

Time Series

John Green

April 2024

Outline

- ① Notes on macroeconomic data
- ② Applications
- ③ Lags
- ④ Autocovariance and autocorrelation
- ⑤ Forecasting

Macroeconomic data

- Volatility clustering
- Seasonal adjustments
- Data revisions and sources
- Unemployment rate
- Labor force participation rate

Uses of time series econometrics

Time series analysis is a highly valuable skill in industry:

- Forecasting (eg interest rates, returns to a stock, GDP growth)
- Dynamic treatment effects: impact of a policy after 1 month, 3 months, 2 years...
- Dynamic system analysis

But, we will have to deal with problems like correlation over time.

Time series data

- Generally similar to panel data, but with only one unit
- Takes the form $\{Y_t\}_{1 \leq t \leq T}$ with covariates $\{X_t\}_{1 \leq t \leq T}$
- For now, we assume timing is evenly spaced and there are no missing periods

Lags

- Lags are very important in time series analysis
- X_{t-1} is often a very good predictor of X_t
- More generally, the j^{th} lag of Y_t is Y_{t-j}
- We might take a difference: $\Delta_j Y_t = Y_t - Y_{t-j}$

Log variables

- Oftentimes we will use *log* variables instead of the raw values
- Two benefits:
 - Makes exponential growth linear
 - Differences in logs are (approximately) equal to the percentage change:

$$\log(1 + \alpha) \approx \alpha$$

So long as α is “small”

- So, $\log X_t - \log X_{t-1} = \log \frac{X_t}{X_{t-1}}$ is roughly the percentage change in X from $t - 1$ to t

Autocorrelation

- Just like with panel data, we will often have correlation between lagged values
- First autocovariance of Y_t : $\text{cov}(Y_t, Y_{t-1})$
- First autocorrelation of Y_t : $\text{corr}(Y_t, Y_{t-1}) \equiv \rho_1$
- Can calculate for any lag j in the same way and get the j^{th} serial correlation coefficient

Forecasting and stationarity

- We will often be interested in making an out-of-sample (OOS) forecast of our variable
- This is similar to what we did before trying to target the oracle prediction
- We no longer care about causal interpretation of coefficients, OVB, etc.; we just want to get an accurate OOS prediction
- This requires that our data be *stationary* (or jointly stationary)
 - Intuitively, our out-of-sample data needs to be similar to our sample Data
 - Technically, we need the joint distribution across time to be independent of the time period we are looking at

Forecasting error

- An s periods ahead forecast, using estimated coefficients based on $t = 1, \dots, T$, is:

$$\hat{Y}_{T+s|T}$$

- The forecast error is:

$$Y_{T+s} - \hat{Y}_{T+s|T}$$

- And we might evaluate our model based on the mean squared forecasting error:

$$MSFE = \mathbb{E} \left[\left(Y_{T+s} - \hat{Y}_{T+s|T} \right)^2 \right]$$

Autoregression

- Regress Y_t on its past (lagged values)
- p^{th} order autoregression is Y_t on Y_{t-1}, \dots, Y_{t-p} , denoted $AR(p)$:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + u_t$$

- Not causal!
- Can easily test if value in period $t - s$ is useful with t-test or multiple periods with F-test
- For the $AR(1)$ model $Y_t = \beta Y_{t-1} + u_t$, where $|\beta| < 1$ and $u = 0$, covariance and correlation between t and $t - s$ are easy:

$$\text{corr}(Y_t, Y_{t-s}) = \beta^s$$

Moving averages

- Moving average model:

$$Y_t = \epsilon_t + \theta \epsilon_{t-1}$$

- Make substitutions:

$$Y_t = \epsilon_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \dots$$

- Some nice properties:

- $\text{Var}(Y_t) = \sigma^2(1 + \theta^2)$
- $\text{cov}(Y_t, Y_{t-1}) = \theta\sigma^2$
- $\text{cov}(Y_t, Y_{t-j}) = 0 \quad \forall j > 0$
- unfortunately, we can't estimate with OLS; have to use NLS
- Box-pierce test for serial correlation: H_0 is the series is white noise

- We can add in other variables (and their lags) to form an *autoregressive distributed lag* (ADL) model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-r} + u_t$$

with p lags of Y and r lags of X , denoted $ADL(p, r)$.

Forecast error

- Sometimes we look at the root mean squared forecast error, which gives us a “typical” error:

$$RMSFE = \sqrt{E \left[\left(Y_{T+1} - \hat{Y}_{T+1|T} \right)^2 \right]}$$

- Again, MSFE has two components:

$$MSFE = E(u_{T+1})^2 + E \left[\left(\hat{\beta}_0 - \beta_0 \right) + \left(\hat{\beta}_1 - \beta_1 \right) Y_T + \left(\hat{\beta}_2 - \beta_2 \right) X_T \right]^2$$

- 1 Unavoidable oracle uncertainty (first term)
- 2 Uncertainty from estimating coefficients (second term)

Forecast error

Three methods for estimating the MSFE:

- 1 If number of regressors is small, use the in-sample regression error, $\widehat{MSFE}_{SER} = s_{\hat{u}}^2 = \frac{SSR}{T-p-1}$ which ignores the uncertainty from $\hat{\beta}$
- 2 If errors are homoscedastic, we can use this approximation for the full MSFE:

$$\widehat{MSFE}_{FPE} = \left(\frac{T+p+1}{T} \right) s_{\hat{u}}^2 = \left(\frac{T+p+1}{T-p-1} \right) \left(\frac{SSR}{T} \right)$$

- 3 Method with fewest assumptions: pseudo out of sample. Create a rolling estimate through $t = s$, calculate error for $t = s+1, \dots, T$, repeat; $\widehat{MSFE}_{POOS} = \frac{1}{P} \sum_{s=T-P+1}^T \tilde{u}_s^2$

