Time series forecasting

Tyler Ransom

Univ of Oklahoma

Apr 4, 2019

Today's plan

1. How to forecast using time series
    1.1 Forecasting
    1.2 Granger Causality
    1.3 Forecast Intervals
    1.4 Advanced Forecasting Methods
2. In-class activity: Practice forecasting

Forecasting

Brief refresher on purposes of time series

- One purpose of investigating time series is to find **causality**
- e.g. does inflation cause lower unemployment?
- Another purpose is **forecasting**
- e.g. what will tomorrow's stock price (or next quarter's GDP) be?
- or even what will our company's web server load be on Black Friday?

Elements of Forecasting

- Any forecast of y_{t+1} requires the following components:
- **Information set:** includes y_t, lags of y_t, and current/lags of other variables
- **Forecast error:** $e_{t+1} = y_{t+1} - f_t$
- **Loss function:** how we choose to weight forecast errors, typically min e_{t+1}^2
- Then, we choose our forecast to minimize the expected loss:

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

Example

- When forecasting, it's best to use only lagged y's and other variables

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

  for example

- Letting T be our sample size, the forecast is

$$\hat{f}_T = \hat{\delta}_0 + \hat{\alpha}_1 y_T + \hat{\gamma}_1 z_T$$

  and the forecast error is

$$\hat{e}_{T+1} = y_{T+1} - \hat{f}_T$$

How to know which variables to include

- Which variables should be included in the model? How many lags?
- Start with an $AR(2)$ model and see if lags are significant
- Number of lags depends on frequency (i.e. annual vs. monthly vs. daily)
- Once we choose an AR model for y, we can do the same for z, etc.

Model selection

- A forecasting problem is one of **model selection**
- This is quite different from causal inference
- In model selection, use **out-of-sample criteria** to measure performance
- Most common out-of-sample statistic is **Root Mean Squared Error (RMSE):**

$$RMSE = \sqrt{\frac{1}{m}\sum_{h=0}^{m-1} \hat{e}^2_{T+h+1}}$$

Note: This is computed on *future* forecast errors

Measuring forecast performance

- The model with the **lowest RMSE** has the best forecast performance
- This is the model that should be used
- Could also use the one with the lowest Mean Absolute Error (MAE):

$$MAE = \frac{1}{m}\sum_{h=0}^{m-1} |\hat{e}^2_{T+h+1}|$$

- There are many other metrics that could be used (AIC, BIC, HQC, ...)
- Forecasting is a key component of **machine learning**; take my other class!

Vector Autoregressive (VAR) Model

- We could also **jointly** forecast y_t and z_t:

$$\begin{cases} y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 z_{t-1} + u_t^y \\ z_t = \eta_0 + \beta_1 y_{t-1} + \rho_1 z_{t-1} + u_t^z \end{cases}$$

- This is known as a **Vector Autoregressive (VAR) model**

- (y, z) form a system of equations (or “vector” of equations)

- VARs are widely used in macroeconomics

- $y =$ AAPL stock price, $z =$ GOOGL stock price

Granger Causality

Granger Causality

- Recall our VAR model from before:

$$\begin{cases} y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 z_{t-1} + u_t^y \\ z_t = \eta_0 + \beta_1 y_{t-1} + \rho_1 z_{t-1} + u_t^z \end{cases}$$

- This allows us to test if lagged z's affect y **holding fixed** lagged y's

- We say z **Granger causes** y if

$$E\left(y_t | I_{t-1}\right) \neq E\left(y_t | J_{t-1}\right)$$

  where I_{t-1} contains past information on y and z but J_{t-1} only has y

- Note: Granger causality says nothing about *contemporaneous* causality!

- Note: named for Sir Clive Granger, Nobel laureate

Forecast Intervals

Forecast Intervals

- From before, we used our model to obtain $\hat{f}_T$
- $\hat{f}_T$ is a **point forecast** of y_{T+1}
- $\hat{e}_{T+1} = y_{T+1} - \hat{f}_T$ is the forecast error
- We can compute the **forecast interval** as follows: (Why?)

$$se\left(\hat{e}_{T+1}\right) = \sqrt{\left[se\left(\hat{f}_T\right)\right]^2 + \hat{\sigma}^2}$$

$$\text{FI} \approx \hat{f}_T \pm 1.96 \cdot se\left(\hat{e}_{T+1}\right)$$

Properties of Forecast Intervals

- $se\left(\hat{f}_T\right)$ is usually much smaller than $\hat{\sigma}$
- Forecast intervals get increasingly wider with time (mostly because of $\hat{\sigma}$)
- Example: Yahoo (now Altaba) stock price tomorrow, next month, next year
- Example: Forecast ETA to OKC when driving from Norman vs. Stillwater

Advanced Forecasting Methods

Forecasting trending, seasonal, and integrated series

- It's easy to forecast with trending/seasonal/integrated processes
- Just plug in the values of the trend or seasonal dummies
- **Don't use a linear trend to forecast a random walk with drift**
- For I(1) processes, can forecast y_{t+1} through Δy_{t+1}
- Or just use a general AR or VAR model to forecast y_{t+1}

ARIMA models

- One well-known model for forecasting is the **ARIMA** model
- ARIMA = Auto Regressive Integrated Moving Average
- Hybrid of $AR(p)$, $I(d)$, and $MA(q)$ processes
- "$ARIMA(p, d, q)$ model" where p, d and q specify the time series process
- With just a few parameters, these models can forecast quite well

ARIMA equation

- The ARIMA model looks like:

$$E\left(y_t | I_t\right) = \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{k=1}^{q} \beta_k e_{t-k}$$

- And if y is $I(1)$, then

$$E\left(\Delta y_t | I_t\right) = \alpha_0 + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} + \sum_{k=1}^{q} \beta_k e_{t-k}$$

GARCH models

- We talked about GARCH models back when we covered serial correlation

- GARCH = Generalized Auto Regressive Conditional Heteroskedasticity

- GARCH models are interested in $V\left(y_t|I_t\right)$

- Allow the variance of y_t and of e_t to differ with t

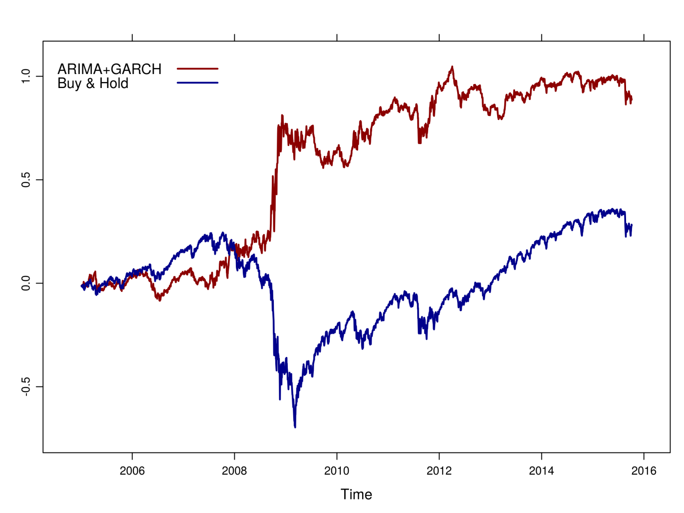
- Particularly useful when forecasting financial asset prices:

- Certain trading periods may have higher volatility

ARMA-GARCH models

- One can get an even better forecast by combining these two
- Use ARMA (or ARIMA) to forecast conditiona mean of y
- Use GARCH to forecast conditional variance of y

ARMA-GARCH vs. Buy-and-Hold since 2005



Source: Quantstart.com