



BANK OF ENGLAND

CENTRE FOR CENTRAL BANKING STUDIES

ECONOMIC MODELLING AND FORECASTING

**(Conditional) point, range and
density forecasting with VARs**

by

Andy Blake, Haroon Mumtaz and Ole Rummel
Centre for Central Banking Studies
Bank of England

10 February 2015

Contents

1	Introduction	3
2	Forecasting with VAR models	3
2.1	Forecasting when the parameters of the VAR are known	3
2.2	Forecasting when the parameters of the VAR are estimated	4
2.3	Forecasting with deterministic and exogenous variables	5
3	The data	6
4	Unconditional forecasts from a reduced-form VAR model	6
5	Conditional forecasting with a reduced-form VAR model	12
6	Stochastic simulation of a reduced-form VAR model	14
6.1	Steps to performing a stochastic simulation: theory	15
6.2	Steps to performing a stochastic simulation: practice	16
7	Policy scenarios using reduced-form VAR models	22
8	Generating a fan chart around a policy projection generated by a (VAR) model	26
8.1	Drawing a probability density function	30
8.2	Presenting a probability density function	31
8.3	Drawing a fan chart	33
	References and further reading	35

1 Introduction

The vector autoregression (VAR) model is one of the most successful, flexible and easy-to-use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series. The VAR model has been proven to be especially useful for describing the dynamic behaviour of economic and financial time series and for **forecasting**. It often provides superior forecasts to those from univariate time-series models and elaborate theory-based simultaneous equation models. Forecasts from VAR models are quite flexible because they can be made **conditional** on the potential future paths of other specified variables in the model.

Consider the basic VAR(p) model:

$$Y_t = \mu + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t \quad (1)$$

where $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$ is a $(n \times 1)$ vector of $t = 1, \dots, T$ observations on n time-series variables, μ is a $(n \times 1)$ vector of intercepts, A_i is a $(n \times n)$ matrix of regression coefficients for the i -th lag, $i = 1, \dots, p$ (with p being the maximum number of lags) and $\varepsilon_t \sim N(0, \Sigma)$,

We generally spend a lot of time on constructing an adequate model for the data-generating process (DGP) of the VAR in equation (1).¹ Once such a model has been found, it may be used for economic analysis as well as **forecasting**. In fact, forecasting is one of the main objectives of multivariate time-series analysis and this exercise provides an illustration of (conditional) point, range and density forecasting with a VAR in EViews.

2 Forecasting with VAR models

With one exception having to do with EViews, to be discussed in Section 4 below, forecasting using a VAR model is completely analogous to forecasting with a univariate AR model.

2.1 Forecasting when the parameters of the VAR model are known

Consider first the problem of forecasting future values of Y_t when the parameters of the A_i matrices of the VAR(p) process in equation (1) are assumed to be known (and, for simplicity, there are no deterministic terms or exogenous variables).

The best linear predictor, in terms of yielding the minimum mean squared error (MSE), of Y_{t+1} (or one-step-ahead forecast) based on information available at time T is:

$$Y_{T+1|T} = A_1 Y_T + \dots + A_p Y_{T-p+1} \quad (2)$$

Forecasts for longer horizons h (h -step-ahead forecasts) may be obtained using the chain-rule of forecasting. The latter exploits the fact that with autoregressive models, including VARs, the $(h + 1)$ -step-ahead forecast is an easily evaluated function of the h -step-ahead forecast. The h -step-ahead forecast of the VAR given in equation (1) is:

$$Y_{T+h|T} = A_1 Y_{T+h-1|T} + \dots + A_p Y_{T+h-p|T} \quad (3)$$

¹ In a nutshell, by adequate we mean a model that contains enough lags such that the residuals, ε_t , have a zero mean, a fixed variance and cannot be predicted from past values of Y_t , i.e., there is no residual serial correlation.

where $Y_{T+j|T} = Y_{T+j}$ for $j \leq 0$. The corresponding h -step-ahead forecast errors therefore are given by:

$$Y_{t+h} - Y_{T+h|T} = \varepsilon_{T+h} + \Psi_1 \varepsilon_{T+h-1} + \dots + \Psi_{h-1} \varepsilon_{T+1}$$

which may be expressed as:

$$Y_{T+h} - Y_{T+h|T} = \sum_{s=0}^{h-1} \Psi_s \varepsilon_{T+h-s} \quad (4)$$

where the matrices Ψ_s are determined by recursive substitution:

$$\Psi_s = \sum_{j=1}^s \Psi_{s-j} A_j \quad s = 1, 2, \dots \quad (5)$$

with $\Psi_0 = I_n$ and $A_j = 0$ for $j > p$. The forecasts are unbiased since all of the forecast errors have expectation zero and the MSE variance-covariance matrix for $Y_{T+h|T}$ is:

$$\Sigma(h) = E[(Y_{T+h} - Y_{T+h|T})(Y_{T+h} - Y_{T+h|T})'] = \sum_{s=0}^{h-1} \Psi_s \Sigma \Psi_s' \quad (6)$$

where Σ is the time-invariant variance-covariance matrix of the VAR in equation (1).

2.2 Forecasting when the parameters of the VAR model are estimated

Now consider the forecast Y_{T+h} when the parameters of the VAR(p) process in equation (1) are estimated using OLS, rather than known *a priori*. This is, of course, the situation we normally find ourselves in. Note that in this case, forecasts are based on conditional expectations assuming an independent white-noise process for the ε_t in equation (1). This is a strong assumption to make in many circumstances and we will return to this point below. The best linear predictor of Y_{T+h} is now:

$$\hat{Y}_{T+h|T} = \hat{A}_1 \hat{Y}_{T+h-1|T} + \dots + \hat{A}_p \hat{Y}_{T+h-p|T} \quad (7)$$

where the \hat{A}_i 's ($i = 1, \dots, p$) are estimated parameter matrices and $\hat{Y}_{t+j|T} = Y_{T+j}$ for $j \leq 0$. The h -step-ahead forecast error will be given by:

$$Y_{T+h} - \hat{Y}_{T+h|T} = (Y_{T+h} - Y_{T+h|T}) + (Y_{T+h|T} - \hat{Y}_{T+h|T}) = \sum_{j=0}^{h-1} \Psi_j \varepsilon_{T+h-j} + (Y_{T+h} - \hat{Y}_{T+h|T}) \quad (8)$$

The first term on the right-hand side involves future residuals, while the second term, $(Y_{T+h} - \hat{Y}_{T+h|T})$, is determined by present and past values (if only past variables have been used for estimation) and captures the part of the forecast error due to estimating the

parameters of the VAR. It follows that the two terms on the right-hand side of equation (8) are independent if ε_t is independent white noise. The MSE variance-covariance matrix of the h -step-ahead forecast is then:

$$\hat{\Sigma}(h) = \Sigma(h) + \text{MSE}(Y_{T+h} - \hat{Y}_{T+h|T}) \quad (9)$$

The variance-covariance matrix in the case of estimated VAR model parameters consists of two parts: the original MSE variance-covariance matrix given by equation (6) plus an additional term that accounts for the estimation uncertainty. This term can get quite complicated. In many – if not most – practical applications, the second term $\text{MSE}(Y_{T+h} - \hat{Y}_{T+h|T})$ is therefore often ignored and $\hat{\Sigma}(h)$ is computed using:

$$\hat{\Sigma}(h) = \sum_{s=0}^{h-1} \hat{\Psi}_s \hat{\Sigma} \hat{\Psi}_s' \quad (10)$$

with $\hat{\Psi}_s = \sum_{j=1}^s \hat{\Psi}_{s-j} \hat{A}_j$. The reason for this is that the difference $(Y_{T+h} - \hat{Y}_{T+h|T})$ is small in probability as the sample size used for estimation gets large and the VAR coefficients are estimated more and more precisely. Thus, as far as the forecast MSE is concerned, estimation uncertainty may be ignored in large samples. In small samples, it may be preferable to include a correction term. Lütkepohl (2005, Chapter 3) gives an approximation to $\text{MSE}(Y_{T+h} - \hat{Y}_{T+h|T})$ which may be interpreted as a finite-sample correction to equation (9).

Assuming normally distributed disturbances, the above results can be used for setting up asymptotic $(1 - \alpha) \cdot 100$ per cent confidence intervals for any linear combination of these forecasts. For the individual elements of $\hat{Y}_{T+h|T}$, forecast intervals are then computed as:

$$[\hat{Y}_{k,T+h|T} - c_{1-\alpha/2} \hat{\sigma}_k(h), \hat{Y}_{k,T+h|T} + c_{1-\alpha/2} \hat{\sigma}_k(h)] \quad (11)$$

where $c_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution, $\hat{Y}_{k,T+h|T}$ denotes the k -th component of $\hat{Y}_{T+h|T}$ and $\hat{\sigma}_k(h)$ is the square root of the k -th diagonal element of $\hat{\Sigma}(h)$.

Finally, we note that if any of the variables in Y_t is in first-differenced form, a forecast of the variable in levels rather than first differences can be obtained by integration, i.e., if $Y_t = \Delta z_t = z_t - z_{t-1}$, then $z_{T+h|T} = Y_{T+h|T} + \dots + Y_{T+1|T} + z_T$.

2.3 Forecasting with deterministic terms and exogenous variables

If deterministic or exogenous variables, or both, are present, we can extend the usual VAR forecasting formulae to allow for such terms.

Deterministic variables need to be available for each forecast period $T + 1, \dots, T + h$. Because the future development of deterministic variables is known by their nature (variables such as (linear) time trends, intercepts and seasonal dummies are, after all, deterministic), they are particularly easy to handle. It is frequently possible to simply

extrapolate future values from the actual deterministic series. Such extrapolation works well for trend, intercept and seasonal dummies.

Exogenous variables, which will also need to be available for each forecast period $T + 1, \dots, T + h$, are somewhat more difficult to deal with. These variables enter the model in the guise of a VAR with exogenous variables, frequently referred to as a VARX model. If their future development were known, they would also be easy to handle. Otherwise they have to be predicted alongside the endogenous variables in the VAR, in which case a model for their DGP is required. Alternatively, if the exogenous variables are under the full control of a policymaker, it may be desirable to forecast the endogenous variables conditionally on a specific future path of the exogenous variables to check the future implications of their specific values.

3 The data

The file **question4.wf1** contains quarterly data on UK GDP growth (`dgdgdp`), UK inflation (`inf`), the Bank of England's Bank rate (`bankrate`) and house price inflation (`dhp`) over the period from 1947 Q1 to 2010 Q4 (although not all series have data over such a long period). To open the EViews workfile from within EViews, choose **File, Open, EViews Workfile...**, select **question4.wf1** from the appropriate folder and click on **Open**. Alternatively, you can double-click on the workfile icon outside of EViews, which will open EViews automatically.

A key consideration before any estimation can be attempted is the form the variables must have when they enter the VAR: should they enter in levels, gaps or first differences? The answer is simply 'it depends'. It depends on what the VAR is going to be used for. For forecasting purposes, we must avoid potential spurious regressions that may result in spurious forecasts. In other words, if we are interested in using the estimated VAR (or VECM) model for forecasting, say, we need to ensure that all variables are either stationary or cointegrated to avoid the spurious regression problem associated with unit roots. For that reason, UK GDP appears as a first difference (i.e., as a growth rate), Bank rate appears in levels (interest rate should be stationary – why?) and house prices appear in first differences again.

4 Unconditional forecasts from a reduced-form VAR model

In contrast to other instances in EViews, forecasts from a VAR (or VECM) are not available from within the estimated VAR object, which means that there is no convenient **Forecast** button that appears in the object window (as it does after estimation using OLS, say). Instead, forecasts have to be obtained by solving a **Model**, which needs to be created from the estimated VAR/VECM.

Question 1: Estimate an **illustrative** VAR(2) using GDP growth, inflation and Bank rate using the sample until 2010 Q4. Using the estimated model, construct a forecast for the three variables over the period from 2011 Q1 to 2011 Q4.

Use **Quick, Estimate VAR...**, select `dgdgdp`, `inf` and `bankrate` in that order as endogenous variables and select 1 2 for the lag interval. An equivalent way of getting EViews to do the estimation is to type the following command in the command window:

```
var var2.ls 1 2 dgdp inf bankrate
```

This command specifies a VAR called **var2** with a lag length of two. The estimation results are given in Table 1.

