Recitation 10

Seung-hun Lee

Columbia University

December 13th, 2021

Time series: Setting up the approach

- Collect data on same observational unit i for multiple time periods.
- Primary uses: Forecasting, modeling risks, and analyzing dynamic causal effects
- Time series differs in that errors are likely to be autocorrelated and thus require different ways to calculate the standard error.
- Let Y_t be the time series data captured at certain period t - GDP
- **Lags** are characterized as Y_{t-1} and **leads** are defined as Y_{t+1}.
- $\Delta Y_t \equiv Y_t - Y_{t-1}$: The **first difference** at time t.

AR and ADL: Number of lags for dependent and independent variables

- $AR(p)$: Y_t is regressed against its own lagged values by p times:

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

  - Each coefficient β_k indicates how past values are useful in forecasting
- $ADL(p, q)$: p lags of dependent variable and q lags for X variable

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + ... + \beta_p Y_{t-p} + \delta_1 X_{t-1} + ... + \delta_q X_{t-q} + u_t$$

AR and ADL: Selecting number of lags

- Right amount of p and q minimizes the following **information criteria**

$$AIC : \ln\left(\frac{SSR(p,q)}{T}\right) + (K)\frac{2}{T} \quad BIC : \ln\left(\frac{SSR(p,q)}{T}\right) + (K)\frac{\ln T}{T}$$

  where $K = 1 + p + q$

- **Granger causality**: Test that helps us see whether X is useful in predicting Y

$$H_0 : \delta_1 = ... = \delta_q = 0, \quad H_1 : \neg H_0$$

  If the null hypothesis is rejected, we say that X *Granger-causes* Y

Forecasting: What is a good forecast?

- **Forecast inteval** captures how accurate your forecasts are.
- Assume an ADL(1,1) type of equation, with forecast error $Y_{T+1} - \hat{Y}_{T+1|T}$, or

$$[(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)Y_T + (\delta_1 - \hat{\delta}_1)X_T] + u_{T+1}$$

- **Mean squared forecast error** (MSFE): standard errors used to create forecast interval

$$\begin{aligned} E[(Y_{T+1} - \hat{Y}_{T+1|T})^2] &= E[u_{T+1}^2] + E\left[\Big((\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)Y_T + (\delta_1 - \hat{\delta}_1)X_T)\Big)^2\right] \\ &= var(u_{T+1}) + var[(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)Y_T + (\delta_1 - \hat{\delta}_1)X_T] \end{aligned}$$

- As sample size increases, the uncertainty part (variance term), converges to 0. So MSFE is approximately equal to $var(u_{T+1})$.

Forecasting: How volatile is the forecast?

- **Root mean squared forecast error** (RMSFE) is just $\sqrt{E[(Y_{T+1} - \hat{Y}_{T+1|T})^2]}$
- Measures the spread of the forecast error distribution (magnitude of an 'error in forecasting')
- In practice, we estimate RMSFE using either SER (from several lectures ago) or based on actual forecast history for $t = t_1, .., T$ and get

$$MSFE = \frac{1}{T - t_1 + 1} \sum_{t=t_1-1}^{T-1} (Y_{t+1} - \hat{Y}_{t+1|t})^2$$

- **Forecast interval**: The bounds for the forecast intervals if size 95% is constructed as

$$\hat{Y}_{T+1|T} \pm 1.96 \times \widehat{RMSFE}$$

where $\widehat{RMSFE}$ is obtained using one of the two methods mentioned.

Stationarity: Is the distribution of the data stable across time or not?

- Stationary: Distribution of $(Y_{t+1}, .., Y_{t+s})$ does not depend on t.
- In other words, the distribution of Y does not change over time
- Nonstationary: When there is a trend or a break in the movement of the data (or any change in underlying parameters),
- Trends
  - **Deterministic trend** is a nonrandom function of time, $(Y_t = \alpha t^2)$
  - **Stochastic trend** is random, and time-variant distribution, such as the random walk $Y_t = Y_{t-1} + u_t$ (You can check that $var(Y_t) = t\sigma_u^2$)
  - Any other case where $\beta_1 > 1$ is also nonstationary

Stationarity: Testing for this in AR(1)

- Eyeball test: Check graphically
- Dickey-Fuller test: Check for the existence of a 'unit root' by testing

$$H_0 : \beta_1 \geq 1, \;\; H_1 : \beta_1 < 1$$

- For a general case with $AR(p)$,

$$\begin{aligned}
Y_t &= \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + u_t \\
&= \beta_0 + \beta_1 Y_{t-1} - \beta_2 Y_{t-1} + \beta_2 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t \\
&\ldots \\
&= \beta_0 + (\beta_1 + \ldots + \beta_p) Y_{t-1} - (\beta_2 + \ldots + \beta_p)\Delta Y_{t-1} - \ldots - \beta_p \Delta Y_{t-p+1} + u_t
\end{aligned}$$

  - Test whether $H_0 : \beta_1 + \ldots + \beta_p \geq 1, \;\; H_1 : \beta_1 + \ldots + \beta_p < 1$

Stationarity: Testing structural breaks

- Assume a ADL(1,1) structure, but that we know when the structural break occurs at year τ
- Let $D_t(\tau) = 1$ if year $t \geq \tau$ and 0 otherwise.
- Then we write the equation as

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} + \gamma_1 D_t(\tau) + \gamma_2 D_t(\tau) Y_{t-1} + \gamma_3 D_t(\tau) X_{t-1} + u_t$$

- To check for structural break, test joint hypothesis of the following form:

$$H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0, \ H_1 : \neg H_0$$

  This is the idea behind the **Chow test**.
- If structural break is unknown, we can do a **Quandt Likelihood Ratio test** that implements multiple Chow tests and finds the point where structural break most likely happened, if it occurred.

Dynamic Causal analysis with distributed lag models

- **Dynamic causal effect** captures the effect of X on Y over time.
- Write the distributed lag model as

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + ... + \beta_p X_{t-p} + u_t$$

- β_0 captures the contemporaneous impact of X on Y, holding past values of X constant.
- $\beta_j, j \in [1, p]$ captures the impact of X from j period(s) ago on Y, holding X from other periods constant

Dynamic Causal analysis: Implementation

- Cumulative effect: Cumulative effect can be captured by summing over multiple β's
- Specifically, we can write

$$
\begin{aligned}
Y_t &= \alpha + \beta_0 X_t + \beta_1 X_{t-1} + u_t \\
&= \alpha + \beta_0 X_t - \beta_0 X_{t-1} + \beta_0 X_{t-1} + \beta_1 X_{t-1} + u_t \\
&= \alpha + \beta_0 \Delta X_t + (\beta_0 + \beta_1) X_{t-1} + u_t
\end{aligned}
$$

- Assumptions
  - (Sequential) Exogeneity: $E[u_t | X_t, X_{t-1}, ..., X_1] = 0$. Or that error terms should not be correlated with current and past values of X
  - Stationarity: Y and X should have stationary distributions and (Y_t, X_t) and (Y_{t-j}, X_{t-j}) becomes independent as j gets large.
  - Y and X has nonzero finite moments
  - There is no perfect multicollinearity

Dynamic Causal analysis: Standard errors

- Given that there is a possibility that autocorrelation can exist, we need a standard error that takes into account autocorrelation and heteroskedasticity.
- This is known as **heteroskedasticity and autocorrelation consistent** errors (HAC errors).
- The takeaway is that standard errors in the typical STATA output can be wrong and we need to take a slightly different approach.
- Use `newey` in STATA, with m lags for standard errors

$$m = 0.75 \times T^{1/3}$$

where T is the total time periods in the data