0 0 ECON0022 Metrics for Macro and Finance Index

Table of Contents

Table of Contents

Week 1: Basic Time Series and AR Models

1 1 Basic TS Notations & Autocorrelations

Notations

- Basic TS Notations
- Lags, First Difference, Percentage Change

$$\%\Delta Y_t \approx 100\Delta \log(Y_t) = 100(\log Y_t - \log Y_{t-1})$$

Autocorrelations

- Population Autocorrelations

$$\begin{aligned}\rho_j &= \frac{Cov(Y_t, Y_{t-j})}{\sqrt{Var(Y_t) \cdot Var(Y_{t-j})}} \\ &= \frac{Cov(Y_t, Y_{t-j})}{Var(Y_t)}\end{aligned}$$

- Sample estimations of Autocorrelations

$$\hat{\rho}_j = \frac{\widehat{Cov}(Y_t, Y_{t-j})}{\widehat{Var}(Y_t)} = \frac{\frac{1}{T}\sum_{t=j+1}^{T}(Y_t - \bar{Y})(Y_{t-j} - \bar{Y})}{\frac{1}{T}\sum_{t=1}^{T}(Y_t - \bar{Y})^2}$$

  where $\bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$
  - This always offers a *consistent* estimator of autocorrelations as long as we have stationarity and mixing, but without enough lags as the actual DGP, we need HAC SE
- (PS1) Partial Autocorrelations: ϕ_p in the regression model:

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t$$

  It measures the correlation between Y_t and Y_{t-p} after controlling for Y_t's correlation with $Y_{t-1} \dots Y_{t-p+1}$

1 2 OLS Estimators in TS Models

- Casual Inference / Forecasting: OLS gives you the best linear prediction ($\min MSFE$) regardless of actual DGP

OLS Estimator in AR(1)

Intro and Estimation

- Intro:

$$Y_t = \mu + \phi Y_{t-1} + \epsilon_t, \ \ t = 2, 3, \ldots, T$$

- Solving backward:

$$Y_t = \mu \sum_{i=0}^{t-2} \phi^i + \phi^{t-1} Y_1 + \sum_{i=0}^{t-2} \phi^i \epsilon_{t-i}$$

- OLS Estimator in AR(1)

$$\hat{\phi}_{AR1} = \frac{\sum_{t=2}^{T} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y}_{-1})}{\sum_{t=2}^{T} (Y_{t-1} - \bar{Y}_{-1})^2}$$

  where $\bar{Y} = \frac{1}{T-1} \sum_{t=1}^{T} Y_t$ and $\bar{Y}_{-1} = \frac{1}{T-1} \sum_{t=2}^{T} Y_{t-1}$
  - $\hat{\phi}_{AR1} \approx \hat{\rho}_1$ (we distinguish $\bar{Y}, \bar{Y}_{-1}$ here, but not in $\hat{\rho}_1$)
  - Note that:

$$\hat{\mu} = \bar{Y} - \hat{\phi} \bar{Y}_{-1} \approx (1 - \hat{\phi}) \bar{Y} \iff \bar{Y} \approx \frac{\hat{\mu}}{1 - \hat{\phi}}$$

- (PS2) A general property for AR(1):

$$\rho_Y(k) = \phi^k$$

Forecasting in AR(1)

- 1-period ahead forecast using population coefficients: $Y_{T+1|T} = \mu + \phi Y_T$
- Feasible 1-period ahead forecast using estimated coefficients: $\hat{Y}_{T+1|T} = \hat{\mu} + \hat{\phi} Y_T$
- 1-period ahead forecast error: $Y_{T+1} - \hat{Y}_{T+1|T}$
  - Forecast errors are out-of-sample (unobserved); residuals are in-sample (observed)

OLS Estimator in AR(p)

Intro and Estimation

- Intro:

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t \quad t = p+1, \ldots, T$$

- OLS Estimator in AR(p)

$$\hat{\phi}_i = \frac{\sum_{t=p+1}^{T} \hat{r}_{i,t} Y_t}{\sum_{t=p+1}^{T} \hat{r}_{i,t}^2}$$

  where $\hat{r}_{i,t} = Y_{i,t} - \sum_{j \neq i} \hat{\alpha}_j Y_{j,t}$ are first-stage residuals (partialling out other lags)

Forecasting in AR(p)

- Feasible 1-period ahead forecast:

$$\hat{Y}_{T+1|T} = \hat{\mu} + \sum_{i=1}^{p} \hat{\phi}_i Y_{T+1-i}$$

Week 2: Large-Sample & Inference in Stationary AR Models

2 1 Asymptotic Theories for Stationary & G.Mixing TS

TS Concepts

Stationarity

- Definition: A time series $\{Y_t : t = 1, 2, \dots\}$ is stationary if its probability distribution does not change over time:

$$(Y_s, Y_{s-1}, \dots, Y_{s-p}) =^d (Y_t, Y_{t-1}, \dots, Y_{t-p}) \;\forall\; s, t = 1, 2, \dots$$

- Implications:
  - $\forall s, t$, Y_s and Y_t have the same joint distribution:

  $$\mathbb{E}[Y_s] = \mathbb{E}[Y_t], Var[Y_s] = Var[Y_t] \;\forall\; s, t$$

  - (Y_s, Y_{s-k}) and (Y_t, Y_{t-k}) have the same joint distribution:

  $$Cov(Y_s, Y_{s-k}) = Cov(Y_t, Y_{t-k}) \;\forall\; s, t, k$$

  - $\bar{Y} = \frac{\sum_{t=1}^{T} Y_t}{T}$ is an unbiased estimator of $\mathbb{E}[Y_t]$, but NOT necessarily consistent (data can be highly autocorrelated $\rightsquigarrow Var(\hat{Y}) \to \infty$)
- Theorem:
  - If a time series $\{Y_t\}$ is stationary:
    - Any transformation $f(Y_t)$ will also be stationary
    - For any $p \geq 1$, $Z_t = (Y_t, Y_{t-1}, \dots, Y_{t-p})$ will also be stationary
- Stationarity allows for (high) autocorrelations, iid is a special case of stationarity
- A pair of time series $\{Z_t\} = \{(X_t, Y_t)\}$ is stationary if $\forall\, p \geq 0$, the joint distribution of $(Z_t, Z_{t-1}, \dots, Z_{t-p})$ does not depend on t

Mixing / Weakly Dependent

- Mixing restricts the dependence in data, requiring $\{Y_t\}$ to have "short memory"
- Implication:
  - Long run forecasts $\to^p$ unconditional mean: for a function of y ($\phi(Y)$):

  $$\phi(Y_{T+k|T}) = \mathbb{E}[\phi(Y_{T+k})|Y_T, Y_{T-1} \dots] \to^p \mathbb{E}[\phi_{T+k}] \text{ as } k \to \infty$$

- Long-run auto-covariance = 0: $Cov(Y_{T+k,T}) \to 0$ as $k \to \infty$

LLN and CLT for Time Series, HAC SE

- Law of Large Numbers (LLN) for Stationary and Mixing TS: suppose that $\{Y_t\}$ is stationary and geometrically mixing with a valid 2nd moment $\mathbb{E}[||Y_t||] < \infty$,

$$as\ T \to \infty,\ \bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t \to^p \mu_Y = \mathbb{E}[Y_t]$$

- Central Limit Theorem (CLT) for Stationary and Mixing TS: suppose that $\{Y_t\}$ is stationary and geometrically mixing with a valid 4th moment $\mathbb{E}\left[||Y_t||^4\right] < \infty$,

$$as\ T \to \infty,\ \sqrt{T}(\bar{Y} - \mu_Y) \to^d \mathcal{N}(0, \bar{\sigma}_Y^2)$$

where $\bar{\sigma}_Y^2$ is the Long-run Variance:

$$\begin{aligned}\bar{\sigma}_Y^2 &= Var(Y_t) + 2\sum_{k=1}^{T-1}\frac{T-k}{T}Cov(Y_t, Y_{t-k}) \\ &= Var(Y_t)\left[1 + 2\sum_{k=1}^{T-1}\frac{T-k}{T}\rho_Y(k)\right]\end{aligned}$$

- Long-run Var can be estimated by HAC SE: Newey-West Estimator

$$\hat{\sigma}_{Y,T}^2 = \hat{\sigma}_Y^2\left[1 + 2\sum_{k=1}^{m-1}\frac{T-k}{T}\hat{\rho}_Y(k)\right]$$

where $1 \le m \le T$ is a truncation parameter (S&W recommends $0.75T^{\frac{1}{3}}$ and Wooldridge recommends $0.025T^{\frac{1}{4}}$)

- A caveat of using CLT/LLN for time-series data: we always use $\frac{1}{\sqrt{T}}$ or $\frac{1}{T}$ because we can freely discard segments of sample in time-series asymptotic analysis: $\frac{1}{T}\sum_{t=k}^{T} Y_t = \frac{1}{T}\sum_{t=1}^{T} Y_t - \underbrace{\frac{1}{T}\sum_{t=1}^{k} Y_t}_{=0\ as\ T\to\infty} = \frac{1}{T}\sum_{t=1}^{T} Y_t$

2 2 Asymptotic Theory for OLS Estimator in Stationary AR Models

- Note that these theories also apply to TS regressions with regressors other than lags of Y_t

Properties of OLS Estimator in Stationary AR(1) Models

Assumptions

- Because TS data are not iid in general, OLS only has well-defined distribution in large samples

- AR(1).1: Errors are Unpredictable:

$$Y_t = \mu + \phi Y_{t-1} + \epsilon_t$$

with:

  - Strong AR(1).1: $\mathbb{E}[\epsilon_t | Y_{t-1}, Y_{t-2} \ldots] = 0$ if this holds, standard SE can be used (because all autocorrelations of ϵ_t are 0), and AR(1) is the optimal model for forecasting
    - Equivalent to $\mathbb{E}[\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, \ldots] = 0$ or $\mathbb{E}[\epsilon_t|Y_{t-1}, \epsilon_{t-1}, Y_{t-2}, \epsilon_{t-2}, \ldots] = 0$ or $\rho_\epsilon(k) = 0 \; \forall \; k > 0$
    - Weaker than saying errors are iid (e.g. allowing for heteroskedasticity)
  - Weak AR(1).1: $\mathbb{E}[\epsilon_t | Y_{t-1}] = 0$ if only this holds (e.g. selected less lags than DGP), then HAC SE will be needed, and the model is not optimal for forecasting
- AR(1).2: $\{Y_t\}$ is Stationary and Geometrically Mixing with a Valid 4th Moment $\mathbb{E}\left[||Y_t||^4\right]$
  - This provides access to LLN & CLT
- AR(1).3: Variation in Regressor $Var(Y_t) > 0$

Asymptotic Analysis

- Asymptotic theory of OLS in AR(1)

- Under AR(1).1-3 (*strong AR(1).1*) $\implies$ AR(1) estimator will be consistent and follow an asymptotic normal distribution, *standard SE* can be used:

$$\sqrt{T}(\hat{\phi} - \phi) \rightarrow^d \mathcal{N}\left(0, \frac{\mathbb{E}\left[(Y_{t-1} - \mu_Y)^2 \epsilon_t^2\right]}{Var(Y_t)^2}\right)$$

and $Avar(\hat{\phi}) = \frac{1}{T} \frac{E\left[(Y_{t-1} - \mu_Y)^2 \epsilon_t^2\right]}{Var(Y_t)^2}$

- *Weaker AR(1).1* $\implies$ AR(1) estimator will still be consistent and follow an asymp. normal distribution, but *HAC SE* needed: let $v_t = (Y_{t-1} - \mu_Y)\epsilon_t$:

$$Avar(\hat{\phi}) = \frac{1}{T} \frac{Var(v_t)\left\{1 + 2\sum_{k=1}^{T-1} \frac{T-k}{T} \rho_{v_t, v_{t-k}}\right\}}{Var(Y_t)^2}$$