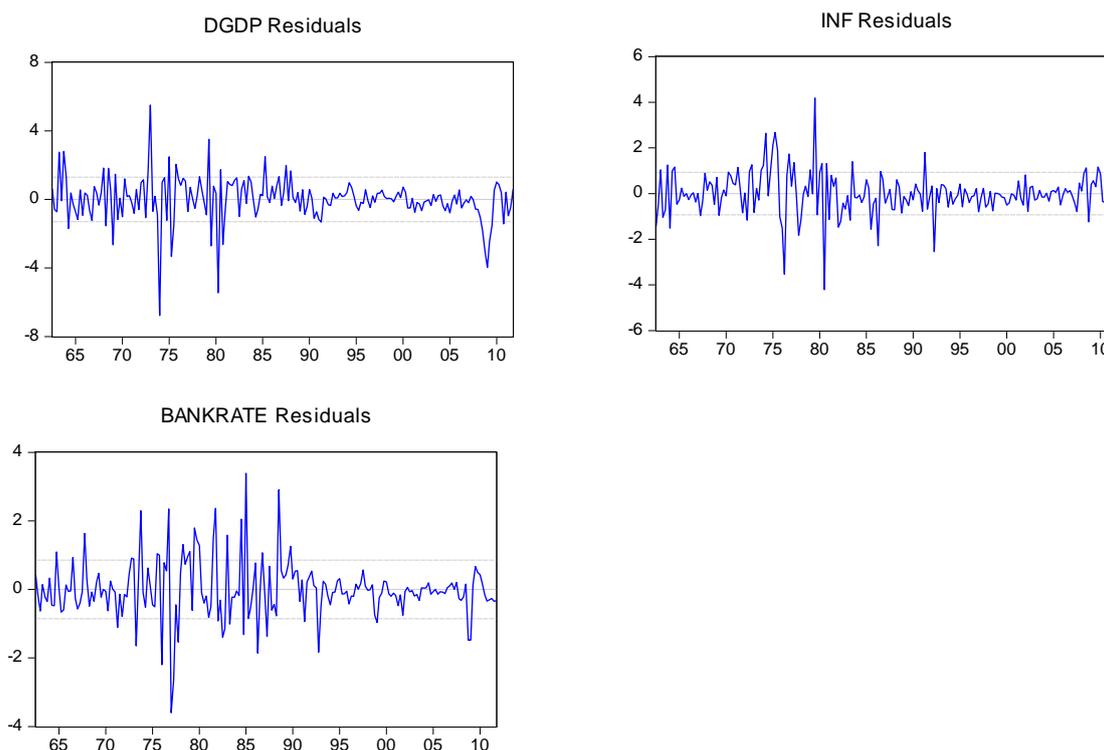
Table 1: VAR(2) estimates, 1962 Q3 – 2010 Q4

Vector Autoregression Estimates
Sample (adjusted): 1962Q3 2010Q4
Included observations: 194 after adjustments
Standard errors in () & t-statistics in []

	DGDP	INF	BANKRATE
DGDP(-1)	0.829407 (0.07605) [10.9066]	0.120573 (0.05501) [2.19173]	0.046216 (0.05041) [0.91681]
DGDP(-2)	-0.038321 (0.07672) [-0.49947]	-0.094153 (0.05550) [-1.69640]	0.057731 (0.05086) [1.13516]
INF(-1)	-0.027178 (0.09092) [-0.29893]	1.409460 (0.06577) [21.4294]	-0.053903 (0.06027) [-0.89437]
INF(-2)	0.012357 (0.09102) [0.13576]	-0.461040 (0.06585) [-7.00182]	0.088695 (0.06034) [1.47001]
BANKRATE(-1)	-0.051485 (0.10737) [-0.47951]	0.254850 (0.07767) [3.28108]	1.226133 (0.07117) [17.2274]
BANKRATE(-2)	0.006879 (0.10888) [0.06318]	-0.214341 (0.07876) [-2.72135]	-0.286326 (0.07217) [-3.96726]
C	0.911070 (0.26812) [3.39801]	-0.094855 (0.19396) [-0.48905]	0.029694 (0.17773) [0.16707]
R-squared	0.675526	0.962159	0.944343
Adj. R-squared	0.665115	0.960945	0.942558
Sum sq. resids	318.7162	166.7922	140.0475
S.E. equation	1.305513	0.944424	0.865400
F-statistic	64.88625	792.4587	528.8158
Log likelihood	-323.4290	-260.6164	-243.6641
Akaike AIC	3.406485	2.758932	2.584166
Schwarz SC	3.524397	2.876845	2.702078
Mean dependent	2.285966	5.505240	7.816011
S.D. dependent	2.255971	4.778910	3.610781
Determinant resid covariance (dof adj.)		1.007803	
Determinant resid covariance		0.902600	
Log likelihood		-815.8820	
Akaike information criterion		8.627650	
Schwarz criterion		8.981387	

A quick visual check of the residuals (**View, Residuals, Graphs**) reveals a very different behaviour of the residuals between the first and the second half of the sample, as illustrated in Figure 1. Residuals seem much more volatile in the first half, with some large outliers, while they appear much less volatile in the second half of the sample. Nevertheless, `dgdp` and `bankrate` in particular display some very large residuals at the end of the sample period.

Figure 1: Estimated residual series from the reduced-form VAR(2) model



Indeed, summary statistics on the third and fourth moments of these series, which are then combined in the Jarque-Bera test for normality, reject the null hypothesis of an underlying normal distribution (this is left as an exercise).² Looking at the individual statistics, we find that it is excessive kurtosis in particular that underlies the rejection of the null hypothesis of normality.

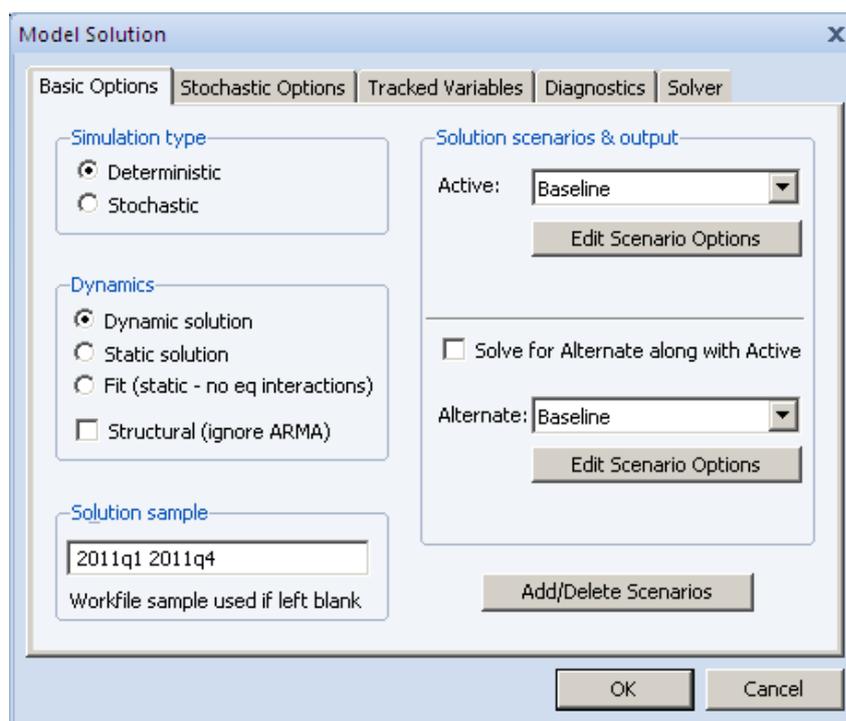
We should be wary of this fact in the stochastic simulations later on in this exercise for two reasons. One, these large residuals may cause a lot of variation in the three variables in the stochastic simulations, potentially leading to quite wide error bounds on the point forecasts and hence also the intervals in the range forecast. Two, EViews offers two different approaches to generating the innovations for the stochastic simulations. As we will see below, if the method is set to **Normal Random Numbers**, the innovations will be generated by drawing a set of random numbers from the standard normal distribution. If, on the other hand, the method is set to **Bootstrap**, the innovations will be generated by drawing randomly (with replacement) from the set of actual innovations observed within a specified sample.

² Hint: **Proc, Make Residuals, View, Descriptive Stats, Common Sample.**

Using bootstrapped innovations may be more appropriate than normal random numbers in cases where the equation innovations do not seem to follow a normal distribution, for example if the innovations appear asymmetric or appear to contain more outlying values than a normal distribution would suggest. This is most likely the case in the model at hand. Note, however, that a set of bootstrapped innovations drawn from a small sample may provide only a rough approximation to the true underlying distribution of the innovations.

Once the VAR(2) has been estimated, click on **Proc, Make Model** from the VAR window toolbar to create an EViews model object from the estimated VAR/VECM. A model in EViews is a set of one or more equations that jointly describe the relationship between a set of variables. Overall, EViews models allow us to combine equations inside a single object, which may then be used to create a deterministic or stochastic joint forecast or simulation for all the variables in the model.³ To keep the model permanently in the workfile, we name the model by clicking on the **Name** button, entering a name such as model01 and clicking **OK**.

Our model specification is now complete. To construct the unconditional forecast (within the **Model** object), either go to **Proc, Solve Model...** or click on the **Solve** button on the model toolbar. EViews will display a tabbed dialog containing the solution options.⁴ For the current forecasting exercise, we will choose the following options: on the **Basic Option** tab, select **Deterministic** under **Simulation type**, **Dynamic solution** in the **Dynamics** box and select 2011q1 2011q4 as the **Solution sample**. Leave everything else unchanged and click on **OK** to start solving, i.e., forecasting, the model. All the required options are shown in the screenshot below.



³ In a deterministic setting, the inputs to the model are fixed at known values and a single path is calculated for the output variables. In a stochastic environment, uncertainty is incorporated into the model by adding a random element to the coefficients, the equation residuals or the exogenous variables.

⁴ This **Basic Option** page contains the most important options for the simulations. While the options on other tabs, such as **Stochastic Options**, can often be left at their default values, the options on this page will definitely need to be set appropriately for the task we are trying to perform.

In a **deterministic** simulation, all equations in the model are solved so that they hold without error during the solution sample, all coefficients are held fixed at their point estimates and all exogenous variables, if present, are held constant at their historical values in the sample. This results in a **single** path for the endogenous variables which can be evaluated by solving the model once.

Options in the **Dynamics** box determine how EViews uses historical data for the endogenous variables when solving the model:

- when **Dynamic solution** is chosen, only values of the endogenous variables preceding the solution sample are used when forming the forecast – a dynamic solution is typically the correct method to use when forecasting values several periods into the future (a multi-step-ahead forecast) or for evaluating how a multi-step-ahead forecast would have performed historically; and
- when **Static solution** is chosen, values of the endogenous variables up to the previous period are used each time the model is solved – a static solution is typically used to produce a set of one-step-ahead forecasts over the historical data to examine the historical fit of the model within the sample period.⁵

Once the calculations are finished, the model window will switch to the **Solutions Messages** view, thereby indicating that the model has been solved successfully. If that should not be the case, an error message will appear in this view, which may (or may not) be helpful in indicating what has gone wrong in the model solution.

Now that we have solved the model, we would like to look at the results. We find that EViews has created three new series for us: `dgdp_0`, `inf_0` and `bankrate_0`, which contain the forecasts from the baseline scenario, i.e., the VAR(2) we have just estimated. Since the newly created series are standard EViews objects, we could use the workfile window to open the series and examine them directly. But the model object provides a much more convenient way to work with the series through a view called the **Variable View**. The easiest way to switch to the **Variable View** is to click on the button labelled **Variables** on the model window button bar.

The **Variable View** brings up a window that lists all the available variables. For example, to look at a graph containing the actual and forecast values for `bankrate`, say, simply select that variable (you can select more than one variable by holding down the control key clicking on the variable names) and then clicking on **Proc, Make Graph...** to enter the dialog window. The dialog window has many options, but for our current purposes, we can leave most settings at their default values. Simply make sure that the **Active: Baseline** and **Compare: Actuals** checkboxes are checked, set the sample for the graph to 2000 Q1 to 2011 Q4 and click **OK**. Once again, the required options are illustrated in the following screenshot.

⁵ As such, a static solution cannot be used to predict more than one observation into the future at a time.

