Session 6
$\mathbb{Q}$uantitative Analysis of Financial Markets
Modeling and Forecasting Trend

Christopher Ting

http://www.mysmu.edu/faculty/christophert/

Christopher Ting

✉: christopherting@smu.edu.sg

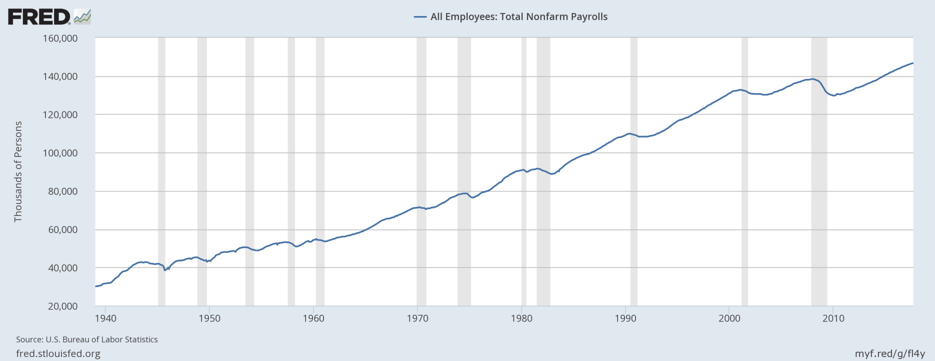
☎: 6828 0364

: LKCSB 5036

November 8, 2018

Table of Contents

An Example of Trend



- Is it a linear, quadratic, or generally nonlinear trend?
- Can the trend be modeled? How?
- How can we use the model to forecast future value?

QA-10 Modeling and Forecasting Trend

Chapter 5.

Francis X. Diebold, Elements of Forecasting, 4th Edition (Mason, Ohio: Cengage Learning, 2006).

- Describe linear and nonlinear trends.
- Describe trend models to estimate and forecast trends.
- Compare and evaluate model selection criteria, including mean squared error (MSE), s^2 (unbiased variance of residuals), the Akaike information criterion (AIC), and the Schwarz information criterion (SIC).
- Explain the necessary conditions for a model selection criterion to demonstrate consistency.

Modeling with TIME

- Model 1: Trend_t is a simple linear function of time

$$\text{Trend}_t = \beta_0 + \beta_1 \text{TIME}_t,$$

- The (deterministic) variable TIME is constructed artificially and is called a **time trend** or **time dummy**.
- $\text{TIME}_t = t$, where $t = 1, 2, \ldots, T$.
- β_0 is the regression intercept; it's the value of the trend at $t = 0$.
- β_1 is the regression slope
  - positive if the trend is increasing
  - negative if the trend is decreasing.

Quadratic

- Model 1: Trend_t is a quadratic function of time

$$\text{Trend}_t = \beta_0 + \beta_1 \text{TIME}_t + \beta_2 \text{TIME}_t^2.$$

- $\beta_1 > 0$ and $\beta_2 > 0$: the trend is monotonically increasing.
- $\beta_1 < 0$ and $\beta_2 < 0$: the trend is monotonically decreasing.
- $\beta_1 < 0$ and $\beta_2 > 0$: the trend has a U shape.
- $\beta_1 > 0$ and $\beta_2 < 0$: the trend has an inverted U shape.

Trend of Constant Growth

- If trend is characterized by constant growth at rate, then we can write
$$\text{Trend}_t = \beta_0 \, e^{\beta_1 \text{TIME}_t}.$$
- The trend is a nonlinear (exponential) function of time in levels, but in logarithms we have
$$\ln\left(\text{Trend}_t\right) = \ln(\beta_0) + \beta_1 \text{TIME}_t.$$
- Thus, $\ln\left(\text{Trend}_t\right)$ is a linear function of time.

Estimation

- Least squares proceeds by finding the argument (in this case, the value of θ) that minimizes the sum of squared residuals.
- Thus, the least squares estimator is the "argmin" of the sum of squared residuals function.

$$\widehat{\theta} = \arg\min_{\theta} \sum_{t=1}^{T} \big(y_t - \mathrm{Trend}_t(\theta)\big)^2,$$

where θ denotes the set of parameters to be estimated.