Number of lags

- How can we choose number of lags p ?
- We can use a bunch of t or F tests... but model will usually be too large
- *Information criterion* helps us select by allowing more bias but lower variance (kind of like with the shrinkage estimators)

$$\text{AIC}(p) = \ln \left(\frac{\text{SSR}(p)}{T} \right) + (p+1) \frac{2}{T}$$
$$\text{BIC}(p) = \ln \left(\frac{\text{SSR}(p)}{T} \right) + (p+1) \frac{\ln T}{T}$$

- BIC is consistent; AIC is not but may be good if you think more lags are more important
- Can generalize to ADL model

Misc. forecast topics

- Forecast evaluation
- Combining forecasts
- Vector autoregression (VAR):

$$\begin{aligned}Y_t &= \beta_{01} + \beta_{11}Y_{t-1} + \dots + \beta_{1p}Y_{t-p} + \gamma_{11}X_{t-1} + \dots + \gamma_{1p}X_{t-p} + u_{1t} \\X_t &= \beta_{02} + \beta_{21}Y_{t-1} + \dots + \beta_{2p}Y_{t-p} + \gamma_{21}X_{t-1} + \dots + \gamma_{2p}X_{t-p} + u_{2t}\end{aligned}$$

- Granger causality
- Direct vs. iterative prediction

Asset returns

- Two features:
 - ① Volatility clustering
 - ② Fat tails
- ARCH model:

$$Y_t = \mu + u_t$$

$$u_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

- Generalized ARCH (GARCH):

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \phi_1 \sigma_{t-1}^2$$

- These capture both of our desired features

Stationarity

- All these models are assuming *stationarity*
- First threat is a *trend*, a persistent long-term movement which might be deterministic or stochastic
 - $Y_t = \beta_0 + Y_{t-1} + u_t$ where u_t is serially correlated and β_0 gives the drift
- If Y_t is a random walk, then ΔY_t is stationary
- Trends may cause:
 - ① Coefficients biased to 0; bad forecast
 - ② t-test doesn't work
 - ③ Spurious correlation between 2 random walks

Dickey-Fuller Test

- Dickey-Fuller test has H_0 of random walk vs. H_A of stationarity
- Use DF table for critical value
 - Depends if model has intercept only or also has a time trend
- If DF fails to reject and has a unit root, use first differences model

Breaks

- Second threat are structural breaks where the model itself (coefficients) changes over time
- If we suspect a break at a specific point we can test this using dummies for before and after the break

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} \\ + \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) \times Y_{t-1}] + \gamma_2 [D_t(\tau) \times X_{t-1}] + u_t$$

- Use a Chow test statistic on γ
- If we don't know when the break happens we can test dates on $\underline{t} \leq t \leq \bar{t}$ and take the largest Chow statistic (QLR statistic)
- The more QLR stats we compute, the larger the critical value

Dynamic causal effects

- So far we've been focused on *forecasting*, where we care about prediction but not inference
- But we might also want to know about the causal effect of a policy in a time series context, where there may be different impacts across time
 - Effect of a tax cut on GDP growth after 1 month, 1 year, 1 decade...
- Problems: we only have one subject, and we can't randomize the treatment
- So, let's think about the same entity receiving a treatment at different dates
 - This is possible if our data is stationary

Distributed lag estimator

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \dots + \beta_r X_{t-r} + u_t$$

- β_1 is the impact of a change in X_t on Y_t , just like normal
- β_2 is the 1-period dynamic multiplier, giving the effect of a change in last period's value X_{t-1} on today's outcome Y_t
- When we are considering changing X_t , we might want to know what its total impact will be over time; this is given by the **cumulative multiplier**:

$$\gamma_r = \beta_1 + \beta_2 + \dots + \beta_r$$

where γ_r gives the total effect over the next r periods from changing the value of X_t in this period

Distributed lag model assumptions

- ① Exogeneity: $\mathbb{E}[u_t | X_t, X_{t-1}, \dots] = 0$
 - Intuition: if we know the full past, we can't predict the error
 - Might require *strict exogeneity* which also conditions on the future, but let's not worry about threat
- ② $\{X_t, Y_t\}$ and $\{X_{t+j}, Y_{t+j}\}$ are independent when j is large
 - Intuition: two events which are far enough apart in time are not connected
- ③ $\{X_t, Y_t\}$ are both stationary (similar to the iid requirement, provides internal + external validity)
- ④ Large outliers are rare (up to 8 finite moments!)
- ⑤ No perfect multicollinearity

Standard errors

- If we have autocorrelation, standard errors will be incorrect
- Use SEs which are robust to heteroskedasticity *and* to autocorrelation, called Newey-West or HAC standard errors
- Depend on how far apart we think autocorrelation extends, denoted by the truncation parameter m
 - No magic for choosing this: should not be too big (then SEs won't work) but not too small (then SEs are not believable)
 - Rule of thumb, $m = 0.75 T^{1/3}$
- We don't need these in AR and ADL models because we include lags of Y itself so that error should be uncorrelated

Cumulative multiplies

- We defined the cumulative effect of a change in X_t after r periods as:

$$\gamma_r = \beta_1 + \beta_2 + \dots + \beta_r$$

- But this does not give us the SEs for γ_r
- Simple solution: rewrite the model in terms of changes in X and γ_i coefficients:

$$Y_t = \gamma_0 + \gamma_1 \Delta X_t + \gamma_2 \Delta X_{t-1} + \dots + \gamma_{q-1} \Delta X_{t-q+1} + \gamma_q X_{t-q} + u_t$$

- Then we can just estimate this thing with OLS and use the HAC SEs