- 如果regressor是X的话，那么分母换成$Var(X_t)^2$
- Proof of theorem:

-

$$
\begin{aligned}
\hat{\phi} &= \frac{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})Y_t}{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})^2} \\
&= \mu \underbrace{\frac{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})}{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})^2}}_{=0} + \phi \underbrace{\frac{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})Y_{t-1}}{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})^2}}_{=1} + \frac{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})\epsilon_{t-1}}{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})^2} \\
&= \phi + \frac{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})\epsilon_{t-1}}{\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})^2} \\
\implies \sqrt{T}(\hat{\phi}-\phi) &= \frac{\frac{1}{\sqrt{T}}\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})\epsilon_{t-1}}{\frac{1}{T}\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})^2}
\end{aligned}
$$

- With Strong AR(1).1:

$$
\mathbb{E}[(Y_{t-1}-\mu_Y)\epsilon_t] =^{LIE} \mathbb{E}\Big[\big(Y_{t-1}-\mu_Y)\underbrace{\mathbb{E}[\epsilon_t|Y_{t-1},Y_{t-2},\ldots]}_{=0}\big)\Big] = 0
$$

- and there will be no covariance terms, so normal SE estimator can be used:

$$
\begin{aligned}
\forall\, k \geq 1 : \mathbb{E}&\Big[[(Y_{t-1}-\mu_Y)\epsilon_t]\cdot[(Y_{t-k-1}-\mu_Y)\epsilon_{t-k}]\Big] \\
&= \mathbb{E}\Big[\big[(Y_{t-1}-\mu_Y)\underbrace{\mathbb{E}[\epsilon_t|Y_{t-1},Y_{t-2},\ldots]}_{=0}\big]\cdot[(Y_{t-k-1}-\mu_Y)\epsilon_{t-k}]\Big] \\
&= 0
\end{aligned}
$$

- Thus:

$$
\begin{cases}
\frac{1}{\sqrt{T}}\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})\epsilon_{t-1} & \to^{d} \mathcal{N}\Big(0, \mathbb{E}\big[(Y_{t-1}-\mu_Y^2\epsilon^2)\big]\Big) \quad \text{by CLT} \\
\frac{1}{T}\sum_{t=2}^{T}(Y_{t-1}-\overline{Y}_{-1})^2 & \to^{p} Var(Y_t) \quad \text{by LLN}
\end{cases}
$$

- By Slutsky's Theorem:

$$
\sqrt{T}(\hat{\phi}-\phi) \to^{d} \mathcal{N}\left(0, \frac{\mathbb{E}[(Y_{t-1}-\mu_Y)^2\epsilon_t^2]}{Var(Y_t)^2}\right)
$$

- Importance of AR.1: Strong AR(1).1 $\implies$ no covariance term (above) $\implies$ normal SE estimator can be used
  - If we only have weak AR(1).1, then replace $\mathbb{E}[(Y_{t-1}-\mu_Y)^2\epsilon_t^2]$ with HAC SE
- See OLS in DL for asymptotic analysis of OLS estimators in DL models

Testing AR Assumptions

- Testing Strong AR.1 (Errors are Unpredictable / Zero Conditional Mean): strong AR.1 implies all autocorrelations in ϵ_t are 0, so we can:
  - Plot residuals and look for outliers
  - Check/test autocorrelation in residuals

- Examine their marginal distributions (ideally close to Normal)
- Testing AR.2 (Stationarity): All roots of the characteristic polynomial lie outside the unit circle (see Stationarity of AR Process)

Properties of OLS Estimator in Stationary AR(p) Models

Assumptions

- AR(p).1: Errors are Unpredictable $Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t$ with
  - *Strong* AR(p).1: $\mathbb{E}[\epsilon_t | Y_{t-1}, Y_{t-2} \ldots] = 0$ if this holds, standard se can be used (because all autocorrelations of ϵ_t are 0)
  - *Weak* AR(p).1: $\mathbb{E}[\epsilon_t | Y_{t-1} \ldots Y_{t-p}] = 0$ if only this holds (e.g. selected less lags than DGP), then HAC SE will be needed
- AR(p).2: $\{Y_t\}$ is Stationary and Geometrically Mixing with a Valid 4th Moment $E\left[||Y_t||^4\right]$
  - This provides access to LLN & CLT
  - Requires all roots lie outside the unit circle and errors are iid PrimitiveConditions
  - No structural breaks
- AR(p).3: Variation in Regressor $Var(Y_t) > 0$ and there is No Perfect Multicollinearity between $Y_{t-1}, \ldots, Y_{t-p}$

Asymptotic Analysis

- Asymptotic theory of OLS in AR(p)
  - Under AR(p).1-3 with:
  - Strong AR(p).1 $\implies$ Standard SE:

$$\sqrt{T}(\hat{\phi}_k - \phi_k) \to^d \mathcal{N}\left(0, \frac{\mathbb{E}\left[r_{k,t}^2 \epsilon_t^2\right]}{Var(r_{k,t})^2}\right), k = 1, \ldots, p$$

    and $Avar(\hat{\phi}_k) = \frac{1}{T} \frac{\mathbb{E}\left[r_{k,t}^2 \epsilon_t^2\right]}{Var(r_{k,t})^2}$
  - Weaker AR(p).1 $\implies$ HAC SE needed: let $v_t = r_{k,t}\epsilon_t$,

$$Avar(\hat{\phi}) = \frac{1}{T} \frac{Var(v_t)\left\{1 + 2\sum_{k=1}^{T-1} \frac{T-k}{T} \rho_{v_t, v_{t-k}}\right\}}{Var(r_{k,t})^2}$$

- Is Assumption AR.1 (Strong) plausible in AR(p) model?
  - Wold's Decomposition: any stationary TS can be well-approximated (i.e. Strong AR(p).1 satisfied / no autocorrelation in ϵ_t) by an AR(p) if p is large enough

---

Week 3: Stationarity of AR Process and ADL Models

3 1 Stationarity of MA & AR Models

Stationarity of Moving Average (MA) Process

MA(1)

- Setup:

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

  where ϵ_t are i.i.d. distributed with $\mathbb{E}[\epsilon_t] = 0, Var(\epsilon_t) = \sigma_\epsilon^2$
- Stationarity of MA(1) process: $(\epsilon_t, \epsilon_{t-1})$ is stationary by construction, Y_t is a function of $(\epsilon_t, \epsilon_{t-1})$, so it's also stationary
- Moments of MA(1) process:

$$\begin{cases} \mathbb{E}[Y_t] & = \mu \\ Var(Y_t) & = (1+\alpha^2)\sigma_\epsilon^2 \\ Cov(Y_t, Y_{t-k}) & = \begin{cases} \alpha\sigma_\epsilon^2 & , k = 1 \\ 0 & , k > 1 \end{cases} \end{cases}$$

MA(∞)

- Setup:

$$Y_t = \mu + \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i}$$

  where $\sum_{i=1}^{\infty} \alpha_i^2 < \infty$ and $\{\epsilon_t\}$ are i.i.d. distributed with $\mathbb{E}[\epsilon_t] = 0, Var(\epsilon_t) = \sigma_\epsilon^2$
- Stationarity of MA process of infinite order: $(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-\infty})$ is stationary by construction, Y_t is a function of $(\epsilon_t, \epsilon_{t-1})$, so it's also stationary
- $\sum_{i=1}^{\infty} \alpha_i^2 < \infty$ ensures MA(∞) has a well-defined 2nd moment:

$$\mathbb{E}\left[\left(\sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i}\right)^2\right] = \sigma_\epsilon^2 \sum_{i=1}^{\infty} \alpha_i^2 < \infty$$

ARMA Models

- ARMA(p, q):

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}$$

Stationarity of AR Process

Stationarity of AR(1)

- Condition: $|\phi| < 1$ and $\{\epsilon_t\}$ are i.i.d.

- Under these conditions, *AR(1) → MA(∞)* by iterating backwards, hence stationary:

$$
\begin{aligned}
Y_t &= \mu + \phi Y_{t-1} + \epsilon_t \\
&= \phi^k Y_{t-k} + \mu \sum_{i=0}^{k-1} \phi^i + \sum_{i=0}^{k-1} \phi^i \epsilon_{t-i} \\
(as\ k \to \infty) &= \underbrace{\phi^k Y_{t-k}}_{\to 0} + \mu \underbrace{\sum_{i=0}^{k-1} \phi^i}_{\to \frac{\mu}{1-\phi}} + \sum_{i=0}^{k-1} \phi^i \epsilon_{t-i} \\
&= 0 + \frac{\mu}{1-\phi} + \underbrace{\sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}}_{MA(\infty)}
\end{aligned}
$$

Stationarity of AR(p)

- Condition:
  - All roots of the Characteristic Polynomial $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ lie outside the unit circle: $|z^*| > 1$
  - $\{\epsilon_t\}$ are i.i.d. ($\implies$ standard SE)
- Under these conditions, *AR(p) → MA(∞)* by iterating backwards, hence stationary:

$$
\begin{aligned}
Y_t &= \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t \\
&= \frac{\mu}{1 - \phi_1 - \ldots - \phi_p} + \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-1}
\end{aligned}
$$

  where $\alpha_1, \alpha_2 \ldots.$ are such that $\sum_{i=0}^{\infty} \alpha_i z^i = \frac{1}{\phi(z)}$ and $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 \ldots \phi_p z^p$

Primitive Conditions for OLS to be Normal in LN

- Primitive conditions
  - Roots of $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 \ldots \phi_p z^p$ are outside of the unit circle ($|z^*| > 1$)
    - Note that "eigenvalues" are reciprocal values of roots, so they need to be within the unit circle
    - Formal test: Dickey-Fuller Test
  - If further: errors ϵ_t are i.i.d., then we can use standard SE. Otherwise, HAC SE
    - Check by residual analysis, see Testing AR Assumptions

3 2 Stationarity of AR Models and ADL Model

Stationarity of AR Models - Detailed Derivation

- Stationarity of AR(p) process
- The role of the characteristic polynomial

Autoregressive Distributed Lag (ADL) Models

- Setup:

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \psi_j X_{t-j} + \epsilon_t$$

- Stationarity of ADL model:
  - All $\{X_t\}$ needs to be stationary (fit AR models to check, same techniques as below)
  - The AR component needs to be stationary (check roots of characteristic polynomial, test by Dickey-Fuller Test)
- Estimation and inference for ADL model: if stationary and mixing, OLS follows an asymptotic Normal distribution
- Granger Causality
  - A F-test: H_0: coefficients of all value of one of the variable $X_{a,t-1}, X_{a,t-2}, \ldots, X_{a,t-q_a}$ are 0
  - It is a test for marginal predictability (not really causality)
- Forecasting

Week 4: Forecasting and Model Selection

4 1 Forecast Uncertainty and Model Selection

Forecast Uncertainty and Forecast Intervals

Forecast Uncertainty - Concept

- Optimality in terms of MSFE:

$$\mathbb{E}[Y|X] = arg \min_{m(X)} \underbrace{\mathbb{E}\left[(Y - m(X))^2\right]}_{MSFE}$$

  and OLS consistently estimates the best linear predictor (if $m(X)$ is linear):

$$\min_{\beta_0, \beta_1} \mathbb{E}[(Y - \beta_0 - \beta_1 X)^2]$$

- MSFE of OLS forecast under *correct* specification:
  - A combination of:
  - forecast error distribution ($\epsilon_{T+1} = Y_{T+1} - Y_{T+1|T}$)
  - uncertainty from using estimated parameter ($Y_{T+1|T} - \hat{Y}_{T+1|T}$)
  - Total Forecast Error:

$$\begin{aligned} e_{T+1} &= Y_{T+1} - \hat{Y}_{T+1|T} \\ &= \underbrace{\left\{Y_{T+1} - Y_{T+1|T}\right\}}_{\epsilon_{T+1}} + \left\{Y_{T+1|T} - \hat{Y}_{T+1|T}\right\} \\ &= \epsilon_{T+1} + (Y_{T+1|T} - \hat{Y}_{T+1|T}) \end{aligned}$$

