

Econ6037: Economic Forecasting

University of Hong Kong

Project #2 – An Evaluation of the Output Growth Forecasts

Due date: Friday, February 21, 11:30p.m. (via the course website)

A note from the instructor

1. Please pay attention to the long instructions below. They are all here for good reasons!
2. This assignment is meant to be completed individually. Communication with, and hence learning from, classmates is strongly encouraged. Caution, however, too much reliance on our classmates for help diminishes the amount of our learning from the assignment. Each student is expected to collect his/her own data, write or modify the R-scripts to suit the purpose of the assignment, conduct his/her own analysis and write up this/her own report. Remember:
Give a man (your classmate) a fish, and you feed him for a day.
Teach a man (your classmate) to fish, and you feed him for a lifetime.
3. To enrich our understanding of the world, no students are allowed to work on the same data set.
4. To some students, it might appear easier and more convenient to use Excel or to use R interactively for the computational part of this assignment. Here, aside from some very basic data manipulation using Excel, students should use R to do most of the calculation as much as possible. Try to write a short program/script of R for the task, with annotations so that readers of the R script will know your programming logic. Points will be deducted if you do not use R to generate the graphs and statistics.
5. Always try to write the report in a self-contained way and in a style that you would be happy to show to your current or potential employer.
6. Start early. This assignment is very demanding, especially if you are not familiar with R!
7. Make sure your report is black-and-white printer friendly. For grading, I almost always print out the reports using a black and white printer. Keep in mind that colors will not show on a black and white printout.

We would like to evaluate the forecast of various simple forecasting models of output growth, including the combined forecast. Most of the models we considered are similar to those described in Stock and Watson (2004)¹. We will work with data at quarterly frequency. Try to choose a country with a longer time series, whenever possible. Data series that start later than 1980s are generally not acceptable.

Pick a country. Indicate your choice of country in “Project #2 Wiki” in our course moodle page. Once a country is taken, other students can only choose some other countries. First come, first served!

The following description (and hence notations) are very similar to the description of Stock and Watson (2004). Let $Y_t = 400\Delta \ln Q_t = 400[\ln Q_t - \ln Q_{t-1}] = 400 \ln(Q_t/Q_{t-1})$, where Q_t is the level of output (real GDP) at quarterly frequency, and let X_t be a candidate predictor, again, at quarterly frequency. Consider at least five such candidate predictors of your choice (e.g., the term spread and unemployment rate). Let Y_{t+h}^h denote output growth over the next h quarters, expressed at an annual rate, that is, let $Y_{t+h}^h = (400/h) \ln(Q_{t+h}/Q_t)$. We will consider regression based forecasting model to produce h -step-ahead forecast of the type:

$$Y_{t+h}^h = \beta_0 + \beta_1(L)X_t + \beta_2(L)Y_t + u_{t+h}^h \quad (1)$$

where u_{t+h}^h is an error term and $\beta_1(L)$ and $\beta_2(L)$ are lag polynomials, each with a maximum degree of 8. Forecasts are computed for $h = 2, 4, 8$ -quarter horizons. Let the forecast of Y_{t+h}^h be denoted as $\hat{Y}_{t+h|t}^h$. Denote these models (five different predictors X_t) with index $i = 1, 2, 3, 4, 5$.

Two univariate benchmark forecasts are considered. The first is a “multi-step autoregressive (AR)” forecast, essentially equation (1) with no X_t predictor. That is,

$$Y_{t+h}^h = \beta_0 + \beta_2(L)Y_t + u_{t+h}^h \quad (2)$$

Denote this model with index $i = 6$. Note that while we follow Stock and Watson (2004) to call equation (2) a multi-step autoregressive (AR) forecasting model (p.408 of Stock and Watson, 2004), it is in fact not an autoregression in the usual sense. Autoregression should include the lags of the dependent variable as the independent variables. It is better to understand this as a special form of regression based forecasting model.

The second is a random-walk-like forecast based on the historical mean, in which $\hat{Y}_{t+h|t}^h = \hat{\mu}_t$, where $\hat{\mu}_t$ is the sample average of Y_s (not $400Y_s$), $s = 1, \dots, t$. That is,

$$\hat{Y}_{t+h|t}^h = \hat{\mu}_t = \frac{1}{t} \sum_{s=1}^t Y_s \quad (3)$$

Denote this model with index $i = 7$.

