Principles of Forecasting

Econometrics II
YTU Economics
Prof. Dr. Hüseyin Taştan

Forecasting economic time series

- Forecasting involves predicting future value(s) of a time series using all available information upto time t.
- Market participants, policy-makers, and even ordinary people routinely make forecasts (systematically or unsystematically)
- Time series models can especially be useful when we form forecasts
- In forecasting, we are not particularly interested in modeling causal relationships, but in predicting future outcomes using currently available information.
- As such, predictive perfomance is more important.

How to form forecasts?

- One can show that the forecasting rule with the minimum expected squared forecasting error is given by the conditional expectation.

One-step ahead forecast of y → $\hat{y}_{t+1} = E(y_{t+1}|\ I_t\ )$ ← All information available up to period t

- Here, we only consider one-step-ahead forecasts, multiple-step-ahead forecasts are similar but more complicated (and also less precise).

Forecast Error

- Let f_t be the forecast for y_{t+1} made at time t using all information upto time t. Then, the forecast error can be computed using

$$e_{t+1} = y_{t+1} - f_t$$

Once we observe the actual value y_{t+1} at time $t + 1$.

- Choosing the forecast to minimize the expected square forecast error given by

$$\mathrm{E}(e_{t+1}^2|I_t) = \mathrm{E}[(y_{t+1} - f_t)^2|I_t].$$

Yields the solution where $\mathrm{E}[y_{t+1}|I_t]$ is the optimal forecast

Example

- Consider a sequence of i.i.d. Random variables with mean zero and a constant variance: $\{y_t : t = 0,1,2,\ldots\}$
- Let the information set be the entire history $\mathrm{I_t} = \{y_t, y_{t-1}, \ldots, y_0\}$
- What's the best forecast? By definition, the future value is independent of the past thus

$$E(y_{t+1}|y_t, y_{t-1}, \ldots, y_0) = 0$$

- In other words, the past does not help us predict the future. The technical term for this is the martingale difference sequence.
- Can you consider a real life example? Stock market returns? (but with positive mean instead of zero).

Example

- Now let us consider another time series where the best forecast is not zero but the last observed value, that is,

$$E(y_{t+1}|y_t, y_{t-1}, \ldots, y_0) = y_t$$

- In other words, the predicted value for the next period is always the last value. The technical term for this is the martingale sequence.
- As an example consider the Random Walk. What is the best one-step-ahead forecast?
- What about random walk with drift?

Types of Forecasts

- In-sample forecast: the forecasted value is within the stimation sample. For example, Predicted y values from an OLS regression are in-sample forecasts of the dependent variable.
- Out-of-sample forecast: the forecasted value is not in the estimation sample. When we say «forecast» we almost always mean «out-of-sample forecast».
- Forecast success of a model can only be determined by evaluating out-of-sample forecasts.
- Static forecast: based on the actual values of time t information.
- Dynamic forecast: based on the forecasted values instead of observed values.

Regression based forecast models

- As an example consider the following static regression model

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

- For simplicity, assume that parameters are known (we can just estimate and plugin those values). Because

$$E(u_{t+1}|I_t) = 0$$

- The best one-step-ahead forecast is

$$E(y_{t+1}|I_t) = \beta_0 + \beta_1 z_{t+1}$$

- This is an example of a conditional forecast because it is conditioned on the value of z_{t+1}
- This simply implies that we need to know z_{t+1} or forecast it. If we use a deterministic model, such as linear or polynomial trend, or seasonal dummies, they are automatically known.

Regression based forecast models

- The previous model was static. It makes more sense to use a dynamic model such as

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 z_{t-1} + u_t$$
$$\mathrm{E}(u_t|I_{t-1}) = 0,$$

- Note that all values on the right-hand side will be observed at time t.
- Our best one-step-ahed forecast will be

$$E(y_{t+1}|I_t) = \delta_0 + \alpha_1 y_t + \gamma_1 z_t$$

- It is very easy to add further variables on the right-hand-side if we think they will be useful to predict future.

