EC508: Econometrics
Time-Series Regression and Forecasting

Jean-Jacques Forneron

Spring, 2023

Boston University

Time-Series in Economics

- Time-Series are a type of data indexed over time: $Y_1, \ldots, Y_T$ or $Y_t, t = 1, \ldots, T$
- e.g. GDP/Unemployment every quarter/year since 1950
- We are interested in mainly two things:
  1. Forecasting
     What is our best prediction for unemployment next quarter? An important problem for setting interest rates in Central Banks
  2. Estimating dynamic causal effects
     What is the effect of an oil price shock on GDP/inflation? To choose a policy (interest or exchange rate) we need to know the effect of the shock we want to counter

Regression Model for Time-Series (SW 14)

- We will mainly be interested in the following regression model:

$$\begin{aligned} Y_t =& \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} \\ & + \delta_{1,1} X_{1,t-1} + \cdots + \delta_{1,q_1} X_{1,t-q_1} \\ & + \cdots + \delta_{k,1} X_{k,t-1} + \cdots + \delta_{k,q_k} X_{k,t-q_k} + u_t \end{aligned}$$

- Y_{t-1} is called a lagged variable: value of Y_t in the previous period; e.g. last quarter
- In practice, usually pick $q_1 = \cdots = q_k = q$
- With X: $ADL(p, q)$ model; without X: $AR(p)$ model
- The past values of Y_t can affect its current value (persistence) but also other factors $X_1, \ldots, X_k$ and their past
- Since the data is correlated over time (Y_t depends on the past), we a concept that replaced i.i.d. sampling

Beyond i.i.d. random sampling

- To allow for the persistence in time-series data, we introduce two concepts which are key to getting asymptotic results (LLN, CLT):
  1. Stationarity:
     The distribution of $(Y_t, Y_{t+1}, \ldots, Y_{t+s}, X_t, \ldots, X_{t+s})$ **does not depend on** t **for any** $s \geq 0$
  2. Weak Dependence:
     $(Y_t, X_{1,t}, \ldots, X_{k,t})$ and $(Y_{t+s}, X_{1,t+s}, \ldots, X_{k,t+s})$ **become independent** as s becomes large
- Under these assumptions Y_t and Y_{t+s} are nearly independent and have the same distribution; this is almost like a random sample.
- This is the key idea to proving LLNs and CLTs in this setting

OLS estimation of the ADL/AR model

- The coefficients β, δ are estimated by OLS

$$
\begin{aligned}
Y_t = &\beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} \\
&+ \delta_{1,1} X_{1,t-1} + \cdots + \delta_{1,q_1} X_{1,t-q_1} \\
&+ \cdots + \delta_{k,1} X_{k,t-1} + \cdots + \delta_{k,q_k} X_{k,t-q_k} + u_t
\end{aligned}
$$

- The estimator is consistent and asymptotically normal under the following assumptions:
    1. $\mathbb{E}(u_t | Y_{t-1}, \ldots, X_{1,t-1}, \ldots) = 0$
    2. Stationarity and Weak Dependence
    3. Large outliers are unlikely $Y_t, X_{1,t}, \ldots, X_{k,t}$ have finite non-zero fourth moments
    4. There is no perfect multicollinearity
- NB: for $AR(p)$ models $\mathbb{E}(u_t | Y_{t+s}) \neq 0$ for $s \geq 0$ in general

The asymptotic variance of the OLS estimator

- Recall that for time-series Y_t can be correlated with its past (this may also be the case for u_t)
- This has implications for the asymptotic variance of sample means (CLT) and, in turn, for the OLS estimator itself
- We'll start with the sample mean: $\bar{Y} = 1/T \sum_{i=1}^{T} Y_t$
- For $T = 2$, we get:

$$
\begin{aligned}
& var([Y_1 + Y_2]/2) \\
& = [var(Y_1) + var(Y_2) + cov(Y_1, Y_2) + cov(Y_2, Y_1)]/4 \\
& = [var(Y_1) + cov(Y_1, Y_2)]/2
\end{aligned}
$$