We will also consider combination forecast. Let $\hat{Y}_{i,t+h|t}^h$ denote the model i 's individual out-of-sample forecast of Y_{t+h}^h , computed at date t . We consider only the equal-weighted combination forecast with the form

$$\hat{Y}_{8,t+h|t}^h = \sum_{i=1}^5 \frac{1}{5} \hat{Y}_{i,t+h|t}^h \quad (4)$$

That is, the forecast is denoted with index $i = 8$.

Suppose we have a total of T observations. We start to produce forecast with the first R observations. Let's make $R/T \approx 1/2$.

We will use the recursive scheme to produce the forecast. For concreteness, let's consider $h = 2$ in the following discussion. The extension to $h = 4$ and $h = 8$ should be straightforward.

¹Stock, James H., and Mark W. Watson (2004): “Combination Forecasts of Output Growth in a Seven-Country Data Set,” *Journal of Forecasting*, 23: 405-430.

- ▷ Start at $t = R$, we will use all first R observations to estimate a model and produce a forecast of Y_{R+2}^2 using the five models with X predictors ($i = 1, 2, 3, 4, 5$), the model without X predictors ($i = 6$), the random-walk like model ($i = 7$) and subsequently the combined forecast ($i = 8$). Of course, for the model $i = 1, 2, 3, 4, 5, 6$, we will need to choose the lags included based on some model selection criteria, say, AIC. Call these $\hat{Y}_{i,R+2|R}^2$, $i = 1, 2, 3, 4, 5, 6, 7, 8$. Be careful, the superscript “2” on Y does not mean square, it is used to denote the horizon of forecast only.
- ▷ Then, at $t = R + 1$, add one more observation to our sample. We will use all first $R + 1$ observations to estimate a model and produce a forecast of Y_{R+3}^2 using the five models with X predictors ($i = 1, 2, 3, 4, 5$), the model without X predictors ($i = 6$), the random-walk like model ($i = 7$) and the combined forecast ($i = 8$). For the model $i = 1, 2, 3, 4, 5, 6$, we will need to choose the lags included AGAIN, say, based on AIC. Call these $\hat{Y}_{i,R+3|R+1}^2$, $i = 1, 2, 3, 4, 5, 6, 7, 8$.
- ▷ Repeat the above with $t = R + 2$ to $t = T - 2$.

Then, for $h = 2$, after we are done with recursive scheme, we should have a total of $T - R - 1$ forecasts. Forecast errors can be calculated.

$$e_{i,t+2|t}^2 = Y_{t+2}^2 - \hat{Y}_{i,t+2|t}^2$$

Note again, the superscript “2” on e does not mean square, it is used to denote the horizon of forecast only. Our major measurement is the mean squared forecast errors (MSFE), sometimes called mean squared prediction errors (MSPE), can be computed.

Assignment Project Exam Help

$$MSFE_i = \frac{1}{T - R - 1} \sum_{t=R}^{T-2} \left(e_{i,t+h|t}^2 \right)^2$$

<https://tutorcs.com>

Repeat with $h = 4$ and $h = 8$. Produce at least a summary table of MSFE like the followings.

Model	$h = 2$	$h = 4$	$h = 8$
Benchmark models			
AR ($i = 6$)
Random Walk ($i = 7$)
Individual models			
ARX ($i = 1$)
ARX ($i = 2$)
ARX ($i = 3$)
ARX ($i = 4$)
ARX ($i = 5$)
Combined ($i = 8$)

We would also like to perform a statistical evaluation of the performance of these models.

- ▷ For each of the forecasting models ($i = 1, 2, 3, 4, 5, 6, 7, 8$), assess the forecastability via a Mincer-Zarnowitz regression.
- ▷ For each of the forecasting models ($i = 1, 2, 3, 4, 5, 6, 8$), test statistically whether any of the forecasts has the same MSFE as the random-walk-like forecast.

When you are conducting these statistical tests, remember to state clearly the hypothesis being tested and how the test is conducted.

Keep in mind that our focus is forecast evaluation, and the hope to conclude a best model we can use in our future work. Write up a report discussing your forecast and your observations from the forecast evaluation.

Also keep in mind that we would like to see how students apply what they have learned from class to conclude the final model and forecast evaluation. Clear statements of DECISIONS and CONCLUSIONS at different juncture with the appropriate supporting evidence is the key.

