \{0.5, 0.4, 0.3, 0.2\}$ and $\gamma = \{0.05, 0.10\}$. Since our evaluation sample with at most 27 observations is rather short, we are unlikely to benefit from online weighting.

References

Assenmacher-Wesche, K. and M. H. Pesaran (2008): "A VECX* Model of the Swiss Economy," *Swiss National Bank Economy Studies*, forthcoming.

Clark, T. E. and M. W. McCracken (2004): "Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts," Federal Reserve Bank of Kansas City, *mimeo*.

Clark, T. E. and M. W. McCracken (2006): "Averaging Forecasts from VARs with Uncertain Instabilities," Federal Reserve Bank of Kansas City, *Research Working Paper* No. RWP-06-12.

Clements, M. P. and D. F. Hendry (1998): "Forecasting Economic Time Series," *Cambridge University Press,* Cambridge.

Clements, M. P. and D. F. Hendry (1999): "Forecasting Non-stationary Economic Time Series," *The MIT Press,* Cambridge (Mass.).

Clements, M. P. and D. F. Hendry (2006): "Forecasting with Breaks," **G. Elliot, C. W. J. Granger and A. Timmermann** (eds), *Handbook of Economic Forecasting*, Elsevier, Amsterdam, 605-657.

Deschamps, P. J. (2007): "Comparing Smooth Transition and Markov Switching Autoregressive Models of US Unemployment," Université de Fribourg, *mimeo*.

Diebold, F. X. and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 134-144.

Elliot, G. and A. Timmermann (2007): "Economic Forecasting," CEPR Discussion Paper 6158.

Garratt, A., K. Lee, M. H. Pesaran and Y. Shin (2003a): "A Long Run Structural Macroeconometric Model of the UK," *Economic Journal* 113, 412-455.

Garratt, A., K. Lee, M. H. Pesaran and Y. Shin (2003b): "Forecast Uncertainties in Macroeconomic Modeling: An Application to the U.K. Economy," *Journal of the American Statistical Association* 98, 829-838.

Garratt, A., K. Lee, M. H. Pesaran and Y. Shin (2006): "Global and National Macroeconometric Modelling: A Long Run Structural Approach," *Oxford University Press*, Oxford.

Geweke, J. and C. H. Whiteman (2006): "Bayesian Forecasting," **G. Elliot, C. W. J. Granger and A. Timmermann** (eds), *Handbook of Economic Forecasting*, Elsevier, Amsterdam, 3-80.

Harvey, D., S. J. Leybourne and P. Newbold (1997): "Testing the Equality

of Prediction Mean Squared Errors," *International Journal of Forecasting* 13, 281-291.

Jordan, T. J. and M. R. Savioz (2003): "Does it Make Sense to Combine Forecasts from VAR Models? An Empirical Analysis with Inflation Forecasts for Switzerland," *Schweizerische Nationalbank, Quartalsheft* 4, 80-93.

Pesaran, M. H., D. Pettenuzzo and A. Timmermann (2006): "Forecasting Time Series Subject To Multiple Structural Breaks," *Review of Economic Studies* 73, 1057-1084.

Pesaran, M. H. and S. Skouras (2002): "Decision-Based Methods for Forecast Evaluation," **M. P. Clements and D. F. Hendry** (eds), *A Companion to Economic Forecasting*, Basil Blackwell, Oxford, 241-267.

Pesaran, M. H. and A. Pick (2007): "Forecasting Random Walk Models under Drift Instability, University of Cambridge, *mimeo.*

Pesaran, M. H. and A. Timmermann (2007): "Selection of Estimation Window in the Presence of Breaks," *Journal of Econometrics* 137, 134-161.

Pesaran, M. H., C. Schleicher, and P. Zaffaroni (2007): "Model Averaging in Risk Management with an Application to Futures Markets," *mimeo.*

Sancetta, A. (2006): "Online Forecast Combination for Dependent Heterogeneous Data," University of Cambridge, *mimeo.*

Smith, J. and K. F. Wallis (2007): "A Simple Explanation of the Forecast Combination Puzzle," University of Warwick, *mimeo.*

Timmermann, A. (2006): "Forecast Combinations," **G. Elliot, C. W. J. Granger and A. Timmermann** (eds), *Handbook of Economic Forecasting*, Elsevier, Amsterdam, 135-196.

Yang, Y. (2004): "Combining Forecasting Procedures: Some Theoretical Results," *Econometric Theory* 20, 176-222.