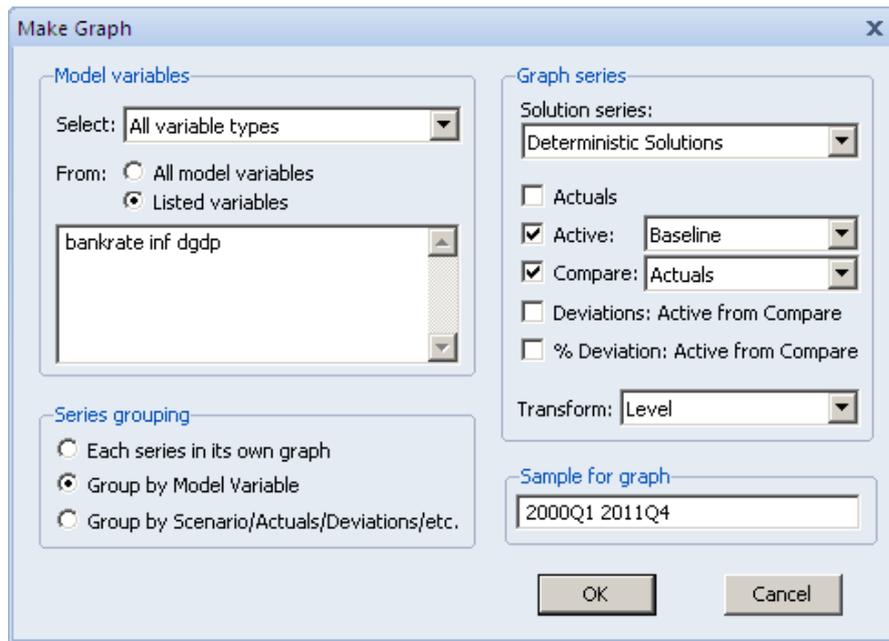
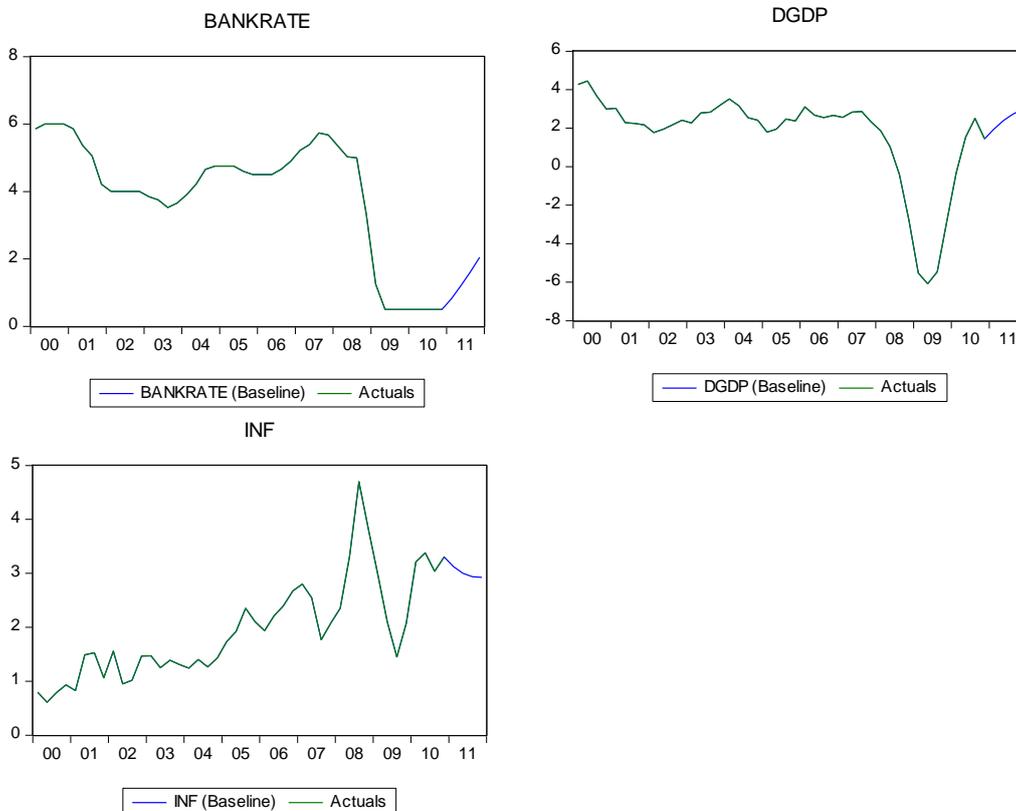


Figure 2 below shows the history of bankrate, dgdg and inf starting in 2000 Q1 as well as the forecasts for these three variables based on the estimated VAR(2) over the forecast period from 2011 Q1 to 2011 Q4.

Figure 2: Actual and forecasted values for bankrate, dgdg and inf from the reduced-form VAR (2) model, 2000 Q1 - 2011 Q4



As illustrated in Figure 2, `bankrate` is forecast to increase from 0.5 to 2 per cent over the course of 2011. GDP growth is forecast to increase to almost 3 per cent by 2011 Q4, while inflation falls slightly from 3.3 to 2.9 per cent over the course of 2011. Is this a good forecast?

5 Conditional forecasting with a reduced-form VAR model

The forecast algorithms considered up to now are unconditional multivariate forecasts. But it is sometimes desirable to obtain forecasts of some variables in the system conditional on knowledge of the future path of other variables in the system. For example, when forecasting multivariate macroeconomic variables using quarterly data from a VAR model, it may happen that some of the future values of certain variables in the VAR model are known, because data on these variables are released earlier than data on the other variables. By incorporating the knowledge of the future path of certain variables, in principle it should be possible to obtain more reliable forecasts of the other variables in the system.

When working with a model, we will often want to compare model predictions under a variety of different assumptions regarding the path of the (exogenous) variables, or with one or more of the equations excluded from the model. This is done using model **scenarios**, which allow us to do these operations without overwriting previous data or changing the structure of the model. The most important function of a scenario is to specify which series will be used to hold the data associated with a particular solution of the model.

Question 2: Assume that Bank rate will be fixed at 0.5 per cent over 2011. Re-calculate the forecast using this **conditioning** assumption.

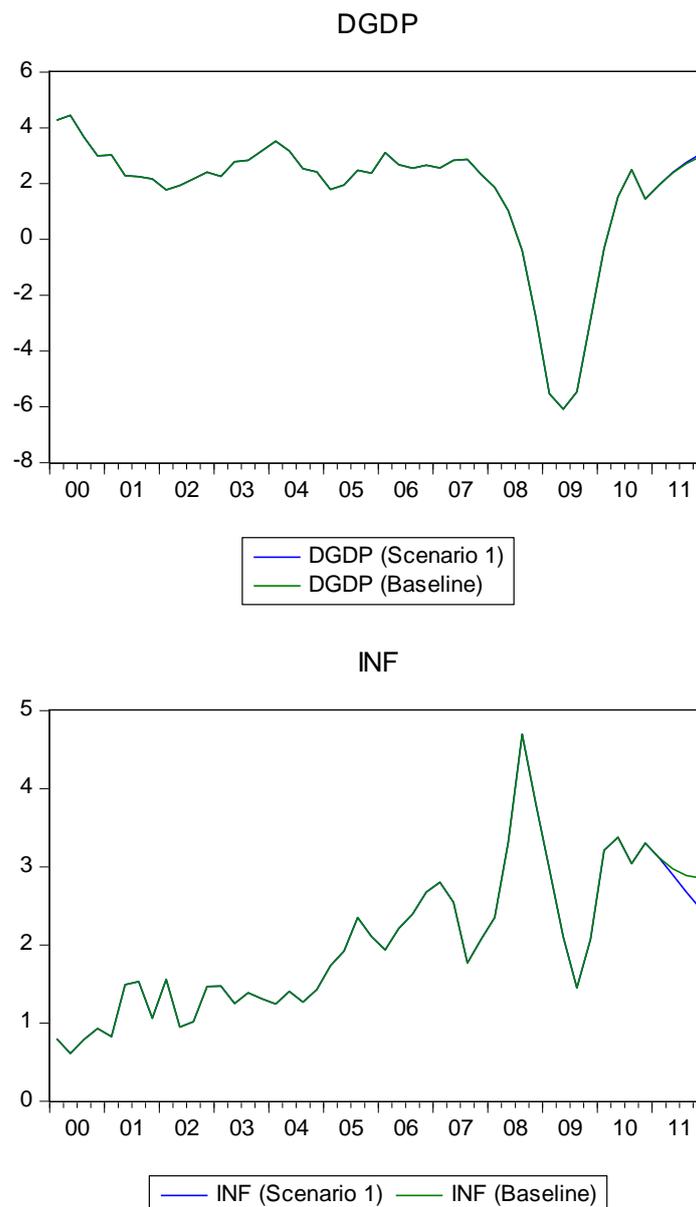
To condition on the Bank of England's Bank rate staying at 0.5 per cent over the forecast sample period, we need to manually change the value of `bankrate` to 0.5 over 2011 in the underlying data series in the workfile. The easiest way of making changes to the data associated with an equation or a system of equations (model) is to open the series or group spreadsheet window containing the data and then edit the data by hand. In other words, we activate the workfile window by clicking on it and then open the series `bankrate` by double-clicking on its icon. We click on **Edit +/-** to be able to manipulate individual cells in the spreadsheet and manually enter 0.5 into each of the four cells for 2011. We then click **Edit +/-** again to 'lock' the spreadsheet and close the series window.

We are now ready to solve the model again. Click on **Solve** to bring up the **Model Solution** object. On the **Basic Options** tab, go to **Solution scenarios & output** and choose **Active: Scenario 1**. Then select **Edit Scenario Options** and click on the **Excludes** tab and type `bankrate` in the window. This operation will keep `bankrate` exogenously fixed at its actual data values in the workfile over the forecast period, i.e., 2011. Click **OK** to return to the **Basic Options** tab. We now solve the model again over the period 2011 Q1 2011 Q4 by clicking on **Solve** in the model object. This new scenario creates the related forecast variables `inf_1`, `dgdgdp_1` and `bankrate_1`.⁶

⁶ To distinguish the data associated with different scenarios, each scenario modifies the names of the model variables. Typically, this will involve adding an underscore followed by a number such as `_0` or `_1` to the variable names of the model. The data for each scenario will then be contained in series in the workfile with the adjusted names.

In essence, this set-up excludes the endogenous variable `bankrate` from the model. When an endogenous variable is excluded, as in this case, the equation associated with that variable is dropped from the model and the values of `bankrate` are taken directly from the workfile series with the same name.⁷ Results for this conditional forecasting exercise can be seen in Figure 3. We once again use **Variable View**. This time, though, we only select `dgdg` and `inf`. We then click on **Proc, Make Graph...** to enter the dialog window. On this occasion, we select **Active: Scenario 1** and **Compare: Baseline**. We once again set the sample for the graph to 2000 Q1 to 2011 Q4 and click **OK**.

Figure 3: Actual and forecasted values for `dgdg` and `inf` from the reduced-form VAR(2) model, 2000 Q1-2010 Q4



⁷ In that sense, excluding an endogenous variable effectively treats the variable as an exogenous variable for the purposes of solving the model.

As shown in Figure 3, GDP growth, $dgdp$, is more or less unchanged across the two scenarios. In both cases, i.e., with $bankrate$ allowed to respond to the other variables in the model and with $bankrate$ fixed over 2011, $dgdp$ returns to its unconditional mean in the sample.

The response of inflation, on the other hand, is somewhat different, with inflation in Scenario 1 falling faster and further than in the baseline scenario.

Do either of these responses make much sense? We would normally expect lower interest rates for longer to boost demand and thus GDP growth, while also having a positive effect on inflation. Instead, we find that another year of $bankrate$ remaining at 0.5 per cent has no discernible positive impact on $dgdp$, while inflation is actually **falling** with interest rates staying fixed at their lower level. This decrease in inflation is faster and further than in the baseline scenario of interest rates increasing to 2 per cent at the end of 2011.

6 Stochastic simulation of a reduced-form VAR

So far, we have been working under the assumption that our stochastic equations in the VAR hold exactly over the forecast period. The (dynamic) solution considered so far is based on a single drawing of the random shocks, ε_{it} ($i = 1, \dots, n$, where n is the size of the VAR). A new drawing will generate a different solution path. In reality, we would expect to see the same sort of errors occurring in the future as we have seen in history. We have also been ignoring the fact that the coefficients in our equations are estimated rather than fixed at known values. We may therefore like to reflect this uncertainty about our coefficients in some way in the results from our model. The most convenient way of doing so is to use **stochastic simulations**.