- Mean Square Forecast Error (MSFE):

$$\begin{aligned}
MSFE : \mathbb{E}[e_{T+1}^2] &= \mathbb{E}[(Y_{T+1} - \hat{Y}_{T+1|T})^2] \\
&= \sigma_\epsilon^2 + \underbrace{\mathbb{E}[(Y_{T+1|T} - \hat{Y}_{T+1|T})^2]}_{\text{Error from Parameter Uncertainty}} \\
(T \to \infty) &= \sigma_\epsilon^2
\end{aligned}$$

- RMSFE: $RMSFE = \sqrt{MSFE}$

Forecast Uncertainty - Estimation

- Three ways to estimate RMSFE
  1. Large sample approximation ($\widehat{RMSFE} \approx \underbrace{\hat{\sigma}_\epsilon}_{SER}$)
  2. Use actual forecast history (not practical)
  3. Pseudo out-of-sample forecasting (POOs)
     - Re-estimate the model every period for $t = t_1, \ldots, T-1$
     - Compute the forecast for $t+1$ using model estimated using data up to time t: $\hat{Y}_{t+1|t}$
     - Compute the poos forecast error: $Y_{t+1} - \hat{Y}_{t+1|t}$
     - POOS MSFE:

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

       - Idea: re-estimate the model every period from t_1 to $T-1$, compute forecast error $Y_{t+1} - \hat{Y}_{t+1|t}$, and calculate $\widehat{MSFE}$

Forecast Intervals

- Using the RMSFE to construct forecast intervals: If ϵ_{T+1} is normally distributed:

$$\begin{cases} 95\%\ FI : \hat{Y}_{T+1|T} \pm 1.96 \times \widehat{RMSFE} \\ 67\%\ FI : \hat{Y}_{T+1|T} \pm \widehat{RMSFE} \end{cases}$$

  - Forecast intervals are not confidence intervals because Y_{T+1} is random, and FI is valid only if ϵ_{T+1} is normal

Multi-Step Forecasts

- Iterated multi-period forecast:

$$\begin{aligned}
&\widehat{Y}_{T+1|T} = \hat{\mu} + \hat{\phi} Y_T \\
\rightsquigarrow\ &\widehat{Y}_{T+2|T} = \hat{\mu} + \hat{\phi} Y_{T+1|T} \\
&\cdots \\
\rightsquigarrow\ &\widehat{Y}_{T+h|T} = \hat{\mu} + \hat{\phi} Y_{T+h-1|T}
\end{aligned}$$

- Direct multi-period forecast: exclude lags of order below h in the model:

$$\begin{aligned} h=1: \widehat{Y}_t &= \hat{\mu}_1 + \hat{\phi}_1 Y_{t-1} \implies \widehat{Y}_{T+1|T} = \hat{\mu}_1 + \hat{\phi}_1 Y_T \\ h=2: \widehat{Y}_t &= \hat{\mu}_2 + \hat{\phi}_2 Y_{t-2} \implies \widehat{Y}_{T+2|T} = \hat{\mu}_2 + \hat{\phi}_2 Y_T \\ \ldots \end{aligned}$$

- Which to use?
    - Iterated method is generally preferred, especially when the original AR(p) model is correctly specified
    - In ADL models, direct forecasting may be more suitable because we do not need to build forecast model for each X

AIC/BIC: Lag Length Selection using Information Criteria

- Typically, we don't shrink model by consecutive F-tests
    - see why not using F-test for details
- Lag Length Selection Using Information Criteria
    - Idea: balance the bias-variance tradeoff: bias/model fit (too few lags) and estimation variance/uncertainty (too many lags)

$$MSFE = \mathbb{E}\Big[\underbrace{(Y_{T+1} - Y^*_{T+1|T})^2}_{\epsilon_T^2}\Big] + \mathbb{E}\Big[(Y^*_{T+1} - \hat{Y}^*_{T+1|T})^2\Big]$$

- First term in AIC/BIC (related to SSR or MSR) is designed to *penalise "bias" / forecast error*, which is $\mathbb{E}\Big[\underbrace{(Y_{T+1} - Y^*_{T+1|T})^2}_{\epsilon_T^2}\Big]$ in MSFE
- Second term (related to number of lags p) is designed to *penalise estimation uncertainty and overfitting*, which is $\mathbb{E}\Big[(Y^*_{T+1} - \hat{Y}^*_{T+1|T})^2\Big]$ in MSFE

Bayes Information Criterion (BIC)

- Bayes Information Criterion (BIC)

$$\begin{aligned} \hat{p}_{BIC} &= arg\min_p BIC(p) \\ &= arg\min_p \left\{ \underbrace{log\left(\frac{SSR(p)}{T}\right)}_{\downarrow \text{ in } p} + \underbrace{(p+1)\frac{log(T)}{T}}_{\uparrow \text{ in } p} \right\} \\ &= arg\min_p \left\{ \underbrace{log\,(MSR)}_{\downarrow \text{ in } p} + \underbrace{(p+1)\frac{log(T)}{T}}_{\uparrow \text{ in } p} \right\} \end{aligned}$$

- If $\{Y_t\}$ is indeed generated by an AR(p), then $\hat{p}_{BIC} \to^p p$

Akaike Information Criterion (AIC)

- Akaike Information Criterion (AIC)

$$\begin{aligned}\hat{p}_{AIC} &= arg\min_p AIC(p) \\ &= arg\min_p \left\{ \underbrace{log\left(\frac{SSR(p)}{T}\right)}_{\downarrow \text{ in } p} + \underbrace{(p+1)\frac{2}{T}}_{\uparrow \text{ in } p} \right\} \\ &= arg\min_p \left\{ \underbrace{log(MSR)}_{\downarrow \text{ in } p} + \underbrace{(p+1)\frac{2}{T}}_{\uparrow \text{ in } p} \right\}\end{aligned}$$

  - $\hat{p}_{AIC} \geq \hat{p}_{BIC}$
  - If the DGP is indeed an AR(p), AIC is not consistent (it *overestimates* lags)
  - Drawback of AIC/BIC and the alternative -- POOs (see Alternative lag length selection)
  - Generalization of BIC to ADL Models

$$BIC(p) = log\left(\frac{SSR(p,q)}{T}\right) + (p+q+1)\frac{log(T)}{T}$$

  - SATAT command `varsoc` restricts $p = q$
  - so alternatively, you can: choose lags of Y by BIC, and decide whether X should be included by the Granger causality test
- Choosing the right lag length: Important for inference
  - For *forecasting*, incorrect lag length $\rightsquigarrow$ HAC SE needed; not optimal forecast; but still consistent as long as the errors are mean-independent
  - For *causal inference / estimating the true parameter of DGP*, incorrect lag length $\rightsquigarrow$ Bias and inconsistency
    - see Week 8: Estimating Dynamic Causal Effects

4 2 [Forecasting and Model Selection II] - A

- Population forecasts: Choosing loss function: we can choose other loss functions $\mathcal{L}(Y_{T+1} - Y_{T+1|T})$ other than MSFE
- Optimal forecast:

$$MSFE = \mathbb{E}\Big[\underbrace{(Y_{T+1} - Y^*_{T+1|T})^2}_{\epsilon_T^2}\Big] + \mathbb{E}\Big[(Y^*_{T+1} - \hat{Y}^*_{T+1|T})^2\Big]$$

  - Bias-variance trade-off: shrinking the complexity of possible forecasts (by choosing finite lags / parametric form) will increase the first term ("bias" $\uparrow$) and decrease the second term ("variance/estimation uncertainty" $\downarrow$)
- Dimension reduction

- Finite # lags
- Parametric form
- Bayes' and Akaike's Information Critera
- Why not use F and t test to choose correct model?
  - It relies on repeated hypothesis testing with small effective sample size -- *size control is very difficult* (for k potential models, we need $k!$ tests)
  - *Overall confidence level can be very low* (every time we have 5% chance of rejecting a correct null, so probability of type-1 errors will accumulate) $\rightsquigarrow$ tends to select too many lags
- Alternative model selection rule: Poos:
  - *AIC and BIC both rely on large sample approximation of the exact MSFE*, which could be imprecise in small/moderate samples. Thus, we may consider POOS for small/moderate samples
  - Select a t_1
  - For each model, re-estimate the model every period from t_1 to $T-1$, compute forecast error $Y_{t+1} - \hat{Y}_{t+1|t}$, and calculate POOS MSFE:

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

  - Compare different models in terms of their $\widehat{MSFE}$
- Forecast interval: If $e_{T,h} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, then $\hat{e}_{T,h} \sim \mathcal{N}(0, MSFE)$ (approximately) so:

$$Y_{T+1} - \hat{Y}_{T+1|T} \sim \mathcal{N}(0, MSFE)$$

  and

$$\begin{cases} 95\% \text{ FI}: & \hat{Y}_{T+1|T} \pm 1.96\widehat{MSFE} \\ 67\% \text{ FI}: & \hat{Y}_{T+1|T} \pm \widehat{MSFE} \end{cases}$$

- Note that FI is only valid when the distribution of error term is Normal!
- Iterative Forecasting in STATA

Week 5: Trends

5 1 [Trends I] - A

- Two important types of non-stationarity
  - Trends / Unit roots (this week)
  - Structural breaks (Week 7: Structural Breaks)

Trends

- Trends and Cycles: Most economic TS can be decomposed into:

$$Y_t = \underbrace{\text{Trend}_t}_{\text{non-stationary}} + \underbrace{\text{Cycle}_t}_{\text{stationary}}$$

- Deterministic and stochastic trends
  - Deterministic trend: non-random function of time (time drift)
  - Stochastic trend: random and varies over time (unit roots)

Random Walk / Unit Root I(1)

UR/RW in AR(a) and Error Correction

- Setup:

$$Y_t = Y_{t-1} + \epsilon_t, \;\; \epsilon_t \text{ are serially uncorrelated}$$

- Two key features of a random walk
  - Best prediction is current value:

$$\mathbb{E}[Y_{T+h|T}] = Y_T$$

  - Unstable/exploding 2nd moment: suppose $Y_0 = 0$,

$$Var(Y_t) = t\sigma_\epsilon^2$$

    (variance grows linearly with time)
- A random walk with drift:

$$Y_t = \mu + Y_{t-1} + \epsilon_t, \;\; \epsilon_t \text{ are serially uncorrelated}$$

  - $\mu \neq 0$ induces a linear trend/drift in Y: $Y_{T+h|T} = \mu h + Y_T$
- Stochastic trends and unit autoregressive roots:
  - Random walk is an AR(1) with
    $\phi = 1 \iff$ charateristic polynomial $\phi(z) = 1 - z \iff \phi(1) = 0$, so First Differencing it makes it stationary (Error-correction Representation of AR(1)):

$$\Delta Y_t = \mu + \underbrace{\delta}_{=0 \text{ if RW}} Y_{t-1} + \epsilon_t$$

UR/RW in AR(2) Error Correction

- Unit roots in an AR(2):