Table 1: RMSFE for long-run restricted VECX*(2,2) model with over-identified β

RMSFE in %	#	y_t	π_t	r_t
Horizon				
1 step ahead	27	0.572	0.272	0.070
2 step ahead	26	0.457	0.155	0.068
3 step ahead	25	0.428	0.122	0.066
4 step ahead	24	0.402	0.101	0.068
8 step ahead	20	0.328	0.069	0.063

Note: Sequential out-of-sample forecasts from 2000Q1 to 2006Q3, estimation period 1965Q4 to 1999Q4. The forecast statistics pertain to forecasts for h steps ahead, divided by the forecast horizon, h. Forecasts of the exogenous variables come from the $\mathfrak{M}_a^*$ marginal model. # indicates the number of point forecasts available to compute the RMSFE.

Table 2: RMSFE for average forecast over different β of VECX*(2,2) model

RMSFE in %	#	y_t	π_t	r_t
Horizon				
1 step ahead	27	0.540	0.236	0.067
2 step ahead	26	0.407	0.113	0.062
3 step ahead	25	0.363	0.082	0.058
4 step ahead	24	0.327	0.066	0.060
8 step ahead	20	0.232	0.039	0.062

Note: Sequential out-of-sample forecasts from 2000Q1 to 2006Q3, estimation period 1965Q4 to 1999Q4. The forecast statistics pertain to forecasts for h steps ahead, divided by the forecast horizon, h. Forecasts of the exogenous variables come from the $\mathfrak{M}_a^*$ marginal model. # indicates the number of point forecasts available to compute the RMSFE.

Table 3: RMSFE for average four-quarter-ahead forecast across different model dimensions

	$\mathfrak{M}_a^*$	$\mathfrak{M}_b^*$	Average	$\mathfrak{M}_c^*$	$\mathfrak{M}_d^*$
y_t					
VECX*(2,2)	0.327	0.318	0.315	0.314	0.335
VECX*(2,1)	0.315	0.307	0.302	0.306	0.330
VECX*(1,2)	0.352	0.336	0.342	0.331	0.271
VECX*(1,1)	0.331	0.313	0.319	0.314	0.299
Average	0.325	0.316	0.316	0.313	0.305
π_t					
VECX*(2,2)	0.066	0.069	0.067	0.065	0.044
VECX*(2,1)	0.068	0.069	0.068	0.066	0.045
VECX*(1,2)	0.072	0.075	0.073	0.073	0.068
VECX*(1,1)	0.075	0.077	0.076	0.075	0.064
Average	0.069	0.071	0.070	0.069	0.052
r_t					
VECX*(2,2)	0.060	0.058	0.058	0.057	0.027
VECX*(2,1)	0.056	0.053	0.054	0.054	0.028
VECX*(1,2)	0.063	0.060	0.061	0.058	0.026
VECX*(1,1)	0.060	0.056	0.058	0.057	0.024
Average	0.059	0.056	0.058	0.056	0.025

Note: Sequential out-of-sample forecasts from 2000Q1 to 2006Q3. The table shows the RMSFE of the average forecast over different β per quarter. $\mathfrak{M}_a^*$ and $\mathfrak{M}_b^*$ indicate the marginal models described in Section 4.1, $\mathfrak{M}_c^*$ and $\mathfrak{M}_d^*$ set the exogenous variables to their sample mean or their realized value, average indicates the average over the $\mathfrak{M}_a^*$ and $\mathfrak{M}_b^*$ marginal models. The marginal models are estimated over the same sample as the conditional model. All results are averaged over the different choices for β.

Table 4: Summary of performance of *AveM* forecast relative to individual forecasts across estimation windows

Window	y_t		π_t		r_t	
	Percent	Exceedence	Percent	Exceedence	Percent	Exceedence
1965 Q4	13.542	0.057	16.667	0.021	31.944	0.006
1966 Q4	13.542	0.057	19.097	0.021	30.903	0.007
1967 Q4	11.458	0.057	15.972	0.022	32.986	0.007
1968 Q4	8.681	0.055	19.444	0.020	30.556	0.008
1969 Q4	13.889	0.061	18.403	0.019	26.736	0.010
1970 Q4	27.431	0.081	16.667	0.021	30.208	0.008
1971 Q4	31.250	0.086	13.194	0.025	42.014	0.008
1972 Q4	29.167	0.085	7.986	0.027	46.181	0.007
1973 Q4	16.667	0.070	18.403	0.040	50.000	0.010
1974 Q4	6.250	0.051	5.556	0.021	30.556	0.010
1975 Q4	14.236	0.025	4.861	0.021	31.944	0.007
1976 Q4	15.972	0.023	7.292	0.020	31.944	0.007
AveAve	10.619	0.054	10.735	0.017	32.060	0.010
AveAve RMSFE		0.313		0.069		0.054

Note: Sequential out-of-sample forecasts from 2000Q1 to 2006Q3. Forecasts are averaged over all models and pertain to the four-quarter-ahead forecast. Percent shows the share of models whose RMSFE is below the *AveW* RMSFE. Exceedence gives the average RMSFE loss of not using those models that perform better than the *AveW forecast*. For comparison, the last row shows the RMSFE of the *AveAve forecast*.

