A Statistical Forecasting Method for Inflation Forecasting: Hitting Every Vector Autoregression and Forecasting under Model Uncertainty

Ippei Fujiwara and Maiko Koga

Typically, when conducting econometric forecasting, estimation is carried out on a forecasting model that is built upon some assumed economic structure. However, such techniques cannot avoid running into the possibility of misspecification, which will occur should there be some error in the assumptions underlying this economic structure. In this paper, in which we concentrate upon inflation forecasting, we present a method of hitting every vector autoregression (VAR) and forecasting under model uncertainty (HEVAR/FMU) that stresses statistical relationships among time-series data, and that makes no structural assumptions, other than to set up the underlying variables. Use of this HEVAR/FMU, in addition to establishing a more objective setting and enabling us to produce forecasts that take uncertainty into account, gives better results when forecasting qualitative movements in inflation. Therefore, we can state that the HEVAR/FMU can also play a valuable role in providing a cross-check for forecasts produced using such structural-type models.

Keywords: Inflation; Forecast; Reduced rank VAR; Kernel smoothing; Mixture distribution; Nonparametric test
JEL Classification: C32, C35, C53, E31

Ippei Fujiwara: Research and Statistics Department, Bank of Japan (E-mail: ippei.fujiwara@boj.or.jp)
Maiko Koga: Research and Statistics Department, Bank of Japan (E-mail: maiko.koga@boj.or.jp)

In producing this paper, we have had the benefit of valuable comments from Munehisa Kasuya and Hideaki Shimizu (both of the Research and Statistics Department, Bank of Japan), as well as from many other staff at the Bank. In carrying out the analysis, we received considerable support from Toshinao Yoshiba (Institute for Monetary and Economic Studies [IMES], Bank of Japan) and Toshiro Nomura (Graduate School of Economics, Keio University, currently at the Financial Research Center, Nomura Securities). We would like to offer them our sincere thanks. In addition, in conducting estimation of the reduced rank VAR model and in calculation of the Pesaran-Timmermann statistics, we made use of the GAUSS program created by George Kapetanios (Bank of England, currently at Queen Mary College, University of London), with some modifications of our own. We also thank an anonymous referee for valuable comments. It remains only to emphasize that the opinions expressed in this paper are our own and do not reflect those of the Bank of Japan, the Research and Statistics Department, or IMES.

I. Introduction

Typically, when using econometric techniques to forecast economic variables, estimation is carried out on a forecasting model that is built upon some assumed economic structure, based upon *a priori* knowledge and economic principles. However, such techniques cannot avoid running into the possibility of misspecification, which will occur should there be some error in the assumptions underlying this economic structure. Even when diagnostic tests have been easily cleared, a small change in the way this structure is set up can induce large differences in the forecast value. In other words, the researcher's subjective choices in setting up the model can have a substantial influence on the estimated forecast.[1]

Given this, rather than using a model that relies on some specific hypothesis above, it may be considered desirable to employ a forecasting technique that is as objective as possible and based purely on statistical relationships additionally.

Of course, however objective the methodology is, it is not feasible to completely remove the chance that some misspecification occurs in writing the model. For this reason, to produce forecasts that reflect observed reality as accurately as possible, it is necessary to make full use of all the information we possess concerning the uncertainty inherent in both the model and the forecast value itself. To achieve this, in making our forecasts we do not look merely at the fluctuations observed in the forecast value of some specific model, but instead take account of the uncertainty inherent in the forecast errors and the choice of model itself, admirably captured in the forecast distribution (illustrated in the fan charts discussed later in this paper).

In full awareness of the above problems, the object of this paper is to use our forecasting of the inflation rate as an illustration[2] with which we present our method of hitting every vector autoregression (VAR) and forecasting under model uncertainty (HEVAR/FMU). This HEVAR/FMU is one example of a forecasting technique that scores highly on objectivity, and that takes appropriate account of uncertainty. Specifically, we produce forecasts using our HEVAR/FMU in accordance with the following process.

To begin with, we construct a VAR forecasting model for the inflation rate. Even for a reduced-form VAR model, which is highly objective, it is not possible to do away entirely with all traces of subjectivity.[3] The forecast result will inevitably depend to some extent upon the choice of data, the combination of adopted variables, and the period for which estimation is carried out. Recognizing that some bias may emerge as

1. Krolzig and Hendry (2001) introduce a process for selecting a model objectively based on the general-to-specific (Gets) concept. However, we do not cover this selection methodology in this paper.
2. Various papers dealing with inflation forecasting have been published by the Bank of Japan's Research and Statistics Department, following sessions such as the "Workshop on Inflation Forecasting Errors" (September 2000); *cf.*, for example, Ban and Saito (2001), Kitagawa and Kawasaki (2001), Kasuya and Shinki (2001), and Fukuda and Keida (2001), among others. Although the analysis in this paper is close to that of Kitagawa and Kawasaki (2001), we differ from them in that we produce the forecast distributions (fan charts) which take account of the uncertainty inherent in the forecast errors and the model selection.
3. There are models other than VAR models that place their main emphasis on statistical relationships; we might for example have chosen to adopt a method such as the neural network approach. However, in the analysis that follows, we have restricted ourselves to the VAR framework, attempting to produce inflation forecasts which take account of the uncertainty inherent in the forecast errors and the model selection.

a result of our choice of data, or of some particular combination of variables, we follow Stock and Watson (1999) and Pesaran and Timmermann (1992, 1995) and carry out estimations of the model for as many different combinations of variables as possible, in the hope that by thus constructing a large number of VAR forecasting models we may improve our results. As our estimation methodology, we employ a VAR model that takes account of rank restrictions among the variables, in other words a reduced rank VAR model (RR-VAR model[4]). As described above, in estimating VAR models based on various different combinations of variables, it is not unreasonable to assume that some quantity of redundant information will be included in the matrix of parameters, since several of the variable combinations are likely to display similar tendencies. To deal with this, we employ rank restrictions on the parameter matrix and are thus able to carry out estimation using a more parsimonious[5] model, which is likely to produce improved results. On this point, by comparing the RR-VAR model with the standard VAR model, Camba-Mendez *et al.* (1999) have demonstrated the performance of the former to be superior to the latter for forecasting purposes.

Having reached this point, we then construct the forecast distribution, or fan chart, which takes account of the uncertainty in both the model and the forecast value. We produce three different constructions of the forecast distribution: (1) the top model distribution; (2) the nonparametric distribution; and (3) the mixed distribution. (1) is the distribution we get when we make use of the forecast error from the model that performs best out of all of the forecasting models which we constructed. (2) is the estimate of future uncertainty that we get when we depict a dispersion of the forecast values (point estimates) obtained from each of our forecasting models. For (3), we take matters a step further: each of the forecast values obtained from each of the forecasting models is assumed to be normally distributed according to parameters extracted from the forecast errors of the respective model; then, using the "forecast error reciprocals" from each of the distributions as the base for constructing weights, we compose a distribution that is able to capture uncertainty.[6] To put it differently: we construct three different forecast distributions, where (1) is based on parametric methodology, (2) on nonparametric methodology, and (3) on mixed distribution methodology.[7] Lastly, by arranging these various constructed distributions into time series, we illustrate the extent of uncertainty in three fan charts.

In addition to the quantitative forecasting question "Roughly what percentage rate of inflation do we expect to see?" we also consider the question "Is future inflation more likely to move upward or downward?" and construct a qualitative forecast (in other words, a simple up-or-down forecast) to examine this.

