Time Series Errors Overview

Time Series Residuals

Model:

$$y_t = \alpha + \beta x_t + e_t$$

Sample Estimates:

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{e}_t$$

- Need standard errors for $\hat{\alpha}$, $\hat{\beta}$.
- If e_t has heteroskedasticity or serial correlation/lag dependence, OLS standard errors are wrong
  - violates assumptions 3 and/or 4 of Gauss Markov
- If $x_t = y_{t-1}$ and e_t correlated with e_{t-1}, $\hat{\beta}$ is also biased.
  - violates assumptions 4 and 5 of Gauss Markov
- What can we do?

This Lecture

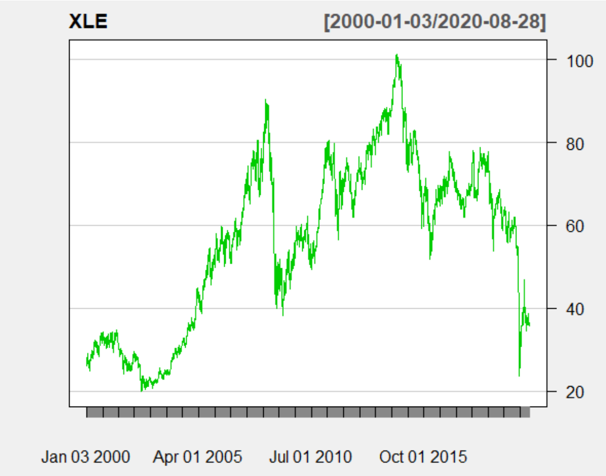
- Stationarity (briefly)
- Discuss OLS standard errors ("plain vanilla").
- 30,000-foot view of options if we have heteroskedasticity and/or serial correlation, with pros and cons of each.
  1. Alter the method of solving for $\hat{\alpha}$, $\hat{\beta}$.
  2. Adjust standard errors after solving for $\hat{\alpha}$, $\hat{\beta}$.
     - Heteroskedasticity-consistent (HC), robust, White standard errors.
     - Heteroskedasticity- and autocorrelation-consistent (HAC) or Newey-West standard errors.

For now assume all variables are *Stationary*

- *Strict Stationarity*: The joint distribution of sequential observations $(r_t, r_{t+1}, ..., r_{t+s})$ is identical to that of $(r_{t+k}, r_{t+k+1}, ..., r_{t+k+s})$ for any arbitrary time shift k.
- *Weak/Covariance Stationarity*: The mean, variance, and covariance with past values are constant through time, i.e., $E(r_t) = \mu$ (a constant) and $Cov(r_t, r_{t-s}) = \gamma_s$ only depends on the distance between the observations s and not time period t.

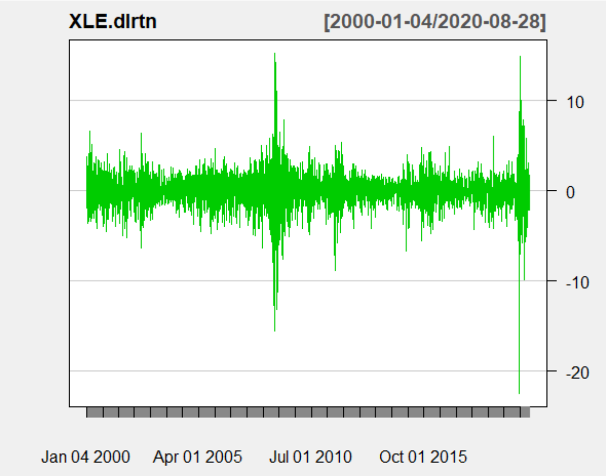
NONSTATIONARY

SPDR Energy ETF prices



STATIONARY

SPDR Energy ETF returns



OLS Standard Errors

Estimated model:

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{e}_t$$

- Assume y_t, x_t, e_t are stationary.
- Formula for OLS standard error of $\hat{\beta}$:

$$\widehat{se(\hat{\beta})} = \sqrt{\frac{\hat{\sigma}_e^2}{T \cdot \hat{\sigma}_x^2}}$$

- Will be small if large sample size (T), low error variance ($\hat{\sigma}_e^2$), large variance in X variable ($\hat{\sigma}_x^2$).
- With heteroskedasticity or autocorrelation, $\hat{\sigma}_e^2$ is underestimated.