As described in more detail in Section 6.1 below, stochastic simulation involves solving the model a (large) number of times (called the number of replications), each time adding a pseudo-random shock to each equation and drawing a new set of shocks for each replication. Consider taking repeated drawings of ε_t in the VAR model, assuming a particular distribution with zero mean and variance σ^2 . In a stochastic simulation, the equations of the model are solved so that they have residuals which match randomly drawn errors and, optionally, the coefficients and exogenous variables of the model are also varied randomly (to account for coefficient uncertainty).⁸ The model solution therefore generates a distribution of outcomes for the endogenous variables in every period. We approximate the distribution by solving the model many times using different draws for the random components in the model and then calculating summary statistics over all the different outcomes.

The average of these drawings will be a consistent estimate of the expected value of the model variables and higher-order moments such as variances, skew and kurtosis can easily be computed. If desired, quantiles of the distribution can also be computed and these can be displayed graphically in the form of a fanchart. This procedure is known as **stochastic simulation** (Blake (1996)) and increasing the number of drawings or **replications** increases its precision. The advantage of stochastic simulation is that, by drawing from the distribution of the estimated residuals, the bands of the fanchart reflect the uncertainty inherent in the estimated model.

⁸ The random errors are chosen such that their average value is zero, which is the appropriate assumption if we have no information as to the errors in our stochastic equations.

6.1 Steps to performing a stochastic simulation: theory

In general, simulation refers to any method that generates repeated replications or trials of a model and seeks to summarise uncertainty in the model forecast in terms of the distribution of these random trials.

For illustrative purposes only, assume that we are using the following simple AR(1) regression model:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t \quad (12)$$

on an observed sample of data $\{y_1, y_2, \dots, y_T\}$.

The steps to performing a stochastic simulation are then as follows.

Step 1: Estimate the model

We estimate the AR(1) model given by equation (12) and store the parameter estimates $\hat{\phi}_0$ and $\hat{\phi}_1$.

Step 2: Solve the model

For each available time period t in the model, use $\hat{\phi}_0$ and $\hat{\phi}_1$ to generate the one-step-ahead forecasts:

$$\hat{y}_{t+1} = \hat{\phi}_0 + \hat{\phi}_1 y_t$$

and then compute and store the one-step-ahead forecast errors:

$$\hat{\varepsilon}_{t+1|t} = \hat{y}_{t+1} - y_{t+1}$$

Step 3: Simulate the model

We then forecast the model forward from time T , but instead of a forecast based solely on the best guesses for the unknowns, the uncertainty is explicitly accounted for by including an error term. The error term is obtained either by drawing from some known parametric distribution (such as the normal distribution) or by taking random draws (with replacement) from the estimated one-step-ahead forecast errors:

$$\begin{aligned} \hat{y}_{T+1}^1 &= \phi_0 + \phi_1 y_T + \tilde{\varepsilon}_{T+1} \\ \hat{y}_{T+2}^1 &= \phi_0 + \phi_1 \hat{y}_{T+1} + \tilde{\varepsilon}_{T+2} \\ &\vdots \\ \hat{y}_{T+h}^1 &= \phi_0 + \phi_1 \hat{y}_{T+h-1} + \tilde{\varepsilon}_{T+h} \end{aligned}$$

where the $\tilde{\varepsilon}_{T+i}$ are all random drawings (with replacement) from the $\hat{\varepsilon}_{t+1|t}$, i.e., the computed one-step-ahead forecast errors from **Step 2**. The series of forecasts $\{\hat{y}_{T+1}^1, \hat{y}_{T+2}^1, \dots, \hat{y}_{T+h}^1\}$ represents **one** repetition of a Monte Carlo simulation of the model.

Step 4: Repeat Step 3

Step 3 is now repeated S times to obtain a set of forecasts:

$$\begin{array}{cccccc}
 \hat{y}_{T+1}^1 & \hat{y}_{T+1}^2 & \hat{y}_{T+1}^3 & \dots & \hat{y}_{T+1}^{S-1} & \hat{y}_{T+1}^S \\
 \hat{y}_{T+2}^1 & \hat{y}_{T+2}^2 & \hat{y}_{T+2}^3 & \dots & \hat{y}_{T+2}^{S-1} & \hat{y}_{T+2}^S \\
 \vdots & \vdots & \vdots & & \vdots & \vdots \\
 \hat{y}_{T+h}^1 & \hat{y}_{T+h}^2 & \hat{y}_{T+h}^3 & \dots & \hat{y}_{T+h}^{S-1} & \hat{y}_{T+h}^S
 \end{array}$$

Step 5: Summarise the uncertainty

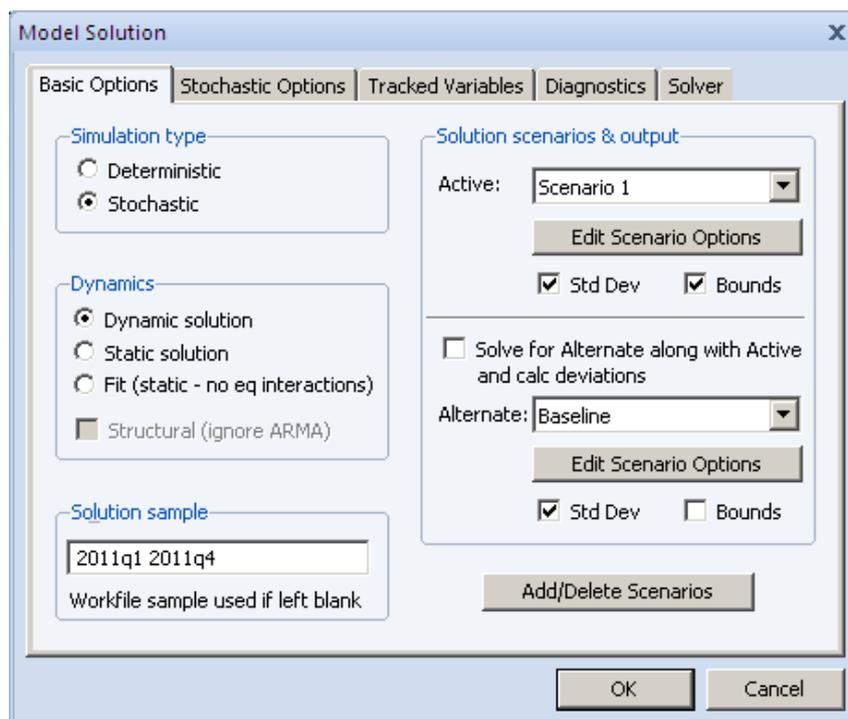
Each column of this set of forecasts is one representation of a possible outcome of the model and therefore the entire set collectively captures the uncertainty of the forecast. In particular, the percentiles of these simulated forecasts in each row – or alternatively, for each forecast period $T + i$ – give an accurate picture of the empirical distribution of the forecasts at that specific forecast horizon. Note that if the disturbances used to generate the forecasts are drawn from the actual one-step-ahead prediction errors and not from a known parametric distribution such as the normal, the forecast uncertainty will reflect any non-symmetry or fat tails present in the estimated prediction errors.

6.2 Steps to performing a stochastic simulation: practice

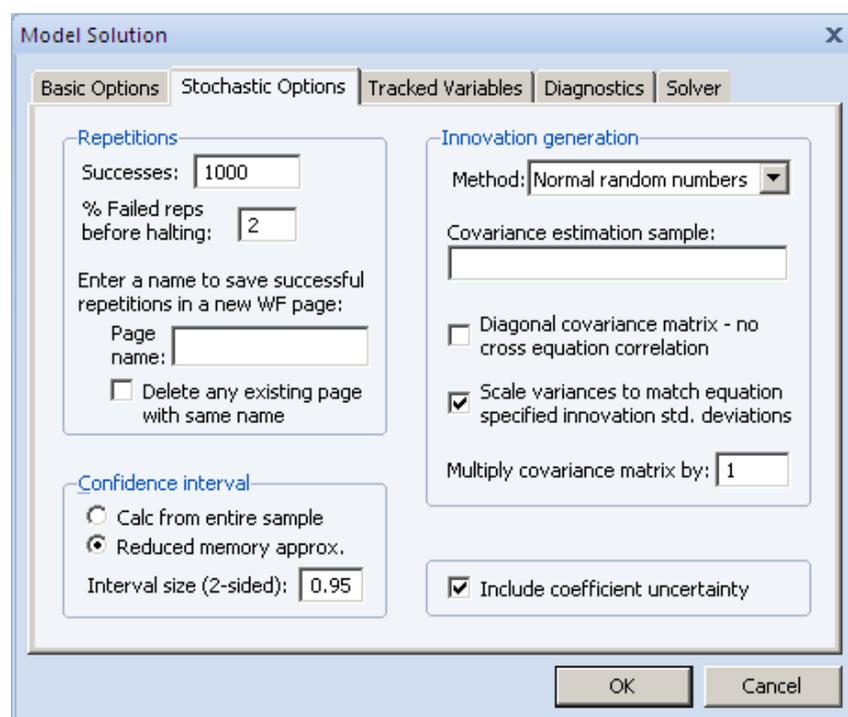
The previous question showed how to generate multivariate forecasts from a fitted VAR model. Since the multivariate forecast errors given by equation (8) are assumed to be asymptotically normally distributed with variance-covariance matrix given by equation (9), the forecasts Y_{T+h} can be simulated by generating multivariate normal random variables with mean zero and variance-covariance matrix given by equation (9). These simulation-based forecasts can be obtained in the following way.

Question 3: Repeat **Question 2** by using a stochastic simulation to calculate the confidence intervals of the forecasts for `dgdgdp` and `inf`.

In case of our VAR(2) with the conditioning assumption that the Bank of England policy interest rate is remains at 0.5 per cent over the course of 2011 (which is still in the data, as we have not removed the values for 2011 from the actual series), we can use a stochastic simulation to provide some measure of the uncertainty of our results by adding error bounds to our predictions. In the model window, click on the **Solve** button. When the **Model Solution** dialog appears, choose **Stochastic** for the **Simulation type** and retain **Dynamic solution** under **Dynamics** and 2011Q1 2011Q4 for the **Solution sample**. In the **Solutions scenarios & output** box, retain (or choose) **Active: Scenario 1** and make sure that either the **St. Dev.** or the **Bounds** (or both) checkbox is checked. The required options are illustrated in the following screenshot.



The **Stochastic Options** dialog box requires a bit of interaction, so it is worth investigating the different options in more detail.



The **Repetitions** box in the top left corner of the dialog allows you to set the number of repetitions that will be performed during the stochastic simulation. A higher number of repetitions will reduce the sampling variation in the statistics being calculated, but will take

more time. The default value of 1000 repetitions is generally adequate to get a good idea of the underlying values, although there may still be some random variation visible between adjacent observations.

Also in the repetitions box is a field labeled **% Failed reps before halting**. Failed repetitions typically result from random errors driving the model into a region in which it is not defined, for example where the model is forced to take the log or square root of a negative number. When a repetition fails, EViews will discard any partial results from that repetition and then check whether the total number of failures exceeds the threshold set in the **% Failed reps before halting** box. The simulation continues until either this threshold is exceeded or the target number of successful repetitions is met.

Note, however, that even one failed repetition indicates that care should be taken when interpreting the simulation results, since it indicates that the model is ill-defined for some possible draws of the random components. Simply discarding these extreme values may create misleading results, particularly when the tails of the distribution are used to measure the error bounds of the system.

We can see that the **Scale variances to match equation specified std. deviations** box is checked by default. This option lets you determine how the variances of the residuals in the equations are determined. If the box is not checked, the variances are calculated from the model equation residuals. If the box is checked, then any equation that contains a specified standard deviation will use that number instead. Note that the sample used for estimation in a linked equation may differ from the sample used when estimating the variances of the model residuals.

In addition, note that there is also an option to add coefficient uncertainty, which allows us to consider the implications of using estimated VARs for the precision of the forecast. EViews does not appear to apply Lütkepohl's (1991) finite-sample correction to the forecast standard errors computed from asymptotic theory. Instead, the **Include coefficient uncertainty** field at the bottom right of the **Stochastic Options** dialog specifies whether or not estimated coefficients in linked equations should be varied randomly during a stochastic simulation. When this option is selected, coefficients are randomly redrawn at the beginning of each repetition, using the coefficient variability in the estimated equation, if possible. During the repetition, errors are generated for each observation in accordance with the residual uncertainty and the exogenous variable uncertainty in the model. At the end of each repetition, the statistics for the tracked endogenous variables are updated to reflect the additional results. This technique provides a method of incorporating uncertainty surrounding the true values of the coefficients into variation in our forecast results.