4. For details, see Velu, Reinsel, and Wichern (1986) and Lütkepohl (1991).
5. In other words, a model that obeys the principles of parsimony.
6. Model uncertainty enters in three possible ways: (1) the average difference between the real values and those forecast by the model (the forecast error); (2) misspecification of the estimating model; and (3) parameter uncertainty (i.e., how the parameters obtained via estimation are distributed). When we carry out stochastic simulations for the parameters, although it is possible to consider (3), it is highly computationally demanding to do so. Thus, for our forecast distribution number (3) we in fact think only about the uncertainty that arises from a mixture of concepts (1) and (2) above.
7. This approach may be considered to follow the thinking of the forecast combination method (for which, see Clements and Hendry [1998], Diebold [1998], and Granger and Newbold [1986]).

II. Forecasting Techniques and Results

Forecasting in our HEVAR/FMU is carried out via three processes: (1) forecasts with the RR-VAR models; (2) construction of forecast distributions and fan charts; and (3) the qualitative up-down directional forecast that makes use of the Pesaran-Timmermann test (for an outline of techniques [1] and [2], see Figure 1). More detail on each of these processes follows.

A. Forecasting with the RR-VAR Model

Here, we first of all construct a large number of RR-VAR models from different combinations of stationary data, and use these to calculate a correspondingly large number of forecast values over several periods. Next, we arrange the forecasting models in order, based on their in-sample dynamic forecast[8] performance over the past two years (where performance is judged according to root mean-squared error [RMSE]), and we select the model that performs best for each forecasting period (one period ahead, . . . , four periods ahead). Stringing together the out-of-sample dynamic forecasts obtained from these best-performing models, we construct the "movement over time of the forecast value obtained from the best-performing forecasting model." In this way, we manage to account for the possibility that, when we compare the performance for each object forecasting period across models, the combination of variables that produces the best forecasts for, for example, one period ahead (i.e., one quarter from now) may differ from the combination of variables that works best when forecasting four periods ahead (one year from now).

Turning to the details, we see that there are seven specific steps involved in calculating the forecast values. These are as follows:

1. Choice of data series

Having selected the appropriate series of representative macroeconomic variables, we carry out the necessary seasonal adjustment[9] and take logs (see Table 1 for data selection and seasonal adjustment procedures).

2. Ensuring data stationarity

We perform unit-root tests[10] on each data series, taking differences until stationarity is ascertained. Making use of their means and standard deviations, we then standardize[11] these now-stationary data series. Having discovered that the consumer price index (CPI), which we are looking to forecast, follows an I(2) process, we employ the twice-differenced series, namely "change in the inflation rate."[12]

3. Choice of explanatory variables

From the data series obtained in step 2, we select the arbitrary number of explanatory variables. Specifically: from N data series we select $\underline{m}$ to M different combinations

8. Clements and Hendry (1998) observe the following: when for example carrying out a four-period-ahead forecast, using only data that precede the object forecasting period by at least four periods to estimate the model directly can produce forecasts whose performance is superior to dynamic forecasts. However, we do not use this technique here.
9. Monthly variables are reconstituted, after seasonal adjustment, in quarterly form.
10. We use the augmented Dickey-Fuller (ADF) test here.
11. We transform them into data series with means of zero and standard deviations of one.
12. Since our data are in logs, the second difference gives the (magnitude of the) change in the rate of change. Having taken differences to ensure stationarity, the combination of the differenced explained variables and the differenced explanatory variables may become difficult to explain from the standpoint of economic theory.

Figure 1 Outline of the Forecasting Method

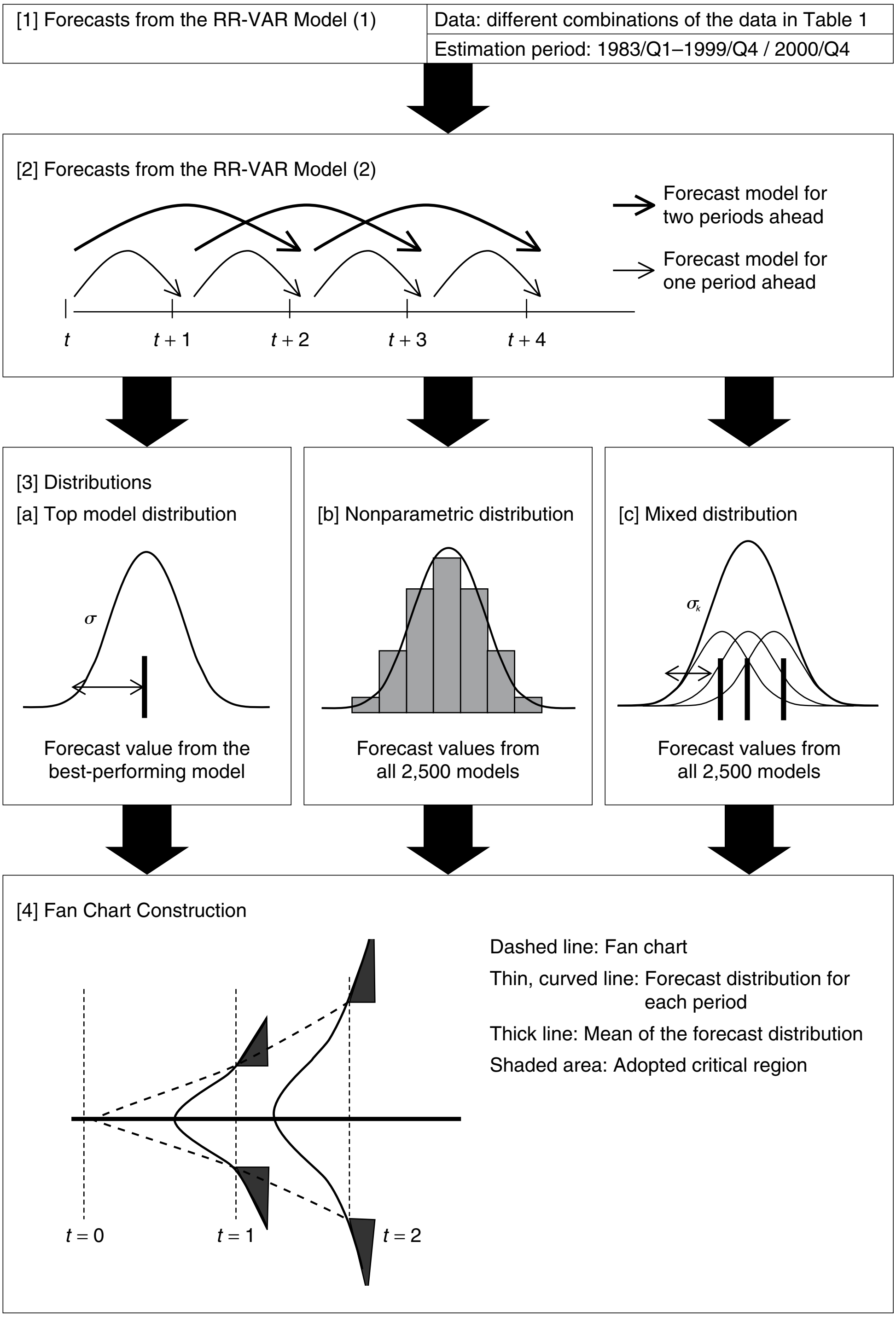


Table 1 Data Series Used in Performing the Estimation

	Data	Method of adjustment
0	CPI (excluding fresh food, package tours overseas, personal computers)	Seasonally adjusted; consumption tax adjusted
1	Domestic wholesale price index	Consumption tax adjusted
2	Import price index	
3	Industrial production index (value-added weights)	Seasonally adjusted
4	Index of tertiary industry activity	Seasonally adjusted
5	SNA private consumption (in real terms)	Seasonally adjusted
6	SNA nonresidential investment (in real terms)	Seasonally adjusted
7	SNA public investment (in real terms)	Seasonally adjusted
8	SNA exports of goods and services (in real terms)	Seasonally adjusted
9	Rate of unemployed in labor force	Seasonally adjusted
10	*Tankan* diffusion index of employment conditions	
11	Hourly contractual cash earnings (contractual cash earnings/total number of working hours)	Seasonally adjusted
12	Monetary aggregate (M2+CDs) (average outstanding)	Seasonally adjusted
13	TOPIX (average closing)	
14	Government bond yield (10-year)	
15	Nominal effective exchange rate	
16	Import penetration of consumption goods (imports of consumption goods/aggregate supply of consumption goods)	