Option 1: Alter your method (don't use OLS).

- Heteroskedasticity: different observations have different variance
- Movie recommendations: you have two friends. One always recommends an OK movie. The other recommends Black Panther, Terminator: Genisys, and Ice Age: Collision Course (some great, some awful).
- Attribute more noise (higher variance) to your second friend. Give him less **weight** in your calculation.
- Minimize sum of weighted squared residuals - "Generalized Least Squares" or GLS instead of OLS.

Option 1: Alter your method (don't use OLS).

OLS:

$$min_{\alpha,\beta} \sum_{t=1}^{T} e_t^2 = min_{\alpha,\beta} \sum_{t=1}^{T} (y_t - \alpha - \beta x_t)^2$$

GLS:

$$min_{\alpha,\beta} \sum_{t=1}^{T} w_t \cdot e_t^2 = min_{\alpha,\beta} \sum_{t=1}^{T} w_t \cdot (y_t - \alpha - \beta x_t)^2$$

- Give your unreliable friends a lower weight based on how noisy their recommendations are.
- Drawback: not clear what to use for the weights. Getting the weights wrong gives biased answers for $\hat{\alpha}$ and $\hat{\beta}$.

Option 1: Alter your method (don't use OLS).

- Serial correlation: residuals are correlated over time.
- Intuition: naive estimate ignores the fact that each new piece of information e_t is partly "old news", ρe_{t-1}.
- We think our model is more precise than it really is.
- Movie recommendations: every time you ask your friend, she recommends the latest comic book hero movie.
- Adjust your model of movie quality for the persistence in your friend's recommendations.

Option 1: Alter your method (don't use OLS).

- Include a specific model for the residual autocorrelation when solving for $\hat{\alpha}$ and $\hat{\beta}$. Use MLE.

$$y_t = \alpha + \beta x_t + e_t$$

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + u_t$$

- After conditioning on e_{t-1}, e_{t-2}, assume u_t is an i.i.d. residual
- Using MLE:

$$max_{\alpha,\beta,\rho_1,\rho_2} \sum_{t=1}^{T} ln\, f(u_t)$$

$$f(u_t) = f(y_t - \alpha - \beta x_t - \rho_1 e_{t-1} - \rho_2 e_{t-2})$$

- Where $f(\cdot)$ is the Normal distribution.

Option 1: Alter your method (don't use OLS).

- Drawback: again, if we get the wrong model for residual autocorrelation we can bias the estimates.
  - However, autocorrelation is often easier to model than heteroskedasticity.
- If we have lagged y_{t-p} in the model, we **must** model the serial correlation, otherwise we get biased coefficients.

$$y_t = \alpha + \beta y_{t-1} + e_t$$

$$e_t = \rho_1 e_{t-1} + u_t$$

- e_t is correlated with y_{t-1} because of e_{t-1}.

Option 2: Alter your standard errors after the fact.

- If only assumptions 3 and 4 fail, coefficients are **unbiased** but **inefficient**.
- Use OLS for coefficients, but then re-weight observations when calculating standard errors.
  - This is the most common approach.
- HC or White standard errors:
  - Run OLS, calculate the residuals.
  - Recalculate the standard errors, weighting by the squared OLS residuals.
- HAC or Newey-West standard errors:
  - Do the same procedure as HC, but adjust the weights for autocorrelation in the OLS residuals.

Option 2: Alter your standard errors after the fact.

Drawbacks:

- In both bases (heteroskedasticity and autocorrelation) naive/vanilla standard errors are too small - makes you overconfident in rejecting null hypotheses.
- But HC and HAC can be too big - you might fail to reject a null hypothesis that should have been rejected. This is a small concern.
- With HAC, still need to decide how much autocorrelation to include in the weights. But there are rules of thumb for this.
- However, if true model has no heteroskedasticity or autocorrelation, HC and HAC standard errors collapse to plain vanilla.
- Takeaway: almost always just use some version of robust standard errors (HC, HAC, or clustered if you have panel data).

Three cases of interest for residuals e_t

1. y_t and x_t are stationary, x_t does NOT include lags of y, and e_t has lag dependence.
2. y_t and x_t are stationary, x_t DOES include lags of y, and e_t has lag dependence.
3. y_t and/or x_t are unit root nonstationary. Interpretation depends on whether e_t is stationary or not.
   - *cointegrated* vs. *spurious*