DISCUSSION PAPER SERIES

IZA DP No. 3071

Assessing Forecast Uncertainties in a VECX Model for Switzerland: An Exercise in Forecast Combination across Models and Observation Windows

Katrin Assenmacher-Wesche
M. Hashem Pesaran

September 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Assessing Forecast Uncertainties in a VECX Model for Switzerland: An Exercise in Forecast Combination across Models and Observation Windows

Katrin Assenmacher-Wesche
Swiss National Bank

M. Hashem Pesaran
Cambridge University, CIMF
USC and IZA

Discussion Paper No. 3071
September 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3071
September 2007

ABSTRACT

Assessing Forecast Uncertainties in a VECX Model for Switzerland: An Exercise in Forecast Combination across Models and Observation Windows*

We investigate the effect of forecast uncertainty in a cointegrating vector error correction model for Switzerland. Forecast uncertainty is evaluated in three different dimensions. First, we investigate the effect on forecasting performance of averaging over forecasts from different models. Second, we look at different estimation windows. We find that averaging over estimation windows is at least as effective as averaging over different models and both complement each other. Third, we explore whether using weighting schemes from the machine learning literature improves the average forecast. Compared to equal weights the effect of the weighting scheme on forecast accuracy is small in our application.

JEL Classification: C53, C32

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

Corresponding author:

M. Hashem Pesaran
Faculty of Economics
University of Cambridge
Sidgwick Avenue
Cambridge, CB3 9DD
United Kingdom
E-mail: hashem.pesaran@econ.cam.ac.uk

* The views expressed in this paper are solely our own and not necessarily shared by the Swiss National Bank. We are grateful to Sylvia Kaufmann, Ron Smith and participants of the Oxford Forecasting Workshop and the BuBa-OeNB-SNB Workshop for helpful comments.

1 Introduction

Forecasting macroeconomic variables is of importance for market participants and policy makers alike. Although in general great care is taken in designing a specific forecasting model, the true forecast uncertainty is often underestimated since various sources of forecasting errors, like model uncertainty or future uncertainty, are not taken properly into account. For a recent review of the literature on forecasting see Elliott and Timmermann (2007).

This paper deals with forecast uncertainty in a long-run structural vector error correction 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, ouput 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 similar models that differ only with respect to these characteristics instead of considering entirely different model types. Statistical procedures used to select the order of the VECX* model or the number of cointegrating relations often give ambiguous results, such that different choices may be justified. Moreover, theory often suggests restrictions that are rejected by the data. Though imposing these restrictions inevitably deteriorates the fit of the model, forecasting performance may improve when imposing them. In addition, different assumptions can be maintained regarding the data-generation process for the exogenous variables. To assess their effects on the forecasts of the endogenous variables, we consider different marginal models for prediction of the exogenous variables. To allow for model uncertainty we apply Bayesian model averaging and combine forecasts from several, equally 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 the model over different observation windows. While estimation is more efficient if all available data are used, the occurrence of a structural break, which often is difficult to detect with statistical methods, might bias the estimates. 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 long-run coefficients, but consider alternative estimates of the short-run coefficients computed over different observation windows starting between 1965 and 1976.

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

We find that averaging over different models is able to reduce the forecast error considerably. In addition, averaging over estimation windows is at least as effective as model averaging in improving forecast precision. Moreover, averaging across these two dimensions complements each other, in the sense that averaging the model average forecasts over different estimation windows leads to a further reduction in the forecast error. By contrast, in our application the effect of different weighting schemes is minor. This is probably due to the fact that we consider a class of closely related models so that the gain from excluding certain variants of the models is not able to change forecasting performance significantly.

The paper proceeds 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 our model. In Section 4 we explore 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 that perform 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).[1]

In the implementation of the long-run structural modelling a number of choices have to be made, see Garratt, Lee, Pesaran and Shin (2006, pp 108-109), among which are the choice of the core endogenous and exogenous variables, their lag orders, the deterministics (namely the choice of intercept and linear trends), and the sample period. 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, real output and the nominal interest rate. Another is the Fisher interest-rate parity which establishes a long-run relation between the interest rate and inflation. Switzerland as a small, open economy can be expected to be subject to influences from the exchange rate. Therefore, purchasing power parity, which links the domestic price level to the nominal exchange rate and the foreign price level, is also included. In addition, we consider the price of 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 output and interest rates. The latter two variables, together with the oil price, are regarded as weakly exogenous variables.

2.1 Econometric methodology

Starting point of the VECX* modelling strategy is a standard vector autoregressive (VAR) model that can be written as

$$\mathbf{A}(L)\mathbf{z}_t = \mathbf{D}\mathbf{d}_t + \mathbf{u}_t, \tag{1}$$

[1] A more detailed documentation of the model can be found in Assenmacher-Wesche and Pesaran (2007).

where L is the lag operator such that $L\mathbf{z}_t = \mathbf{z}_{t-1}$, $\mathbf{A}(L) = \mathbf{I}_0 - \mathbf{A}_1 L - ... - \mathbf{A}_p L^p$ and $\mathbf{z}_t = (\mathbf{x}_t', \mathbf{x}_t^{*\prime})'$ consists of a $k_x \times 1$ vector of endogenous variables, $\mathbf{x}_t$, and a $k_{x^*} \times 1$ vector of exogenous variables, $\mathbf{x}_t^*$, with $k_x + k_{x^*} = k$. The vector $\mathbf{d}_t$ is a vector of deterministic variables, such as intercept and trend, and the error term, $\mathbf{u}_t$, is distributed as $iid(0, \boldsymbol{\Sigma})$.

The VECX* model starts with an explicit formulation of the long-run relationships between the variables in the model, derived from macroeconomic theory. These long-run relations are then incorporated in an otherwise unrestricted VAR. The cointegrating VECX* embeds the structural long-run relations as the steady-state solutions while the short-run dynamics, about which economic theory in general is silent, is estimated from the data without restrictions.

In error-correction form the VAR model in equation (1) can be written as

$$\Delta \mathbf{z}_t = -\boldsymbol{\Pi} \mathbf{z}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{z}_{t-i} + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{u}_t, \qquad (2)$$

where the matrix $\boldsymbol{\Pi}$ is a $k \times k$ matrix of long-run multipliers and the matrices $\{\boldsymbol{\Gamma}_i\}_{i=1}^{p-1}$ contain the short-run responses; $\mathbf{a}_0$ denotes a vector of constants and $\mathbf{a}_1$ a vector of trend coefficients. To partition the system into a conditional model for the endogenous variables, $\Delta \mathbf{x}_t$, and a marginal model for the exogenous variables, $\Delta \mathbf{x}_t^*$, the parameter matrices and vectors $\boldsymbol{\Pi}$, $\boldsymbol{\Gamma}_i$, $\mathbf{a}_0$, $\mathbf{a}_1$ and the error term $\mathbf{u}_t$ are partitioned conformably with $\mathbf{z}_t = (\mathbf{x}_t', \mathbf{x}_t^{*\prime})'$ as $\boldsymbol{\Pi} = (\boldsymbol{\Pi}_x', \boldsymbol{\Pi}_{x^*}')'$, $\boldsymbol{\Gamma}_i = (\boldsymbol{\Gamma}_{xi}', \boldsymbol{\Gamma}_{x^*i}')'$, $i = 1, ..., p-1$, $\mathbf{a}_0 = (\mathbf{a}_{x0}', \mathbf{a}_{x^*0}')'$, $\mathbf{a}_1 = (\mathbf{a}_{x1}', \mathbf{a}_{x^*1}')'$, and $\mathbf{u}_t = (\mathbf{u}_{xt}', \mathbf{u}_{x^*t}')'$. The variance matrix of $\mathbf{u}_t$ can be written as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx^*} \\ \boldsymbol{\Sigma}_{x^*x} & \boldsymbol{\Sigma}_{x^*x^*} \end{pmatrix},$$

so that

$$\mathbf{u}_{xt} = \boldsymbol{\Sigma}_{xx^*} \boldsymbol{\Sigma}_{x^*x^*}^{-1} \mathbf{u}_{x^*t} + \boldsymbol{\upsilon}_t,$$

where $\boldsymbol{\upsilon}_t \sim iid(0, \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xx^*} \boldsymbol{\Sigma}_{x^*x^*}^{-1} \boldsymbol{\Sigma}_{x^*x})$ is uncorrelated with $\mathbf{u}_{x^*t}$ by construction. By assuming that the process $\{\mathbf{x}_t^*\}_{t=1}^{\infty}$ is weakly exogenous with respect to the parameters in the matrix $\boldsymbol{\Pi}$, we have $\boldsymbol{\Pi}_{x^*} = \mathbf{0}$. This means that the information available from the model for $\Delta \mathbf{x}_t^*$ is redundant for efficient conditional estimation of the parameters in the model for $\Delta \mathbf{x}_t$.[2] The restrictions $\boldsymbol{\Pi}_{x^*} = \mathbf{0}$ also imply that the variables $\mathbf{x}_t^*$ are $I(1)$ and not cointegrated.

[2] See the Appendix in Dees, di Mauro, Pesaran and Smith (2007).

The system then can be written as a conditional model for the endogenous variables,

$$\Delta \mathbf{x}_t = -\mathbf{\Pi}_x \mathbf{z}_{t-1} + \mathbf{\Lambda} \Delta \mathbf{x}_t^* + \sum_{i=1}^{p-1} \mathbf{\Psi}_i \Delta \mathbf{z}_{t-i} + \mathbf{c}_0 + \mathbf{c}_1 t + \boldsymbol{\upsilon}_t, \tag{3}$$

and a marginal model for the exogenous variables,

$$\Delta \mathbf{x}_t^* = \sum_{i=1}^{p-1} \mathbf{\Gamma}_{x^*i} \Delta \mathbf{z}_{t-i} + \mathbf{a}_{x^*0} + \mathbf{u}_{x^*t}, \tag{4}$$

where $\mathbf{\Lambda} \equiv \mathbf{\Sigma}_{xx^*} \mathbf{\Sigma}_{x^*x^*}^{-1}$, $\mathbf{\Psi}_i \equiv \mathbf{\Gamma}_{xi} - \mathbf{\Sigma}_{xx^*} \mathbf{\Sigma}_{x^*x^*}^{-1} \mathbf{\Gamma}_{x^*i}$, $i = 1, ..., p-1$, $\mathbf{c}_0 \equiv \mathbf{a}_{x0} - \mathbf{\Sigma}_{xx^*} \mathbf{\Sigma}_{x^*x^*}^{-1} \mathbf{a}_{x0}$ and $\mathbf{c}_1 \equiv \mathbf{a}_{x1} - \mathbf{\Sigma}_{xx^*} \mathbf{\Sigma}_{x^*x^*}^{-1} \mathbf{a}_{x^*1}$, see Garratt, Lee, Pesaran and Shin (2006, pp. 135-136). In the forecasting exercise below we will consider the effects on the forecasts of the endogenous variables of choosing different marginal models for the exogenous variables, $\Delta \mathbf{x}_t^*$.