Upload a zip file containing the whole folder of your work related to this project to Assignment corresponding to project #2. The zip file should include the report (pdf format), the R script, the data file, and the Word file, etc.

Often, students are tempted to write a lot. Please don't. Try to write concisely, yet precisely. When you are writing up the report, you should assume a reader from the industry (say, Economist Intelligence Unit). Always ask: "We know what we are doing but do the readers know what we are doing?" "Is the report too long such that readers will find it boring?" In your report, try to include the following sections:

1. An introduction. (One to two pages?)
 - ▷ What we plan to do in the paper and why we want to do it.
2. A brief description of the data. (One to two pages?)
 - ▷ A brief description of the variables.
 - ▷ Data source: the URLs or tickers or acronyms from the database such as Bloomberg, Datastream; the definitions, the original source of the data, etc.
 - ▷ Sample period, and data frequency.
 - ▷ Reason(s) for the choice of the country.
 - ▷ Brief reason(s) for choice of the predictors.
3. Estimation. (Three to five pages?)
 - ▷ A brief description of the models and modelling strategies.
 - ▷ How we arrive at the chosen model, with supporting evidence.
 - ▷ The recursive scheme, with a reference to the current forecasting project.
4. Major findings of forecast comparison (Three to five pages?)
 - ▷ How the forecast comparison is done.
 - ▷ How the tests are done.
 - ▷ Our observations from the tables of statistics and plots.
5. Concluding remarks (One to two page?)
 - ▷ Major conclusion, policy implication (if any) and potential improvement of the analysis.
 - ▷ Take the best model to produce a forecast of 2019 and 2010, and discuss the outlook of the economy.
6. Reference section (One page?)

The report should have less than 16 pages, with at least 12 pt fonts, at least 1.5 line spacing, and at least 2 cm of margins on each side. Page numbering, figure numbering and table numbering should be included. Some students feel obliged to fill up all 16 pages. Please don't. A shorter report is always preferred. It is about how to present the idea and analysis to the readers clearly. For the same content and same clarity, readers always prefer shorter reports.

R: R is a free software environment for statistical computing and graphics, available at <http://www.r-project.org/>.

Bloomberg: Bloomberg is available from our computer lab on 10/F of KK Leung Building. Students are welcome to explore other reliable databases. Nonetheless, Bloomberg is preferred, and familiarity with Bloomberg is a valuable assets in the business/research field.

DataStream: DataStream is available from our University Main Library. Familiarity with DataStream is a valuable assets in the business/research field.

FRED: Federal Reserve Economic Data is available at <https://fred.stlouisfed.org/>

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Objectives of this assignment:

- ▷ To practice how to forecast with simple time series models.
- ▷ To practice the forecast evaluation.
- ▷ Writing up the report: tighten up the logic of discussion (why we are doing this and that).
- ▷ Widen our horizon to see what happen in other countries (students have to work on a diverse set of countries).
- ▷ To see the advantages of different models and modelling strategies.

Grading rubrics (the following items may carry different weights):

Grading is mainly based on the report. The other materials are referred only when necessary. On some items below, as an illustration, we highlight the points that are commonly deducted on common mistakes.

- ▷ Cover page: title of the report, the name and student ID number, and date. (5 points deducted if missing)
- ▷ Basic formatting: page numbering, equation number, table numbering, figure numbering; table title, figure title. (5 to 10 points deducted if missing)
- ▷ Discussion associated with plots or tables. If you include a plot, make sure you discuss it. (up to 50 points deducted if inadequate)
- ▷ Whether the R script and data file are adequate to regenerate the results used in the paper. (up to 50 points deducted if inadequate)
- ▷ Data description / Data sources. (up to 10 points deducted if missing)
- ▷ Properly labeled tables and figures (clear titles); whether notes to tables / figures are included. (up to 10 points deducted if missing)
- ▷ Adequate guidance to readers in understanding the paper. (up to 50 points deducted if inadequate)
- ▷ Writing: Grammar, organization, transition from one paragraph to the next, etc. (up to 30 points deducted if inadequate)
- ▷ Proper citations and references. (up to 20 points deducted if inadequate)
- ▷ Motivation / Policy implications / Potential use of the analysis. (up to 20 points deducted if inadequate)
- ▷ Are claims properly supported with evidence and statistical logic? (up to 50 points deducted if inadequate)
- ▷ Discussion of the linkage of the paper to policy implications. (up to 20 points deducted if inadequate)