Note: Seasonal adjustment of the CPI is carried out after exclusion of (1) fresh food, (2) package tours overseas, and (3) personal computers. This is because each of these items (1) is subject to large seasonal variability depending on changes in climate; (2) is subject to large seasonal fluctuations (however, since this is newly adopted from the 2000 base, the sample size of this item is insufficient for seasonal adjustment); and (3) has exhibited a downward trend movement since 2000, because of the adoption from the 2000 base. As a result, these items have caused disturbance in the overall CPI time series.

Sources: Cabinet Office, *National Accounts*; Ministry of Health, Labour and Welfare, *Monthly Labour Survey*; Ministry of Economy, Trade and Industry, *Indices of Industrial Production, Indices of Industrial Domestic Shipments and Imports, Indices of Tertiary Industry Activity*; Bank of Japan, *Balance of Payments Monthly, Wholesale Price Index*, Tankan—*Short-Term Economic Survey of Enterprises in Japan, Money Supply*, etc.

of explanatory variables that are assumed to be included in each model. The total number of combinations of variables turns out to be $K = \sum_{m=\underline{m}}^{M}(N!/m!(N-m)!)$.

In this paper, there are 16 data series from which we select combinations of two to four explanatory variables. Thus, we produce a total of exactly 2,500 different models.

4. Estimation of the RR-VAR models

We carry out estimation using three to five dimensional RR-VAR models, made up of the CPI that we are seeking to forecast and the explanatory variables selected in step 3. Where $\boldsymbol{y}_t$ is a multivariate stationary time series made up of the group of variables included in some given model k, the RR-VAR model is set up as follows. Note that the lag length (L), which is to take one to four, is determined where the Akaike information criterion (AIC) is the smallest.[13] The following example illustrates the case when the estimation period for the parameters is 1983/Q1–1999/Q4.

13. We do not consider the use of AIC corrected for finite sample (usually called AICC, *cf.* Hurvich and Tsai [1989]), although it is known as a more appropriate criterion to determine the lag length of the VAR system given the limited length of economic time series.

$$\boldsymbol{y}_t = \sum_{l=1}^{L} \boldsymbol{B}_l \boldsymbol{y}_{t-l} + \boldsymbol{\eta}_t = \boldsymbol{B}\boldsymbol{x}_t + \boldsymbol{\eta}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_t + \boldsymbol{\eta}_t, \quad t = 1983/\text{Q1}, \ldots, 1999/\text{Q4}.$$

Here $\boldsymbol{B}_l$ is an (m, m) matrix. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_l$ are (m, r^*) matrices ($r^* \leq m$) when they are of rank r^*. Also, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_L')'$, $\boldsymbol{x}_t = (\boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-L}')'$, and $\boldsymbol{B} = \boldsymbol{\alpha}\boldsymbol{\beta}'$.

In carrying out estimation using the RR-VAR model, we first estimate the matrix of parameters that we get using a standard VAR model; then, taking the rank r^* of this matrix as a base, we reduce the matrix into two more parsimonious matrices without any loss of information (i.e., having estimated the rank of the parameter matrix $\boldsymbol{B}$, we then reduce this into the form of the two matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$).

We follow Bartlett (1947) in estimating the rank. Bartlett's rank test is presented below in the form of a likelihood ratio test.

$$\text{H}_0\text{: rank } (\boldsymbol{B}) = r^* \quad \text{against} \quad \text{H}_1\text{: rank } (\boldsymbol{B}) = r > r^*,$$

$$l_c(\text{H}_0) - l_c(\text{H}_1) = \frac{T}{2} \sum_{i=r^*+1}^{m} ln(1 + \hat{\lambda}_i^2) \xrightarrow{d} \chi^2[(m - r^*)(mL - r^*)].$$

l_c are the concentrated log likelihoods,[14] while $\hat{\lambda}_i$ are the eigenvectors of the matrix $T^{-1}\boldsymbol{\Sigma}_\eta^{-1/2}\boldsymbol{Y}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{X}')^{-1}\boldsymbol{X}\boldsymbol{Y}'\boldsymbol{\Sigma}_\eta^{-1/2}$, where $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T]$, $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T]$. $\boldsymbol{\Sigma}_\eta = E(\boldsymbol{\eta}\boldsymbol{\eta}')$ denotes the error variance-covariance matrix, estimated by the quasi-maximum likelihood method.[15]

Under the null hypothesis H_0, the concentrated log likelihood ratio $l_c(\text{H}_0) - l_c(\text{H}_1)$ is known to converge to the χ^2 distribution described above, and comparing it with the significance points (corresponding to the adopted significance level) for the χ^2 distribution determines the rank r^*.

Having calculated the rank in this way, we then use this to estimate the RR-VAR parameter matrices (in other words, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$) as[16]

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}_\eta^{1/2}\hat{\boldsymbol{V}}, \quad \hat{\boldsymbol{\beta}}' = \hat{\boldsymbol{V}}'\boldsymbol{\Sigma}_\eta^{-1/2}\boldsymbol{Y}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{X}')^{-1}.$$

Here $\hat{v}_r$ is the standardized eigenvector corresponding to the r-th largest eigenvalue $\hat{\lambda}_r$ (where r is the rank calculated above), and $\hat{\boldsymbol{V}}$ describes the matrix $\hat{\boldsymbol{V}} = [\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \ldots, \hat{\boldsymbol{v}}_{r^*}]$.

5. Calculation of the in-sample forecasts

For each of our 2,500 models, we carry out a series of in-sample dynamic forecasts,[17] ranging from one-period-ahead forecasts to four-period-ahead forecasts. The evaluation period for gauging forecast performance stretches across two years of past data (1998/Q1–1999/Q4). In other words, we produce forecasts recursively, shifting the

14. In cases when the log likelihood is made up of several parameters, this is concentrated so as to be able to express just the parameters of interest. Specifically, we concentrate the log likelihood function by sequential replacement of the first-order condition in the log likelihood function for all parameters other than the parameters of interest.
15. In other words, we calculate $T^{-1}\boldsymbol{Y}\boldsymbol{Y}' - T^{-1}\boldsymbol{Y}\boldsymbol{X}'(T^{-1}\boldsymbol{X}\boldsymbol{X}')^{-1}(T^{-1}\boldsymbol{Y}\boldsymbol{X}')'$. No structural assumption is made. Therefore, estimated variance-covariance matrices are full ranked.
16. For more details, see Velu, Reinsel, and Wichern (1986) or Lütkepohl (1991).
17. Here, having ranked models based on their in-sample dynamic forecast performance, we produce out-of-sample dynamic forecasts. Since our interest lies ultimately in the out-of-sample forecasts, there is an argument to suggest that we should also rank models based on their out-of-sample forecast performance (e.g., Clark [2000]). However,

starting period for the dynamic forecast each time, and obtaining values for the one-period-ahead forecast ($\hat{\boldsymbol{y}}_{\tau|\tau-1}$) to the four-period-ahead forecast ($\hat{\boldsymbol{y}}_{\tau|\tau-4}$), where τ is taken across the eight periods from 1998/Q1 to 1999/Q4 (see Figure 1 [2]).[18] Since we have 2,500 models and we calculate forecast values for one to four periods ahead for each of the eight past periods, we produce a total of 80,000 forecasts (2,500 × 4 × 8).