If the model includes an unrestricted linear trend, in general there will be quadratic trends in the level of the variables when the model contains unit roots. To avoid this, the trend coefficients are restricted such that

$$\mathbf{c}_1 = \mathbf{\Pi}_x \boldsymbol{\gamma},$$

where $\boldsymbol{\gamma}$ is an $k \times 1$ vector of free coefficients, see Pesaran, Shin and Smith (2000) and Garratt, Lee, Pesaran and Shin (2006, p.136). The nature of the restrictions on $\mathbf{c}_1$ depends on the rank of $\mathbf{\Pi}_x$. In the case where $\mathbf{\Pi}_x$ is full rank, $\mathbf{c}_1$ is unrestricted, whilst it is restricted to be equal to $\mathbf{0}$ when the rank of $\mathbf{\Pi}_x$ is zero. Under the restricted trend coefficients the VECX* model can be written as

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

where

$$\tilde{\mathbf{c}}_0 = \mathbf{c}_0 + \mathbf{\Pi}_x \boldsymbol{\gamma}.$$

Note that $\tilde{\mathbf{c}}_0$ remains unrestricted since $\mathbf{c}_0$ is not restricted.

2.2 Data on the core variables

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 domestic variables are the logarithm of real money balances (M2 definition) denoted by m_t, the logarithm of real gross domestic product (GDP), y_t, the 3-month LIBOR rate, r_t, and the quarterly rate of inflation, π_t.[3] These variables are treated as endogenous. Further endogenous variables are the logarithm of the nominal exchange rate, e_t, and the log ratio of the domestic to the foreign price level, $p_t - p_t^*$. The exogenous variables are the logarithm of foreign real GDP, y_t^*, the foreign 3-month interest rate, r_t^*, and the the logarithm of the oil price, p_t^{oil}. 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.

The foreign (star) variables 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 weights. The trade weights are based on Switzerland's 15 largest trade partners. The weights 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. Trade to these 15 countries on average covers about 82 percent of total Swiss foreign trade. Germany is the most important trading partner of Switzerland—accounting for a trade share of about 30 percent—followed by France, Italy and the United States. Out of the 15 trading partners, 11 are European economies that account for as much as 83 percent of the Swiss trade. The trade share of the US in the Swiss economy is around 9 percent, with Asian countries picking up the rest. The exchange rate and the foreign interest rate are averages of a reduced number of countries, given that financial markets are dominated by developments in the euro area, the UK, the US, and Japan. Initially all estimations and tests were carried out over the period 1965Q4 to 1999Q4, since we reserve the rest of the available data to investigate the forecasting performance of the model.

[3] A detailed description of the variables, their sources, and the construction of the foreign variables is given in the appendix.

2.3 Unit root test results

We first need to consider the unit root properties of the core variables in the VECX* model, which is needed if we are to make a meaningful distinction between long-run and short-run properties of the VECX* model. Since there is considerable evidence that price levels might be $I(2)$, in order to avoid working with mixtures of $I(1)$ and $I(2)$ variables, instead of p_t and p_t^* we shall consider $\pi_t = p_t - p_{t-1}$ and $p_t - p_t^*$, and test if the latter are all $I(1)$. In this way, at least in principle, we could have both the Fisher equation and the PPP holding simultaneously.

Since the power of unit root tests is often low, in addition to the standard Augmented Dickey-Fuller (ADF) test, we shall also apply the generalized least squares version of the Dickey-Fuller test (ADF-GLS) proposed by Elliot, Rothenberg, and Stock (1996) and the weighted symmetric ADF test (ADF-WS) of Park and Fuller (1995), which have been shown to have better power properties than the ADF test. All tests are conducted with up to four lags. When plotting the variables it becomes apparent that e_t, $p_t - p_t^*$, m_t, y_t, y_t^* and p_t^{oil} show a trending behaviour whereas r_t, π_t, r_t^* do not show a clearly visible trend. The regressions in levels therefore include a trend and an intercept for the former group of variables and an intercept only for the latter group. All ADF regressions applied to the first differences include an intercept. The results for the regressions in first differences are reported in Table 1, and for the levels they are summarized in Table 2. Entries in boldface show the lag length which was selected by the Akaike information criterion (AIC). The sample period for the computation of all the ADF statistics runs from 1966Q3 to 1999Q4, so that the AIC relates to a common sample for all tests.

In establishing the unit root properties of the core variables we shall first check whether their first differences are in fact stationary. The unit root tests applied to the levels, to be discussed subsequently, will be valid if their first differences are stationary. The unit root test results for the first differences reject the presence of unit roots in all the series, except for the ratio of domestic to foreign prices, when the lag length recommended by the AIC is used. With fewer lags, however, the test statistic is close to or below the 5 percent critical value for the ADF and the ADF-WS test, while the ADF-GLS test indicates stationarity only when no lags are included. Since the exchange rate and the price ratio should be of the same order of integration, and there is clear indication of stationarity for changes in the exchange rate, we proceed with the assumption that all the first differences are stationary.

Turning to the level of the variables, the unit root test results in Table 2 suggest that the unit root hypothesis can not be rejected when the lag length recommended by the AIC is used in all the variables with the exception of real money and the domestic and foreign interest rates. Generally this result continues to hold for different choices of the augmentation order of the ADF tests. For inflation the ADF and the ADF-GLS tests do not reject the unit root null, whereas the null is rejected if one uses the ADF-WS with the lag length selected by the AIC. Economic theory suggests that interest rates and inflation should be of the same order of integration. Given the mixed statistical evidence obtained for inflation and in order not to run into theoretical inconsistencies, we shall suppose that inflation and all the other series under considerations are best approximated by $I(1)$ processes. The error of falsely assuming that inflation is $I(0)$ and the nominal interest rate is $I(1)$ is likely to be more serious than assuming that both series are $I(1)$.

2.4 Lag lengths and deterministic components

The first stage in the empirical analysis is to determine the lag order of the underlying unrestricted VAR in equation (1). Table 3 shows the results from the application of different lag order selection criteria: the Akaike information criterion (AIC), the final prediction error (FPE) (see Lütkepohl 1993), the Hannan-Quinn (HQ) criterion and the Schwarz information criterion (SIC). The maximum lag length considered is four because we use quarterly data. Considering a higher number of lags does not seem appropriate as with the number of lags the number of coefficients estimated in a VAR rises quickly. The AIC and the FPE criterion point to a lag order of two, whereas the HQ and the SIC favor a lag of order one. We proceed with a lag length of $p = 2$, because overestimating the order of the VAR is much less serious than underestimating it, see Kilian (2002). As deterministic variables a constant and a trend are included, since trends seem to be present in the long-run output relationship and possibly also in the PPP relation. The trend is restricted to lie in the cointegration space, which ensures that there will be no quadratic trends in the model.

2.5 The long-run structural model

Starting point for the estimation is the conditional vector error correction model in equation (3). 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, $\mathbf{x}_t^* = \{y_t^*, r_t^*, p_t^{oil}\}$.

Having decided on $p = 2$ for the order of the underlying VAR(p), we now need to consider the determination of the number of cointegrating relations. When there are r cointegrating relations among the variables $\mathbf{z}_t$, the matrix $\mathbf{\Pi}_x$ has rank $r < k$ and can be written as

$$\mathbf{\Pi}_x = \boldsymbol{\alpha}_x \boldsymbol{\beta}', \tag{5}$$

where $\boldsymbol{\alpha}_x$ ($k_x \times r$) is a matrix of error correction coefficients and $\boldsymbol{\beta}$ ($k \times r$) is a matrix of long-run coefficients. The null hypothesis of no cointegration is investigated by testing the rank of $\mathbf{\Pi}_x$ using the sample 1965Q4-1999Q4. Table 4 shows the eigenvalues as well as the λ-max and the trace statistics together with their simulated critical values. While the trace statistic only marginally rejects the hypothesis that $r = 4$ at the 10% level of significance, the maximum eigenvalue (λ-max) suggests that $r = 3$. However, Assenmacher-Wesche and Pesaran (2007) using data over a more recent sample (1976 to 2004) find five cointegrating relations that is more in line with the long run theory. In what follows we also consider five long run relations, as economic theory suggests, but investigate the effect of dropping cointegrating relations later when dealing with model uncertainty.

To exactly identify r long-run relations, r restrictions (including a normalization restriction) must be imposed on each of the r cointegrating relations. The cointegrating vectors obtained by exact identification are not presented here, since they do not have an economic interpretation. 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 per capita 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 are as follows:

$$\begin{aligned} \text{PPP:}\;\; & e_t - (p_t - p_t^*) = b_{10} + 0.009t + \xi_{1t}, \\ \text{MD:}\;\; & m_t - y_t = b_{20} - 24.89 r_t + \xi_{2t}, \\ \text{GAP:}\;\; & y_t - y_t^* = b_{30} - 0.0036t + \xi_{3t}, \\ \text{UIP:}\;\; & r_t - r_t^* = b_{40} + \xi_{4t}, \\ \text{FIP:}\;\; & r_t = b_{50} + b_{55}\pi_t + \xi_{5t}. \end{aligned} \tag{6}$$

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 since it will be re-estimated in the recursive out-of-sample forecasting exercise.