Table 5: Predictive accuracy of *AveAve* forecast

	Against long-run restricted VECX*(2,2) model					
	Squared loss			Absolute loss		
Horizon	y_t	π_t	r_t	y_t	π_t	r_t
1	**-1.954**	-0.766	**-1.793**	**-2.737**	-0.462	**-2.317**
2	**-2.088**	**-2.167**	**-3.160**	**-1.900**	-1.422	**-3.632**
3	-1.507	-1.670	**-2.316**	-1.131	-1.469	**-3.651**
4	-1.334	-1.421	**-1.829**	-1.273	-1.536	**-2.449**
	Against AR(1) model					
	Squared loss			Absolute loss		
Horizon	y_t	π_t	r_t	y_t	π_t	r_t
1	-0.251	-1.626	**-3.173**	0.546	**-1.905**	**-4.801**
2	-0.743	**-3.197**	**-2.407**	-0.558	**-4.102**	**-2.930**
3	-0.130	**-3.356**	**-1.900**	-0.372	**-3.685**	**-2.538**
4	0.287	**-4.594**	-1.657	-0.249	**-35.001**	**-2.423**

Note: Significant test statistics at the 5 percent level are denoted in boldface. A negative entry indicates that the *AveAve* forecast outperforms the alternative model. The AR(1) model is estimated over the longest estimation window.

Table 6: RMSFE for inflation in percent for *AveM* forecast using different weights

Estimation window	Equal weights	AIC weights	Yang (2004)	Online weights
1965 Q4	0.068	0.071	0.086	0.067
1966 Q4	0.069	0.072	0.079	0.068
1967 Q4	0.068	0.070	0.077	0.067
1968 Q4	0.069	0.072	0.085	0.068
1969 Q4	0.069	0.074	0.092	0.069
1970 Q4	0.068	0.074	0.076	0.068
1971 Q4	0.068	0.072	0.074	0.067
1972 Q4	0.069	0.071	0.078	0.067
1973 Q4	0.070	0.072	0.087	0.069
1974 Q4	0.080	0.073	0.091	0.076
1975 Q4	0.078	0.093	0.085	0.075
1976 Q4	0.077	0.082	0.082	0.075
AveAve	0.069	0.073	0.076	0.069

Note: Sequential out-of-sample forecasts from 2000Q1 to 2006Q3. The table shows the RMSFE for the *AveM forecast, estimated over different estimation windows.* Forecasts are averaged over the $\mathfrak{M}_a^*$ and $\mathfrak{M}_b^*$ marginal models, applying equal weights. The marginal models are estimated over the same window as the conditional model.