Regression based forecast models

- Typical forecast models predict a variable in a linear regression using lagged values of the variable and lagged values of other variables.

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \gamma_1 z_{t-1} + \ldots + \gamma_p z_{t-p} + u_t$$

$$E(u_t|I_{t-1}) = E(u_t|y_{t-1}, z_{t-1}, y_{t-2}, z_{t-2}, \ldots) = 0$$

- One may include the lagged value of arbitrarily many other variables.
- If enough lags have been included, the model is dynamically complete and there is no serial correlation in the error (but may be heteroskedasticity).
- OLS inference methods can be used if the error is conditionally normal.

Example: Forecasting the unemployment rate

$$\widehat{unem}_t = \underset{(.577)}{1.572} + \underset{(.097)}{.732}\ unem_{t-1}$$

$$n = 48,\ \bar{R}^2 = .544$$

$$\widehat{unem}_t = \underset{(.490)}{1.304} + \underset{(.084)}{.647}\ unem_{t-1} + \underset{(.041)}{.184}\ inf_{t-1}$$

← Lagged inflation significantly helps to predict current unemployment

$$n = 48,\ \bar{R}^2 = .677$$

Note that these regressions are not meant as causal equations. The hope is that the linear regressions approximate well the conditional expectation.

Vector autoregressive model (VAR)

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \gamma_1 z_{t-1} + \ldots + \gamma_p z_{t-p} + u_t$$

$$z_t = \eta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \rho_1 z_{t-1} + \ldots + \rho_p z_{t-p} + v_t$$

- The model above is a VAR(p), a vector autoregression model with order p.
- It involves 2 variables, y and z, and two equations. We can add new variables into the system easily.
- The first equation relates the current value of y with the past values of itself and the past values of z.
- Similarly, the second equation relates the current value of z with its own past and the past of y.

Vector autoregressive model (VAR)

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \gamma_1 z_{t-1} + \ldots + \gamma_p z_{t-p} + u_t$$

$$z_t = \eta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \rho_1 z_{t-1} + \ldots + \rho_p z_{t-p} + v_t$$

- VAR models can easily be estimated using OLS.
- We can just simply apply OLS to each equation separately.
- All variables (y and z above) must be I(0)
- In effect, we are modeling short run dynamic relationships among the variables.

Granger Causality

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \gamma_1 z_{t-1} + \ldots + \gamma_p z_{t-p} + u_t$$

$$z_t = \eta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \rho_1 z_{t-1} + \ldots + \rho_p z_{t-p} + v_t$$

- Granger causality has a distinct meaning. It should not be confused with the regular meaning of causality.
- Basically, the Granger causality is a statement on the predictive performance of one variable on the other.
- We say that «the variable z Granger-causes the variable y, if and only if adding the lags of z in the equation of y improves the one-step ahead forecast»
- In other words, z is helpful in forecasting y.

Granger Causality

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \gamma_1 z_{t-1} + \ldots + \gamma_p z_{t-p} + u_t$$

- Granger causality can easily be tested using an F test.
- Let's consider the case whether x Granger-causes y. The null hypothesis is:

H0: z does not Granger-cause y, or, all coefficient on the lagged z are jointly zero**.**

- H1: z Granger-causes y
- Note that under the null hypothesis, the equation above will have only lagged y's. (so lagged z is not helpful).

Granger Causality

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} \quad \text{EXCLUDED UNDER H0} \quad + u_t$$

- The model above is the restricted model that we obtained under the null hypothesis.
- If the F test statistic is sufficiently large, we will reject the null hypothesis.
- If we reject H0 then we can say «z Granger-causes y» or «z is helpful in reducing the variance of one-step ahead forecasts»
- We can similarly test if y Granger-causes z. For this, we need to use the equation for z, and test the joint significance of lagged y's.