Since analytical standard errors are valid only asymptotically and may give a wrong impression of the coefficient's significance, we bootstrap confidence intervals for the coefficients. The reported confidence intervals are obtained by a non-parametric bootstrap with 1000 replications. The estimate of the interest-rate elasticity of money demand is significantly negative with a point estimate of -24.89 and a lower 95 percent confidence bound of -32.22 and an upper 95 percent bound of -18.00. The estimate of the trend coefficient in the PPP equation is 0.009 with bootstrapped confidence bounds of 0.0015 and 0.0004, implying a trend appreciation of the real Swiss franc exchange rate. The trend coefficient in the output-gap equation is -0.0036 with a 95% confidence bounds of -0.0029 to -0.0042, showing that the Swiss economy has grown less over the sample period than its trading partners.

A likelihood ratio (LR) test of the over-identifying restrictions gives a test statistic of 106.21, which is asymptotically distributed as a χ^2 variate with 22 degrees of freedom. Because the asymptotic distribution tends to over-reject, we obtain the critical values from a non-parametric bootstrap with 1000 replications. This gives a critical value for the LR test statistic of 61.66 for the five percent level of significance and of 71.04 for the one percent significance level, as compared to the test statistic of 106.21. 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 look at 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.

2.6 Error correction equations

Table 5 displays the estimates of the reduced-form error correction equations and some diagnostic statistics. The deviations from the long-run relations, or the error correction (EC) terms, enter in most equations with high levels of significance. The EC term associated with PPP helps explain the variations in the exchange rate and real domestic output. The EC term for real money balances has a statistically significant effect in the money demand equation. The deviation of domestic from foreign output is significant in the equations for real money balances, inflation, and the price-differential equation, while the deviation of the domestic interest rate from the foreign interest rate contains information for the change in the exchange rate, domestic output, the domestic interest rate, inflation and the price differential. The EC term of the Fisher parity has an influence on the change in the exchange rate, domestic output, inflation and the price differential. Apart from the error correction coefficients (except for deviations from PPP and money demand) the change in the inflation rate is mainly influenced by changes in the foreign interest rate. Inflation is also significantly affected by oil prices, although the effect of oil-price changes on inflation is quantitatively less than that of the foreign interest rate.

The $\overline{R}^2$ of the error-correction equations ranges from 0.25 for the exchange rate equation to 0.71 for the money demand equation. The inflation equation fits somewhat less well with a $\overline{R}^2$ of 0.34 for the change in the inflation rate. Serial correlation is absent except in the equation for Δy_t and $\Delta(p_t - p_t^*)$. The test for functional form does not reject for any of the equations. The indicated departures from normality for e_t and r_t are mainly due to large outliers in 1978/79 when the Swiss National Bank switched to an exchange rate target to counter the rapid appreciation of the Swiss franc. For y_t and π_t residuals show outliers at the time of the first oil-price shock in 1974, which are likely to cause the rejection of normality. Output and inflation are also subject to heteroscedasticity since both series were more volatile before 1974 than thereafter. Whilst it would be possible to model some of these features by adding more lags and by introducing dummy variables, we do not believe that such a strategy would be of much help in forecasting. Most likely, it could involve over-fitting and *ad hoc* specifications that could be counterproductive in forecasting.

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, parameter (for a given model), and model uncertainties.[4] 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

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

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 are performed, so that different, equally 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, 2006) emphasize, structural breaks are often the main source of forecast failure and represents 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 of forecasts from the same model but estimated over different sample periods, as well as by pooling of 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.[5] The pooling of forecasts from 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. See also Pesaran and Timmermann (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. In pooling of 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.[6] We also consider

[5] Timmermann (2006) surveys the literature on forecast combinations, while Geweke and Whiteman (2006) discuss forecast combinations in a Bayesian setting.

[6] 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

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

To construct the forecasts we need both the conditional and the marginal models as set out in equations (3) and (4). 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 $\mathbf{z}_t = (\mathbf{x}_t', \mathbf{x}_t^{*\prime})'$, $\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_x \\ \mathbf{0} \end{pmatrix}$, $\mathbf{\Phi}_1 = \mathbf{I}_m - \boldsymbol{\alpha}\boldsymbol{\beta}' + \mathbf{\Gamma}_1$, $\mathbf{\Phi}_i = \mathbf{\Gamma}_i - \mathbf{\Gamma}_{i-1}$, $i = 2, ..., p-1$, $\mathbf{\Phi}_p = -\mathbf{\Gamma}_{p-1}$. 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 $\mathbf{a}_1 = \begin{pmatrix} \mathbf{\Pi}_y \gamma \\ \mathbf{0} \end{pmatrix}$. In order to avoid deterministic trends in interest rates, $\mathbf{a}_{x^*0}$ is set to zero in the foreign interest-rate equation.

Our strategy for forecast evaluation is as follows: Every model is estimated to the end of 1999Q4 and dynamic one- to eight-quarter-ahead forecasts are then produced for 2000Q1 to 2002Q4. 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 2003Q1. 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 observations for the one-quarter-ahead forecast to 20 for the eight-quarter-ahead forecast.

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,

forecasts using a cross-validation approach. However, such a procedure is data intensive and does not seem suitable for quarterly macroeconometric forecasting.

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. At the same time, interest rates were low compared to 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 would 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.

3.1 Forecast statistics

We evaluate the forecasts in terms of their "bias", "root mean squared forecast error" (RMSFE), and their directional accuracy, or "hit rate".

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

$$\hat{x}_t(h) = \left[\hat{z}(t+h,t) - z_t\right]/h,$$

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

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

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

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

For convenience, we report the RMSFE in percent. Starting our sequential out-of-sample forecasts in 2000Q1, we can evaluate 27 one-step-ahead forecasts until 2006Q3. The first two-step-ahead forecast is for 2001Q2, so that for the two-step-ahead forecasts we have 26 forecast errors and so on; thus ending up with 20 forecast errors for the evaluation of the eight-step-ahead forecasts.

The bias measures how far the mean of the forecast is from the mean of the actual series. Since it is dependent on the scale of the variable, we divide it by the mean of the actual series during the forecast period:

$$Bias = \frac{(n-h)^{-1} \sum_{t=T+1}^{T+n-h} e_t(h)}{(n-h)^{-1} \sum_{t=T+1}^{T+n-h} x_t(h)}.$$

A positive (negative) number thus indicates that the forecast systematically under-predicts (over-predicts) the actual values.

The third criterion of interest in evaluating forecasts is the ability of the forecasts to track the turning points. We therefore look at the proportion of correctly predicted directions of change in the variable, which we call the hit rate. For the non-trended variables in $\mathbf{z}_t$, (i.e., r_t and π_t), the event of interest is whether the variable rises or falls over the next period (i.e., the probability that the predicted change in the variable has the same sign as the actual change, $\Pr(\Delta(\pi_t) > 0)$). The h-step ahead forecast sign indicator equals unity if $x_t(h) \times \hat{x}_t(h) \geq 0$ and zero otherwise. We report the proportion of times that the sign indicator is greater than or equal to zero.

For output as a trended variable we consider whether output *growth* rises or falls (i.e., $\Delta^2 y_t > 0$). We will report the cumulative hit ratio since it is not relevant whether the right direction of change between, say, three quarters and four quarters ahead, has been predicted but whether the model was able to forecast the right direction of change between now and four quarters ahead. More precisely, this means that we report the proportion of times that $[x_t(h) - x_{t-1}(h)] \times [\hat{x}_t(h) - \hat{x}_{t-1}(h)] \geq 0$.

3.2 Forecast results for the VECX*(2,2) model

We consider forecast horizons of up to eight quarters ahead since this is the relevant time horizon for central banks when setting interest rates. Table 6 shows the RMSFE, the bias and the hit rate of 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.[7] The first panel of Table 6 shows the average RMSFE per quarter in percent for the long-run restricted VECX*(2,2) model. The average RMSFE per quarter decreases with a longer forecast horizon. The reason is that we focus on the average change per

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

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 0.06 percent per quarter.

The bias in the second panel of Table 6 shows that the mean of the forecasted changes is fairly close to the actual mean change for output, whereas the inflation forecasts and, to a lesser extent, the interest-rate forecasts show larger biases. In particular, the VECX*(2,2) model tends to over-predict across the board, which is perhaps not surprising given the recession in 2002 and the low interest rates in 2003 and 2004, as compared to their historical levels.

The last panel of Table 6, for each variable, displays the percentage of forecasts that correctly predict the direction of changes. Since output is a trending variable, the calculations are based on output growth rates, whereas for inflation and the interest rate we focus on correct prediction of the sign of inflation and interest rate changes. For a random walk without a drift one would expect to predict the change in the variable correctly in 50 percent of the cases. For output growth the model shows a poor performance, which matches with the finding of Ruoss and Savioz (2002) that also professional forecasters produce wide margins of error when forecasting Swiss GDP. By contrast, results for the interest rate and particularly inflation forecasts are more encouraging, at least at horizons of about one year. For inflation over the one-year horizon the model predicts the right direction of change in 78 percent of the cases.

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.[8] The problem of interest can be described as

[8] The advantages of model averaging for forecasting Swiss inflation are documented in Jordan and Savioz (2003).

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, $\mathbf{Z}_{w,T} = (\mathbf{z}_{T-w+1}, ..., \mathbf{z}_T)$, 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

$$\widehat{\Pr}(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{w,T}) = \sum_{m=1}^{M} \widehat{\Pr}(\mathfrak{M}_m \mid \mathbf{Z}_{w,T}) \widehat{\Pr}(\mathbf{Z}_{T+1,h} \mid \mathbf{Z}_{w,T}, \mathfrak{M}_m), \tag{7}$$

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 observations 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 would be to consider a number of alternative windows, starting from a minimum window size to the largest permitted by the available data set. 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 underlie 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 which 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 forecasts averages that weight all forecasts equally, before we investigate other weighting schemes that have been proposed in the literature.[9]

4.1 Models to be considered in the averaging process

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, p and q, in the VECX*(p,q) specification. Third, we shall consider different specifications for the model we use to forecast the exogenous

[9] The weighting schemes are discussed in Appendix A.3.

variables.[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*. Moreover, the average over models and estimation windows will be referred to as the *AveAve* forecast.