The latter step is necessary since the coefficients of the VAR(2) process are not known with certainty but have been estimated. We have to take account of this fact when calculating forecast uncertainty bounds, which will have to reflect both the uncertainty coming from the initial estimation of the VAR coefficients and the uncertainty coming from the forecasting procedure. While estimation uncertainty may be ignored in large samples, it may be preferable to include a correction term in small samples. Further details on possible correction terms can be found in Lütkepohl (2005, Chapter 3) for the stationary case and Basu and Sen Roy (1987), Engle and Yoo (1987) and Reimers (1991) for non-stationary processes.

But the key selection in this regard is the **Innovation generation** method, of which there are two.⁹ As mentioned above, if the method is set to **Normal Random Numbers**, the

⁹ Ideally, we should be using an antithetic shock generation method. Antithetic shocks are pairs of shocks with the same magnitude but opposite sign. The use of antithetic stochastic shocks not only forces the distribution of

innovations will be generated by drawing a set of random numbers from the standard normal distribution. If on the other hand the method is set to **Bootstrap**, the innovations will be generated by drawing randomly (with replacement) from the set of actual innovations observed within a specified sample. Asymptotically, bootstrapped shocks will have the same distribution as the empirical distribution of the single equation residuals. Furthermore, and much more important, this method does not assume normality. Using bootstrapped innovations may therefore be more appropriate than normal random numbers in cases where the equation innovations do not seem to follow a normal distribution, for example if the innovations appear asymmetric or appear to contain more outlying values than a normal distribution would suggest. This is most likely the case in the model at hand. Note, however, that a set of bootstrapped innovations drawn from a small sample may provide only a rough approximation to the true underlying distribution of the innovations.

For the time being, we will choose **Method: Normal random numbers** for the **Innovation generation** under the **Stochastic Options** tab. Click on **OK** to begin the simulation.

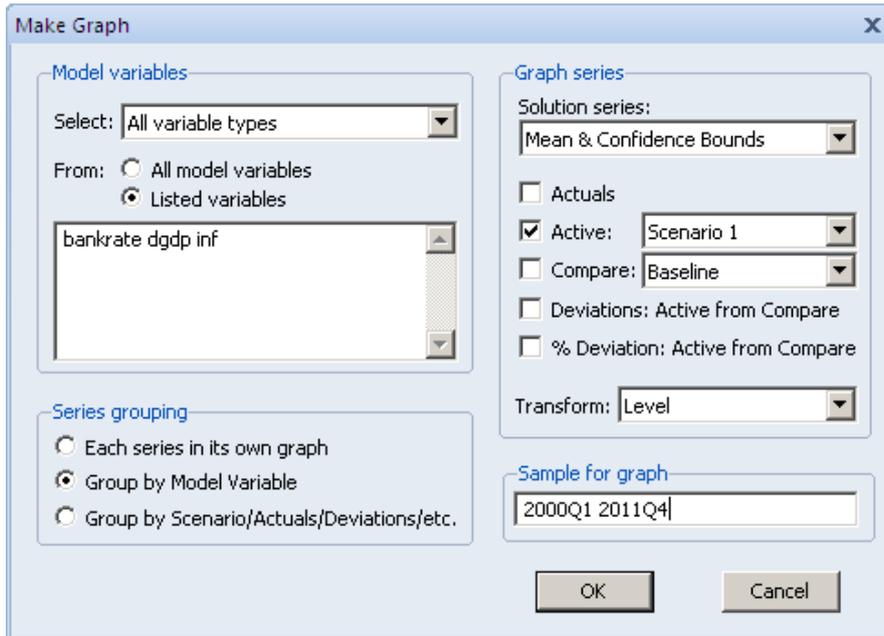
When the simulation is complete, we find that EViews has once again created new series for us. Depending on whether you selected the standard deviation or the bounds option, you should find variables with the suffixes `_1h`, `_1l`, `_1m` and `_1s` in the workfile. These new variables represent the upper (or higher) bound, the lower bound, the mean forecast and the standard error of the forecast respectively.

We can return to the **Variable View**, select the variables `bankrate`, `dgdp` and `inf` as discussed above and then click on **Proc, Make Graph...** When the **Make Graph...** dialog appears, select either:

- the option **Mean +/- 2 Std. Deviations** in the **Solution Series** list box in the **Graph Series** area on the right of the dialog – if you have selected the St. Dev. option; or
- the option **Mean & Confidence Bounds** in the **Solution Series** list box in the **Graph Series** area on the right of the dialog – if you have selected the Bounds option

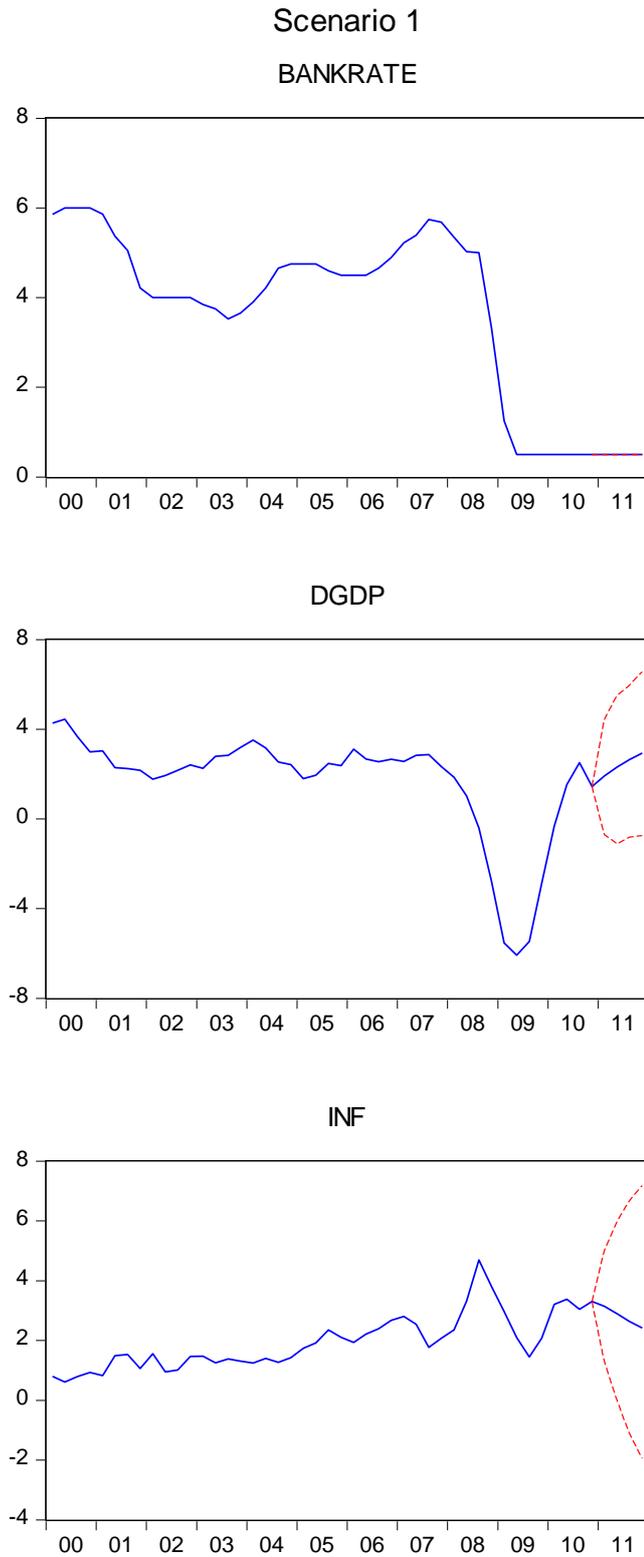
Set the sample to 2000 Q1 to 2011 Q4 and click on **OK**. The required options are illustrated in the following screenshot.

generated shocks to be (exactly) symmetric but also increase the precision of a simulation. In a linear model, a simulation with antithetic shocks will have a mean exactly equal to the deterministic solution of the model. But in a non-linear model, this method will not guarantee that the distribution of the model variables will be symmetric.



The h -step-ahead forecasts ($h = 1, \dots, 4$) for *dgdg*, *inf* and *bankrate* using the stochastic simulation method are given in Figure 4 below. The error bounds in the resulting output graph show that we should be reluctant to place too much weight on the point forecasts for our model, since the bounds are quite wide on *inf* and *dgdg*. But much of that uncertainty is probably due to the large residual in the GDP growth equation, which is creating a lot of variation in the remaining equations of the model in the stochastic simulation.

Figure 4: Forecasted values for *dgdp* and *inf* with confidence bounds from the reduced-form VAR(2) model, 2000 Q1-2010 Q4



7 Policy scenarios using reduced-form VAR models

Another use of conditional forecasting is the generation of forecasts conditional on different **policy scenarios**. These scenario-based conditional forecasts allow one to answer the question: if something happens to some variables in the system in the future, how will it affect forecasts of other variables in the future?

Question 4: Re-estimate the **illustrative** VAR(2) model using GDP growth, inflation, Bank rate and house price inflation up until 2010 Q4. In the next step, we assume that the future path of inflation and GDP growth is given by the Bank of England's February 2011 *Inflation Report* forecasts for the **mode** of GDP growth and inflation (shown on the next page). What is the conditional forecast for house price growth over 2011 if inflation and GDP growth follow the projected path?

CPI inflation projections based on market interest rate expectations and £200 billion asset purchases														
	2011 Q1	2011 Q2	2011 Q3	2011 Q4	2012 Q1	2012 Q2	2012 Q3	2012 Q4	2013Q1	2013Q2	2013Q3	2013Q4	2014Q1	
Mode	4.08	4.37	4.45	4.11	2.70	2.27	1.99	1.77	1.62	1.62	1.64	1.67	1.72	
Median	4.12	4.42	4.51	4.18	2.78	2.41	2.20	2.05	1.94	1.93	1.95	1.99	2.04	
Mean	4.13	4.43	4.53	4.20	2.80	2.45	2.26	2.13	2.02	2.02	2.04	2.07	2.12	
Uncertainty	0.61	0.88	1.11	1.27	1.34	1.37	1.39	1.42	1.48	1.49	1.50	1.52	1.52	
Skewness	0.05	0.07	0.08	0.10	0.10	0.18	0.26	0.35	0.40	0.40	0.40	0.40	0.40	
Probability that distribution will be above or below particular values														
Pr. {<1.0%}	<5%	<5%	<5%	<5%	9%	15%	19%	23%	26%	27%	26%	26%	25%	
Pr. {1.0% - 1.5%}	<5%	<5%	<5%	<5%	8%	10%	12%	12%	12%	12%	12%	12%	12%	
Pr. {1.5% - 2.0%}	<5%	<5%	<5%	<5%	11%	13%	14%	13%	13%	13%	13%	13%	13%	
Pr. {2.0% - 2.5%}	<5%	<5%	<5%	<5%	14%	14%	14%	13%	12%	12%	12%	12%	12%	
Pr. {2.5% - 3.0%}	<5%	<5%	5%	8%	15%	14%	12%	11%	11%	11%	11%	11%	11%	
Pr. {>3.0%}	97%	95%	92%	83%	43%	34%	29%	27%	26%	26%	26%	27%	28%	
Pr. {>Mode}	52%	52%	52%	52%	52%	54%	56%	58%	58%	58%	58%	58%	58%	
Pr. {<Mode}	48%	48%	48%	48%	48%	46%	44%	42%	42%	42%	42%	42%	42%	

* Because of the difficulties in precisely quantifying low probability events, probabilities of less than 5% are not shown in this table.