$$\hat{\boldsymbol{y}}_{\tau|\tau-h} = [\boldsymbol{\alpha\beta}']^h \boldsymbol{x}_{\tau-h}, \quad \tau = 1998/Q1, \ldots, 1999/Q4, \quad h = 1, ., 4.$$

6. Calculation of the RMSE

We calculate the RMSEs,[19] which indicate the relative performance of our respective inflation forecasts. However, since for the CPI our estimate is of the change in the inflation rate,[20] we produce forecasts of the inflation rate itself by adding as appropriate the estimated changes in the inflation rate to the inflation rate from the last estimation period of the model. The RMSEs are calculated based on the differences between the actual values and the forecasts produced in this way.

In other words, denoting the actual value of the inflation rate in period τ as R^{π}_{τ}, and the forecast value as $\hat{y}^{\pi}_{\tau|\tau-h}$, we get

$$RMSE_h = T^{-1}\sum_{\tau=1}^{T}\epsilon^2_{\tau,h}, \quad \tau = 1998/Q1, \ldots, 1999/Q4, \quad h = 1, ., 4,$$

$$\epsilon_{\tau,h} = R^{\pi}_{\tau} - \hat{y}^{\pi}_{\tau|\tau-h}.$$

Carrying out the above calculations for every model, we obtain $\{RMSE_{h,k}\}^{2,500}_{k=1}$.

7. Calculation of the out-of-sample forecasts

Having obtained the $RMSE_{h,k}$, we use these to select, for each respective forecasting period, the best-performing model (i.e., the model with the smallest RMSE). Having transformed into inflation rates the out-of-sample dynamic forecasts ($\hat{\boldsymbol{y}}_{\tau|\tau-h}$) obtained from these best-performing models, we produce a time series for our object forecast period (i.e., four periods into the future) by linking these estimates together and constructing the "movement over time of the forecast value obtained from the best-performing models." This is depicted as the central line (the line showing

for evaluations based on out-of-sample forecasts, since the estimation period for performance evaluation is set to be a certain number of periods (here, eight periods) prior to the first forecast period, it becomes impossible to use all the information potentially available to us when carrying out the estimation. Here, therefore, we choose to evaluate forecast performance as we do, based on the conjecture that the out-of-sample performance will be superior if we conduct our estimations using all of the information available to us.

18. Note that the estimation period itself remains unchanged throughout.
19. As pointed out in Clements and Hendry (1998), there is potentially a problem here since we do not consider possible serial correlation in the forecast errors when calculating the RMSE. There are two possible ways that can be considered for dealing with this: (1) we could set the forecasting periods so that there is no overlap between the forecasts (i.e., we could set the data frequency so that, for example, it is quarterly for a one-period-ahead forecasting model or yearly for a four-period-ahead forecasting model); or (2) we could use instrumental variables. However, whichever of these methods we use, we run into problems, such as a loss of degrees of freedom or the difficulty of choosing the right instruments. For this reason, although the RMSE represents extremely important information for constructing the forecast distributions, we do not revise them to correct for the possibility that they are excessive because of serial correlation in the forecast errors.
20. Strictly speaking, since as we pointed out in Section II.A.2 the estimated forecast values are normalized, we carry out these calculations after first reversing the normalization procedure.

movement of the mean over time) in the fan chart (Figure 2) for the "top model distribution" (see below for further explanation).

$$\hat{\boldsymbol{y}}_{\tau|\tau-h} = [\boldsymbol{\alpha\beta}']^h \boldsymbol{x}_{\tau-h}, \quad \tau = 2000/\mathrm{Q}1, \ldots, 2000/\mathrm{Q}4, \quad h = 1, ., 4.$$

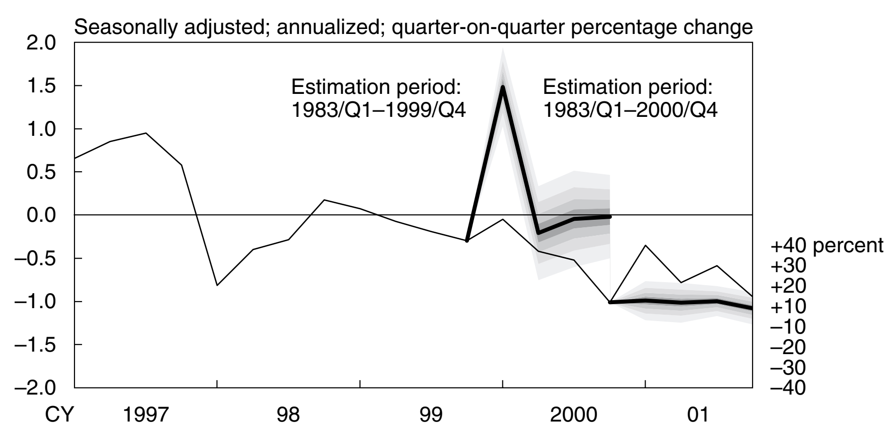
B. Construction of the Forecast Distribution and Fan Charts

The movements over time in the forecast values obtained from the best-performing models and calculated in the preceding section may be considered, in themselves, to be highly useful as forecasts based on a technique that is as objective as possible. However, as discussed earlier, we are interested here in capturing the uncertainty inherent in the forecast, and this is well accounted for in the forecast distribution. Thus, we now demonstrate how we produce this forecast distribution, making use of the out-of-sample forecast values and the forecast errors.

As shown in Figure 1 [3] a–c, the forecast distribution captures in distributional form the uncertainty inherent in the forecast value. Setting the desired significance level for the forecast distribution at each point in time, and then linking the resulting significance points as a time series, gives us our fan charts (Figure 1 [4]).

Here, we depict first of all (1) the "top model distribution," which uses information from the best-performing models alone (their forecast values and forecast errors). Next, we depict (2) the "nonparametric distribution," which makes use of the forecast values from the full set of models. Lastly, we produce (3) the "mixed distribution," which makes use of the forecast values and the forecast errors from the full set of models. In contrast with the "top model distribution," which reflects only the information gleaned from the best-performing model, the "nonparametric distribution" and the "mixed distribution" capture uncertainty about the future: specifically, the former does

Figure 2 Fan Charts Based on the Top Model Distribution



Note: The thinner line depicts the actual value, and the thicker line the mean of the forecast distribution. The fan charts illustrate 10 percent increments in the width of the band around the mean, using gradations in the shading.

so through a dispersion of forecast values taken from the full set of models (i.e., it addresses the possibility of model misspecification); while the latter takes into account in addition the forecasting errors inherent in each forecast value (i.e., it addresses the issues of both forecasting errors and model misspecification).

1. Top model distribution

First, for the best-performing models selected above, we create, for each forecast period, a normal distribution with the forecast value as its mean and the RMSE as its standard deviation. We then produce fan charts with starting periods of 2000/Q1 and 2001/Q1,[21] by joining together as time series the significance points attained for each 10 percent increment in the one-sided significance level for this forecast distribution.