4.2 Average over different model specifications (AveM)

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.[11] In addition, we consider models with one to five exactly identified cointegrating vectors. This gives a total of 36 different model specifications.

Averaging over different specifications of the long-run restrictions generally improves forecasts over the VECX*(2,2) model. Table 7 shows the forecast statistics

[10] 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.

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

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.[12] 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. Also the bias is reduced considerably for all variables. For inflation and the interest rate the hit rate improves and beats the random walk up to the one-year horizon.

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 AIC and the Schwarz criterion pointed to the inclusion of two lags whereas the HQ criterion 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 over all models is likely to improve forecast performance further for output and interest rates. In the following, we present the average RMSFEs for the forecast up to four quarters ahead. The results pertain to the averages over the 36 different specification of the long-run relations. The first column in Table 8 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.

We next investigate the effect of different marginal models for the exogenous variables on the forecasting performance of the VECX*. 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 (4). 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 influence on foreign variables. We refer to this marginal model as $\mathfrak{M}_b^*$ model.

[12] 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.

For forecasting the marginal model is estimated sequentially over the same sample period as the conditional model in the case of both marginal models, $\mathfrak{M}_a^*$ and $\mathfrak{M}_b^*$. 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.[13] As at the time of forecasting the realized values of $\mathbf{x}^*$ are unknown, these forecasts are not feasible and are provided 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 8 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 8. We compute averages only over the $\mathfrak{M}_a^*$ and $\mathfrak{M}_b^*$ models 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 8 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 8. 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 estimation windows are taken into account; an issue that we will explore next.

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

4.3 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.[14] 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.[15] 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 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 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 15^{th} percentile and the 85^{th} percentile of the RMSFEs, while the lower end of the box marks the 25^{th} percentile and the upper end the 75^{th} percentile. The line inside the box represents the median. RMSFEs falling outside the 15^{th} and the 85^{th} 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* does not perform particularly well, whereas 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*.

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

[15] The constants in equation (6) are re-estimated together with the short-run coefficients.

4.4 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 all forecast horizons, we have 3456 models for each 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 over 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 over different windows seems a useful practical way of dealing with this uncertainty.

4.5 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 9 provides these summary statistics for the performance of the AveM forecast relative to the individual forecasts. For inflation and output growth, less than 25 percent of the models are able to beat the AveM forecast, while the performance of the AveM forecast is worse for about 40 percent of the individual models for the interest rate. 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 AveAve forecast, whereas for the interest rate this figure rises to 32 of the individual models. In terms of the RMSFE, the average gain of using the better performing models 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 *ex ante* not available.

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.[16] To assess whether the improvement in forecasting accuracy is significant, we apply the test of predictive accuracy proposed by Diebold and Mariano (1995) and its modification suggested by Harvey, Leybourne and Newbold (1997). 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 $\mathcal{L}^s_{ij,t} = (x_t - \widehat{x}_{t,h,i})^2 - (x_t - \widehat{x}_{t,h,j})^2$ and the absolute loss, $\mathcal{L}^a_{ij,t} = |x_t - \widehat{x}_{t,h,i}| - |x_t - \widehat{x}_{t,h,j}|$, where i is the AveAve forecast and j the forecast from the long-run restricted VECX*(2,2) model (or a univariate AR(1) model), respectively. 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 consider only forecasts up to four steps ahead since for longer horizons the number of independent observations becomes too small to expect significant results.

Table 10 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

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

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 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 also be necessary in this case. Indeed if a correction for first-order autocorrelation is applied, the test statistic becomes -1.914 for the Diebold-Mariano (DM) test (-1.807 for the Harvey-Leybourne-Newbold (HLN) test) and thus become 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.

Table 11 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. Again, the one-step-ahead test statistics for the squared loss for inflation become significant if serial correlation is allowed for (-4.097 for the DM test and -3.869 for the HLN test). 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, which subsequently led to a recession, and a steep rise in the oil price in 2004 that coincided with an economic recovery. Table 12 confirms that the results remain unchanged when the average over estimation windows (AveW) for the AR(1) model is considered.

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

4.6 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 Appendix A.3. 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 weights 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 13 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.[17] 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.

[17] This result is often found in the forecast combination literature though not completely understood yet (see Timmermann 2006).

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 is able 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 that case a portfolio of models is likely to cope better with unexpected future events when it comes to forecasting.

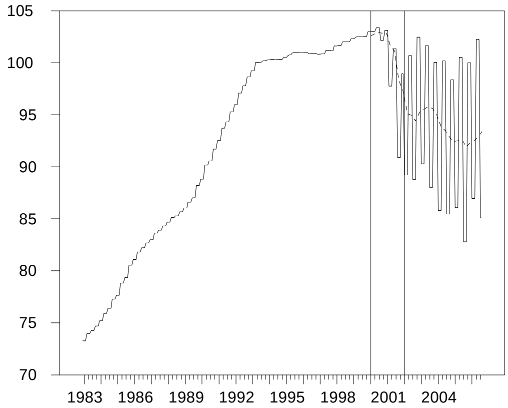
A Appendix: Sources and Construction of Data

A.1 Swiss Data

All Swiss data are from the data base of the Swiss National Bank (SNB). Money is M2 in the definition of 1995, excluding Liechtenstein. The short-term interest rate is the end-of-month three-month London Interbank Offered Rate (3M LIBOR) for Swiss francs, denoted by R. The interest rate is expressed as $0.25 \ln(1 + R/100)$, so that it is in the same unit of measurement as the inflation rate. The price level is the consumer price index (CPI) with the base of December 2005 = 100. Output is the seasonally adjusted quarterly real gross domestic product (GDP) computed by the SECO (Secrétariat d'Etat à l'économie) from 1981 onward. Quarterly output estimates before 1981 were interpolated from the official annual data by the SNB.

For the CPI an adjustment was made to overcome breaks due to new data collection procedures at the Swiss Federal Statistical Office. From 2000 on the CPI includes end-of-season sales. This introduces substantial seasonality into the sub-index for clothing and footwear, as can be seen in Figure A.1. In addition, the data

Figure A.1: Price index clothing and footwear



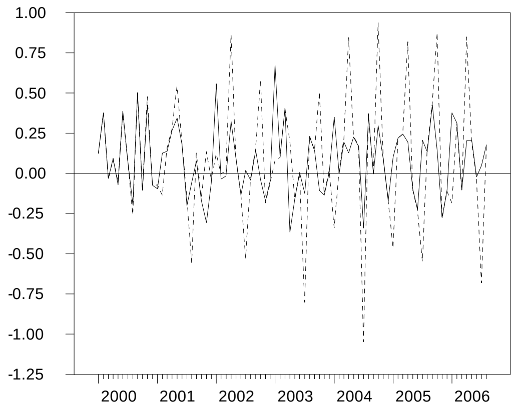
collection had been shifted from the end of the month to the beginning of the month in January 2002, which introduces another break into the series. We adjust for these changes by shifting the series by one month backward between January 2000 and January 2002, the period indicated by the vertical lines in Figure A.1. The resulting missing value is filled by inserting the December 2001 value of the sub-index. The series is smoothed by computing a 12-month backward moving average. The smoothed sub-index is added to the CPI without clothing and footwear, using the weight of this sub-index in the CPI. Figure A.2 shows the original and the adjusted CPI series. Though the weight of the clothing-and-footwear sub-index is less than 5 percent since 2000, it is clearly visible that the adjustment considerably reduces the seasonal variability of the inflation rate since 2002.

Monthly data for M2, the CPI and the 3M LIBOR are aggregated into quarterly averages of monthly figures. Inflation is the quarterly per cent difference of the CPI.

A.2 Foreign Data

The foreign price level, the exchange rate and foreign GDP are constructed using trade weighted data from Switzerland's 15 most important trading partners. These are (in the order of their importance) Germany, France, Italy, the United States, the United Kingdom, Austria, the Netherlands, Japan, Belgium, Spain, Sweden, Hongkong, China, Ireland and Denmark. Trade data are from the Eidgenössische Zollverwaltung. Trade is defined as the sum of imports and exports from and to

Figure A.2: Monthly inflation rate without (solid line) and with adjustment (dashed line) of the CPI



a specific country. The countries considered have an average share of at least 1 percent in total Swiss foreign trade during 1974 to 2006. Together, the 15 countries considered account for about 82 percent of total Swiss foreign trade. For Ireland, Hongkong and China, trade data were not available before 1988. For these countries, the trade shares were set to the January 1988 value for the period before 1987. This avoids level effects that would otherwise appear if the trade weights for these countries were set to zero over the time where data are not available. The trade weights used in the aggregation are three-year moving averages of the trade share of the respective country in Switzerland's total trade with these 15 countries.

The foreign price level is the trade-weighted aggregate of the consumer price indices, and foreign GDP is the trade-weighted aggregate of the real GDP indices of the 15 main trading partners. The CPI and real GDP data are from the Main Economic Indicator data base of the OECD. Missing data have been supplemented with IFS and BIS data. For countries where the GDP data were not seasonally adjusted at the source, the X12 procedure was used to seasonally adjust the original series. When quarterly data were not available, annual data were interpolated.[18] The resulting GDP series was converted to an index with the base year 2000 and then aggregated using the three-year moving averages of the trade weights. This avoids the use of exchange rates to convert GDP into a common currency.

The exchange rate is the weighted average of the exchange rate of the Swiss Franc vis-a-vis Switzerland's 15 main trading partners. After the transition to European

[18] This was the case for the Netherlands and Denmark until 1976, for Belgium until 1979, for Ireland and Hong Kong until 1985, and for China until 1999.

monetary union, the exchange rate for the members of the European Monetary Union are replaced by the Euro exchange rate, converted with the official conversion rates of the national currencies to the Euro at the start of the European Monetary Union in 1999.