Estimation of Linear Trends

- Linear trend

$$\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = \underset{\widehat{\beta}_0, \widehat{\beta}_1}{\arg\min} \sum_{t=1}^{T} \left(y_t - \beta_0 - \beta_1 \mathsf{TIME}_t\right)^2.$$

- Quadratic trend

$$\left(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2\right) = \underset{\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2}{\arg\min} \sum_{t=1}^{T} \left(y_t - \beta_0 - \beta_1 \mathsf{TIME}_t - \beta_2 \mathsf{TIME}_t^2\right)^2.$$

- Log linear trend

$$\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = \underset{\widehat{\beta}_0, \widehat{\beta}_1}{\arg\min} \sum_{t=1}^{T} \left(\ln y_t - \ln \beta_0 - \beta_1 \mathsf{TIME}_t\right)^2.$$

Forecast

- The linear trend model, which holds for any time t, is

$$y_t = \beta_0 + \beta_1 \mathsf{TIME}_t + \epsilon_t.$$

- At time $T + h$, the future time of interest, the forecast made at T is

$$y_{T+h} = \beta_0 + \beta_1 \mathsf{TIME}_{T+h} + \epsilon_{T+h}.$$

- Key idea: TIME_{T+h} is known at time T, because the artificially-constructed time variable is perfectly predictable; specifically,

$$\mathsf{TIME}_{T+h} = T + h.$$

- The point forecast is for time $T + h$ and is made at time T.

$$y_{T+h,T} = \beta_0 + \beta_1 \mathsf{TIME}_{T+h}.$$

Forecast in Practice

- Replace unknown parameters with their least squares estimates, yielding the point estimate:

$$\widehat{y}_{T+h,T} = \widehat{\beta}_0 + \widehat{\beta}_1 \mathsf{TIME}_{T+h}.$$

- To form an interval forecast with 95% confidence, we use $\widehat{y}_{T+h,T} \pm 1.96\widehat{\sigma}$, where $\widehat{\sigma}$ is the standard error of the trend regression.

- To form a density forecast, we again assume that the trend regression disturbance is normally distributed

$$N\big(\widehat{y}_{T+h,T}, \widehat{\sigma}^2\big).$$

Motivating Questions and Problems

- How do we select among them when fitting a trend to a specific series?
- What are the consequences, for example, of fitting a number of trend models and selecting the model with highest R^2?
- It turns out that model-selection strategies such as selecting the model with highest R^2 do not produce good out-of-sample forecasting models.
- **In-sample overfitting and data mining**
  Including more variables in a forecasting model won't necessarily improve its out-of-sample forecasting performance

Mean Squared Error (MSE)

- Fitted Model 1

$$\widehat{y}_t := \widehat{\beta}_0 + \widehat{\beta}_1 \mathsf{TIME}_t.$$

- Residual

$$\widehat{e}_t := y_t - \widehat{y}_t,$$

- Definition of MSE

$$\mathsf{MSE} := \frac{\displaystyle\sum_{t=1}^{T} \widehat{e}_t^2}{T},$$

where T is the sample size.

R^2 and Mean Squared Error

- MSE's connection with R^2

$$R^2 = 1 - \frac{\displaystyle\sum_{t=1}^{T} \widehat{e}_t^2}{\displaystyle\sum_{t=1}^{T} \left(y_t - \overline{y}\right)^2}.$$

- Mean squared error corrected for degrees of freedom

$$s^2 = \frac{\displaystyle\sum_{t=1}^{T} \widehat{e}_t^2}{T - K},$$

where K is the total number of degrees of freedom used in model fitting.