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

  with its characteristic polynomial $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ has 1 root at $z = 1$:

$$\phi(1) = 1 - \phi_1 - \phi_2 = 0 \iff \phi_1 + \phi_2 = 1$$

- Error-correction representation:
  - Derivation（最好上来先左右同时减Y_{t-1}，然后再从后往前迭）：

$$
\begin{aligned}
Y_t &= \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t \\
&= \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} - \phi_2 Y_{t-2} + \phi_2 Y_{t-2} + \epsilon_t \\
&= \mu + (\phi_1 + \phi_2) Y_{t-1} + \phi_2 \underbrace{(Y_{t-1} - Y_{t-2})}_{\Delta Y_{t-1}} + \epsilon_t \\
&= \mu + (\phi_1 + \phi_2) Y_{t-1} + \phi_2 \Delta Y_{t-1} + \epsilon_t \\
\underbrace{Y_t - Y_{t-1}}_{\Delta Y_t} &= \mu + \underbrace{(\phi_1 + \phi_2 - 1)}_{\delta} Y_{t-1} + \underbrace{\phi_2}_{\gamma_1} \Delta Y_{t-1} + \epsilon_t \\
\Delta Y_t &= \mu + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \epsilon_t
\end{aligned}
$$

  - Result - Error-correction form of AR(2):

$$\Delta Y_t = \mu + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \epsilon_t$$

  where $\delta = \phi_1 + \phi_2 - 1, \gamma_1 = -\phi_2$
  - Unit root: $\delta = 0$
  - If indeed $\delta = 0$, then:
  - ΔY_t is stationary $\iff$ $|\gamma_1| < 1$
  - If $|\gamma_1| = 1$, then ΔY_t is still non-stationary (the process is integrated of order 2)

UR/RW in AR(p) Error Correction

- Unit roots and error correction form in AR(p) models
- Setup:

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t$$

  with $\sum_{i=1}^{p} \phi_i = 1$
- Error-correction form of AR(p):

$$\Delta Y_t = \mu + \delta Y_{t-1} + \underbrace{\gamma_1 \Delta Y_{t-1} + \ldots + \gamma_{p-1} \Delta Y_{t-p-1}}_{\sum_{i=1}^{p-1} \gamma_i \Delta Y_{t-i}} + \epsilon_t$$

  where

$$
\begin{cases}
\delta = (\sum_{i=1}^{p} \phi_i) - 1 \text{ (= 0 if unit root)} \\
\gamma_1 = -\sum_{i=2}^{p} \phi_i \\
\gamma_2 = -\sum_{i=3}^{p} \phi_i \\
\ldots \\
\gamma_{p-1} = -\phi_p
\end{cases}
$$

Problems of Trends

- What problems are caused by trends?

- Biasedness: OLS of AR coefficient will be strongly biased towards 0 $\rightsquigarrow$ poor forecasts
- Not well-behaving: t-stat does not follow standard Normal distribution, even in large samples $\rightsquigarrow$ standard inference tools cannot be used
- If both Y and X have stochastic trends, we may find a "spurious relationship"

Dickey-Fuller (DF) Test for Stochastic Trends

- How do you detect stochastic trends?
  - Eyeball: plot the data and see whether it is highly persistent
  - Check autocorrelations: slow decay in autocorrelations?

Dickey-Fuller Test

- Fit the error-correction form
  - If the original model is AR(1), then:

$$\hat{\delta} = \frac{\sum_{t=1}^{T}(\Delta Y_t - \overline{\Delta Y})(Y_{t-1} - \overline{Y})}{\sum_{t=1}^{T}(Y_{t-1} - \overline{Y})^2}$$

- Test the null (RW/UR):

$$\begin{cases} H_0 : \delta = 0 & \text{Has RW/UR (Non-Stationary)} \\ H_1 : \delta < 0 & \text{No RW/UR (Stationary)} \end{cases}$$

- This is a *one-side test*! $\{Y_t\}$ is stationary only when all roots are outside the unit circle)
- Compare the $t_{\hat{\delta}} = \frac{\hat{\delta}}{se(\hat{\delta})}$ with the critical values of the Dickey-Fuller distribution (because regressor Y_{t-1} is non-stationary under H_0):

TABLE 14.5 Large-Sample Critical Values of the Augmented Dickey-Fuller Statistic

Deterministic Regressors	10%	5%	1%
Intercept only	−2.57	−2.86	−3.43
Intercept and time trend	−3.12	−3.41	−3.96

- When should you include a time trend in the DF test?:
- Error-correction forms of AR(p) with/out time trend

$$\begin{cases} \text{Intercept only :} & \Delta Y_t & = \mu + \delta Y_{t-1} + \sum_{i=1}^{p-1} \gamma_i \Delta Y_{t-i} + \epsilon_t \\ \text{Intercept \& Time Trend :} & \Delta Y_t & = \mu + \alpha t + \delta Y_{t-1} + \sum_{i=1}^{p-1} \{\gamma_i \Delta Y_{t-i}\} + \epsilon_t \end{cases}$$

  - Intercept only: H_0: UR/RW; H_1: $\{Y_t\}$ is stationary around a constant
  - Intercept & Time Trend: H_0: UR/RW; H_1: $\{Y_t\}$ is stationary around a linear time trend (overall still non-stationary), i.e. the TS has long-term growth
- How to address/mitigate problems raised by unit roots:

- A UR/RW in $\{Y_t\}$ $\implies$ First-differencing (i.e. $\Delta Y_t = Y_t - Y_{t-1}$) will be stationary

5 2 [Trends II] - A

- Random walk with drift
  - Random walk with drift are explosive
  - RW with drift:

$$Y_t = \mu + Y_{t-1} + \epsilon_t$$

  - Solving backwards:

$$Y_t = Y_0 + \mu t + \sum_{i=1}^{t} \epsilon_i \implies \begin{cases} \mathbb{E}[Y_t|Y_0] & = Y_0 + \mu t \rightarrow \infty \text{ as } t \rightarrow \infty \\ Var(Y_t|Y_0) & = t\sigma_\epsilon^2 \rightarrow \infty \text{ as } t \rightarrow \infty \end{cases}$$

  - Thus, the MA weights are not square summable $\implies$ $MA(\infty)$ representation will not be available $\implies$ non-stationary
- Testing Random Walk hypothesis
  - Analysis of t statistic
  - Dickey—Fuller (DF) distribution of t-stat

Integrated Stochastic Trends (Multiple Unit Roots)

- General Stochastic Trends
- Integrated processes - multiple unit roots:
  - $I(1)$ process (RW/UR): AR process with 1 unit root in its characteristic polynomial; ΔY_t is stationary:

$$Y_t = \mu + Y_{t-1} + \epsilon_t \iff \Delta Y_t = \mu + \epsilon_t$$

  - $I(2)$ process: AR process with 2 unit roots in its characteristic polynomial; ΔY_t still contains a RW/UR; $\Delta^2 Y_t$ is stationary:

$$\begin{aligned} Y_t = \mu + 2Y_{t-1} - Y_{t-2} + \epsilon_t &\iff \Delta Y_t = \mu + \Delta Y_{t-1} + \epsilon_t \\ &\iff \Delta^2 Y_t = \mu + \epsilon_t \end{aligned}$$

  - $I(d)$ process: AR process with d unit roots in its characteristic polynomial; we have to difference d times to remove the unit roots:
  - $\Delta^d Y_t$ is stationary
  - $\Delta^i Y_t$ is non-stationary $\forall\, i < d$
- Testing for I(0) versus I(1) versus I(2)
  - Sequential testing procedure:
    1. Test ($I(1)$) whether $\delta = 0$ (UR) against $\delta < 0$ (stationary) in FD:

$$\Delta Y_t = \mu + \delta Y_{t-1} + \sum_{i=1}^{p-1} \gamma_i \Delta Y_{t-i} + \epsilon_t$$

2. If accept, test ($I(2)$) whether $\delta = 0$ (UR) against $\delta < 0$ (stationary) in SD:

$$\Delta^2 Y_t = \mu + \delta \Delta Y_{t-1} + \sum_{i=1}^{p-2} \gamma_i \Delta^2 Y_{t-i} + \epsilon_t$$

3. ...

- Summary: detecting and addressing unit roots/stochastic trends

---

Week 6: Asset Return Predictability

6 1 [Asset Return Predictability I] - A

Basics

- Asset Prices and Returns
  - Continuously compounded return or log-return

$$r_t = log(1 + R_t) = p_t - p_{t-1}$$

  where $p_t = log(P_t)$
  - and $r_t \approx R_t$ for small price changes
  - Log-compounding formula

$$r_t(h) = log(1 + R_t(h)) = \sum_{i=0}^{h-1} r_{t-i}$$

  - Excess return:

$$\overline{R}_t = R_t - R_{0,t},\ \overline{r}_t = r_t - r_{0,t}$$

    - where $R_{0,t}, r_{0,t}$ are risk-free returns
  - Risk premium = Expected excess return:

$$\mathbb{E}[\bar{R}_{t+1} | \mathcal{I}_t] = \mathbb{E}[R_{t+1} | \mathcal{I}_t] - R_{f,t+1}$$

Asset Return Predictability & Market Efficiency Hypothesis

- Asset Return Predictability
- Martingale Hypothesis: current price contains all relevant information (any assets have *0 expected payoff* / no RP and risk-free rate is 0)

$$\mathbb{E}[r_{t+1} | \mathcal{I}_t] = 0 \iff \mathbb{E}[p_{t+1} | \mathcal{I}_t] = p_t$$

  - Stochastic processes with this property are known as Martingales

- Random Walk Hypothesis: *allows risk premium* captured by μ:

$$H_{RW}: \mathbb{E}[r_{t+1}|\mathcal{I}_t] = \mu \iff \mathbb{E}[p_{t+1}|\mathcal{I}_t] = p_t + \mu$$

  or

$$p_t = \mu + p_{t-1} + \epsilon_t, \mathbb{E}[\epsilon_t|\mathcal{I}_{t-1}] = 0$$

  - $\mathcal{I}_t$ is the information set used to forecast future price
  - ϵ_t is the forecast error, capturing surprise movements and new information
  - μ is expected log-return (RP)
  - Riskier assets have higher variance (σ_ϵ^2) and higher risk premium μ
- Efficient Market Hypothesis: H_{RW} above with different information sets:
  - Weak-form efficiency: $\mathcal{I}_t$ includes only prices
  - Semi-strong efficiency: $\mathcal{I}_t$ includes all publicly available information
  - Strong efficiency: $\mathcal{I}_t$ includes all information (even private information)
  - EMH *Allows a predictable component (risk premium) in excess return if investors are risk-averse*
    - EMH only requires that a trading strategy cannot earn abnormal returns relative to a benchmark model for the market

Trading Rules

- "Technical" Trading Rules
  - Moving average (MA) rules
  - Do They "Work"?
  - Data Snooping - a multiple testing problem
    - e.g. testing at 5% level, then 5% of useless trading rules will be considered significantly effective
    - Reality-check test quantifies the effect of data snooping by evaluating the performance of the best trading rules in the context of the full "universe" of rules
      - Sullivan et al, 1999 found the best-performing rules still out-perform the benchmark

6 2 [Asset Return Predictability II] - A

- Some Stylized Facts of Returns - the S&P 100
- Random Walk Hypothesis
- Testing the RW Hypothesis - choice of information set and alternative

Testing Weak Form H_{RW} against AR(p)