In contrast to the CPI, the exchange rate and GDP, the foreign interest rate is a weighted average of the three-month interest rate in only three areas, namely the euro area, the UK and the US. Before the existence of European Monetary Union, the euro area interest rate is proxied by a weighted average of the short-term interest rates of the countries that entered the EMU and are among Switzerland's 15 main trading partners. The weights are the shares of the EMU countries, the UK and the US in Swiss foreign trade with these countries. While the EMU countries receive a share of about 82 percent, the weight of the UK is 6 percent and the US financial variables make up 10 percent of the total. The interest rates are from the BIS data base. Like the domestic interest rate, the foreign interest rate is expressed as $0.25\ln(1+R^*/100)$, where R^* is the foreign interest rate per annum in percent.

Also for the foreign variables, the monthly series for the CPI, the interest rate and the exchange rate are aggregated with monthly trade weights and then transformed into quarterly averages.

A.3 Weighting schemes

Let f_{mth} be the m^{th} 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.[19]

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

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

An h-step ahead version can be written as

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

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

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

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

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

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

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

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

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

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

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 $\gamma = 0.05$.[20]

[20] 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 $\gamma = \{0.05, 0.10\}$.

References

Assenmacher-Wesche, K. and M. H. Pesaran (2007): "A VECX* Model of the Swiss Economy," *Swiss National Bank Economy Studies,* forthcoming.

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.

Dees, S., F. di Mauro, M. H. Pesaran and L.V. Smith (2007): "Exploring the International Linkages of the Euro Area: A Global VAR Analysis," *Journal of Applied Econometrics* 22, 1-38.

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., T. J. Rothenberg and J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica* 64, 813-836.

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.

Kilian, L. (2002): "Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order," *Journal of Forecasting* 20, 161-179.

Lütkepohl, H. (1993): "Introduction to Multiple Time Series Analysis," 2^{nd} ed., *Springer*, Berlin.

Park, H. J. and W. A. Fuller (1995): "Alternative Estimators and Unit Root Tests for the Autoregressive Process," *Journal of Time Series Analysis* 16, 415-429.

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., Y. Shin and R. J. Smith (2000): "Structural Analysis of Vector Error Correction Models with Exogenous I(1) Variables," *Journal of Econometrics* 97, 293-343.

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. 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.*

Ruoss, E. and M. R. Savioz (2002): "How Accurate Are GDP Forecasts? An Empirical Study for Switzerland," *Quartalsheft, Schweizerische Nationalbank* 2002, 42-63.

Sancetta, A. (2006): "Online Forecast Combination for Dependent Heterogeneous Data," University of Cambridge, *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: Unit root tests for the first differences

ADF

	Δe	Δm	Δy	Δr	$\Delta \pi$	Δp-Δp^{*}	Δy^{*}	Δr^{*}	Δp^{oil}	C
0	**-9.34**	-4.15	-9.97	**-9.30**	-15.86	-4.71	-8.34	-6.87	**-10.41**	-2.86
1	-7.99	-4.29	-7.69	-6.55	**-12.84**	-3.33	-5.46	-6.32	-8.53	-2.86
2	-6.38	-3.97	**-4.71**	-5.75	-7.99	-2.55	-4.27	-6.00	-6.11	-2.88
3	-6.28	**-4.64**	-4.23	-5.62	-7.52	-2.56	-3.27	-4.55	-5.61	-2.88
4	-6.08	-4.12	-4.37	-5.69	-6.42	**-2.12**	**-3.69**	**-5.30**	-5.73	-2.88

ADF-GLS

	Δe	Δm	Δy	Δr	$\Delta \pi$	Δp-Δp^{*}	Δy^{*}	Δr^{*}	Δp^{oil}	C
0	**-9.02**	-4.09	-9.45	**-9.65**	-14.36	-2.95	-7.67	-6.18	**-10.43**	-2.02
1	-7.58	-4.21	-7.12	-6.31	**-10.46**	-2.01	-4.94	-5.54	-8.54	-2.04
2	-5.95	-3.89	**-4.27**	-5.49	-6.05	-1.46	-3.80	-5.08	-6.11	-2.07
3	-5.74	**-4.52**	-3.79	-5.30	-5.23	-1.46	-2.86	-3.75	-5.61	-1.99
4	-5.42	-4.00	-3.86	-5.28	-4.15	**-1.13**	**-3.16**	**-4.25**	-5.73	-2.02

ADF-WS

	Δe	Δm	Δy	Δr	$\Delta \pi$	Δp-Δp^{*}	Δy^{*}	Δr^{*}	Δp^{oil}	C
0	**-9.59**	-4.35	-10.23	**-9.60**	-16.16	-4.57	-8.56	-7.29	**-10.69**	-2.51
1	-8.22	-4.49	-7.98	-6.73	**-13.08**	-3.17	-5.62	-6.43	-8.78	-2.51
2	-6.50	-4.18	**-4.93**	-5.95	-8.16	-2.28	-4.36	-6.13	-6.32	-2.51
3	-6.49	**-4.85**	-4.45	-5.81	-7.72	-2.31	-3.32	-4.64	-5.83	-2.53
4	-6.29	-4.33	-4.58	-5.88	-6.60	**-1.78**	**-3.69**	**-5.42**	-5.94	-2.50

Note: *ADF* denotes the Augmented Dickey-Fuller Test, *ADF-GLS* the generalized least squares version of the ADF test, and *ADF-WS* the weighted least squares ADF test. The first column shows the number of lags included when computing the test statistics. All regressions include an intercept. The sample period runs from 1966Q4 to 1999Q4. The column C shows the 95% simulated critical values. Entries in boldface denote the lag length selected by the AIC criterion.

Table 2: Unit root tests for the levels

ADF											
	e	m	y	r	π	p-p^*	y^*	r^*	p^{oil}	T	C
0	-1.41	-1.46	-2.57	-2.40	-4.52	0.05	-2.86	-1.79	**-1.56**	-3.49	-2.86
1	**-1.86**	**-3.57**	-2.73	-3.10	-3.39	-0.90	-2.67	-3.36	-1.68	-3.42	-2.86
2	-1.66	-3.36	-2.67	**-3.61**	**-2.47**	-1.27	-2.64	-3.24	-1.50	-3.48	-2.88
3	-1.72	-3.50	**-3.13**	-3.74	-2.71	**-1.69**	-2.60	-3.07	-1.62	-3.39	-2.88
4	-1.49	-2.94	-3.20	-3.56	-2.38	-1.65	**-2.62**	**-3.84**	-1.51	-3.34	-2.88
ADF-GLS											
	e	m	y	r	π	p-p^*	y^*	r^*	p^{oil}	T	C
0	-1.52	-1.15	-1.24	-2.43	-3.86	-0.27	0.15	-1.80	**-1.26**	-3.0	-2.02
1	**-1.93**	**-3.15**	-1.44	-3.12	-2.83	-1.03	-0.25	-3.18	-1.41	-3.00	-2.04
2	-1.74	-2.93	-1.40	**-3.61**	**-2.00**	-1.39	-0.53	-3.05	-1.23	-2.97	-2.07
3	-1.81	-3.04	**-1.92**	-3.74	-2.20	**-1.84**	-0.72	-2.88	-1.39	-2.94	-1.99
4	-1.59	-2.50	-1.99	-3.55	-1.88	-1.80	**-1.03**	**-3.53**	-1.29	-2.95	-2.02
ADF-WS											
	e	m	y	r	π	p-p^*	y^*	r^*	p^{oil}	T	C
0	-1.69	-1.38	-1.24	-2.62	-4.68	0.19	1.32	-2.08	**1.57**	-3.31	-2.51
1	**-2.11**	**-3.55**	-1.55	-3.30	-3.56	-0.82	0.51	-3.46	-1.71	-3.25	-2.51
2	-1.92	-3.33	-1.49	**-3.82**	**-2.67**	-1.18	0.07	-3.44	-1.54	-3.28	-2.51
3	-1.99	-3.50	**-2.23**	-3.95	-2.92	**-1.63**	-0.25	-3.28	-1.68	-3.22	-2.53
4	-1.79	-2.90	-2.36	-3.77	-2.59	-1.59	**-0.75**	**-4.04**	-1.59	-3.25	-2.50

Note: *ADF* denotes the Augmented Dickey-Fuller Test, *ADF-GLS* the generalized least squares version of the ADF test, and *ADF-WS* the weighted least squares ADF test. The first column shows the number of lags included in the test. The regressions include a trend and an intercept for e, $p - p^*$, m, y, $y^*$and p^{oil}, and an intercept only for r, π, and r^*. The sample period runs from 1966Q3 to 1999Q4. The column T gives the 95% simulated critical values for the test with intercept and trend, the column C the 95% simulated critical values for the test including an intercept only. Entries in boldface denote the lag length selected by the AIC criterion.

Table 3: Lag order selection criteria

Lag length	*AIC*	*Log(FPE)*	*HQ*	*SC*
1	-61.61	-61.60	-60.85	-59.73
2	-61.95	-61.92	-60.69	-59.86
3	-61.89	-61.82	-60.14	-57.71
4	-61.82	-61.67	-59.58	56.31

Note: *AIC* is the Akaike information criterion, *FPE* is the final prediction error, *HQ* the Hannan-Quinn criterion and *SC* the Schwarz criterion. The sample period is 1965Q4 to 1999Q4.

Table 4: Cointegration tests

Rank	*Eigenvalue*	*Trace statistic*	*Critical value 90%*	*λ-max statistic*	*Critical value 90%*
0	0.492	261.26	171.50	92.83	57.52
1	0.359	168.43	131.73	60.87	50.54
2	0.286	107.56	95.29	46.21	42.93
3	0.192	61.34	66.30	29.23	35.84
4	0.117	32.10	40.35	17.08	27.28
5	0.104	15.03	19.63	15.02	19.63

Note: The sample period is 1965Q4 to 1999Q4. Critical values simulated using 1000 replications.