With regard to forecasts based on this top model distribution, when we consider that there may not have been such large differences between the RMSE of the top-ranking models, it may well be that we are relying too heavily on just the information gleaned from one model, picked out from its peers on the basis of infinitesimal differences. Putting it another way, when we take into account the possibility that the model may have been misspecified, it brings into question the appropriateness of evaluating our inflation rate forecasts in light of only the forecasts and the expression of forecast uncertainty taken from the top model distribution alone.

As econometric principles, sharpening up an existing model or searching out the best-possible model may be thought important. However, here, where interest lies in the construction of a practical forecasting model, we are searching for a forecasting technique that takes into account the forecasts obtained from a number of models. Thus, in what follows we no longer restrict our attention to the best-performing model, and instead focus on building forecast distributions that make full use of the information contained in the forecasts and concomitant forecasting errors from the other models.

2. Nonparametric distribution

Having produced forecasts from a variety of models for the inflation rate of one particular period, it seems reasonable to assume that, should it be the case that these forecast values are roughly the same whichever forecasting model is used, then the actual value will also lie within this neighborhood. For this reason, we regard a dispersion of the forecast values obtained from a number of models as an indicator of the uncertainty that attends a forecast of the future.

Turning to the risk of misspecification, it becomes important to make full use of the information gained from several models and not look just at the information contained in one particular model.

In addition, because we made the assumption of normality when looking at the uncertainty in the top model distribution, it was not possible to represent any skewness

21. The estimation periods for the parameters are as follows: for fan charts whose starting period is 2000/Q1, the estimation period runs from 1983 through 1999; while for fan charts starting in 2001/Q1, it runs from 1983 through 2000. Likewise, the forecast performance evaluation periods are 1998 and 1999, and 1999 and 2000, respectively. The forecast distribution is drawn based on the RMSE calculated during the evaluation period (i.e., the prior two-year period specified above), and in cases (such as 1998–99) when movements in the actual inflation rate for the relevant period are not easily explained by the model (in other words, there is a large error), the spread of the forecast distribution tends to become wider. We can attribute the wider spread of the fan chart in 2000, compared with that in 2001, to this.

in the distribution. With the nonparametric distribution, it becomes possible to represent asymmetry in the distribution of the inflation rate with regard to its movement in an upward versus a downward direction.

In accordance with this line of thought, and following the two steps outlined below, in this subsection we first of all carry out the estimation of the nonparametric distribution, which makes use only of the information contained in the dispersion of forecast values.

a. Estimation of the nonparametric distribution

The aim of producing a fan chart with these nonparametric distributions[22] is to construct forecast distributions using just the information contained in the dispersion of 2,500 forecast values[23] we have for each period. What we are producing here is not a forecast distribution built assuming some particular shape of probability distribution and using parameters calculated from the forecast values themselves, such as their means and variances. Instead, we elicit a probability distribution from the frequency distribution that derives from all 2,500 forecast values, thus constructing our distribution in a nonparametric manner (i.e., neither estimating any parameters, nor placing any *a priori* restrictions upon the shape of the probability distribution). The nonparametric distribution is expressed using the kernel density estimator below. $\kappa(u)$ is the kernel weighting function (here, a Gaussian kernel), ϕ is the smoothing parameter that determines bandwidth, and the formula for the probability density function (pdf) is written as follows:[24]

$$pdf(q) = \frac{1}{K\phi}\sum_{k=1}^{K}\kappa\left(\frac{q - y_k^{\pi}}{\phi}\right),$$

$$\kappa(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2},$$

$$\phi = 0.9K^{-\frac{1}{5}}\min\{S, A/1.34\}.$$

Here, $\boldsymbol{y}^{\pi} = \{y_1^{\pi}, \ldots, y_k^{\pi}, \ldots, y_K^{\pi}\}$ is the set of forecast values, K is the number of forecast values (2,500 of them), S is the standard deviation of $\boldsymbol{y}^{\pi}$, A is the interquartile range of $\boldsymbol{y}^{\pi}$, and q denotes the estimation point of the density function.

22. For detailed treatments of the nonparametric methodology in general, see Silverman (1986), Härdle and Linton (1994), or Pagan and Ullah (1999).
23. We could consider building the forecast distribution not from the full set of sample values (2,500 values), but from a subset of, for example, the top 5 percent in terms of their performance (125 values). Since such a forecast distribution could prove extremely useful, we would like to examine, in the future, issues such as the most suitable sample size and so on, through an extensive comparison with actual values. Of course, since here our primary point of focus is the introduction of techniques, and since the only strategy that we can conceive for approaching the question of how to choose the sample size is an extensive comparison with actual values to be carried out later, in our production below of the forecast distributions, including the mixed distributions, we use the whole sample.
24. According to Silverman (1986), there are two methods for calculating the bandwidth ϕ: he refers to the method on which the expression used in the text is based as the "subjective method"; the alternative "objective method" sets out to minimize the "cross-validation" criterion (an MSE estimator that measures the difference between the actual value and the estimate obtained with a given ϕ). Although the latter method gets rid of some arbitrariness, in cases such as in this paper, where we confirm the spread of the distribution in the fan charts, it is not desirable to have a bandwidth that varies greatly depending on the forecasting period. For this reason, placing emphasis on the goal of getting a rough idea of the shape of the distribution, we use the former method here.

b. Construction of the fan charts

For the forecast distributions estimated for each period in Section II.B.2.a, we set the significance level, which we vary in one-sided 10 percent increments, and by joining together as time series the respective significance points attained for each significance level, we produce a fan chart.

The forecast distributions produced in this way are illustrated in Figure 3. In contrast to the top model distribution, because the fan chart is constructed with reference to the forecast values from all 2,500 models, it is not overly affected by the movements in the explanatory variables of one particular model (i.e., the best-performing model).

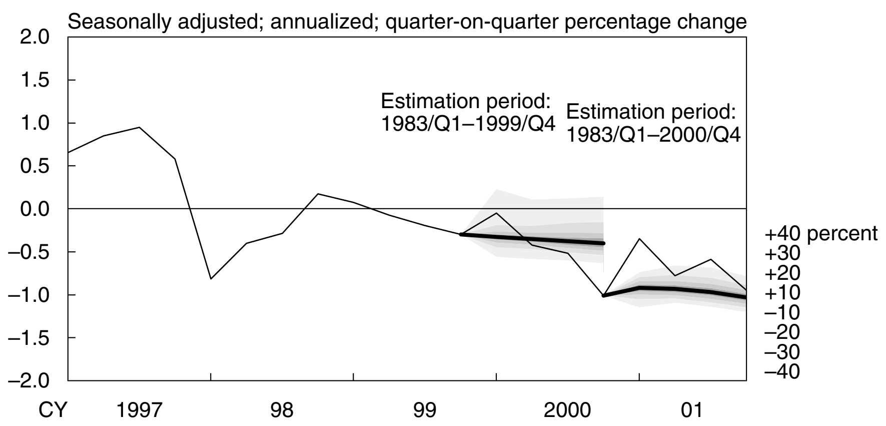
It is also noteworthy that, if we look at for example the forecasts for mid-2000, the upper half of the forecast distribution can be seen to have a long tail. In other words, the mode is situated below the mean, reflecting the fact that, restricting attention to frequency, there are a large number of forecasts predicting inflation rates lower than the mean.