Evaluating forecast quality of one-step-ahead forecasts

- One can measure how good the forecasted values fit the actual values over the whole sample (in-sample criteria, e.g. R-squared).
- It is better, however, to evaluate the forecasting performance when forecasting out-of-sample values (out-of-sample criteria).
- For this purpose, use first n observations for estimation, and the remaining m observations to calculate forecast errors e_{n+h+1}.
- There are different forecast evaluation measures, e.g.

$$MAE = m^{-1} \sum_{h=0}^{m-1} \left| \hat{e}_{n+h+1} \right|, \quad RMSE = \left(m^{-1} \sum_{h=0}^{m-1} \hat{e}^2_{n+h+1} \right)^{1/2}$$

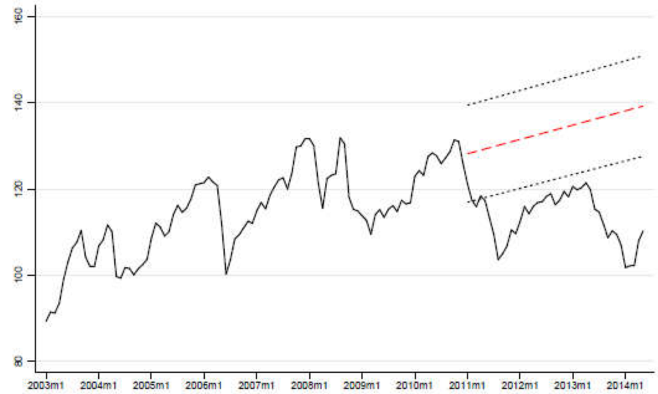
Mean absolute deviation (↑ MAE) Root mean square error (↑ RMSE)

Further topics in forecasting time series

- Multiple-step forecasts are possible, but necessarily less precise.
- Forecasts may make use of deterministic trends, but the error made by extrapolating time trends too far into the future may be large.
- Similarly, seasonal patterns may be incorporated into forecasts.
- Forecasting integrated time series is either based on regressions in levels, or on adding predicted changes (which are I(0)) to base levels.
- It is possible to calculate confidence intervals for the point forecasts.
- Forecast intervals for nonintegrated series converge to the unconditional variance, whereas for integrated series, they are unbounded.

Forecasts from a Linear Trend

$$\widehat{redk} = 108.88 + 0.08t$$

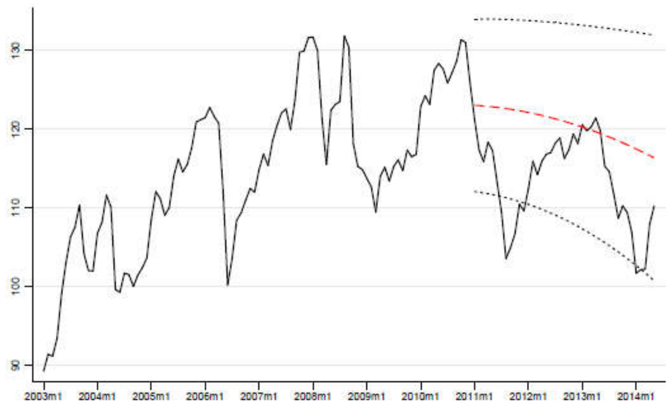


- Forecasts may make use of deterministic trends, but the error made by extrapolating time trends too far into the future may be large.
- Trend forecasts in the graph (red dotted line) starting at 2011.m1 overpredict systematically.

Forecasting Real Effective Exchange Rate (redk) using Linear Trend

Quadratic Trend Forecasts are not any better

$$\widehat{redk} = 95.53 + 0.656t - 0.0042t^2$$



- Red dotted line is the quadratic trend forecasts together with 95% confidence intervals.
- We must be especially cautious when we use polynomial trend forecasts

Forecasting Real Effective Exchange Rate using Quadratic Trend