Table 5: Reduced-form error-correction equations

Equation	Δe_t	Δm_t	Δy_t	Δr_t	$\Delta \pi_t$	$\Delta(p_t - p_t^*)$
$\hat{\xi}_{1,t-1}$	-0.178^* (0.053)	-0.004 (0.029)	0.066^* (0.024)	-0.002 (0.005)	-0.006 (0.011)	0.011 (0.010)
$\hat{\xi}_{2,t-1}$	-0.018 (0.034)	-0.077^* (0.019)	-0.008 (0.016)	-0.001 (0.004)	0.001 (0.007)	-0.004 (0.007)
$\hat{\xi}_{3,t-1}$	0.016 (0.071)	-0.100^* (0.039)	-0.060 (0.033)	0.001 (0.007)	0.042^* (0.014)	0.040^* (0.014)
$\hat{\xi}_{4,t-1}$	-2.640^* (0.917)	0.973 (0.504)	-0.951^* (0.424)	-0.222^* (0.096)	-0.800^* (0.188)	-0.368^* (0.176)
$\hat{\xi}_{5,t-1}$	2.506^* (0.632)	0.837 (0.347)	0.611^* (0.292)	0.028 (0.066)	0.614^* (0.129)	0.330^* (0.122)
Δe_{t-1}	0.272^* (0.089)	-0.157^* (0.049)	0.045 (0.041)	0.032^* (0.009)	0.019 (0.018)	0.012 (0.017)
Δm_{t-1}	-0.026 (0.132)	0.537^* (0.073)	-0.023 (0.061)	-0.008 (0.014)	-0.006 (0.027)	-0.032 (0.025)
Δy_{t-1}	-0.280 (0.207)	0.039 (0.114)	-0.237^* (0.096)	0.004 (0.022)	-0.011 (0.042)	0.034 (0.040)
Δr_{t-1}	-2.204^* (0.836)	-0.470 (0.514)	0.299 (0.433)	0.124 (0.097)	-0.086 (0.192)	-0.344 (0.180)
$\Delta \pi_{t-1}$	-0.374 (0.487)	0.097 (0.268)	-0.152 (0.225)	-0.054 (0.051)	-0.085 (0.100)	-0.170 (0.094)
$\Delta(p_{t-1} - p_{t-1}^*)$	1.953^* (0.525)	0.287 (0.289)	0.292 (0.242)	0.029 (0.055)	-0.089 (0.108)	0.648^* (0.101)
Δy_t^*	0.024 (0.306)	-0.271 (0.168)	0.994^* (0.141)	0.055 (0.032)	0.081 (0.063)	0.068 (0.059)
Δy_{t-1}^*	-0.505 (0.343)	-0.081 (0.188)	0.177 (0.158)	-0.017 (0.036)	0.072 (0.070)	-0.043 (0.066)
Δr_t^*	3.482^* (1.125)	-2.596^* (0.618)	-0.056 (0.520)	0.720^* (0.117)	0.535^* (0.230)	0.174 (0.216)
Δr_{t-1}^*	1.034 (1.443)	0.900 (0.793)	0.646 (0.667)	-0.222 (0.150)	-0.207 (0.295)	0.221 (0.278)
Δp_t^{oil}	0.004 (0.012)	-0.010 (0.007)	0.008 (0.006)	0.002 (0.001)	0.005^* (0.002)	0.001 (0.002)
Δp_{t-1}^{oil}	0.027^* (0.012)	0.001 (0.006)	0.013^* (0.005)	0.0002 (0.001)	-0.002 (0.002)	-0.002 (0.002)
Constant	-0.213 (0.871)	1.327^* (0.479)	0.739 (0.403)	-0.016 (0.091)	-0.515 (0.178)	-0.478^* (0.168)
$\bar{R}^2$	0.25	0.71	0.39	0.37	0.34	0.69
SC: $\chi^2(4)$	3.34	4.81	18.79	3.79	7.22	11.95
FF: $\chi^2(1)$	0.49	0.11	0.56	0.65	1.49	0.02
N: $\chi^2(2)$	98.73	2.01	11.38	25.81	11.71	5.97
HS: $\chi^2(1)$	0.03	0.01	11.61	2.98	12.77	0.87

Note: The error correction terms, ξ_i, are defined in eq. (2.5). An asterisk denotes significance at the 5% level. *SC* is a test for serial correlation, *FF* a test for functional form, N a test for normality and *HS* a test for heteroscedasticity. Critical values are 3.84 for $\chi^2(1)$, 5.99 for $\chi^2(2)$ and 9.49 for $\chi^2(4)$. Constant not shown. The sample period is 1965Q4 to 1999Q4.

Table 6: Forecast statistics for long-run restricted VECX*(2,2) model with over-identified β

Horizon	#	y_t	π_t	r_t
RMSFE in %				
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
Bias				
1 step ahead	27	0.462	16.096	13.247
2 step ahead	26	0.481	14.225	5.245
3 step ahead	25	0.468	18.285	3.128
4 step ahead	24	0.456	17.771	2.605
8 step ahead	20	0.370	17.940	2.134
Hit rate				
1 step ahead	27	42.31	76.92	50.00
2 step ahead	26	40.00	76.00	52.00
3 step ahead	25	41.67	62.50	58.33
4 step ahead	24	34.78	78.26	52.17
8 step ahead	20	26.32	47.37	5.26

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 7: Average forecast over different β of VECX*(2,2) model

Horizon	#	y_t	π_t	r_t
RMSFE in %				
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
Bias				
1 step ahead	27	0.291	7.140	10.429
2 step ahead	26	0.239	5.678	4.209
3 step ahead	25	0.139	6.551	2.592
4 step ahead	24	0.050	6.488	2.198
8 step ahead	20	-0.294	8.280	1.905
Hit rate				
1 step ahead	27	38.46	73.08	57.69
2 step ahead	26	44.00	72.00	60.00
3 step ahead	25	50.00	62.50	62.50
4 step ahead	24	43.48	78.26	52.17
8 step ahead	20	31.58	47.37	10.53

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 8: RMSFE for forecast average across different model dimensions

Horizon	$\mathfrak{M}_a^*$	$\mathfrak{M}_b^*$	*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 average RMSFE per quarter for the four-quarter-ahead forecast. $\mathfrak{M}_a^*$ 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 9: Summary of performance of Ave forecast relative to individual forecasts across estimation windows

	y_t		π_t		r_t	
Window	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 model average RMSFE. Exceedence gives the average RMSFE loss of not using those models that perform better than the model average. For comparison, the last row shows the RMSFE of the model average.

Table 10: Predictive accuracy of AveAve forecast against long-run restricted VECX*(2,2) model

Horizon	y_t		π_t		r_t	
	DM	*HLN*	*DM*	*HLN*	*DM*	*HLN*
			Squared loss			
1	**-1.992**	**-1.954**	-0.780	-0.766	**-1.827**	**-1.793**
2	**-2.216**	**-2.088**	**-2.300**	**-2.167**	**-3.355**	**-3.160**
3	**-1.674**	-1.507	**-1.857**	**-1.670**	**-2.574**	**-2.316**
4	-1.562	-1.334	**-1.664**	-1.421	**-2.142**	**-1.829**
			Absolute loss			
1	**-2.789**	**-2.737**	-0.471	-0.462	**-2.361**	**-2.317**
2	**-2.016**	**-1.900**	-1.509	-1.422	**-3.855**	**-3.632**
3	-1.257	-1.131	-1.633	-1.469	**-4.057**	**-3.651**
4	-1.491	-1.273	**-1.799**	-1.536	**-2.868**	**-2.449**

Note: *DM* indicates the Diebold-Mariano (1995) test statistic, *HLN* the modified test statistic as proposed by Harvey, Granger and Newbold (1997). Significant test statistics at the 5% level are denoted in boldface. A negative entry indicates that the AveAve forecast outperforms the alternative model.

Table 11: Predictive accuracy of AveAve forecast against AR(1) model

Horizon	y_t		π_t		r_t	
	DM	*HLN*	*DM*	*HLN*	*DM*	*HLN*
			Squared loss			
1	-0.255	-0.251	**-1.657**	-1.626	**-3.233**	**-3.173**
2	-0.789	-0.743	**-3.393**	**-3.197**	**-2.555**	**-2.407**
3	-0.144	-0.130	**-3.730**	**-3.356**	**-2.111**	**-1.900**
4	0.337	0.287	**-5.380**	**-4.594**	**-1.940**	**-1.657**
			Absolute loss			
1	0.556	0.546	**-1.941**	**-1.905**	**-4.893**	**-4.801**
2	-0.592	-0.558	**-4.354**	**-4.102**	**-3.110**	**-2.930**
3	-0.414	-0.372	**-4.095**	**-3.685**	**-2.821**	**-2.538**
4	-0.291	-0.249	**-40.988**	**-35.001**	**-2.837**	**-2.423**

Note: The AR(1) model is estimated over the longest estimation window. *DM* indicates the Diebold-Mariano (1995) test statistic, *HLN* the modified test statistic as proposed by Harvey, Granger and Newbold (1997). Significant test statistics at the 5% level are denoted in boldface. A negative entry indicates that the AveAve forecast outperforms the alternative model.

Table 12: Predictive accuracy of AveAve forecast against AveW of AR(1) model

Horizon	y_t		π_t		r_t	
	DM	*HLN*	*DM*	*HLN*	*DM*	*HLN*
			Squared loss			
1	-0.273	-0.268	-1.540	-1.511	**-3.000**	**-2.944**
2	-0.528	-0.498	**-3.007**	**-2.833**	**-2.415**	**-2.275**
3	-0.153	-0.138	**-3.246**	**-2.921**	**-1.980**	**-1.781**
4	0.048	0.041	**-4.895**	**-4.180**	**-1.768**	**-1.509**
			Absolute loss			
1	0.598	0.587	**-1.844**	**-1.810**	**-4.581**	**-4.495**
2	-0.410	-0.386	**-3.514**	**-3.311**	**-2.734**	**-2.576**
3	-0.171	-0.154	**-3.388**	**-3.048**	**-2.514**	**-2.262**
4	-0.184	-0.158	**-30.328**	**-25.897**	**-2.586**	**-2.208**

Note: *DM* indicates the Diebold-Mariano (1995) test statistic, *HLN* the modified test statistic as proposed by Harvey, Granger and Newbold (1997). Significant test statistics at the 5% level are denoted in boldface. A negative entry indicates that the AveAve forecast outperforms the alternative model.