- Random Walk Hypothesis

- "Weak test" of the RW Hypothesis with parametric alternative (AR(p))
    - Test RW/UR:

$$H_{RW} : \mathbb{E}[p_t | p_{t-1}, p_{t-2}, \ldots] = \mu + p_{t-1}$$

    - Assume p_t follows an AR(p) process:

$$p_t = \mu + \sum_{i=1}^{p} \phi_i p_{t-i} + \epsilon_t$$

    - Use the Error Correction Form of AR(p)

$$r_t = \mu + \delta p_{t-1} + \sum_{i=1}^{p-1} \gamma_i r_{t-i} + \epsilon_t$$

        1. Use the DF test to test Unit Root (with drift) in $\{p_t\}$: $H_0 : \delta = 0; H_1 : \delta < 0$
        2. (If accept) Set $\delta = 0$ and test $H_{RW} : \gamma_1 = \gamma_2 = \ldots = \gamma_{p-1} = 0$
        - If both are accepted, then we accept H_{RW}
- is AR(p) the right alternative?
    - Model misspecification (non-linear predictability)
    - Choose p

Model-Free Test of Weak Form H_{RW}

Ljung-Box Q Test

- The Q (or Ljung-Box) Testbased on *autocorrelations*:

$$H_{RW}/H_0 : \rho_r(h) = corr(r_t, r_{t+h}) = 0 \; \forall \; h \in \mathbb{N}$$

- Q-statistic

$$Q_m = T(T+2) \sum_{h=1}^{m} \frac{1}{T-h} \hat{\rho}_r^2(h) \sim \chi_m^2 \text{ under } H_{0/\mathrm{RW}} \text{ (stationary and mixing returns)}$$

- Size control (choosing m): we need to choose m to achieve a balance between determining power (large $m \rightsquigarrow$ more likely to find autocorrelation) and size of the test (small $m \rightsquigarrow$ smaller estimation uncertainty)

Variance-Ratio (VR) Test

- The Variance-Ratio (VR) Test by Lo and MacKinlay (1988) based on *autocovariances*:
- Denote the variance of h-step return as

$$V(h) = Var(r_t(h)) = Var(r_t + r_{t-1} + \ldots + r_{t-h+1})$$

- H_{RW} implies that all autocovariance are zero, so:

$$V(h) = \sum_{i=0}^{h-1} Var(r_{t-i}) = h \times V(1) \iff VR(h) = \frac{V(h)}{hV(1)} = 1$$

- Sample Analogue:

$$\widehat{VR}(h) = \frac{\hat{V}(h)}{h\hat{V}(1)} \text{ where } \hat{V}(h) = \underbrace{\frac{T}{(T-h)(T-h+1)}}_{\approx \frac{1}{T}} \sum_{t=h+1}^{T} \underbrace{\left[r_t(h) - \overline{r}(h)\right]^2}_{v_t(h)}$$

under H_{RW}:

$$t_{VR}(h) = \frac{\widehat{VR}(h) - 1}{\hat{\sigma}_{VR}(h)} \to^d \mathcal{N}(0,1)$$

where $\hat{\sigma}^2_{VR}(h) = \frac{1}{T}\frac{\hat{\Omega}_{VR}(h)}{h^2\hat{V}^2(1)}$ and $\hat{\Omega}_{VR}(h)$ is the HAC estimator of the long-run variance of $VR(h)$

- Size control (choosing h):
  - we need to choose h to achieve a balance between determining power (large $h \rightsquigarrow$ more likely to reject correctly) and size of the test (small $h \rightsquigarrow$ smaller estimation uncertainty)
  - In principle we should check all possible h by constructing a F-test
- In finite samples, the distribution of VR-stat can be far from normal

Testing Semi-Strong H_{RW} by Regressions

Test based on DL Model

- Regression-based Tests: adding more variables to $\mathcal{I}_t$
- simple distributed lag (DL) model:

$$r_{t+h}(h) = p_{t+h} - p_t = \beta_{h,0} + \beta_{h,1}X_t + \epsilon_{h,t}$$

where $h \geq 1$ is the forecast horizon, X_t are potential predictors, and $\epsilon_{h,t}$ is forecast error
  - When h is large, this is know as a long-horizon regression
- In-Sample Evidence of Predictability
- Out-of-sample Evidence of Predictability

Issues

- Some Critical Issues
  - Persistent Regressors
    - If $\{X_t\}$ is highly persistent, then OLS estimator will be a mixture of Augmented Dickey-Fuller (ADF) distribution and $\mathcal{N}(0,1)$:

$$t = \frac{\beta_{h,1}}{\hat{\sigma}_{\beta_{h,1}}} \sim \left[\lambda ADF + (1-\lambda)\mathcal{N}(0,1)\right]$$

- Long-Horizon Regressions
  1. Size control: As h becomes larger, there will be fewer observations left for estimation (effective sample size become smaller)
  2. $r_{t+h}(h)$ (or $\epsilon_t(h)$) becomes more persistent as h gets larger, because it is a rolling summation of $r_t \rightsquigarrow$ OLS will have a Dickey-Fuller type distribution and we need a HAC-type estimator for the SE; the finite-sample distribution will be far from normal
  3. Lack of strict exogeneity of predictors -- a shock to returns will affect future values of predictors
- DL with a single predictor is too weak. We should use an ADL with more predictors

---

Week 7: Structural Breaks (Another Type of Non-Stationarity)

7 1 [Structural Breaks I] - A

Structural Breaks

- Structural break: another type of non-stationarity
  - our model

$$\begin{cases} Y_t = \mu + \phi Y_{t-1} + \psi X_{t-1} + \epsilon_t & , t = 1, \ldots, \tau \\ Y_t = \mu' + \phi' Y_{t-1} + \psi' X_{t-1} + \epsilon_t & , t = \tau, \ldots, T \end{cases}$$

- bias-variance trade-off

Testing for Structural Break when Break Date is Known

- Test for structural break when break date is known
- New variables
  - Dummy: $D_t(\tau) = \mathbb{1}\{t > \tau\} = \begin{cases} 0, t \le \tau \\ 1, t > \tau \end{cases}$
  - Deltas: $\delta_0 = \mu' - \mu, \delta_1 = \phi' - \phi, \delta_2 = \psi' - \psi$
- Rewrite our model

$$Y_t = \mu + \phi Y_{t-1} + \psi X_{t-1} + \delta_0 D_t(\tau) + \delta_1 Y_{t-1} D_t(\tau) + \delta_2 X_{t-1} D_t(\tau) + \epsilon_t$$

- Chow test of structural break
  - $H_0 : \delta_0 = \delta_1 = \delta_2 = 0; H_1 : \text{at least one } \delta \neq 0$
  - Use Chow test: (heteroskedastic-robust) F-stat $\sim F_{q,\infty}$ *when data is stationary in each of the segments* ($|\phi| < 1, |\phi'| < 1, \{X_t\}$ is stationary + mixing) and q is the number of restriction

Testing for Structural Breaks when Break Date is Unknown

Quandt Likelihood (QLR) Test for One Break

- Quandt Likelihood Ratio (QLR) Statistic: test for

$$\begin{cases} H_0 & : \text{No break in } \{\tau_0, \ldots, \tau_1\} \\ H_1 & : \text{One break in } \{\tau_0, \ldots, \tau_1\} \end{cases}$$

- convention is to use the mid 70% data
- QLR statistic:

$$QLR = \max\{F(\tau_0), F(\tau_0 + 1), \ldots, F(\tau_1)\}$$

  where $F(.)$ is the Chow F-stat testing H_0 : no break at τ
- Critical values of QLR statistic
  - Critical values of QLR-stat are *larger* than that of individual F-stats ($F_{q,\infty}$) because it depends on multiple F-distributions:

$$QLR \sim \max_{s_0 \le s \le s_1} \left\{ \frac{1}{q} \sum_{i=1}^{q} \frac{B_i^2(s)}{s(1-s)} \right\}$$

    if $CLR > cv_\alpha$, we reject H_0 and our estimated break date is

$$\hat{\tau} = arg \max_{\tau_0 \le \tau \le \tau_1} F(\tau)$$

- QLR Stat for 15% trimming (keep the middle 70%)

TABLE 14.6 Critical Values of the QLR Statistic with 15% Trimming

Number of Restrictions (q)	10%	5%	1%
1	7.12	8.68	12.16
2	5.00	5.86	7.78
3	4.09	4.71	6.02
4	3.59	4.09	5.12
5	3.26	3.66	4.53
6	3.02	3.37	4.12
7	2.84	3.15	3.82
8	2.69	2.98	3.57
9	2.58	2.84	3.38
10	2.48	2.71	3.23

- Limits of QLR:
  - unable to detect breaks in the last 15% of the sample
  - *allows for only one break*; hard to handle multiple breaks (Discussed in part II below)

Testing Structural Breaks at the End of the TS

- pseudo out-of-sample (poos)
  - Want to check whether there is a break from τ_1 to T:
  1. Found a break at τ using QLR test
  2. Discard all data from $t = 1 \ldots \tau$
  3. Choose some "burn-in" samples from P to τ_1 where $\tau < P < \tau_1$
  4. For $t = P, \ldots, T$:
     1. Estimate the model using data from τ to t
     2. Calculate the in-sample $\widehat{SER}$
     3. Compute $\hat{Y}_{t+1|t}$ and $\widehat{MSFE} = \frac{1}{T-P} \sum_{t=P}^{T} (Y_{t+1} - \hat{Y}_{t+1|t})^2$
  5. Compare the $\widehat{RMSFE}$ with the in-sample $\widehat{SER}$ from the "burn in" regression: if they are similar, then it is likely that there's no break
- Rolling-window estimates
  - Pros: 1. can detect multiple breaks, 2. this is more flexible (no parametric assumption on coefficients); Cons: not a formal statistical test
  1. Choose a fixed time window $H \geq 1$
  2. Estimate the model while rolling the window with H samples included:
     - For $t = H + 1, \ldots, T$: estimate the model using data $Y_t, \ldots, Y_{t+H}; X_t, \ldots X_{t+H}$
     - Store the estimates $\hat{\mu}_t, \hat{\phi}_t, \hat{\psi}_t$
  3. Plot the time series of estimates $\hat{\mu}_t, \hat{\phi}_t, \hat{\psi}_t$ together with their CIs and compare them with overall estimates

7 2 [Structural Breaks II] - A

Testing for Structural Breaks when Break Date is Unknown: QLR Test Part II

- Chow and QLR test of structural break

Quandt Likelihood (QLR) Test for Multiple Breaks

- Multiple breaks
- Method 1: Sequential One-Break QLR Test:
  1. Test for single break using QLR Test for One Break and obtain $\hat{\tau}$
  2. Split the sample at $\hat{\tau}$ and conduct QLR Test for One Break for each subsample:

$$QLR_1 = \max_{\tau_0 \leq \tau \leq \hat{\tau}} F(\tau), QLR_2 = \max_{\hat{\tau}+1 \leq \tau \leq \tau_1} F(\tau)$$

  3. If $QLR_k > cv_\alpha$, we conclude that SB also occurs in the k^{th} subsample
  4. Repeat
  - Pros: computationally easy; Cons: lose size control - the common problem with multiple hypothesis testing

- Method 2: Single Joint QLR Test
    1. If we suppose there are k breaks, the we use k dummies to reconstruct the equation. e.g. when we suppose SB for intercepts:

$$Y_t = \mu + \delta_0 D_t(\tau_1) + \delta_1 D_t(\tau_2) + \cdots + \delta_k D_t(\tau_k)$$

    2. Compute QLR:

$$QLR = \max_{\tau_1,...,\tau_k} \{F(\tau_1, \ldots, \tau_k)\}$$

    - Pros: less statistical problem; Cons: computationally intensive due to the large number of possible combinations of τ

Rolling-window estimates (brief)

Week 8: Estimation of Dynamic Causal Effects

8 1 [Dynamic Causal Effects I] - A

Dynamic Causal Effects

- Dynamic causal effects is the effect of a change in X on Y over time
- Ideal setting: Randomized experiment

