FEB22009(X,S)
Case 1: Predicting house prices

Andreas Pick

May 2016

In this case, you will predict the house price index for a region of the UK. Different groups will predict the index for different regions and time periods. For the predictions you will use different models, obtain the forecasts from these models, and compare their forecast accuracy. There are a series of models that you are expected to use and an optional one.

The different models are:

- Univariate time series models.
- Time series models with additional regressors that are economically meaningful.
- Factor augmented time series models.
- Optionally, spatial models.

Univariate time series model One of the simplest model for forecasting is a univariate time series model from the ARIMA family. These models offer relatively little in terms of economic intuition. However, often they provide some of the most precise forecasts. Select the model that describes the times series the best and use it to construct forecasts.

Time series model with exogenous regressors Add exogenous regressors to a time series model from the ARIMA family in order to improve forecasts. An example of such a model would be an ARDL model. The exogenous regressors should be economically sensible. Examples include GDP and consumer prices, or their respective growth rates. However, you are free to add other variables if you can argue for their importance economically and empirically.

Factor augmented time series model In addition to the variables above, the information from the house prices in other regions in the UK may be useful for the predictions. Stock and Watson (2002) describe the application of factor models for forecasting but a good start is to remember

factor models from Voortgezette Statistiek/Advanced Statistics. You will need to decide on the number of factors and how to incorporate them into the model.

Spatial models [This is a "stretch goal" in kickstarter methodology, that is adding a spatial model can increase your grade but it does not substitute the above models.] A spatial model is a model of correlation across space. An introduction can be found in LeSage (1999), which is freely available on the internet.

Forecasting The time series that you are using will likely contain 113 observations (assuming no more than 48 groups) for one of 12 regions in the UK. Use 90 observation as your initial estimation sample and calculate forecasts from the remaining 28 observations using expanding estimation sample windows. This means that your first forecast is based on the parameters estimated with a sample of 90 observations. The next forecast is based on 91 observations, the initial 90 plus the observation that you were forecasting in the previous iteration. Continue adding one observation at a time.

You will construct one step-ahead and four steps-ahead forecasts, that is, one quarterand one year ahead forecasts. For the four step-ahead forecasts you need to consider your forecasting procedure. The options are direct forecasts or iterated forecasts. Marcellino et al. (2006) discuss these forecast procedures. Choose one method. As a stretch goal you can use both approaches and compare the forecast accuracy for these two approaches in this data set.

Issues to consider

- Statistical nature of the time series (e.g. stationarity). Note that you should only report statistics that are important for the forecasting exercise.
- Implementation of each model.
- Constructing pseudo-out-of-sample forecasts: for any given forecast, future data cannot be used in the estimation (even though the observation is available in the data set). This is an easy mistake to make. You have to imagine that you are in real time and have no information about future exchange rate movements.
- Remember that you are forecasting the house price index and not any derivative thereof.

- Evaluating the forecasts: You should use an econometric method to evaluate the forecasts. As the benchmark use the random walk forecast: $P_{t+1} = P_t$.

Data

A csv file with house prices is available on Blackboard. You will need to use additional regressors that are not provided on Blackboard and which you will have download yourself from the internet. A good source for British data is the Office of National Statistics (ONS) and an alternative is eurostat.

[A table with regions and samples will be placed here once groups have been formed.]

Econometric software

You are free to choose any programming language. However, I would recommend Matlab, Octave, or R. Please report your code in an appendix to your report (it will not count towards the allowed space). Publicly available packages (think CRAN or the Matlab econometrics toolbox) can be used and these functions do not need to be included in the report but you need to make clear which packages you use.

Grading

Beyond the grading rules mentioned in the Course Information document note that in this part a report that does not attempt any of the stretch goals can attain a maximum of 9. Well implemented, documented, and described stretch goals will then allow you to achieve a 10.

Also note that in this case particular attention will be placed on whether your report would allow us to replicate your findings. Adding the code does not substitute a thorough description, and a researcher should be able to replicate your findings even without your code.

References

LeSage, J. P. (1999). *The Theory and Practice of Spatial Econometrics*.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135(1–2):499–526.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.