Table 13: RMSFE for inflation in per cent for Ave forecast

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 different estimation windows and different models. 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 sample period as the conditional model.

Figure 1: RMSFE for output growth across estimation windows and AveAve RMSFE

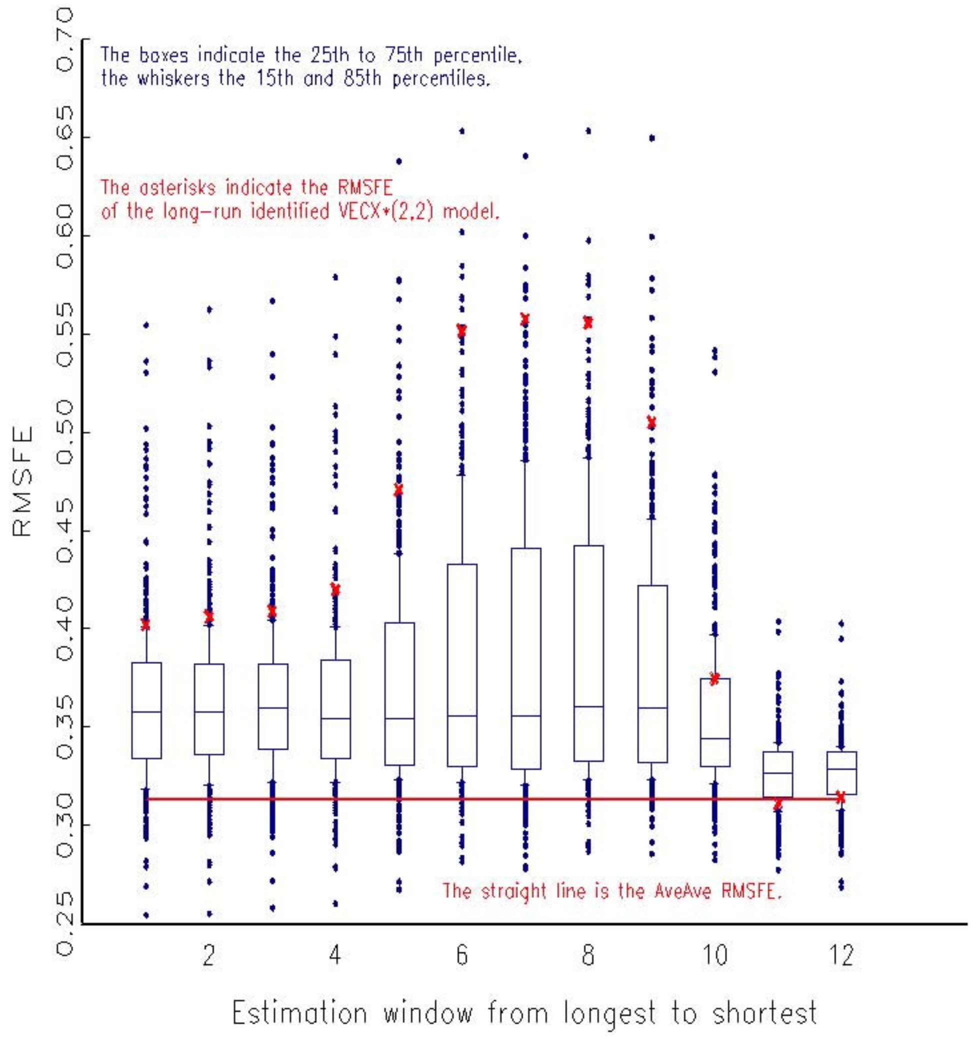


Figure 2: RMSFE for inflation across estimation windows and AveAve RMSFE

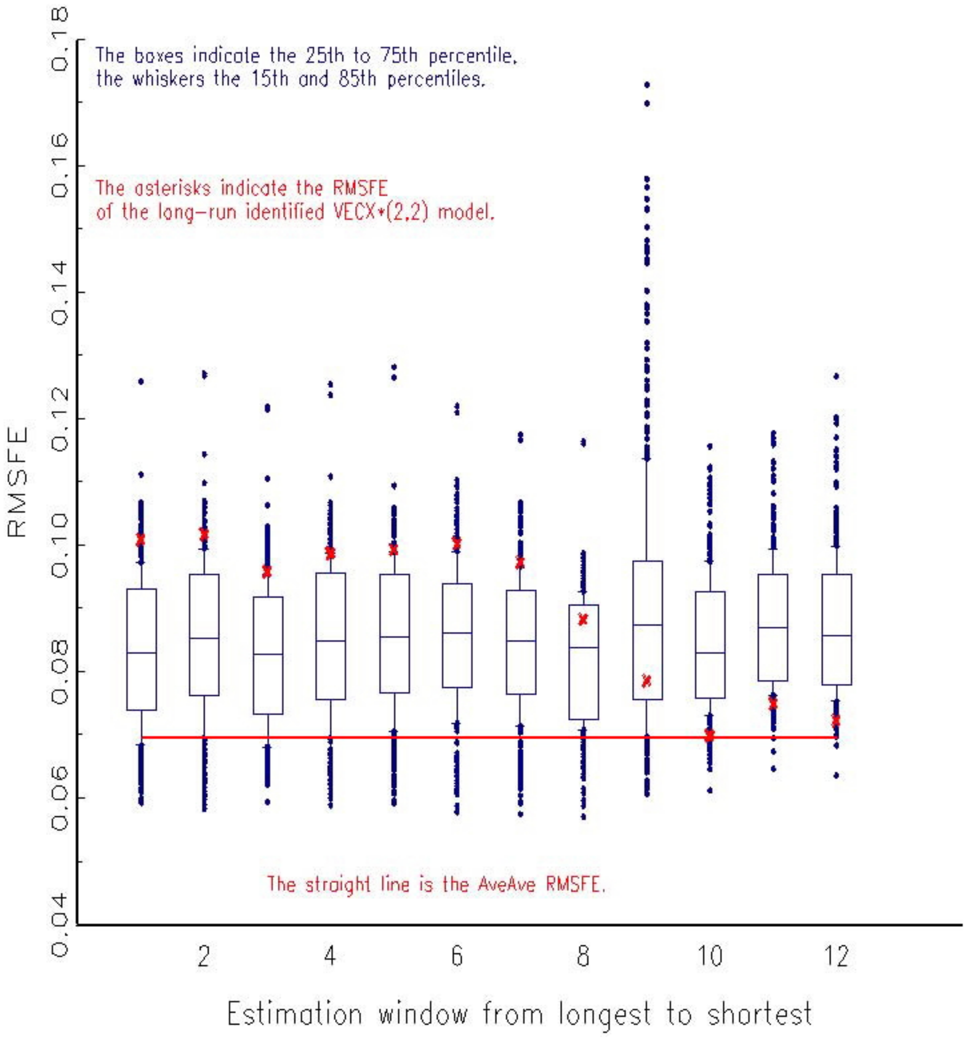