Figure 1: RMSFE for output growth across estimation windows (starting between 1965Q4 and 1976Q4) and AveAve RMSFE
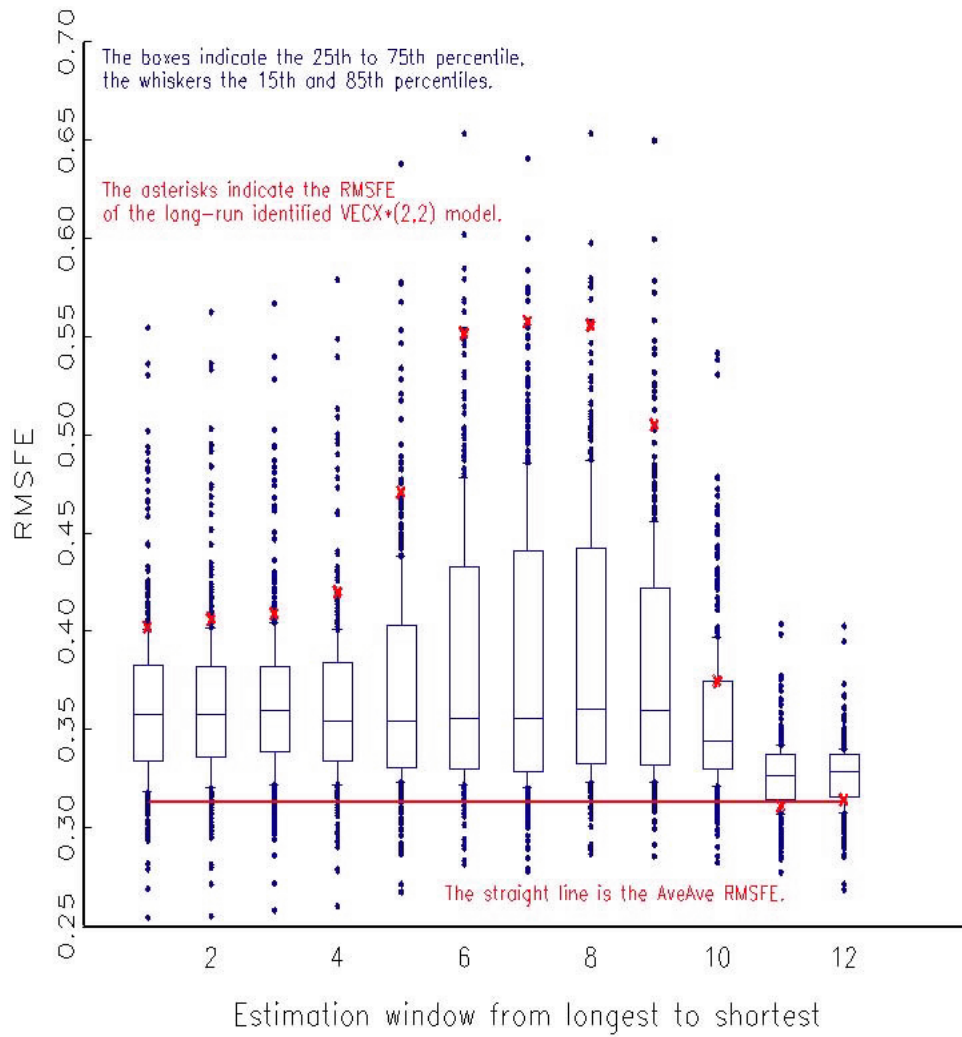
Figure 2: RMSFE for inflation across estimation windows (starting between 1965Q4 and 1976Q4) and AveAve RMSFE