3. The mixed distribution

In the technique based on the nonparametric distribution, we sought to obtain the forecast uncertainty from the dispersion of forecast values. Of course, in this method we take no account at all of the uncertainty inherent in each individual forecast value, nor of the information gleaned about the relative performance of models in the in-sample dynamic forecast. Thus, it is possible that we are not giving due weight to the forecast values from the top-performing models. For this reason, in the development of the mixed distribution that follows, we produce a forecast distribution that takes appropriate account of such information.

To begin with, we construct a normal distribution for each forecast value based on its RMSE. Then we take a weighted linear combination of these normal distributions,

Figure 3 Fan Charts Based on the Nonparametric Distribution



Note: The thinner line depicts the actual value, and the thicker line the mean of the forecast distribution. The fan charts illustrate 10 percent increments in the width of the band around the mean, using gradations in the shading.

with the respective weights in proportion to the inverse RMSE of each forecast distribution.[25] Although admittedly using the inverse RMSEs as weights is somewhat *ad hoc*, this is based on the notion that those models whose performance has been good over the past two years may be considered to provide relatively high forecast accuracy in the future as well.[26] Specifically, the mixed distribution is constructed following the three steps outlined below.

a. Construction of a normal distribution for each forecast value

For a forecast h periods ahead using model k, we construct a normal distribution with the forecast value $\mu_{h,k}$ as its mean, and a standard deviation $\sigma_{h,k} = RMSE_{h,k}$. The probability density function $f_{h,k}$ can then be written as follows:

$$f_{h,k}(z_{h,k};\ \mu_{h,k},\ \sigma_{h,k}) = \frac{1}{\sqrt{2\pi}\sigma_{h,k}} e^{-\frac{(z_{h,k}-\mu_{h,k})^2}{2\sigma_{h,k}^2}},$$

$$\int f_{h,k}(z_{h,k})dz_{h,k} = 1, \quad 0 \le f_{h,k}(z_{h,k}).$$

We then convert the probability density function into a cumulative distribution function:

$$F_{h,k}(z_{h,k};\ \mu_{h,k},\ \sigma_{h,k}) = \int_{-\infty}^{z_{h,k}} f_{h,k}(\zeta_{h,k};\ \mu_{h,k},\ \sigma_{h,k})d\zeta_{h,k}, \quad 0 \le F_{h,k}(z_{h,k}) \le 1.$$

b. Construction of a weighted linear combination of cumulative distributions

The cumulative distribution function for h periods ahead is expressed as F_h, a weighted linear combination of K cumulative distribution functions $F_{h,k}$. The weights $w_{h,k}$ are based on the inverse RMSE and are constructed so as to sum to unity (thus, the area under the weighted linear combination of probability density function f_h is also one).

25. Although the application of the weights is different, distributions that capture uncertainty by compounding other distributions in this way can be found in the field of finance. For an example of existing research in the area, we would suggest Melick and Thomas (1997).

In addition, there is a similar research stream in statistics. The study in forming a mixture predictive density can be traced back to Aitchison (1975). Akaike (1977) proposed the use of exp(-AIC$_j$/2) (say, for the j-th model) to construct the weight of candidate models. Furthermore, model averaging has been actively investigated since the middle of the 1990s, mainly from a Bayesian viewpoint. See Hoeting *et al.* (1999) for an excellent survey.

26. This is indeed just one example of a number of possible ways of applying weights. Creating the same type of forecast distribution of the period for which we have actual values, we also tried the grid search method, in which weights are chosen to bring the mean of the forecast distribution closer to the actual value. However, even with this method, we are ultimately unable to confirm whether or not we are extracting the appropriate information from the forecast values. Among examples of the semiparametric methods, rather than carrying out estimation of the distribution using weights fixed in advance as above, it is common to estimate the weights, the mean, and the variance, by using the expectation-maximization (EM) algorithm, or something similar, to maximize log likelihoods based on a linear combination of distributions. However, in this method we are unable to make use of the information contained in the RMSE.

In addition to the above, there are further alternative methods, as introduced in Clements and Hendry (1998): (1) the "regression method," in which weights are calculated by regressing the actual values on a number of the forecast values such that the forecast errors from the combination forecasts are minimized; and (2) the "variance-covariance approach," in which the weights are calculated such that the forecast error variance is minimized. However, in the case of the former, not only are we unable to secure a large-enough sample to calculate the requisite parameters for 2,500 forecast values, but there are also a number of other definitional problems that emerge, as summed up in Diebold (1998): whether to include a constant term, whether to estimate time-variant weights, and whether to allow for serial correlation in the errors in the constructed forecasts. For the latter method, it is unclear whether, with a small sample, we would be able to obtain reliable estimates of the variances and covariances.

For these reasons, we choose to adopt the inverse RMSE as weights here, as a straightforward and uninvolved method.

$$F_h(z_h;\ \mu_h,\ \sigma_h) = \sum_{k=1}^{K} w_{h,k} F_{h,k}(z_{h,k};\ \mu_{h,k},\ \sigma_{h,k}),\quad 0 \le F_h(z_h) \le 1,$$

$$\sum_{k=1}^{K} w_{h,k} = 1,\quad 0 \le w_{h,k} \le 1,$$

$$f_h(z_h;\ \mu_h,\ \sigma_h) = \sum_{k=1}^{K} w_{h,k} f_{h,k}(z_{h,k};\ \mu_{h,k},\ \sigma_{h,k}),\quad 0 \le f_h(z_h),$$

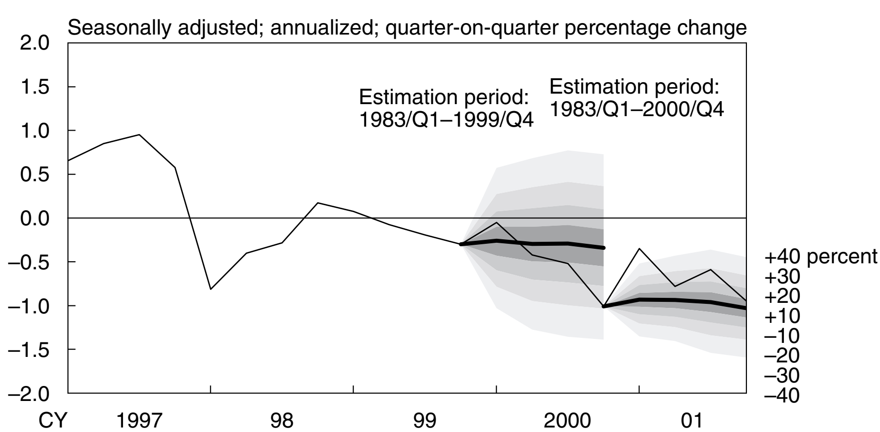
$$\int f_h(z_h) dz_h = 1.$$

c. Construction of the fan chart

For the weighted linear combination of cumulative density function estimated for each period in Section II.B.3.b, we set the significance level, which we vary in one-sided 10 percent increments, and calculate the resultant significance points. Joining these together as a time series, we produce a fan chart.

Comparing the fan chart constructed from the mixed distribution (Figure 4) with that constructed from the nonparametric distribution (Figure 3), we observe that although the movement over time of the means of the forecast distributions is very similar for each period, the spread of the distribution is more pronounced in the former (i.e., there is more uncertainty). We can hypothesize two reasons for this. First, (1) while the nonparametric distribution focuses on "pinpoint" forecast values from each model, the mixed distribution by contrast focuses on the distribution centered on the forecast value (i.e., a spread), calculated from each model. In addition, we may point out (2) while in the nonparametric distribution all the forecast values are given the same weight, so that forecasts from models that performed well during the period over which

Figure 4 Fan Charts Based on the Mixed Distribution



Note: The thinner line depicts the actual value, and the thicker line the mean of the forecast distribution. The fan charts illustrate 10 percent increments in the width of the band around the mean, using gradations in the shading.