Distributed Lag Model

DL Model (A Dynamic Model)

- DL estimators are OLS estimator in distributed lag model:

$$Y_t = \mu + \psi_0 X_t + \cdots + \psi_q X_{t-q} + \epsilon_t$$

  where:
  - ψ_0 is the Immediate Impact Effect of change in X_t: effect of change in X_t on Y_t, holding $(X_{t-1}, \ldots, X_{t-q})$ constant (under weak exogeneity)
  - ψ_1 is the 1-period Dynamic Multiplier: effect of change in X_{t-1} on Y_t, holding $(X_t, X_{t-2}, \ldots, X_{t-q})$ constant (under weak exogeneity)
  - ...
  - h-period Dynamic Multiplier:

$$\psi_h = \frac{\partial \mathbb{E}[Y_t | X_t, X_{t-1}, \ldots]}{\partial X_{t-h}} = \frac{\partial \mathbb{E}[Y_{t+h} | X_{t+h}, X_{t+h-1}, \ldots]}{\partial X_t}$$

  - k-period Cumulative Dynamic Multipliers measure the accumulated effects $= \sum_{i=1}^{k} \psi_i$
  - (PPT II) Estimators of cumulative multipliers and their se's: rewrite the model so that coefficients are the cumulative multipliers

- Example, a DL(2) Model:

$$
\begin{aligned}
Y_t &= \mu + \psi_0 X_t + \psi_1 X_{t-1} + \psi_2 X_{t-2} + \epsilon_t \\
&= \mu + \psi_0 (X_t - X_{t-1}) + \underbrace{(\psi_0 + \psi_1)}_{\delta_1}(X_{t-1} - X_{t-2}) + \underbrace{(\psi_0 + \psi_1 + \psi_2)}_{\delta_2} X_{t-2} + \epsilon_t
\end{aligned}
$$

DL Assumptions

- Distributed Lag Model Assumptions

1. X is exogenous: $\mathbb{E}[\epsilon_t | X_t, X_{t-1}, X_{t-2} \dots] = 0$
    - This allows autocorrelations in ϵ, but if so, HAC SE will be needed (see below)
2. $\{Y_t, X_t\}$ is stationary and geometrically mixing
3. Eighth moments exist: $\mathbb{E}[Y^8] < \infty$ and $\mathbb{E}[X^8] < \infty$
4. No perfect multicollinearity

Properties of OLS under DL Assumptions

- W2: Properties of OLS under DL assumptions
- OLS estimators of $\mu, \psi_0, \dots, \psi_q$ are
    - *consistent (but biased)*
    - have *asymptotic Normal distribution* (so corresponding t/F statistics can be used), but variance needs to be adjusted for autocorrelations and heteroskedaisticity ↓
- DL OLS with autocorrelated errrors
  - *Math derivation* of Distribution with Weak Exogeneity:
  - Model:

$$
Y_t = \mu + \psi_0 X_t + \epsilon_t, \begin{cases} \mathbb{E}[\epsilon_t | X_t, X_{t-1}, X_{t-2} \dots] & = 0 \\ \mathbb{E}[\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2} \dots] & \neq 0 \end{cases}
$$

  - Estimator:

$$
\hat{\phi}_0 = \frac{\sum_{t=1}^{T} (X_t - \overline{X}) Y_t}{\sum_{t=1}^{T} (X_t - \overline{X})^2} = \phi_0 + \frac{\sum_{t=1}^{T} (X_t - \overline{X}) \epsilon_t}{(X_t - \overline{X})^2}
$$

  - Use LLN for stationary and mixing TS:

$$
\frac{1}{T} \sum_{t=1}^{T} (X_t - \overline{X})^2 \to^p Var(X_t) = \sigma_X^2
$$

  - Denote $v_t = (X_t - \mu_X) \epsilon_t$

- Use CLT for stationary and mixing TS and $\mathbb{E}[v_t] = 0$:

$$\frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{X})\epsilon_t \approx \frac{1}{T}\sum_{t=1}^{T} v_t \to^d \mathcal{N}\left(0, \frac{\overline{\sigma}_v^2}{T}\right)$$

$$\text{where } \overline{\sigma}_v^2 = \sigma_v^2 \left\{1 + 2\sum_{k=1}^{T-1} \frac{T-k}{T}\rho_{(v_t, v_{t-k})}\right\} \text{ (see below for estimator)}$$

- *Result*: Distribution with Standard Exogeneity: in the simple model $Y_t = \mu + \psi_0 X_t + \epsilon_t$ with $\mathbb{E}[\epsilon_t | X_t, X_{t-1}, X_{t-2}, \ldots] = 0$:

$$\hat{\psi_0} = \psi_0 + \frac{\sum_{t=1}^{T}(X_t - \bar{X})\epsilon_t}{\sum_{t=1}^{T}(X_t - \bar{X})^2} \sim^a \mathcal{N}\left(\psi_0, \frac{1}{T}\frac{\bar{\sigma}_v^2}{\sigma_X^4}\right)$$

where $v_t = (X_t - \mu_X)\epsilon_t$ and $\bar{\sigma}_v^2$ can be estimated by the Newey-West Estimator:

$$\hat{\bar{\sigma}}_v = \hat{\sigma}_v^2 + 2\sum_{k=1}^{m}\frac{T-k}{T}\widehat{Cov}(v_t, v_{t-k})$$

and autocovariance functions can be estimated by:

$$\widehat{Cov}(v_t, v_{t-k}) = \frac{1}{T}\sum_{t=k}^{T}\hat{v}_t \hat{v}_{t-k}$$

- With Strong Exogeneity Restriction (Dynamically Complete), we can use standard SE: in the simple model $Y_t = \mu + \psi_0 X_t + \epsilon_t$ with $\mathbb{E}[\epsilon_t | X_t, \epsilon_{t-1}, X_{t-1}, \epsilon_{t-2}, X_{t-2}, \ldots] = 0$:

$$\hat{\psi_0} = \psi_0 + \frac{\sum_{t=1}^{T}(X_t - \bar{X})\epsilon_t}{\sum_{t=1}^{T}(X_t - \bar{X})^2} \sim^a \mathcal{N}\left(\psi_0, \frac{1}{T}\frac{\mathbb{E}[(X_t - \mu_X)^2\epsilon_t^2]}{\sigma_X^4}\right)$$

Different Kinds of Exogeneity (PPT II)

- Exogeneity in time series regression
- Standard Exogeneity restriction: error is unpredictable with past and present values of the predictor:

$$\mathbb{E}[\epsilon_t | X_t, X_{t-1}, X_{t-2} \ldots] = 0$$

  - Interpretation of dynamic multiplier:

$$\psi_h = \frac{\partial \mathbb{E}[Y_t | X_t, X_{t-1}, \ldots]}{\partial X_{t-h}} = \frac{\partial \mathbb{E}[Y_{t+h} | X_{t+h}, X_{t+h-1}, \ldots]}{\partial X_t}$$

- Weaker Exogeneity restriction:

$$\mathbb{E}[\epsilon_t | X_t, X_{t-1}, \ldots, X_{t-q}] = 0$$

- With this, we can only interpret ψ_h as a *finite-horizon* effect:

$$\psi_h = \frac{\partial \mathbb{E}[Y_t | X_t, X_{t-1}, \ldots, X_{t-q}]}{\partial X_{t-h}}$$

- Stronger Exogeneity restriction:

$$\mathbb{E}[\epsilon_t | X_t, \epsilon_{t-1}, X_{t-1}, \epsilon_{t-2}, X_{t-2}, \ldots] = 0$$

  - With this, the model is dynamically complete: first q lags of X explains all the dynamics of Y (and HAC SE is not needed)
- Strict Exogeneity restriction: error is unpredictable with past, present, and future values of the predictor:

$$\mathbb{E}[\epsilon_t | \ldots X_{t+2}, X_{t+1}, X_t, X_{t-1}, X_{t-2} \ldots] = 0$$

  - Estimation of Dynamic Causal Effects with Strictly Exogenous Regressors
    - With strict exogeneity, there are more efficient ways to estimate dynamic causal effects than DL estimators
      - GLS estimation
      - ADL estimation
- When is Exogeneity Plausible?
  - Think about whether $Cov(\epsilon_{t+h}, X_t) = 0 \; \forall \; h \geq 0$
  - Reasons that exogeneity fails:
    - Simultaneity: $\begin{cases} Y_t = \mu + \psi_0 X_t + \epsilon_t \\ X_t = \mu_X + \psi_X Y_t + \epsilon_{X,t} \end{cases}$
    - Feedback: $Y_{t-q} \rightarrow X_t$

Standard Errors in ADL Models

- In ADL/AR models, with enough lags of Y_t, the error term cannot be predicted using past Y_t's, which is equivalent to that ϵ_t cannot be predicted by past ϵ_t's (i.e. ϵ_t is not serially correlated)
- But if there're not enough lags, ϵ_t will still be serially correlated, so always check autocorrelations for residuals

Estimators of Cumulative Multipliers and Their SE

- Rewrite the regression with coefficients equal to cumulative multipliers
- Example: DL(2):

$$\begin{aligned} Y_t &= \mu + \psi_0 X_t + \psi_2 X_{t-1} + \psi_2 X_{t-2} + \epsilon_t \\ &= \mu + \psi_0 (X_t - X_{t-1}) + \underbrace{(\psi_0 + \psi_1)}_{\delta_1} (X_{t-1} - X_{t-2}) + \underbrace{(\psi_0 + \psi_1 + \psi_2)}_{\delta_2} X_{t-2} + \epsilon_t \end{aligned}$$

Week 9: CAPM and APT

9 1 [CAPM] - A

Setup, Notations, and Assumptions

- Mean-Variance Portfolio Selection: use variance to measure risk
- The Security Market
    - Risk-free asset: $P_{t+1} = (1 + R_{0,t+1})P_{0,t}$
    - Risky asset: $P_{t+1} = (1 + R_{t+1})P_t$
    - Set $P_{0,t} = 1$
- Investor's Portfolio
    - Units of risk-free / risky asset: $w_{0,t}, w_t$
    - Acquisition cost $V_t = w_{0,t} + w_t P_t$
    - Value of portfolio 1 period later $V_{t+1} = w_{0,t}(1 + R_{0,t+1}) + w_t P_{t+1}$
    - Conditional expected value $\mathbb{E}_t[V_{t+1}] = w_{0,t}(1 + R_{0,t+1}) + w_t \mathbb{E}_t[P_{t+1}]$
    - Conditional variance $\sigma_t^2(V_{t+1}) = w_t^2 \sigma_t^2(P_{t+1})$
- Investor's Preferences - assumption
  - Expected utility is quadratic:

$$\begin{aligned} U_t &= E_t[V_{t+1}] - \frac{A}{2}\sigma_t^2(V_{t+1}) \\ &= w_{0,t}(1 + R_{0,t+1}) + w_t E_t[P_{t+1}] - \frac{A}{2} w_t^2 \sigma_t^2(P_{t+1}) \end{aligned}$$

- Investor's Decision Problem
    - Initial wealth at t: W_t
    - Maximisation:

$$\max_{w_{0,t}, w_t} \left\{ w_{0,t}(1 + R_{0,t+1}) + w_t E_t[P_{t+1}] - \frac{A}{2} w_t^2 \sigma_t^2(P_{t+1}) \right\} s.t.\ w_{0,t} + w_t P_t = W_t$$

    - Solution:$W_t^* = \frac{1}{AP_t} \frac{E_t[\bar{R}_{t+1}]}{\sigma_t^2(\bar{R}_{t+1})}$, $w_{0,t}^* = W_t - w_t^* P_t$ where $\bar{R}_{t+1} = R_{t+1} - R_{0,t+1}$ is the excess return of the asset