GDP growth projections based on Bank estimates of past growth and market interest rate expectations & £200 billion asset purchases														
	2011 Q1	2011 Q2	2011 Q3	2011 Q4	2012 Q1	2012 Q2	2012 Q3	2012 Q4	2013Q1	2013Q2	2013Q3	2013Q4	2014Q1	
Mode	2.36	1.77	1.74	2.80	2.75	3.01	3.11	3.14	3.12	3.09	3.08	3.10	3.14	
Median	2.31	1.68	1.59	2.60	2.56	2.76	2.81	2.79	2.73	2.70	2.68	2.70	2.74	
Mean	2.30	1.65	1.55	2.55	2.50	2.70	2.73	2.70	2.62	2.59	2.58	2.60	2.64	
Uncertainty	1.14	1.29	1.47	1.59	1.63	1.64	1.66	1.69	1.72	1.76	1.80	1.84	1.85	
Skewness	-0.06	-0.13	-0.19	-0.25	-0.25	-0.31	-0.38	-0.44	-0.50	-0.50	-0.50	-0.50	-0.50	
Probability that distribution will be above or below particular values														
Pr. {<0%}	<5%	10%	15%	6%	7%	6%	6%	7%	8%	9%	9%	10%	9%	
Pr. {0% - 1%}	10%	20%	20%	11%	11%	10%	10%	10%	10%	11%	11%	11%	10%	
Pr. {1% - 2%}	26%	30%	26%	19%	19%	17%	17%	16%	16%	16%	16%	16%	16%	
Pr. {2% - 3%}	34%	25%	23%	24%	24%	23%	22%	21%	21%	21%	20%	20%	20%	
Pr. {3% - 4%}	21%	12%	12%	22%	21%	22%	22%	22%	21%	21%	20%	20%	20%	
Pr. {>4%}	7%	<5%	<5%	18%	18%	22%	23%	24%	23%	23%	23%	24%	25%	
Pr. {>Mode}	48%	47%	46%	45%	45%	44%	43%	42%	41%	41%	42%	42%	42%	
Pr. {<Mode}	52%	53%	54%	55%	55%	56%	57%	58%	59%	59%	58%	58%	58%	

* Because of the difficulties in precisely quantifying low probability events, probabilities of less than 5% are not shown in this table.

Before we embark on the estimation, we may or may not want to undo the earlier changes to `bankrate`. In case we do, we activate the workfile window by clicking on it and then open the series `bankrate` by double-clicking on its icon. We click on **Edit +/-** to be able to manipulate individual cells in the spreadsheet and manually change the existing entry of 0.5 to NA in each of the four cells for 2011. We then click **Edit +/-** again to ‘lock’ the spreadsheet and close the series window. Use **Quick, Estimate VAR...**, select `dgdp`, `inf`, `bankrate` and `dhp` in that order as endogenous variables, select 1 2 for the lag interval and estimate the model. Results from the VAR estimation are given in Table 2. Remember that the estimation sample has to end in 2010 Q4.

Table 2: Reduced-form VAR(2) estimation results for `dgdp`, `inf`, `bankrate` and `dhp`, 1962 Q3 – 2010 Q4

Vector Autoregression Estimates
Sample (adjusted): 1962Q3 2010Q4
Included observations: 194 after adjustments
Standard errors in () & t-statistics in []

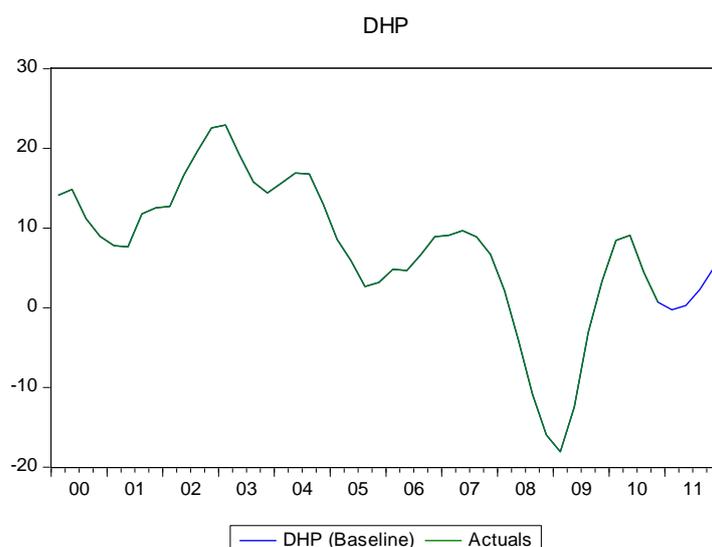
	DGDP	INF	BANKRATE	DHP
DGDP(-1)	0.754682 (0.07697) [9.80444]	0.110654 (0.05751) [1.92398]	-0.006274 (0.05091) [-0.12325]	-0.172423 (0.12792) [-1.34792]
DGDP(-2)	-0.025844 (0.07689) [-0.33613]	-0.098015 (0.05745) [-1.70611]	0.054880 (0.05085) [1.07922]	0.303109 (0.12778) [2.37219]
INF(-1)	-0.059704 (0.08944) [-0.66756]	1.402826 (0.06683) [20.9923]	-0.081627 (0.05915) [-1.37998]	-0.139215 (0.14863) [-0.93665]
INF(-2)	0.018156 (0.08860) [0.20493]	-0.459226 (0.06620) [-6.93702]	0.094968 (0.05860) [1.62072]	0.229522 (0.14724) [1.55885]
BANKRATE(-1)	-0.125385 (0.10847) [-1.15590]	0.240347 (0.08105) [2.96544]	1.164342 (0.07174) [16.2299]	-0.027496 (0.18027) [-0.15253]
BANKRATE(-2)	0.110253 (0.11095) [0.99372]	-0.197795 (0.08290) [-2.38596]	-0.207769 (0.07338) [-2.83147]	-0.086627 (0.18438) [-0.46983]
DHP(-1)	0.106789 (0.03412) [3.12952]	0.007386 (0.02550) [0.28968]	0.060721 (0.02257) [2.69058]	1.585166 (0.05671) [27.9534]
DHP(-2)	-0.075812 (0.03448) [-2.19870]	-0.000793 (0.02576) [-0.03077]	-0.033738 (0.02280) [-1.47947]	-0.690284 (0.05730) [-12.0466]
C	0.699838 (0.26893) [2.60229]	-0.110302 (0.20094) [-0.54893]	-0.092175 (0.17786) [-0.51824]	0.987780 (0.44692) [2.21018]

R-squared	0.696576	0.962251	0.948192	0.945375
Adj. R-squared	0.683455	0.960618	0.945951	0.943013
Sum sq. resids	298.0397	166.3894	130.3642	823.1074
S.E. equation	1.269262	0.948368	0.839447	2.109320
F-statistic	53.08852	589.4668	423.2326	400.2141
Log likelihood	-316.9228	-260.3819	-236.7141	-415.4612
Akaike AIC	3.360029	2.777133	2.533135	4.375889
Schwarz SC	3.511631	2.928735	2.684736	4.527491
Mean dependent	2.285966	5.505240	7.816011	8.612548
S.D. dependent	2.255971	4.778910	3.610781	8.835943
<hr/>				
Determinant resid covariance (dof adj.)		3.932523		
Determinant resid covariance		3.252005		
Log likelihood		-1215.486		
Akaike information criterion		12.90191		
Schwarz criterion		13.50832		
<hr/>				

We should also remember to convert the estimated VAR object into an EViews model object after estimation. I have called this new model `model02`. To condition on the Bank of England's forecasts for `dgdg` and `inf` over the forecast sample period, manually change the value of `dgdg` and `inf` in the underlying data series in the workfile to the values given to the **modal** forecast in the Bank of England's *Inflation Report*. In the **Model** object, go to **Solve, Solution scenarios & output** and choose **Scenario 1**. Then select **Edit Scenario Options** and click on the **Excludes** tab and type `dgdg inf` in the menu. This operation will keep `dgdg` and `inf` exogenously fixed at their actual data values in the workfile over the forecast period, i.e., 2011. We now solve the model again by clicking on **Solve** in the model object and choosing **Scenario 1** under **Active** scenario in the **Solution scenarios & output**. This new scenario creates the related forecast variables `inf_1`, `dgdg_1`, `bankrate_1` and `dhp_1`, which is shown in Figure 5.¹⁰

¹⁰ It is interesting to compare `bankrate_0` with `bankrate_1`. In order to have GDP growth and inflation follow the forecast path given on p. 21, the model-consistent monetary policy interest rate falls below 0.5 per cent in the first three quarters of 2011.

Figure 5: Forecasted values for dhp conditional on dgdp and inf, 2011 Q1-2010 Q4



But this forecast for 2011 is once again a **deterministic** forecast. It would be nice if we could produce a measure of uncertainty around this forecast. One very useful graphical tool in this respect is the fan chart. An example of how we can generate a fan chart around the forecast we just made is provided in the next section.

8 Generating a fan chart around a policy projection generated by a (VAR) model

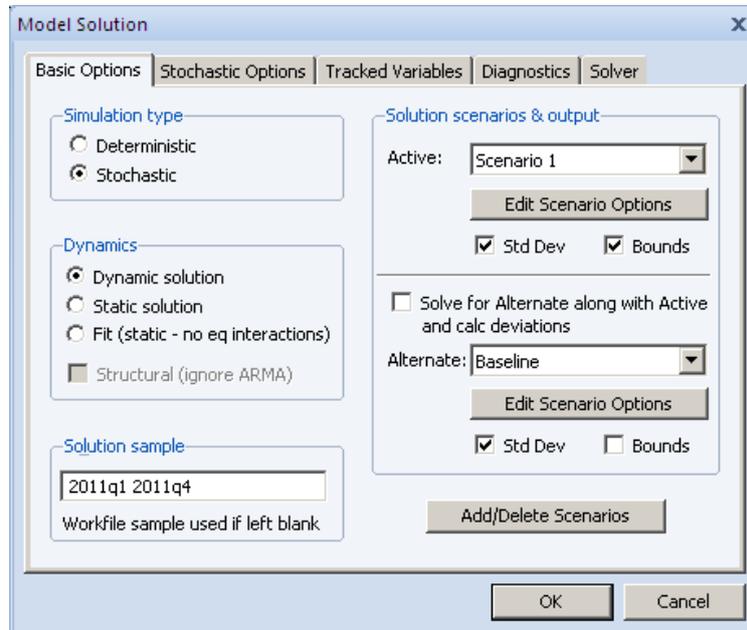
A very useful feature of EViews is its ability to produce fan charts by stochastic simulation. Fan charts have been popularised by, *inter alia*, the Bank of England, as a way of displaying the uncertainty around model forecasts in the form of quantiles from a distribution. The way that the Bank of England does this is to impose an *ad hoc* skewed distribution, namely a two-part normal, around a deterministic central point forecast. EViews, by contrast, can estimate the quantiles of the true empirical distribution of the forecast variables through the construction of a fan chart for the time-path of the variables of interest from stochastic simulation of the (non-linear) equations of the model.

The easiest way of generating the fanchart in Figure 7 below is to run the EViews program called **modelforecast.prg**, which can be found in the current data folder. All that is required at this point is for a particular (VAR) model to exist – we will be using `model02` in the workfile for this purpose. In order to illustrate the individual steps involved, though, we can go through the individual steps, which is what we will do now.

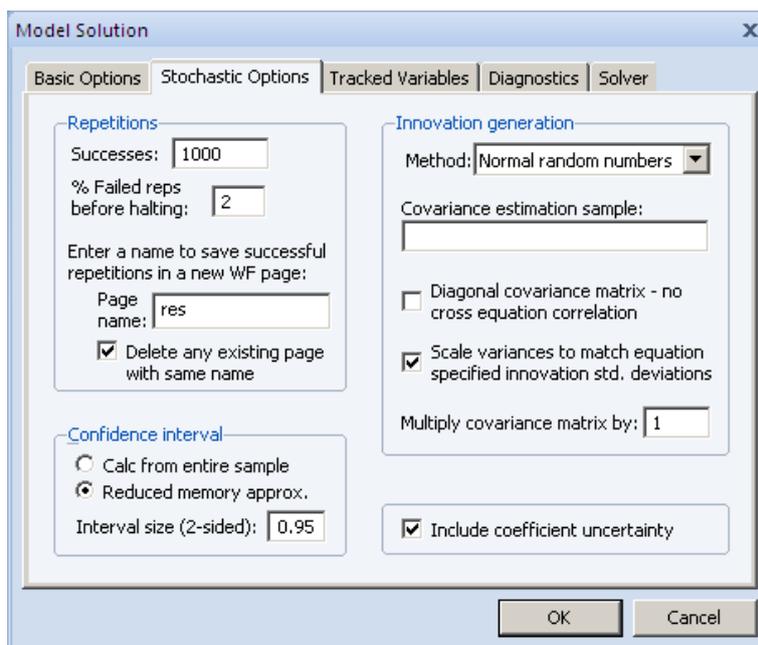
We have seen in Section 6 above how to stochastically simulate an estimated VAR model. We will follow the same steps as above, but making sure that we (a) save all of our draws to calculate the appropriate quantiles, and (b) take account of the conditional nature of the forecast.