RMSEs were calculated are treated no differently from others, in the mixed distribution on the other hand, greater importance is attached to forecast values (and their concomitant distributions) taken from models that performed relatively well over the same period. The result of this is that, in cases when these better-performing forecast values lie apart from the majority of the other forecasts, they may cause the spread of the forecast distribution to widen.[27]

Although we are ultimately judging only from relative performances during 2000 and 2001, the mixed distribution appears to offer the most insightful guide when we are considering at roughly in what range the forecast value will fall.

C. Qualitative Up-Down Forecast Using the Pesaran-Timmermann Test

The aim of the forecast distribution is to obtain a measure of the forecast value and the size of the associated uncertainty (i.e., a quantitative forecast). However interest may lie in the question of "whether the next period's inflation rate will be higher or lower than that observed in this period," and when thinking about the future inflation rate, this kind of forecast of the direction of change (i.e., a qualitative forecast) may also be considered important. Thus, we seek here to produce a simple up-down forecast of the direction of the inflation rate, using the large number of VAR models that we constructed in Section II.A. We evaluate the respective performances of the qualitative forecasts that emerge from each model using the Pesaran-Timmermann test, but the qualitative forecast that makes use of all the models is produced by following the two steps outlined below.

1. Calculation of the Pesaran-Timmermann statistic for each forecast value

The Pesaran-Timmermann test examines whether the directional movements of the real and forecast values are in step with one another, or to put it another way, it checks how well rises and falls in the forecast value follow actual rises and falls: the larger the Pesaran-Timmermann statistic, the better the match.[28]

The way the test works can be simply explained via the following example. Forecasting of coin tosses is carried out according to some model. Whenever heads comes up (or is forecast to come up) the value of +1 is assigned, while whenever tails comes up (or is forecast to come up) the value of –1 is assigned. In each instance, the resulting values from the actual coin toss and from the forecast from the model are multiplied together. At this point, if, for example, the mean of this product is close to unity, it suggests that the performance of the forecasting model is high. The Pesaran-Timmermann statistic is a statistic that indicates qualitative forecasting performance based on this line of reasoning. Substituting a rise (+1) or a fall (–1) in the inflation rate for the heads or tails of the coin toss, we can apply the Pesaran-Timmermann test to our inflation forecasts.

More specifically, denoting the in-sample dynamic forecast calculated in Section II.A.5 (i.e., the forecast value of the magnitude of the change in the inflation rate) as $\hat{y}_t$,[29] and the actual value recorded for the change in the inflation rate as R_t (i.e.,

27. For a discussion of issues underlying the wider base of the forecast distribution in 2000 compared to 2001, see Footnote 21.
28. It takes no account, however, of the differences between the respective magnitudes of rises and falls.
29. As explained in Footnote 20, strictly speaking, the estimated value has been normalized, so that we need first to reverse this process.

$\hat{y}_t = E(R_t \mid \Omega_{t-1})$, where Ω_{t-1} sums up all the information available at $t - 1$), we first of all define the following quantities:

$$\begin{aligned} X_t &= 1 \quad \text{if } R_t > 0 \\ &= 0 \quad \text{otherwise,} \\ Y_t &= 1 \quad \text{if } \hat{y}_t > 0 \\ &= 0 \quad \text{otherwise,} \\ Z_t &= 1 \quad \text{if } R_t \hat{y}_t > 0 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Now the Pesaran-Timmermann statistic S_n defined below follows, asymptotically, a standard normal distribution, under the null hypothesis that R_t and $\hat{y}_t$ are independent random variables.[30]

$$S_n = \frac{\hat{P} - \hat{P}_*}{[\hat{V}(\hat{P}) - \hat{V}(\hat{P}_*)]^{\frac{1}{2}}} \xrightarrow{d} N(0, 1),$$

$$\hat{P}_X = \sum_{t=1}^{n} X_t / n, \quad \hat{P}_Y = \sum_{t=1}^{n} Y_t / n, \quad \hat{P} = \sum_{t=1}^{n} Z_t / n,$$

$$\hat{P}_* = \hat{P}_X \hat{P}_Y + (1 - \hat{P}_X)(1 - \hat{P}_Y).$$

Here, $\hat{V}(\hat{P})$ and $\hat{V}(\hat{P}_*)$ denote the variances of $\hat{P}$ and $\hat{P}_*$, respectively, and are defined as follows:

$$\begin{aligned} \hat{V}(\hat{P}) &= n^{-1} \hat{P}_* (1 - \hat{P}_*), \\ \hat{V}(\hat{P}_*) &= n^{-1} (2\hat{P}_Y - 1)^2 \hat{P}_X (1 - \hat{P}_X) + n^{-1} (2\hat{P}_X - 1)^2 \hat{P}_Y (1 - \hat{P}_Y) \\ &\quad + 4n^{-2} \hat{P}_Y \hat{P}_X (1 - \hat{P}_Y)(1 - \hat{P}_X). \end{aligned}$$

Having calculated the Pesaran-Timmermann statistic for each model from the forecast values and real values for the change in the inflation rate over the object forecasting period (one to four periods ahead), we go on to select those models that reject the null hypothesis described above at the 5 percent significance level[31] (i.e., those models for which the Pesaran-Timmermann statistic lies on the right-hand-side critical region).

2. Construction of the qualitative up-down forecast

Using the models selected in Section II.C.1, we now calculate out-of-sample dynamic forecasts for the object forecasting period. Construction follows that of the fan charts above in that forecasting of the four periods from 2000/Q1–Q4 is carried out using

30. For more detail, see Pesaran and Timmermann (1995). We should note that, although the test is based on asymptotic theory, the sample size in the example in this paper is small.

31. Models that clear a left-sided 5 percent significance test may contain important information regarding, for example, inverse correlation. However, since here we are seeking to pick out those models whose directional movements are in step with actual movements in the inflation rate, we only select those models that clear a right-sided 5 percent significance test.

Table 2 Qualitative Up-Down Forecast Based on the Pesaran-Timmermann Test

Percent

		2000/Q1	Q2	Q3	Q4	2001/Q1	Q2	Q3	Q4
(1)	+	66	57	29	12	53	12	42	51
(2)	–	34	43	71	88	47	88	58	49
(3)	Forecast value	+	+	–	–	+	–	–	+
(4)	Actual value	+	–	–	–	+	–	+	–

Note: Out of the group of models for which the Pesaran-Timmermann statistic was on the right-hand-side 5 percent critical region, (1) describes the proportion for which the forecast value was positive (i.e., the inflation rate rose) in that respective period, while (2) describes the proportion for which the forecast value was negative (i.e., the inflation rate fell). By illustrating the case when a larger proportion of models has a positive rather than negative value with a plus sign, and the reverse case with a minus sign, we express in (3) whether a relatively larger number of the forecasts of the inflation rate from the selected models suggest a rise or fall. In (4), for the respective period we look at the sign of the change in the actual value of the inflation rate from the value recorded in the previous period, illustrating a rise with a plus sign, and a fall with a minus sign.

data from an estimation period from 1983/Q1–1999/Q4; while the estimation period for forecasting of the four periods 2001/Q1–Q4 is from 1983/Q1–2000/Q4.