CAPM

- The Capital Asset Pricing Model (CAPM)
  - Investor $i \in \{1, \dots, M\}$ has risk aversion A_i, initial wealth $W_{i,t}$, and chooses his/her portfolio according to the above mean-var analysis
  - The Market Portfolio:

$$\text{Total Unit Held} = \begin{cases} \bar{w}_t = \sum_{i=1}^{M} w_{i,t}^* = \frac{1}{\bar{A}P_t} \frac{E_t[\bar{R}_{t+1}]}{\sigma_t^2(\bar{R}_{t+1})} \\ \bar{w}_{0,t} = \sum_{i=1}^{M} W_{i,t} - \bar{w}_t P_t = \bar{W}_t - \bar{w}_t P_t \end{cases}$$

where $\bar{A} = \left[\sum_{i=1}^{M} \frac{1}{A_i}\right]^{-1}$

- General CAPM:

$$\mathbb{E}_t[\bar{R}_{i,t+1}] = \beta_{i,t}\mathbb{E}_t[\bar{R}_{m,t+1}]$$
$$\text{where} \begin{cases} \beta_{i,t} = \frac{Cov_t[\bar{R}_{i,t+1}, \bar{R}_{m,t+1}]}{Var_t[\bar{R}_{m,t+1}]}, \ \forall\, i \in \{1, \ldots, n\} \\ \bar{R}_{m,t+1} = \sum_{i=0}^{N} \pi_i \bar{R}_{i,t+1} \text{ with } \pi_i = \text{share of total wealth in asset } i \end{cases}$$

- "Bar" means excess return here
- $\bar{R}_{m,t+1}$ and $\beta_{i,t}$'s' fully explain individual assets' expected excess returns
- Expected returns are linear to betas
- Market RP / expected excess return $\mathbb{E}_t[\bar{R}_{m,t+1}] > 0$: in equilibrium, asset prices must be such that all assets are held, and risk-averse investors will only hold risky assets if they have positive excess returns

Estimation and Testing of CAPM

- Collecting data

CAPM with Constant Betas

- CAPM with Constant Betas
  - Rewrite the CAPM as a regression model:

$$\bar{R}_{i,t} = \beta_{i,t}\bar{R}_{m,t} + \epsilon_{i,t}, \quad \mathbb{E}[\epsilon_{i,t}|\mathcal{I}_{t-1}] = 0$$

  - Assume *time-unvarying* $\beta_{i,t} = \beta_i$, this can be estimated using CAPM Time Series Regression:

$$\bar{R}_{i,t} = \alpha_i + \beta_i\bar{R}_{m,t} + \epsilon_{i,t}, \quad i = 1, \ldots, n$$

  which is a system of distributed lag models

  - CAPM suggests: $H_0 : \alpha_i = 0 \; \forall\, i \in \{1, \ldots, n\}$, which is a cross-equation restriction. We need to account for covariance across assets, so we stack all CAPM equations into a vector DL model:

$$\bar{R}_t = \alpha + \beta'\bar{R}_{m,t} + \epsilon_t$$

  where $\bar{R}_t = (\bar{R}_{1,t}, \ldots, \bar{R}_{n,t})', \alpha = (\alpha_1, \ldots, \alpha_n)', \beta = (\beta_1, \ldots, \beta_n)', \epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{n,t})'$

    - This can be estimated by OLS:

$$\hat{\beta} = \frac{\sum_{t=1}^{T}(\bar{R}_{m,t} - \hat{\mu}_{\bar{R}_m})(\bar{R}_t - \hat{\mu}_{\bar{R}})}{\sum_{t=1}^{T}(\bar{R}_{m,t} - \hat{\mu}_{\bar{R}_m})^2}, \hat{\alpha} = \hat{\mu}_{\bar{R}} - \hat{\beta}\hat{\mu}_{\bar{R}_m}$$

where $\hat{\mu}_{\hat{R}_m}, \hat{\mu}_{\hat{R}}$ are sample means, and the var-cov matrix is:

$$\hat{\Omega}_\epsilon = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t', \ \hat{\epsilon} = \hat{R}_t - \hat{\alpha} - \hat{\beta}\bar{R}_{m,t}$$

- We can test CAPM: $H_0 : \alpha_i = 0, i = 1, \ldots, n$ using an F-stat with HAC SE

Reasons for Rejection

- Possible Explanations for Rejection
    - Unrealistic assumptions (model is too simple)
        - Mean-var preference does not hold
        - Asymmetric information
        - Behaviour
        - Market incompleteness
    - Time-varying betas
    - Data collection: Roll's critique (stock index is not complete representation of the market portfolio), data snooping, etc.

9 2 CCAPM and APT - A

Conditional CAPM (Varying Beta)

- Estimation and Testing of Conditional CAPM

Setup

- Allows for $\beta_{i,t}$ to vary over time:

$$\mathbb{E}_t[\bar{R}_{i,t+1}] = \beta_{i,t+1}\mathbb{E}_t[\bar{R}_{m,t+1}]$$

or

$$\bar{R}_{i,t} = \beta_{i,t}\bar{R}_{m,t} + \epsilon_{i,t}, \quad \mathbb{E}[\epsilon_{i,t}|\mathcal{I}_{t-1}] = 0$$

Estimation Method 1: Instrumenting the Betas (Not IV)

- Instrumenting the Betas: specify the underlying structures of β
- Suppose variations in $\beta_{i,t}$ are caused by a single underlying factor Z_t:

$$\beta_{i,t} = b_{0,i} + b_{1,i}Z_t$$

- If we can identify and observe Z_t, we can run the following regression:

$$\bar{R}_{i,t} = \alpha_i + b_{0,i}\bar{R}_{m,t} + b_{1,i}(Z_t\bar{R}_{m,t}) + \epsilon_{i,t}$$

and test $H_0 : \alpha_i = 0$ for $i = 1, \ldots, n$

- Shortcomings:
  - Require correct choice of Z_t's, which is hard
  - Many standard choices of Z_t's are only available at low frequencies, so we have small effective sample size

Estimation Method 2: Model Time-Variation as Structural Breaks

- Model time-variation as structural change
- We can model CCAPM as a CAPM regression with structural breaks (treat $\alpha_{i,t}, \beta_{i,t}$ as potentially changing coefficients)
  - and we could use QLR Test to identify structural changes in $\beta_{i,t}$
- Advantage: more robust because it's model-free (no misspecification issues about Z_t)
- Disadvantage: less efficient and provides no explanation for time-variation

Estimation Method 3: Rolling-Window Estimator

- Rolling-window estimator
- Setup:

$$\bar{R}_{i,t} = \beta_{i,t}\bar{R}_{m,t} + \epsilon_{i,t}$$

  and

$$\begin{cases} \hat{\beta}_{i,t} = \frac{\sum_{s=1}^{T} w_{s,t}(\bar{R}_{m,s} - \hat{\mu}_{\bar{R}_{m,t}})\bar{R}_{i,s}}{\sum_{s=1}^{T} w_{s,t}(\bar{R}_{m,s} - \hat{\mu}_{\bar{R}_{m,t}})^2} \\ \hat{\alpha}_{i,t} = \hat{\mu}_{\bar{R}_{i,t}} - \hat{\beta}_{i,t}\hat{\mu}_{\bar{R}_{m,t}} \\ \hat{\mu}_{\bar{R}_{i,t}} = \sum_{s=1}^{T} w_{s,t}\bar{R}_{i,s}, \hat{\mu}_{\bar{R}_{m,t}} = \sum_{s=1}^{T} w_{s,t}\bar{R}_{m,s} \end{cases}$$

  where weights satisfy $\sum_s w_{s,t} = 1$. For example:

$$w_{s,t} = \begin{cases} \frac{1}{M} & , t - M \le s \le t \\ 0 & , otherwise \end{cases}$$

- The choice of M is critical, and we can use POOs method.

Arbitrage Pricing Theory (APT)

Setup

- Arbitrage Pricing Theory (APT)
- CAPM assumes the only source of risk is from the covariance with the overall market portfolio; APT is a multi-factor model allowing for multiple sources of risk
- Does not rely on specific assumptions of utilities, but simply on no-arbitrage
- A financial market:

$$\bar{R}_{i,t} = \alpha_i + \beta_{i,1}F_{1,t} + \cdots + \beta_{i,K}F_{K,t} + \epsilon_{i,t} \tag{APT}$$

where

- $F_{1,t}, \dots, F_{K,t}$ are a set of common factors
- Each asset has a set of factor loadings $\beta_i = (\beta_{1,i} \dots \beta_{K,i})'$ quantifying the impact of each factor on the excess return of asset i
- $\epsilon_{i,t}$ are idiosyncratic noise/risk satisfying $\mathbb{E}[\epsilon_{i,t}|F_{1,t}, \dots, F_{K,t}] = 0$

- or:

$$\mathbb{E}[\bar{R}_{i,t}] = \beta_{i,1}\lambda_1 + \dots + \beta_{i,K}\lambda_K$$

where:

- $\lambda_k = \mathbb{E}[F_{k,t}]$ is the risk premium associated with the kth factor (price of risk)
- $\beta_{i,k}$ is the exposure of asset i to risk factor k (quantity of risk)

Choice of factors:

- Choice of Factors
- Macro factors: variables that links "real economy" (production) to nominal one (financial markets); e.g. growth rate of industrial production (business cycle), spread between corporate and government bond yields (changes in aggregate level of risk)
- Returns of traded portfolios: Factors chosen as returns of a set of portfolios; e.g. Fama-French factor model
- Observable characteristics (aka "fundamentals"): Factors are unobserved while beta's are modelled as observable functions of firm-specific characteristics; e.g. size of firm, past performance

Testing

- Any testing of multi-factor models are conditional on having chosen the correct factors
- Under no-arbitrage:

$$\mathbb{E}[\bar{R}_{i,t}] = \beta_{i,1}\lambda_1 + \dots + \beta_{i,K}\lambda_K$$

where

- $\lambda_k = \mathbb{E}[F_{k,t}]$ is the risk premium associated with the kth factor (price of risk)
- $\beta_{i,k}$ is the exposure of asset i to risk factor k (quantity of risk)

- Therefore, we can test APT by running regression (APT) and test $H_0 : \alpha_i = 0 \, \forall \, i$
- Testing Observed factor models:
  - Suppose we have data on n different assets/portfolios
  - Collect their individual excess returns in $\bar{R}_t = (R_{1,t}, \dots, R_{n,t})'$
  - Rewrite joint APT model as a vector regression:

$$\bar{R}_t = \alpha + \beta' F_t + \epsilon_t, \quad \mathbb{E}[\epsilon_t|F_t] = 0$$

  - Estimate the α's and β's jointly
  - Test $H_0 : \alpha_1, \dots, \alpha_n = 0$

- Testing Characteristic-based factor models:
  - Setup:

$$\bar{R}_{i,t} = \alpha_i + \beta_{i,1}F_{1,t} + \cdots + \beta_{i,K}F_{K,t} + \epsilon_{i,t}$$

    but we don't know the true factors. Instead, we assume:

$$\beta_{i,j} = b_{i,j}X_i$$

    where X_i are observed characteristics of firm i and $b_{i,j}$ are loadings of these characteristics
  - Rewrite the model (treat $F_t^* = b_i'F$ as unknown and use X_i as regressors):

$$\bar{R}_{i,t} = \alpha_i + (b_{i,1}F_{1,t} + \cdots + b_{i,K}F_{K,t})X_i + \epsilon_{i,t} = \alpha_i + F_t^* X_i + \epsilon_{i,t}$$