R^2 and $\overline{R}^2$

- s^2 is just the usual unbiased estimate of the regression disturbance variance. That is, it is the square of the usual standard error of the regression.
- Connection with adjusted R^2

$$\overline{R}^2 = 1 - \frac{\dfrac{\displaystyle\sum_{t=1}^{T}\widehat{e}_t^2}{T-k}}{\dfrac{\displaystyle\sum_{t=1}^{T}\left(y_t-\overline{y}\right)^2}{T-1}} = 1 - \frac{s^2}{\dfrac{\displaystyle\sum_{t=1}^{T}\left(y_t-\overline{y}\right)^2}{T-1}}$$

- Note that the denominator depends only on the data want to fit, not the particular model fit.

Penalizing Degrees of Freedom

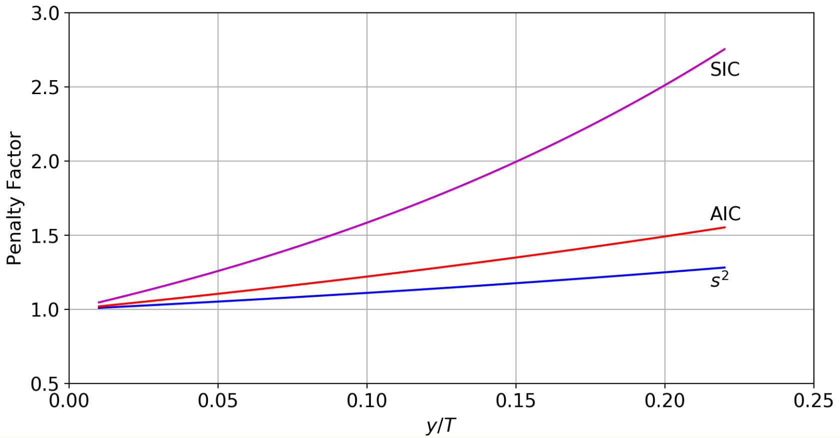
- We need to correct somehow for degrees of freedom K when estimating out-of-sample MSE on the basis of in-sample MSE.
- To highlight the degree-of-freedom penalty, rewrite s^2 as a penalty factor times the MSE,

$$s^2 = \left(\frac{T}{T-K}\right) \frac{\sum_{t=1}^{T} \widehat{e}_t^2}{T} = \frac{1}{1-\frac{K}{T}} \frac{\sum_{t=1}^{T} \widehat{e}_t^2}{T} = \frac{1}{1-\frac{K}{T}} \text{MSE}.$$

- Two very important such criteria are the Akaike information criterion (AIC) and the Schwarz information criterion (SIC). Their formulas are

$$\text{AIC} = e^{\frac{2K}{T}} \text{MSE} \quad \text{and} \quad \text{SIC} = T^{\frac{K}{T}} \text{MSE}.$$

Comparison of Information Criteria



Definition of Consistency

A model selection criterion is consistent if the following conditions are met:

A when the true model—that is, the **data-generating process (DGP)**—is among the models considered, the probability of selecting the true DGP approaches 1 as the sample size gets large

B when the true model is not among those considered, so that it's impossible to select the true DGP, the probability of selecting the best approximation to the true DGP approaches 1 as the sample size gets large.

Which Criterion Is Consistent?

- MSE is inconsistent, because it doesn't penalize for degrees of freedom.
- As T increases, s^2 becomes MSE, thus it is not a consistent model selection procedure.
- The AIC penalizes degrees of freedom more heavily than s^2, but it too remains inconsistent, even as the sample size gets large.
- The AIC selects models that are too large ("overparameterized").
- The SIC, which penalizes degrees of freedom most heavily, is consistent.

Asymptotic Efficiency

- What if both the true DGP and the best approximation to it are much more complicated than any model we fit?
- An **asymptotically efficient model selection criterion** chooses a sequence of models, as the sample size gets large, whose 1-step-ahead forecast error variances approach the one that would be obtained using the true model with known parameters at a rate at least as fast as that of any other model selection criterion.
- The AIC, although inconsistent, is asymptotically efficient, whereas the SIC is not.

Takeaways

- Trend as a function of time
- Forecast: point estimate, interval, density
- Model of in-sample fit: MSE, R^2
- Model selection criteria: $\overline{R}^2$, s^2, AIC, SIC
- Main desired property of a criterion: Consistency