Having produced these forecasts, we then examine, for each period, the proportion of models that produce positive forecasts of the change in the inflation rate relative to those that produce negative forecasts. We note which is greater with the requisite sign.

The results are shown in Table 2, where, comparing the respective signs of the forecasts versus the realized values, we see that rises or falls in the inflation rate are correctly predicted in five out of eight periods. In particular, it is worth noting the consistent accuracy of the qualitative forecast in periods in which there is a conspicuous difference in the proportion of models registering either negative or positive changes[32] (i.e., 2000/Q1, Q3, Q4; 2001/Q2).

While the relative success of this qualitative up-down forecast clearly still requires some verification through many more examples in terms of its practical applicability, it may nevertheless be deemed useful as a supplement to the information provided by the forecast distribution and associated fan charts.

III. Conclusion

In this paper, using forecasting of the inflation rate as an illustration, we presented our HEVAR/FMU as an example of a technique for producing forecasts that not only scores highly on objectivity, but also takes account of uncertainty. The construction process can be outlined as follows. First of all, we built a large number of VAR forecasting models; we then arranged these in order of forecasting performance (in sample), selected the model that performed best for each of our objective forecasting periods, and carried out (out-of-sample) forecasting accordingly. This technique was then supplemented by the production of fan charts that considered the uncertainty

32. In a related point, for the three periods in which the forecast was out, it is worth mentioning that the proportions of models predicting positive versus those predicting negative values were almost the same.

embodied in the forecast errors and indeed inherent in the choice of model itself. We also suggested a way of producing a qualitative up-down forecast.

The three forecast distributions that we produced (the top model distribution, the nonparametric distribution, and the mixed distribution) offer information that may be deemed valuable when thinking quantitatively about the extent of the uncertainty involved in inflation forecasting. In addition, our qualitative up-down forecast roughly follows rises and falls in the actual inflation rate. Although we still need to confirm the practical value of our HEVAR/FMU by carrying out further comparison between forecast and actual values, it is nevertheless fair to say that this HEVAR/FMU both offers valuable forecasting information which would be impossible to extract from a single structural forecasting model, and also can play a useful role in providing a cross-check on the forecasts produced using such structural models.

Without a more comprehensive set of forecast and real values, we cannot judge which of our three forecast distributions is to be preferred.[33] For this reason, at the present time, our HEVAR/FMU comes into its own when we are seeking to evaluate the extent to which a given inflation forecast from some structural model may actually be realized in the future. In such a case, we can put to work both the fan charts, produced using three different techniques, and the results of the qualitative up-down forecast, and in approaching the task of evaluation from several directions at once, demonstrate the value of our HEVAR/FMU as a tool for practical forecasting.

33. In other words, we are unable to rank in any simple way the parametric, nonparametric, and mixed distribution methodologies.

References

Aitchison, John, “Goodness of Prediction Fit,” *Biometrika*, 62 (3), 1975, pp. 547–554.

Akaike, Hirotugu, “On Entropy Maximization Principle,” in P. R. Krishnaiah, ed. *Applications of Statistics*, Amsterdam: North-Holland, 1977, pp. 27–41.

Ban, Kanemi, and Makoto Saito, “Makuro Keiryo Moderu ni yoru Infureritsu Yosoku Gosa no Bunseki (Analysis on the Inflation Rate Forecasting Errors in a Macroeconometric Model),” Bank of Japan Research and Statistics Department Working Paper 01-12, Bank of Japan, 2001 (in Japanese).

Bartlett, Maurice S., “Multivariate Analysis,” *Journal of the Royal Statistical Society*, Series B, 9, 1947, pp. 176–197.

Camba-Mendez, Gonzalo, George Kapetanios, Richard J. Smith, and Martin R. Weale, “Tests of Rank in Reduced Rank Regression Models,” National Institute of Economic and Social Research Discussion Paper No. 150, National Institute of Economic and Social Research, 1999.

Clark, Todd E., “Can Out-of-Sample Forecast Comparisons Help Prevent Overfitting?” *Federal Reserve Bank of Kansas City Research Working Papers*, 00-05, Federal Reserve Bank of Kansas City, 2000.

Clements, Michael P., and David F. Hendry, *Forecasting Economic Time Series*, Cambridge: Cambridge University Press, 1998.

Diebold, Francis X., *Elements of Forecasting*, Cincinnati: South-Western College Publishing, 1998.

Fukuda, Shin-ichi, and Masayuki Keida, “Infure Yosoku ni okeru Jissho Bunseki no Tembo—Phillips Kyokusen no Nihon ni okeru Yosokuryoku wo Chushin ni (Prospects for Empirical Analysis on Inflation Forecasts: The Predictive Power of Phillips Curves in Japan),” Bank of Japan Research and Statistics Department Working Paper 01-21, Bank of Japan, 2001 (in Japanese).

Granger, Clive W. J., and Paul Newbold, *Forecasting Economic Time Series*, San Diego: Academic Press, 1986.

Härdle, Wolfgang, and Oliver Linton, “Applied Nonparametric Methods,” in R. F. Engle and D. L. McFadden, eds. *Handbook of Econometrics*, Vol. IV, Amsterdam: North-Holland, 1994, pp. 2295–2339.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky, “Bayesian Model Averaging,” *Statistical Science,* 14, 1999, pp. 382–401.

Hurvich, Clifford M., and Chih-ling Tsai, “Regression and Time Series Model Selection in Small Samples,” *Biometrika*, 76 (2), 1989, pp. 297–307.

Kasuya, Munehisa, and Kazuhiko Shinki, “Bukka Hendo no Tenkanten Yosoku ni tsuite (Forecasting Turning Points in Price Movements),” Bank of Japan Research and Statistics Department Working Paper 01-20, Bank of Japan, 2001 (in Japanese).

Kitagawa, Genshiro, and Yoshinori Kawasaki, “Jikeiretsu Moderu ni yoru Infureritsu Yosoku Gosa no Bunseki (Analysis on the Inflation Rate Forecasting Errors from a Time-Series Model),” Bank of Japan Research and Statistics Department Working Paper 01-13, Bank of Japan, 2001 (in Japanese).

Krolzig, Hans-Martin, and David F. Hendry, “Computer Automation of General-to-Specific Model Selection Procedures,” *Journal of Economic Dynamics and Control*, 25, 2001, pp. 831–866.

Lütkepohl, Helmut, *Introduction to Multiple Time Series Analysis*, New York: Springer-Verlag, 1991.

Melick, William R., and Charles P. Thomas, “Recovering an Assets Implied PDF from Option Prices: An Application to Crude Oil during the Gulf Crisis,” *Journal of Financial and Quantitative Analysis*, 32 (1), 1997, pp. 91–115.

Pagan, Adrian R., and Aman Ullah, *Nonparametric Econometrics*, Cambridge: Cambridge University Press, 1999.

Pesaran, Hashem M., and Allan Timmermann, “A Simple Nonparametric Test of Predictive Performance,” *Journal of Business and Economic Statistics*, 10 (4), 1992, pp. 561–565.

———, and ———, “Predictability of Stock Returns: Robustness and Economic Significance,” *Journal of Finance*, 50, 1995, pp. 1201–1228.

Silverman, Bernard W., *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall, 1986.

Stock, James H., and Mark W. Watson, "Forecasting Inflation," *Journal of Monetary Economics*, 44, 1999, pp. 293–335.

Velu, Rajabather P., Gregory C. Reinsel, and Dean W. Wichern, "Reduced Rank Models for Multiple Time Series," *Biometrika*, 73 (1), 1986, pp. 105–118.