In order to do a stochastic simulation of the forecast for UK house price inflation from the VAR(2) model over 2011, we open the `model12` model object by double-clicking on its name in the workfile. In the model window, click on the **Solve** button. When the **Model Solution** dialog appears, choose **Stochastic** for the simulation type on the **Basic Options** tab. The **Solution sample** should be set from 2011 Q1 to 2011 Q4. On the **Basic Options** tab, go

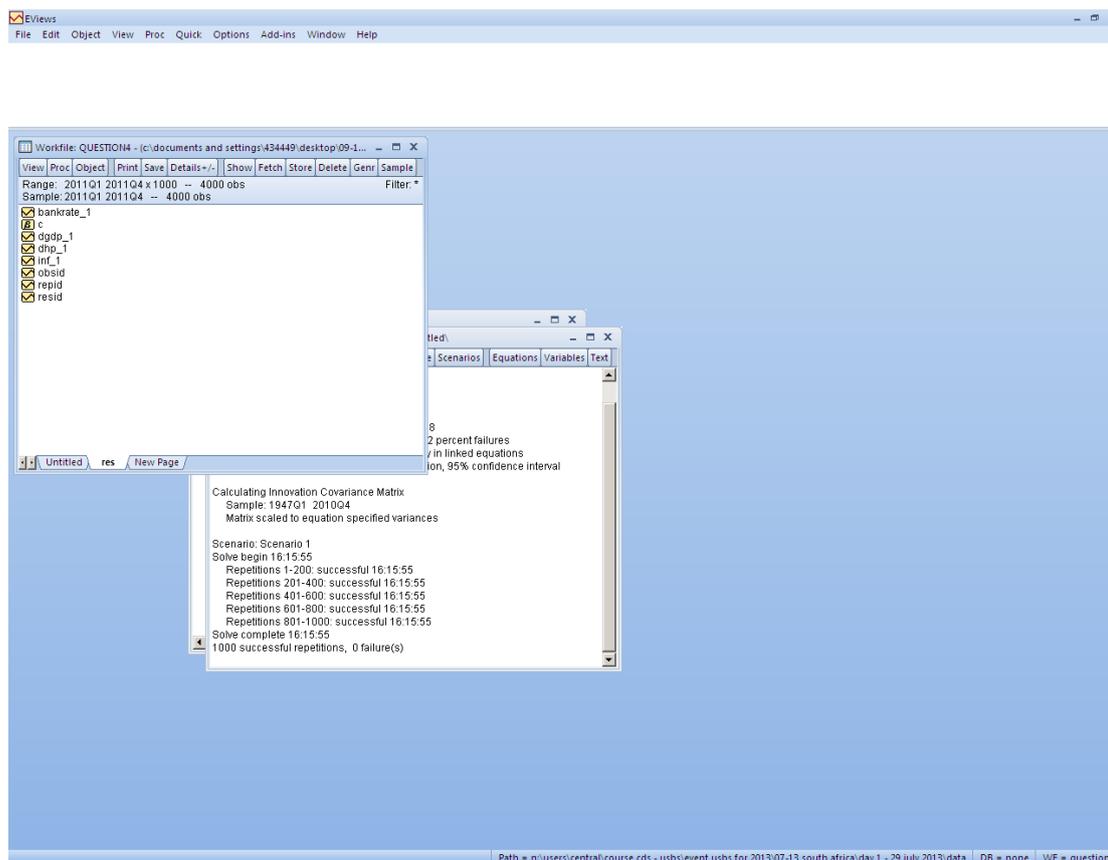
to **Solution scenarios & output** and choose **Active: Scenario 1**. Then select **Edit Scenario Options** and click on the **Excludes** tab and type `dgdp inf` in the window. This operation will keep GDP growth as well as inflation exogenously fixed at their actual data values in the workfile over the forecast period, i.e., 2011. Click **OK** to return to the **Basic Options** tab. Finally, tick either the **St. Dev.** or the **Bounds** (or both) checkboxes.



We then move on to the **Stochastic Options** tab. We select 1,000 in the **Successes** box under **Repetitions**. The successful repetitions that we will do need to be stored somewhere, so we create a new page in the workfile for that purpose. In my case, I have called it **Page Name: res**. By ticking the box below this dialog, we ensure that we delete any existing page with the same name. This will ensure that we only ever have one page open with the same workfile page name. For the time being, we will choose **Method: Normal random numbers** for the **Innovation generation** under the **Stochastic Options** tab. The other option is the **Bootstrap**, and – time permitting – we may use both innovation generation methods to see whether they are much different. Click on **OK** to begin the simulation.



After having done the stochastic simulations to generate 1,000 sets of four forecasts each, we find that EViews has created the new workpage denoted `res` in the workfile:



When the simulations are done, we can have a look at the results, which are stored in the new workfile page called `res`. The 1,000 simulations, which are organised in a panel

format, will be stored as `bankrate_1`, `inf_1`, `dgdgdp_1` and `dhp_1`. Altogether, we have generated 4,000 forecasts. They appear as a stacked $(4 \times 1000) \times 1$ vector, in the order of the first set of four simulated forecasts, the second set of four simulated forecasts and so on all the way down to the thousandth set of four simulations.

To perform the remainder of the required tasks, we have written an EViews program (**farchnew.prg**). The contents of **farchnew.prg** are as follows:

```
smp1 2011q1 2011q4

pageselect RES

for !i = 1 to 4
  smp1 if @obsid = !i
  series forgarch = dhp_1
  genr q05 = @quantile(forgarch, .05)
  genr q20 = @quantile(forgarch, .20)
  genr q35 = @quantile(forgarch, .35)
  genr q65 = @quantile(forgarch, .65)
  genr q80 = @quantile(forgarch, .80)
  genr q95 = @quantile(forgarch, .95)
next

pageunstack(page = "Unstack") repid @date @ q*

smp1 2008q1 2011q4
```

Let us take a closer look at what the program does. The second line of the code directs EViews to the workspace `res` and instructs it to carry out the operations that follow in the program on that particular workspace. The next part of the code:

```
for !i = 1 to 4
  smp1 if @obsid = !i
  series forgarch = dhp_1
  genr q95 = @quantile(forgarch, .05)
  genr q80 = @quantile(forgarch, .20)
  genr q65 = @quantile(forgarch, .35)
  genr q35 = @quantile(forgarch, .65)
  genr q20 = @quantile(forgarch, .80)
  genr q05 = @quantile(forgarch, .95)
next
```

successively moves through each of the four ($i = 1, 2, 3, 4$) forecast periods to calculate quantiles of the 1,000 draws per time period. We do this in order to generate a fan chart around the forecast. By moving from the fifth to the 95th percentile, we cover 90 per cent of all possible outcomes. As we will see, this set-up of the fan chart translates into three bands of increasing coverage. Obviously, the number of bands can be scaled up or down at will.

The second to last line **permanently** moves any series whose name starts with *q* from the **temporary** workpage `res` into another newly created workpage of the workfile called `unstack`.

Rather than just the one workpage, entitled `UK_data`, that we started off with, our workfile **question4.wf1** after running the code now has three workpages: `UK_data`, `res` and `unstack`.

8.1 Drawing a probability density function

A not very elegant way of generating a probability density function (PDF) of our forecast in 2011 Q4 is the following. Select the `res` page by clicking on it. Then enter the next two lines into the command window – one line at a time. Pressing return at the end of each line in turn will generate a new series called `forselect` that contains only data for 2011 Q4:

```
smpl if @obsid = 4
series forselect = dhp_1
```

These two lines are a brute force way of extracting all the 1,000 simulations for 2011 Q4. By sampling every fourth observation, we pick out all the observations in 2011 Q4. Why 4? Remember that $h = 4$ refers to our final forecast horizon. Having done so, the new series will be stored as `forselect` in the workpage `res`.

Opening `forselect` on workpage `res` and selecting **View, Descriptive Statistics & Tests, Stats Table** brings up a table showing summary statistics for the series under investigation. These are reported and discussed in Table 3 below. As we are doing stochastic simulations, these numbers need not be the same across different simulations and will vary slightly from stochastic simulation to stochastic simulation.

Table 3: Summary statistics of 1,000 stochastic simulations for 2011 Q4 (innovations are drawn from the normal distribution)

Mean	5.01
Median	5.10
Maximum	31.46
Minimum	-17.21
Standard deviation	7.36
Skewness	-0.04
Kurtosis	3.07
Observations	1000

Summary statistics of the estimated distribution show that the simulated distribution is (very) slightly skewed to the left and has a small amount of excess kurtosis. In addition, the result of the Jarque-Bera test provides no evidence against the null hypothesis that the series is normally distributed (this is left as an optional exercise).

8.2 Presenting a probability density function

Regardless of how density forecasts are generated, it is important to pay attention to their presentation. Tay and Wallis (2000, p. 246) make the observation that the way densities are presented and communicated may hide some features of the forecast and emphasise others. In other words, an inappropriate presentation of the forecast will reduce the forecast's usefulness for certain forecast users and, worse, may be misleading or misinterpreted, resulting in improper use of the forecast.

One very uncontentious way of assessing the density forecast is to investigate possible asymmetries in the future distribution of the forecast variable. In order to do so, we present the density forecast as a (smoothed) density estimate. The next step, therefore, is to smooth the N ($= 1,000$) simulated values of dhp for 2011 Q4, denoted by $p^{(i)}$, $i = 1, \dots, 1000$, to provide a kernel density estimate. The kernel density estimator is an adjusted histogram in which the 'boxes' of the histogram are replaced by 'bumps' that are smooth (Silverman (1986)). Smoothing is done by putting less weight on observations that are further away from the point being evaluated. Specifically, the kernel density estimate of a series $p^{(i)}$ at a point x is estimated by:

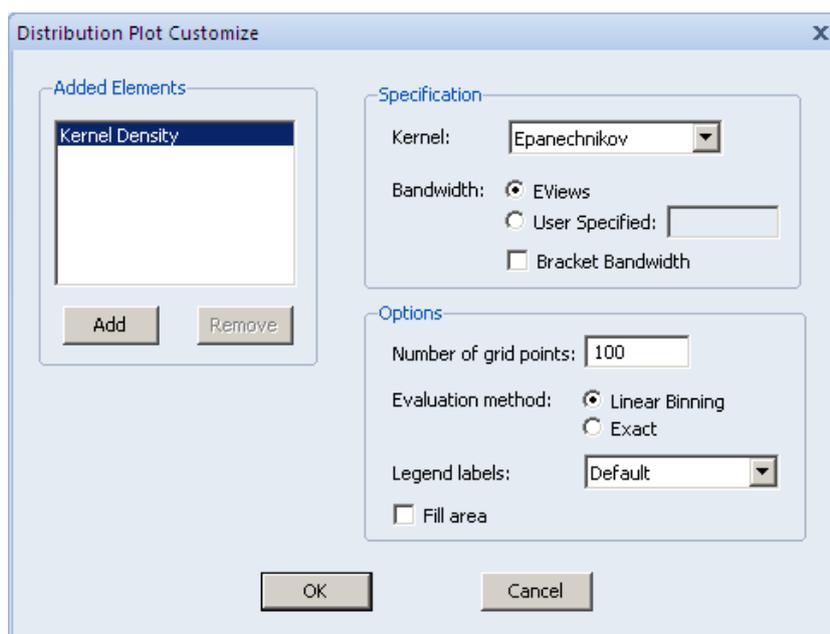
$$f(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - p^{(i)}}{h}\right) \quad (13)$$

where N is the number of observations, h is the bandwidth (or smoothing parameter) and $K(\bullet)$ is a kernel weighting function that integrates to one. The EViews default for $K(\bullet)$ is the Epanechnikov function. The Silverman option, which is the default for bandwidth selection in EViews, uses an automatic data-based method:

$$h = 0.9kN^{-1/5} \min(s, IQR/1.34) \quad (14)$$

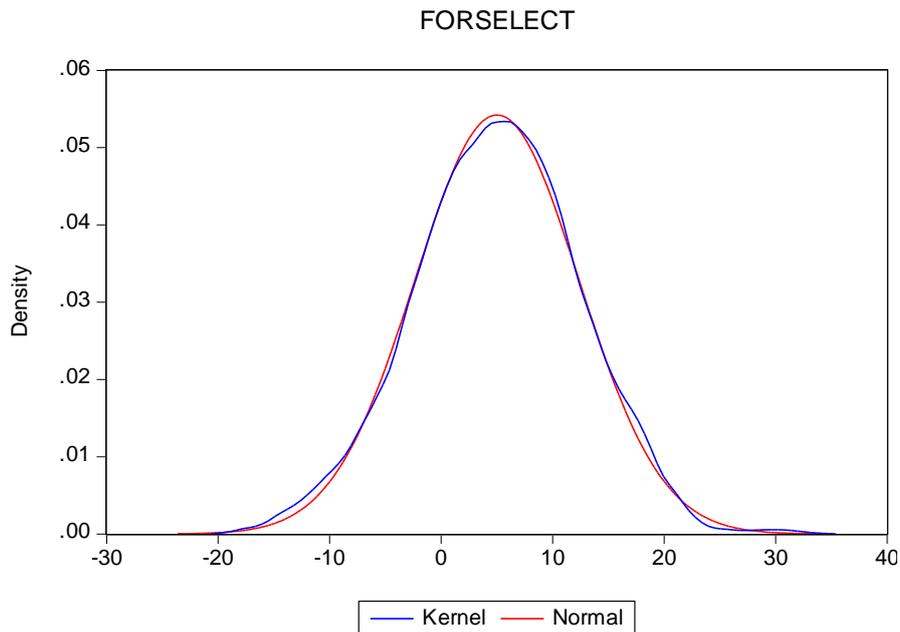
where N is the number of observations, s is the standard deviation and IQR is the interquartile range of the series (Silverman (1986), equation (3.31)). We should note that EViews includes many more kernel density estimation methods than the Epanechnikov function.