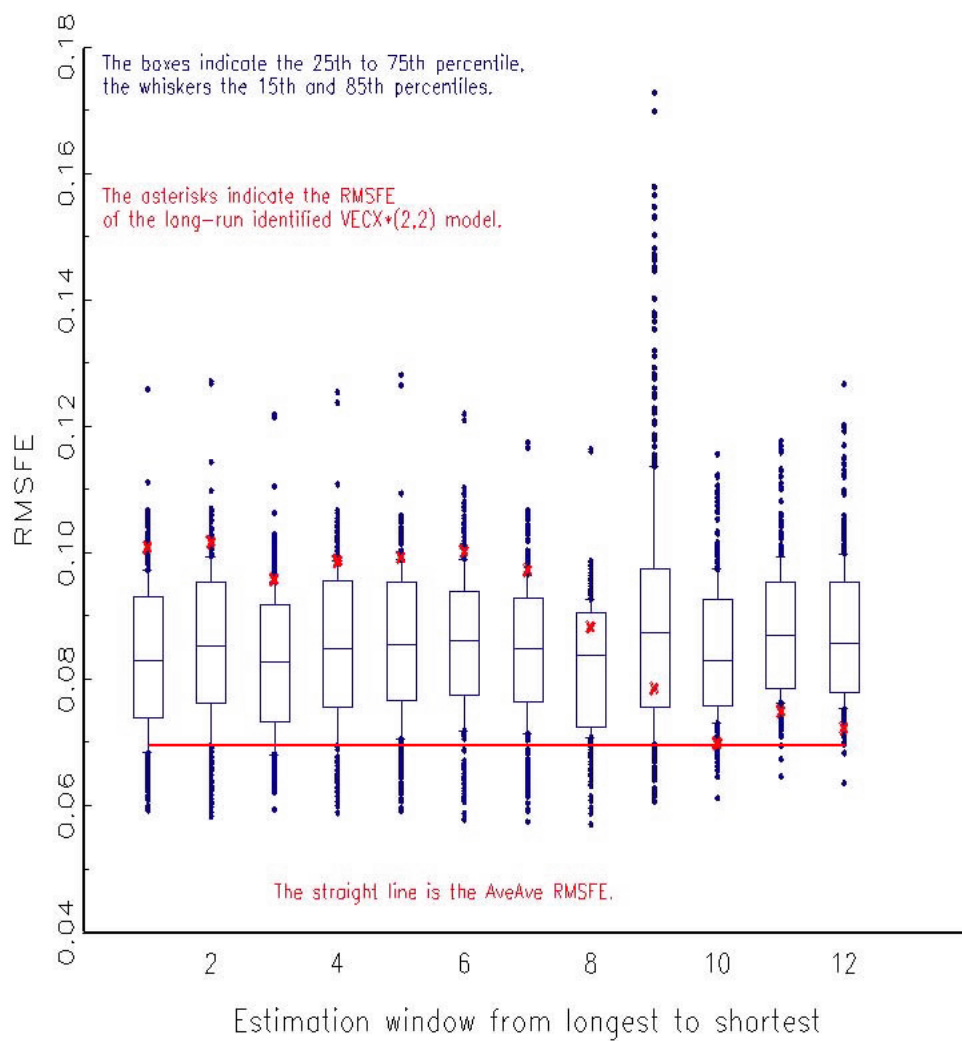
Figure 3: RMSFE for interest rate across estimation windows (starting between 1965Q4 and 1976Q4) and AveAve RMSFE
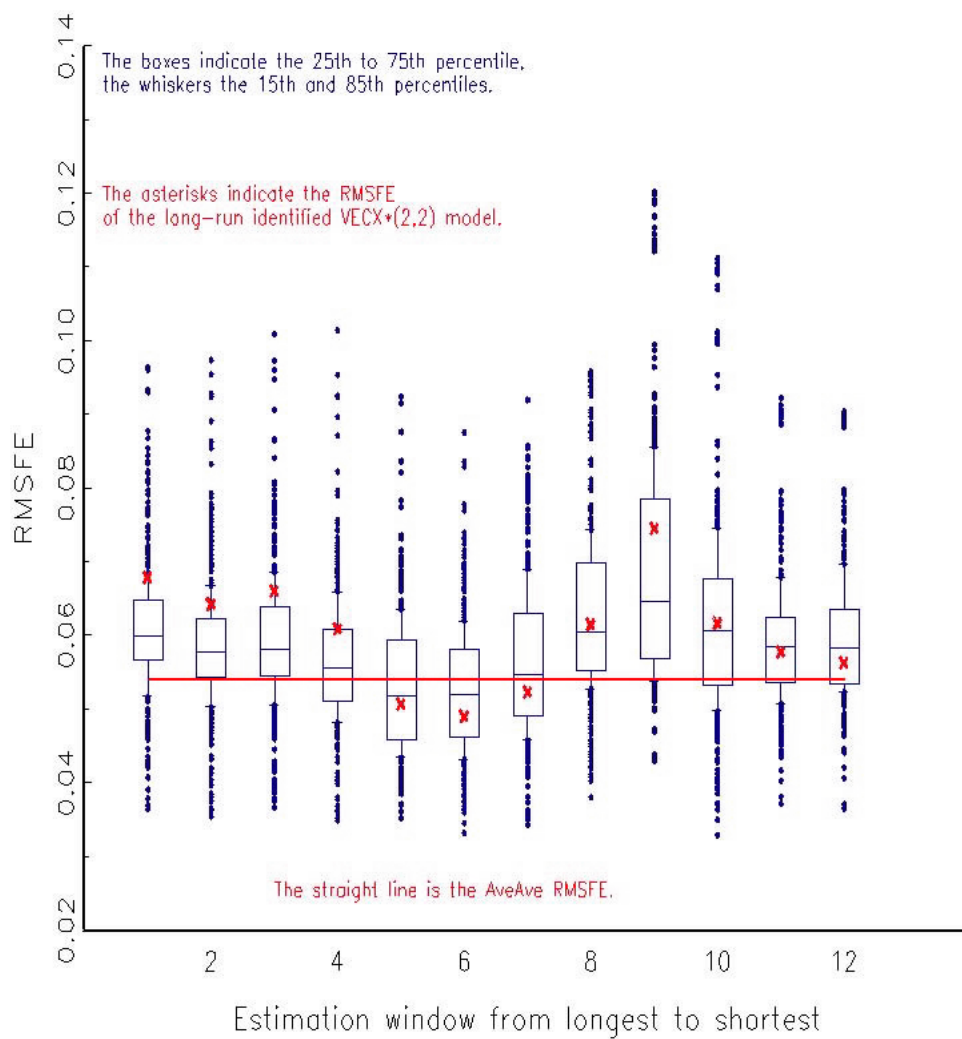