- the last equality comes from the stationarity of Y_t

The asymptotic variance of the OLS estimator, cont'd

- For general T, we get:

$$var(\bar{Y}) = \frac{1}{T} var(Y_t) + \frac{2}{T} \sum_{s=1}^{T-1} \left(\frac{T-s}{T} \times cov(Y_t, Y_{t+s}) \right)$$

- Remark: if $cov(Y_t, Y_{t+s}) = 0$ for all $s \geq 1$ then we get the usual i.i.d. variance formula
- Under stationarity and weak dependence, we have:

$$\sqrt{T}(\bar{Y} - \mathbb{E}(Y_t)) \xrightarrow{d} \mathcal{N}(0, V_{LR})$$

- Where V_{LR} is the Long-Run Variance:

$$V_{LR} = var(Y_t) + 2 \sum_{j=1}^{+\infty} cov(Y_t, Y_{t+s})$$

- $\sum_{j=1}^{+\infty} cov(Y_t, Y_{t+s}) > 0$ implies that the asymptotic variance of the estimator is larger than in the i.i.d. case

Estimating the Long-Run Variance: HAC

- Since V_{LR} is unknown, it needs to be estimated
- A popular estimator is the HAC estimator;
  Heteroskedasticity and Autocorrelation Consistent estimator
- A simple version in the Newey-West HAC estimator:

$$\hat{V}_{LR} = \hat{V}_0 + 2\sum_{j=1}^{m-1} \frac{m-j}{m} \hat{V}_j$$

- where $1 \leq m \leq T$; rule of thumb (SW): $m \simeq 0.75 \times T^{1/3}$ and

$$\hat{V}_j = \frac{1}{T-j} \sum_{t=1}^{T-j} (Y_t - \bar{Y})(Y_{t+j} - \bar{Y})'$$

is an estimator of $cov(Y_t, Y_{t+j})$

OLS: Consistency and Asymptotic Normality

- We'll use the following notation $X_t' = (1, Y_{t-1}, \ldots, X_{t-p}, X_{1,t-1}, \ldots, X_{k,t-q_k})$ and $\beta = (\beta_0, \ldots, \delta_{k,q_k})$; for simplicity $p \geq q_j$, $j = 1, \ldots, k$
- We have the following:

$$\hat{\beta} - \beta = \left(\frac{1}{T} \sum_{t=1+p}^{T} X_t X_t' \right)^{-1} \frac{1}{T} \sum_{t=1+p}^{T} \underbrace{X_t u_t}_{=v_t}$$

- Same as usual: *CLT* for $\bar{v}$ and LLN for lhs yields:

$$\sqrt{T} \left(\hat{\beta} - \beta \right) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

- where $\Sigma = Q_{XX}^{-1} V_{LR} Q_{XX}^{-1}$ and V_{LR} is the LR variance of v_t

Application: Economic Report of the President 1984

```r
library(sandwich)
library(lmtest)
data(Investment)

model <- lm(RealInv ~ RealGNP + RealInt,
   data = Investment)
coeftest(model,vcov=NeweyWest,lag=2)
```

Application: Economic Report of the President 1984

	Estimate	Std. Error	t value	Pr(>—t—)
(Intercept)	−12.534	25.326	−0.495	0.627
RealGNP	0.169	0.024	6.991	0.000
RealInt	−1.001	3.499	−0.286	0.778

- Replicate a result from:
  Executive Office of the President (1984), Economic Report of the President. US Government Printing Office, Washington, DC.
- Conclusions: Investment driven by real gross national product, real interest rate had no significant effect on aggregate real investments

Granger Causality

- Tests proceed the same way: t and Wald statistics allow to perform inference on the coefficients
- Granger Causality: we say that X_j Granger Causes Y_t if $(\delta_{j,1}, \ldots, \delta_{j,q_j}) \neq 0$ i.e. X_j has predictive content on future observations of Y_t, after controlling for past $Y_{t-1}, \ldots$
- Testing for Granger Causality, implemented using a Wald test:

$$H_0 : \delta_{j,1} = \cdots = \delta_{j,q_j} = 0 \text{ vs. } \delta_{j,\ell} \neq 0$$

for some ℓ in $\{1, \ldots, q_j\}$