To compute and display a kernel density estimate for a single series or for each series in a group, choose and open `forselect` and then select **View/Graph...** from the series or group menu. After that, we choose **Distribution** in the **Specific** listbox. The right-hand side of the dialog page will change to provide a **Distribution** combo box prompting you to choose a distribution graph. You should select **Kernel Density**. The different types of kernel density estimators are selected by clicking on the **Options** box next to **Distribution**. The required options are illustrated in the following screenshot.



As we are choosing EViews' default options, we can leave all of the settings as they are. The resulting probability density function is shown in Figure 6 below. In order to assess any asymmetries in the future distribution of the forecast variable, a plot of the density estimate with an appropriate symmetric density function superimposed may be more helpful than the presentation of the forecast by itself. In practice, the normal distribution is conventionally taken as the symmetric distribution of choice. We have therefore imposed the corresponding normal distribution to highlight the presence of asymmetries or excess kurtosis in the data. A comparison between the two curves confirms the results in Table 3, i.e., that the density function is approximately normally distributed. Increasing the number of simulations from 1,000 to a higher number obviously results in a smoother probability density function.

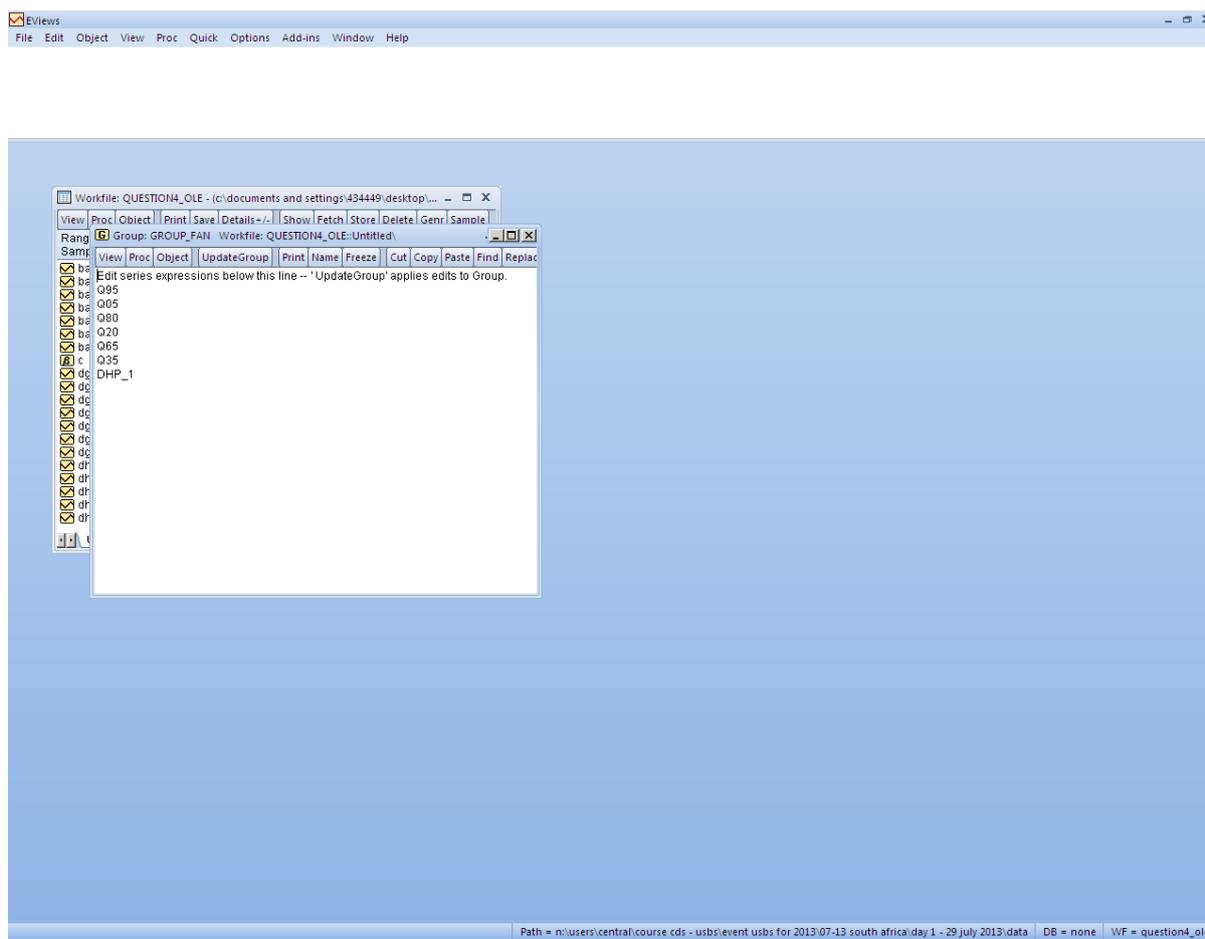
Figure 6: House price inflation density from an estimated VAR(2) model for 2011Q4



8.3 Drawing a fan chart

In the final step, we manually copy and paste the six quantile series (q_{05} , q_{20} , q_{35} , q_{65} , q_{80} and q_{95}) from the workpage `unstack` into the workpage `UK_data`. Dragging and dropping them onto the workpage `UK_data` also works. Note that these six series now have the right sample period for that page.

We now select the seven series q_{95} , q_{05} , q_{80} , q_{20} , q_{65} , q_{35} and `dhp_1` **in that order** and open them as a group. I have given the name of `group_fan` to the resulting group. By going to **View, Group Members** we should find that the series have been selected in this order.



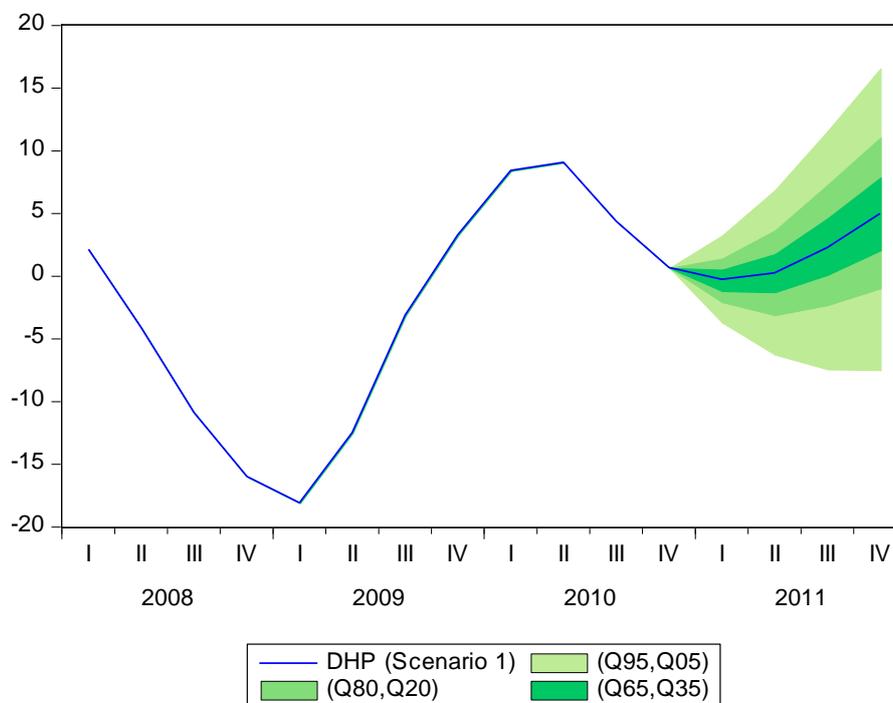
The reason for selecting them in this order is that we will be using the **Area Band** graphing option in EViews. The area band graph is used to display the band formed by pairs of series, filling in the area between the two. While they may be used in a number of settings, band graphs are most often used to display forecast bands or error bands around forecasts; in other words, fan charts. EViews will construct bands from successive pairs of series in the group. If there is an odd number of series in the group, the final series will, by default, be plotted as a line. This will be the case in our example. The ordering of the variables means that EViews will shade the area between the 95th and the fifth quantile (the outermost fan), the area between the 80th and the 20th quantile (the middle fan) and the area between the 65th and the 35th quantile (the innermost fan). The seventh series, `dhp_1`, which represents both the historical data and the point forecasts, will appear as a line in the resulting graph.

Before we can go and graph the series as an **Area Band**, we have to do one more manual operation in the **Group** window. Select **View/Spreadsheet** and press **Edit +/-** to manipulate the underlying data in the spreadsheet. Copy the historical data from the column for `dhp_1`, i.e., the data from 2008 Q1 to 2010 Q4, and copy that data into each of the historical periods for the other six series. All the columns should now be filled in. Press **Edit +/-** again to 'lock' the spreadsheet.

You may now display an area band graph for any group object containing our seven time series. Select **View/Graph...** from the group menu, and then choose **Area Band** in the **Specific** listbox. The **Fill Areas** and **Lines & Symbols** dialogs under the **Graph Elements** group may be used to modify the characteristics of the lines and shades in your graph. In the

example at hand, the outermost (top) colour should be the lightest, the innermost (bottom) colour should be the darkest and the middle colour should be in-between the other two. After setting the appropriate options, the results should be the ‘fan chart’ shown in Figure 7.

Figure 7: Fan chart forecast for UK house price inflation from 2008 Q1 to 2011 Q4



As we can see, the outermost, or 90 per cent, fan indicates that UK house price inflation could fall anywhere between -5 and 16 percent.

References and further reading

- Basu, A K and Sen Roy, S (1987)**, ‘On asymptotic prediction problems for multivariate autoregressive models in the unstable nonexplosive case’, *Calcutta Statistical Association Bulletin*, Vol. 36, No. 149-150, pages 29-37.
- Blake, A P (1996)**, ‘Forecast error bounds by stochastic simulation’, *National Institute of Economic and Social Research Review*, Vol. 156, No. 1, pages 72-79.
- Cuthbertson, K, Hall, S G and Taylor, M P (1992)**, *Applied econometric techniques*, Hemel Hempstead, Harvester Wheatsheaf.
- Engle, R F and Yoo, B S (1987)**, ‘Forecasting and testing in co-integrated systems’, *Journal of Econometrics*, Vol. 35, No. 1, pages 143-59.
- Lütkepohl, H (2005)**, *New introduction to multiple time series analysis*, Berlin, Springer-Verlag.
- Reimers, H-E (1991)**, *Analyse kointegrierter Variablen mittels vektorautoregressiver Modelle*, Heidelberg, Physica-Verlag.

- Samaranayake, V A and Hasza, D P (1988)**, ‘Properties of predictors for multivariate autoregressive models with estimated parameters’, *Journal of Time Series Analysis*, Vol. 9, No. 4, pages 361-83.
- Silverman, B W (1986)**, *Density estimation for statistics and data analysis*, London, Chapman & Hall.
- Tay, A S and Wallis, K F (2000)**, ‘Density forecasting: a survey’, *Journal of Forecasting*, Vol. 19, No. 4, pages 235-54.