Figure 4: Distribution of RMSFEs for output growth across forecast horizons
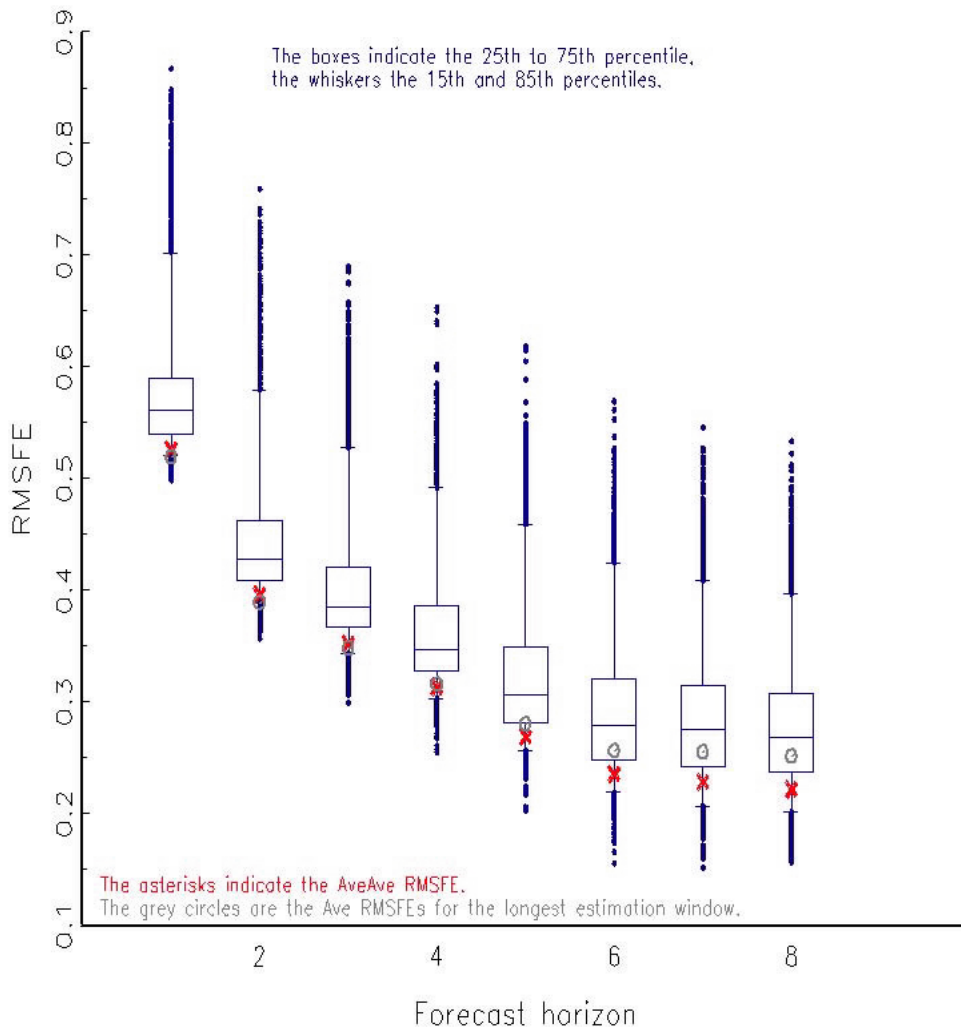
Figure 5: Distribution of RMSFEs for inflation across forecast horizons
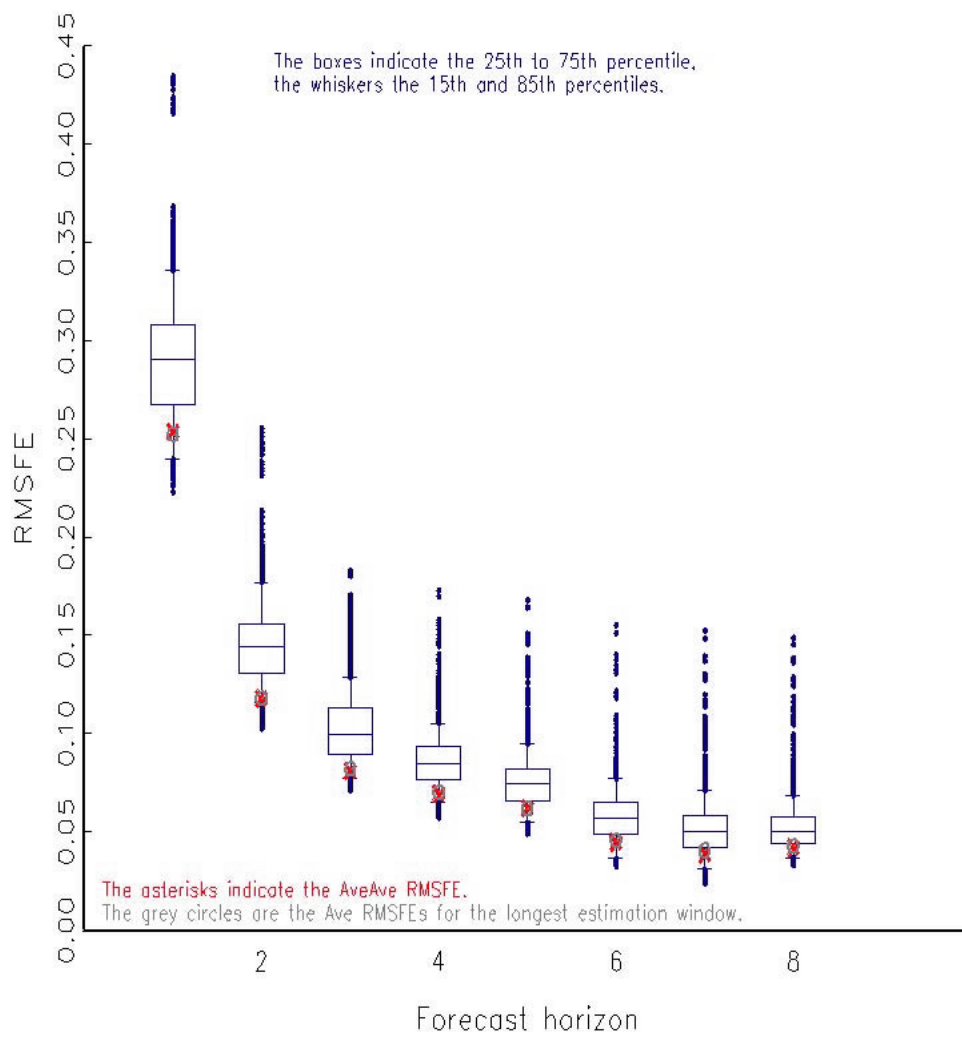