Figure 3: RMSFE for interest rate across estimation windows and AveAve RMSFE

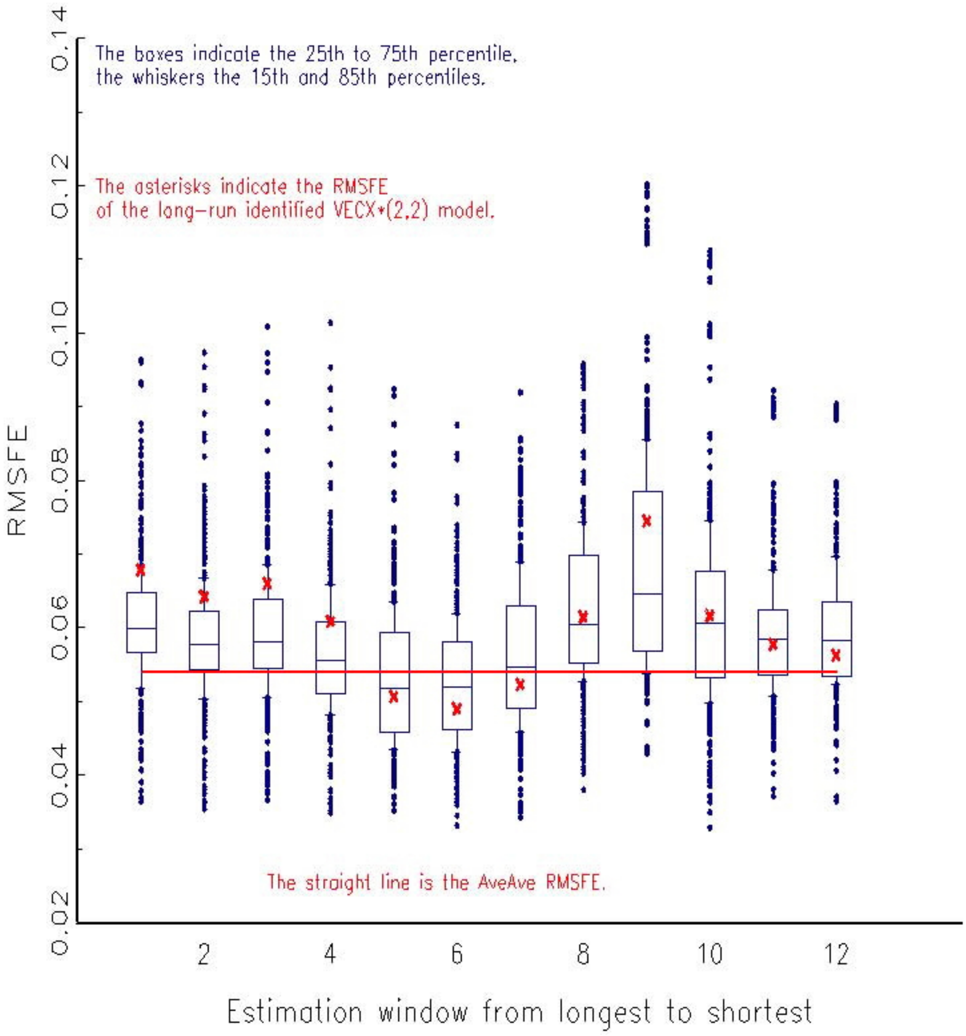


Figure 4: Distribution of RMSFEs for output growth across forecast horizons

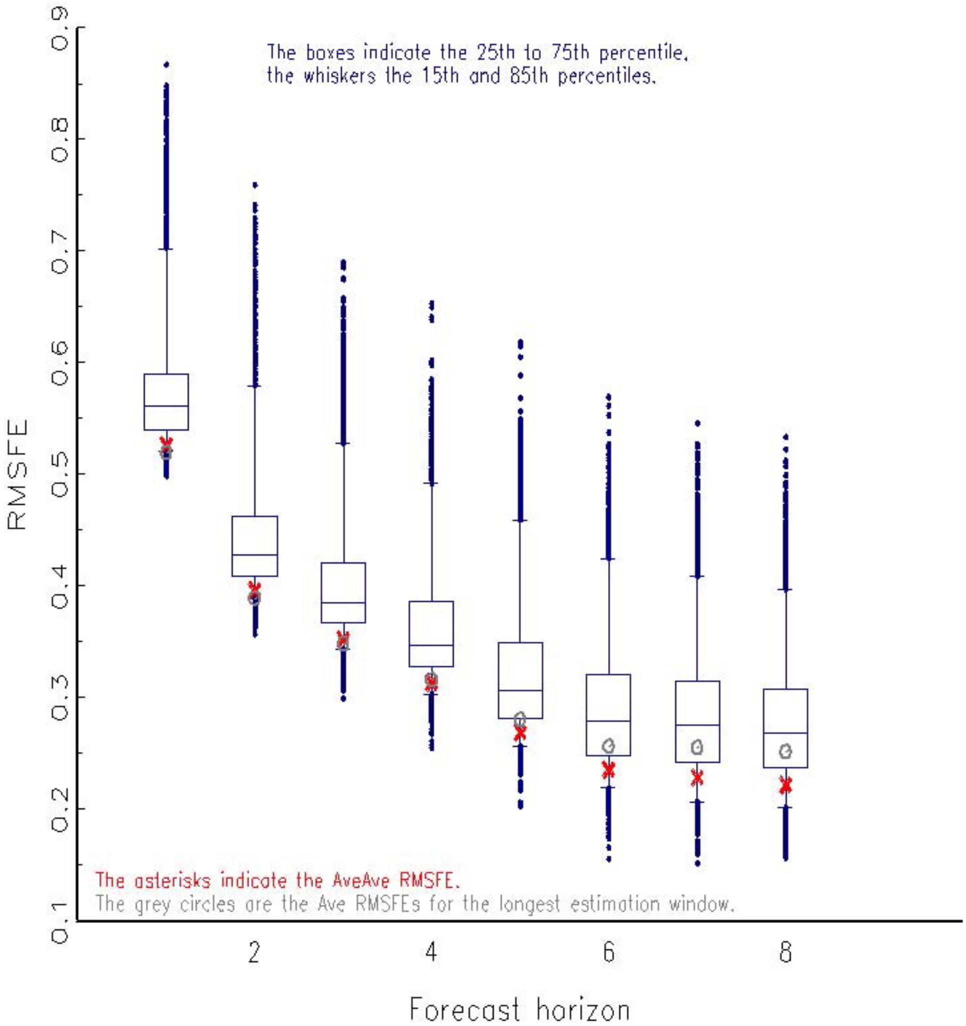


Figure 5: Distribution of RMSFEs for inflation across forecast horizons

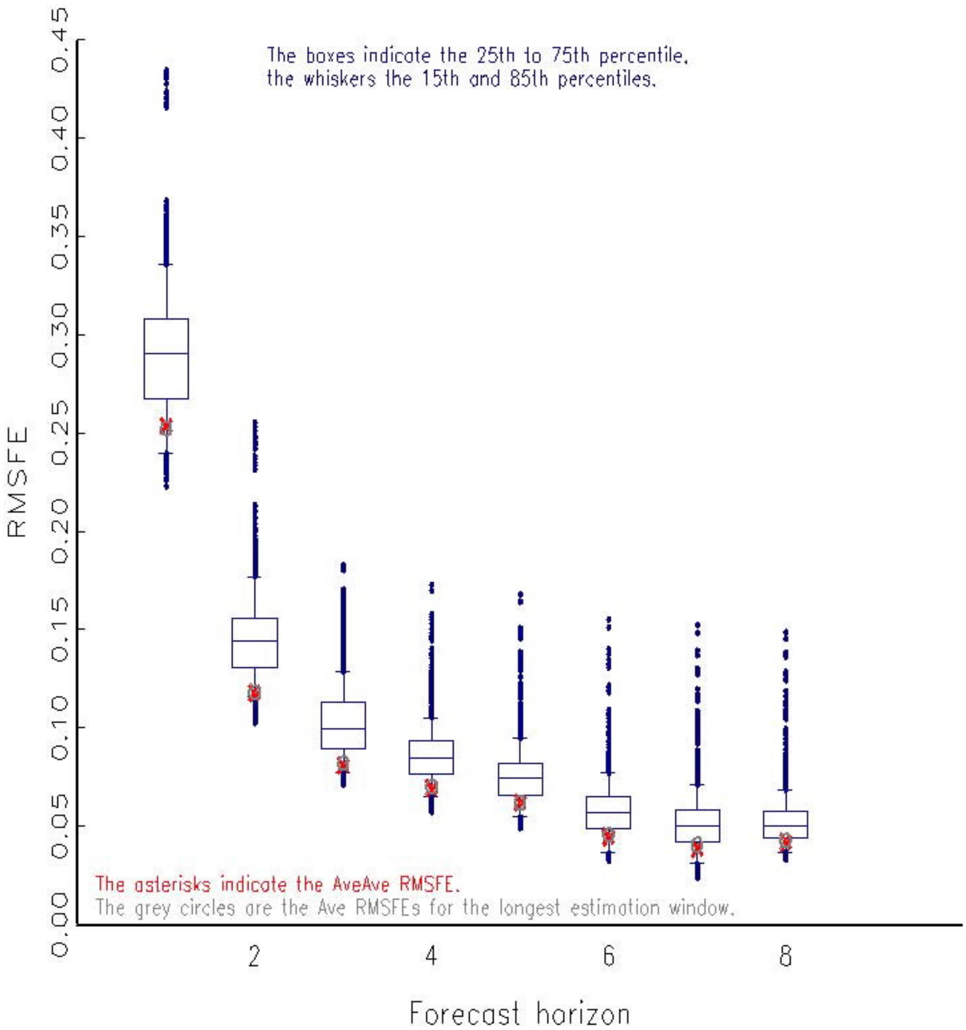


Figure 6: Distribution of RMSFEs for interest rate across forecast horizons

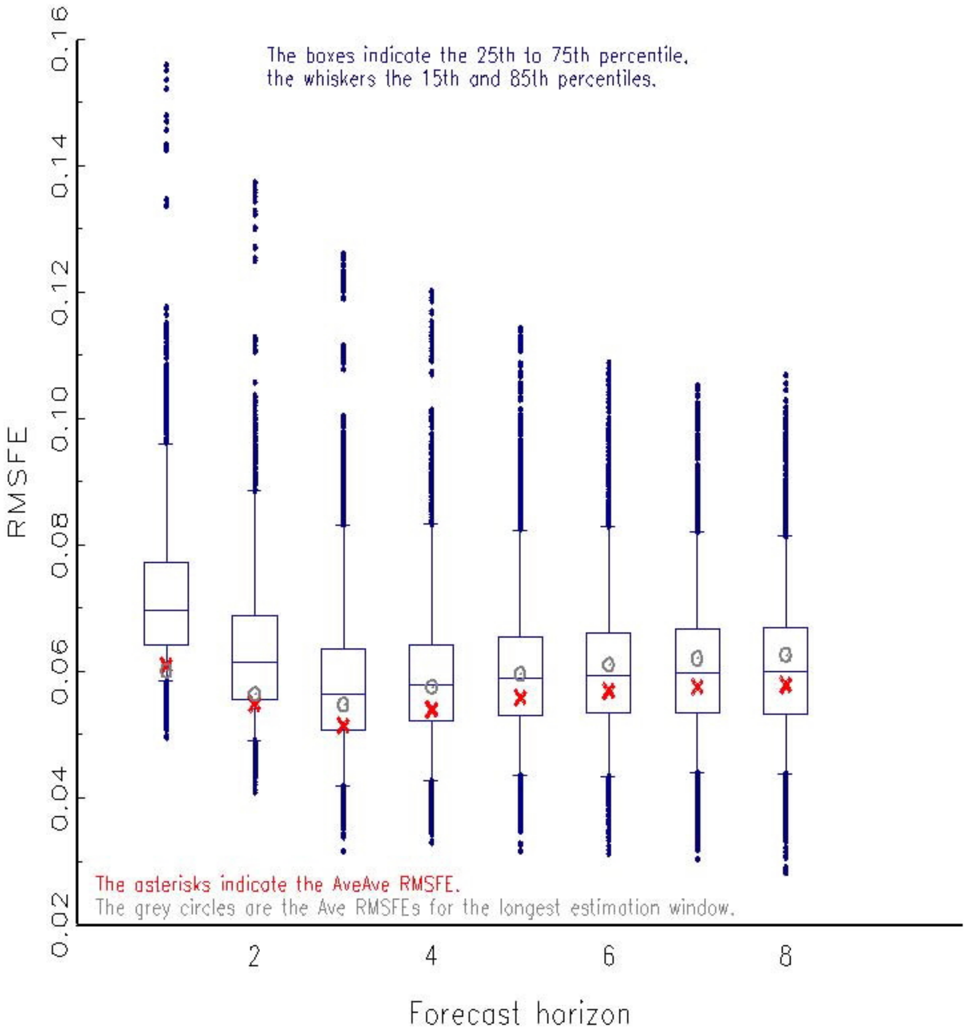


Figure 7: Evolution of weights over forecasting period