Figure 6: Distribution of RMSFEs for interest rate across forecast horizons
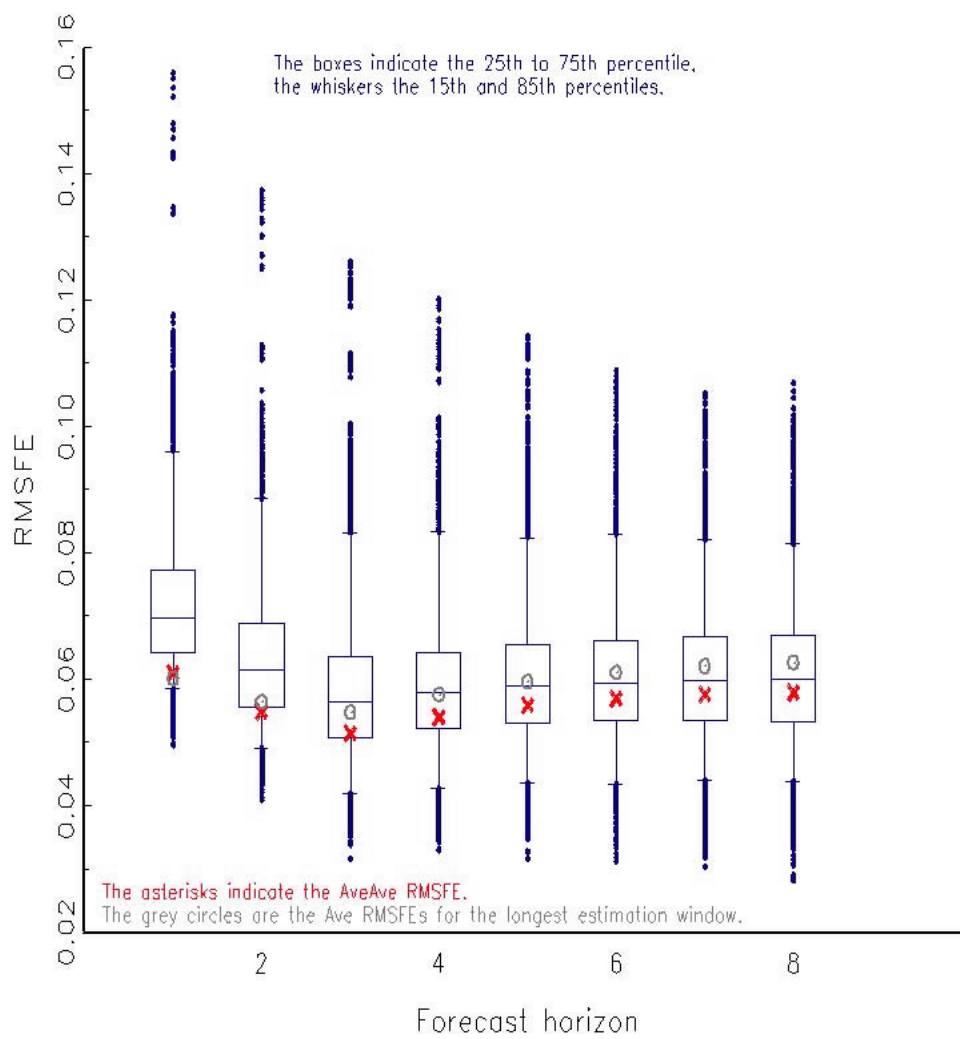
Figure 7: Evolution of weights over forecasting period
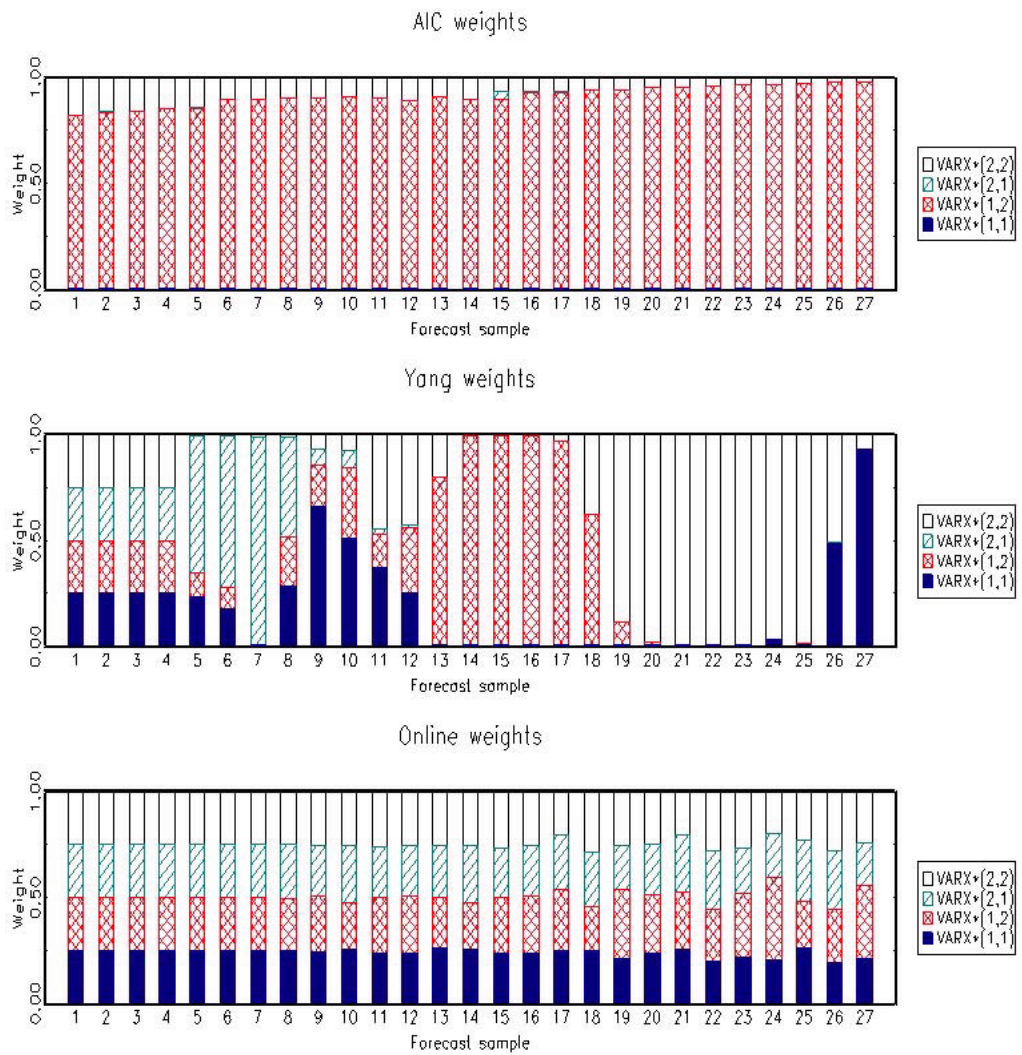
Swiss National Bank Working Papers published since 2004:

2004-1   Samuel Reynard: Financial Market Participation and the Apparent Instability of
          Money Demand

2004-2   Urs W. Birchler and Diana Hancock: What Does the Yield on Subordinated
          Bank Debt Measure?

2005-1   Hasan Bakhshi, Hashmat Khan and Barbara Rudolf: The Phillips curve under
          state-dependent pricing

2005-2   Andreas M. Fischer: On the Inadequacy of Newswire Reports for Empirical
          Research on Foreign Exchange Interventions

2006-1   Andreas M. Fischer: Measuring Income Elasticity for Swiss Money Demand:
          What do the Cantons say about Financial Innovation?

2006-2   Charlotte Christiansen and Angelo Ranaldo: Realized Bond-Stock Correlation:
          Macroeconomic Announcement Effects

2006-3   Martin Brown and Christian Zehnder: Credit Reporting, Relationship Banking,
          and Loan Repayment

2006-4   Hansjörg Lehmann and Michael Manz: The Exposure of Swiss Banks to
          Macroeconomic Shocks – an Empirical Investigation

2006-5   Katrin Assenmacher-Wesche and Stefan Gerlach: Money Growth, Output Gaps and
          Inflation at Low and High Frequency: Spectral Estimates for Switzerland

2006-6   Marlene Amstad and Andreas M. Fischer: Time-Varying Pass-Through from Import
          Prices to Consumer Prices: Evidence from an Event Study with Real-Time Data

2006-7   Samuel Reynard: Money and the Great Disinflation

2006-8   Urs W. Birchler and Matteo Facchinetti: Can bank supervisors rely on market data?
          A critical assessment from a Swiss perspective

2006-9   Petra Gerlach-Kristen: A Two-Pillar Phillips Curve for Switzerland

2006-10   Kevin J. Fox and Mathias Zurlinden: On Understanding Sources of Growth and
          Output Gaps for Switzerland

2006-11   Angelo Ranaldo: Intraday Market Dynamics Around Public Information Arrivals

2007-1   Andreas M. Fischer, Gulzina Isakova and Ulan Termechikov: Do FX traders in
          Bishkek have similar perceptions to their London colleagues? Survey evidence of
          market practitioners' views