  - Estimate the model and test $\alpha_i = 0 \; \forall \; i$
- *If rejected, reasons*:
  - Consumers are irrational / EMH violated
  - Mis-specification: wrong factors have been included / important ones have been excluded

Critical Issues

- Some Critical Issues
- In addition to CAPM, APT is more subject to data-snooping: If a factor is included using standard statistical tools (e.g. t-statistic), then with 5% probability a given factor will be included even though it is insignificant

---

Week 10: Volatility Models

10 1 [Volatility Models] - A

Heteroskedasticity & Notations

- Heteroskedasticity
- Conditional variance:

$$\sigma_t^2 = \sigma^2(\mathcal{I}_{t-1}) = Var(\epsilon_t|\mathcal{I}_{t-1}) = Var(Y_t|\mathcal{I}_{t-1})$$

- We use ϵ_t instead of Y_t as inputs (conditional on $\mathcal{I}_{t-1}$, they are the same: knowing $\epsilon_t \iff Y_t$)

Multiplicative Volatility Models (including all we discussed after)

- Multiplicative volatility model

- rescaled error term:

$$z_t = \frac{\epsilon_t}{\sigma_t} \iff \epsilon_t = \sigma_t z_t$$

and z_t captures *surprise movements* in ϵ_t after controlling for the level of volatility σ_t
- By construction: $\mathbb{E}[z_t|\mathcal{I}_{t-1}] = 0, Var(z_t|\mathcal{I}_{t-1}) = 1$

ARCH Model

- The ARCH Model: autoregressive conditional heteroskedastic model
- *Assumptions*:
  - ARCH.1: Volatility σ_t^2 depends on a finite number of lags p:

$$\sigma_t^2 = \sigma^2(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-p})$$

  - ARCH.2: $\sigma^2(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-p})$ has a simple linear form. In ARCH(1), this is just

$$\sigma^2(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-p}) = w + \alpha\epsilon_{t-1}^2$$

ARCH(1)

- Setup of ARCH(1):

$$\epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = w + \alpha\epsilon_{t-1}^2 \tag{ARCH(1)}$$

with $w > 0, \alpha \geq 0$
- Conditional variance in ARCH(1):

$$Var(Y_t|\mathcal{I}_{t-1}) = Var(\epsilon_t|\mathcal{I}_{t-1}) = \mathbb{E}[\sigma_t^2 z_t^2|\mathcal{I}_{t-1}] = \sigma_t^2 = w + \alpha\epsilon_{t-1}^2$$

- Time series properties:
  - Rewrite σ_t^2 as an *AR(1)* process:

$$\sigma_t^2 = w + \alpha\sigma_{t-1}^2 + u_t, \quad u_t = \alpha\sigma_{t-1}^2(z_{t-1}^2 - 1)$$

    where $\mathbb{E}[u_t|\sigma_{t-1}^2] = 0$ and $\mathbb{E}[u_t u_{t+h}] = 0$
    - $\rightsquigarrow$ we can use all properties from AR(1), such as:

$$\rho_{\sigma_t^2}(k) = \alpha^k$$

  - Long-run/Unconditional variance: taking expectation on both sides, we get:

$$\sigma_\epsilon^2 = \frac{w}{1-\alpha}$$

    - *If $\alpha < 1$, then σ_ϵ^2 exists and $\epsilon_t = \sigma_t z_t$ is stationary*
    - This is not the unconditional variance of Y_t, which needs to be analysed with the model of Y_t itself

ARCH(q)

- Setup:

$$\epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = w + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2 \tag{ARCH(q)}$$

- Long-run/Unconditional variance:

$$\sigma_Y^2 = \sigma_\epsilon^2 = \frac{w}{1 - \sum_{i=1}^{q} \alpha_i}$$

  - *If $\sum_{i=1}^{q} \alpha_i < 1$, ARCH(q) process is stationary with a well-defined 2nd moment*

GARCH Model

- GARCH(1,1):

$$\epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = w + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{GARCH(1,1)}$$

- GARCH(1,1) = restricted ARCH(∞): by backward recursion （注意只iterate σ，不要iterate ϵ）：

$$\begin{aligned}
\sigma_t^2 &= w + \alpha \epsilon_{t-1}^2 + \beta(w + \alpha \epsilon_{t-2}^2 + \beta \sigma_{t-2}^2) \\
&\cdots \\
&= \cancelto{\frac{w}{1-\beta}}{w \sum_{i=0}^{\infty} \beta_i} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} \epsilon_{t-i}^2 + \cancelto{0}{\beta^{\infty} \sigma_{t-\infty}^2} \\
&= \frac{w}{1-\beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} \epsilon_{t-i}^2
\end{aligned}$$

  - It allows all lags of ϵ to affect current volatility

- Long-run/Unconditional Variance:

$$\mathbb{E}[\sigma_t^2] = \frac{w}{1 - \alpha - \beta}$$

  - $\implies$ Requirement for a valid (non-exploding & positive) long-run variance:

$$w > 0, \alpha \geq 0, \beta \geq 0, \alpha + \beta < 1$$

Forecasting Volatility and Evaluation

One-Step Forecast

- ARCH/GARCH deliver exact one-step aead forecasts ($\sigma_{T+1|T}^2$). Ignoring parameter uncertainty, there is *no forecast error* ($\sigma_{T+1|T}^2 = \sigma_{T+1}^2$)
- Plugging in estimated parameters, we get the feasible forecast. Example: for ARCH(1):

$$\hat{\sigma}_{T+1}^2 = \hat{w} + \hat{\alpha} \epsilon_T^2$$

Multi-Step Forecast

- The optimal square-loss forecast:

$$\sigma^2_{T+h|T} = \mathbb{E}\left[\sigma^2_{T+h}|\mathcal{I}_T\right]$$

- Example: ARCH(1):

$$\sigma^2_{T+h|T} = w + \alpha\mathbb{E}[\epsilon^2_{T+h-1}|\mathcal{I}_T] = w + \alpha\sigma^2_{T+h-1|T} = \dots$$

 - Iterating backward, this is equal to

$$w(1 + \alpha + \cdots + \alpha^{h-1}) + \alpha^h\epsilon^2_T$$

 - Plugging in estimated parameters, we get the feasible h-step forecast:

$$\hat{\sigma}^2_{T+h|T} = \hat{w}(1 + \hat{\alpha} + \cdots + \hat{\alpha}^{h-1}) + \hat{\alpha}^h\epsilon^2_T$$

Forecast Errors

- Arises only in multi-step forecast $h > 1$
- Volatility forecast error (*no parameter uncertainty*):

$$e_{T+h} = \sigma^2_{T+h} - \sigma^2_{T+h|T}$$

- Volatility forecast error (*with parameter uncertainty*):

$$\hat{e}_{T+h} = \sigma^2_{T+h} - \hat{\sigma}^2_{T+h|T} = e_{T+h} + (\sigma^2_{T+h|T} - \hat{\sigma}^2_{T+h|T})$$

 - Parameter uncertainty vanishes as $T \to \infty$, so we typically ignore it:

$$\lim_{T\to\infty} \hat{e}_{T+h} = e_{T+h}$$

- Example: 2-step ahead forecast error of ARCH(1):

$$e_{T+2} = \sigma^2_{T+2} - \sigma^2_{T+2|T} = (w + \alpha\epsilon^2_{T+1}) - (w + \alpha\sigma^2_{T+1}) = \alpha\sigma^2_{T+1}(z^2_{T+1} - 1)$$

  - This implies our forecast is unbiased:

$$\mathbb{E}[e_{T+2}|\mathcal{I}_T] = 0,\ \underbrace{\mathbb{E}[e^2_{T+2}|\mathcal{I}_T] = \alpha^2\kappa_4\sigma^4_{T+1}}_{MSFE}$$

  where $\kappa_4 = \mathbb{E}\left[(z^2_{T+1} - 1)^2\right]$
  - MSFE:

$$MSFE = \alpha^2\kappa_4\mathbb{E}[\sigma^4_{T+1}]$$

Forecast Evaluation

- We don't observe the actual volatility σ_t^2. Instead, we use a noisy measure of volatility ϵ_t^2 to proxy it:

$$\epsilon_t^2 = \sigma_t^2 + \underbrace{(\epsilon_t^2 - \sigma_t^2)}_{u_t}$$

where $u_t = \sigma_t^2(z_t^2 - 1)$ is the measurement error:

$$\mathbb{E}[u_t|\mathcal{I}_{t-1}] = 0, \ \ \mathbb{E}[u_t^2|\mathcal{I}_{t-1}] = \sigma_t^4\kappa_4$$

where $\kappa_4 = \mathbb{E}[(z_{t+1}^2 - 1)^2]$
  - Thus, ϵ_t^2 is a noisy but unbiased proxy for σ_t^2 with error variance $\sigma_t^2\kappa_4$
- Forecast evaluation with proxy: For a given forecast $\hat{\sigma}_{t+1|t}^2$, the mean square forecast error can be calculated by POOs:

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

10 2 [Volatility Models II] - A

OLS as Gaussian MLE in AR

- Gaussian MLE in AR

Gaussian MLE of ARCH(1): Estimation

- Gaussian MLE is more efficient than OLS in estimating ARCH
- Example: ARCH(1)
- Let:

$$\begin{cases} Y_t = \mu + \epsilon_t \\ \epsilon_t = \sigma_t z_t, \ \ \sigma_t^2 = w + \alpha\epsilon_{t-1}^2 \end{cases}$$

and suppose $z_t \sim^{iid} \mathcal{N}(0,1) \implies Y_t|Y_{t-1} \sim \mathcal{N}(\mu, \sigma_t^2)$
- The conditional density (likelihood) of individual observation is:

$$f_\theta(Y_t|Y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left[-\frac{(Y_t - \mu)^2}{2\sigma_t^2}\right]$$

- Then the log-likelihood of a single observation is:

$$\log f_\theta(Y_t|Y_{t-1}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{(Y_t - \mu)^2}{2\sigma_t^2}$$

where $\sigma_t^2 = w + \alpha(Y_{t-1} - \mu)^2$

- Gaussian MLE of $\theta = (\mu, w, \alpha)$ maximised the joint log-likelihood:

$$
\begin{aligned}
\hat{\theta}_{MLE} &= arg\max_{\theta} \sum_{t=1}^{T} \log f_{\theta}(Y_t|Y_{t-1}) \\
&= arg\max_{\theta} \sum_{t=1}^{T} \left\{ -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{(Y_t-\mu)^2}{2\sigma_t^2} \right\}
\end{aligned}
$$

Inference in ARCH Models Estimated by MLE

Asymptotic Distribution

- Inference in ARCH Model
- If data is stationary and mixing, the MLE satisfies:

$$
\hat{\theta}_{MLE} \sim^a N\left(\theta, \underbrace{\frac{1}{T}H^{-1}\Omega H^{-1}U}_{\text{Robust SE}}\right)
$$

and if indeed $z_t \sim N(0,1)$ then $\Omega = H$:

$$
\hat{\theta}_{MLE} \sim^a N\left(\theta, \frac{1}{T}H^{-1}\right)
$$

Hypothesis Testing

- Hypothesis Testing
- To test individual hypothesis, we use t-statistics:

$$
t = \frac{\hat{\theta}_{MLE,i} - \theta_i}{\hat{\sigma}_{ii}}
$$

where $\hat{\sigma}_{ii}$ is the (i,i)th element of $\frac{1}{T}\hat{H}^{-1}\hat{\Omega}\hat{H}^{-1}$

- Joint hypothesis test can be done with Likelihood-Ratio (LR) statistic:

$$
LR = 2T\left\{\underbrace{\log f_{\hat{\theta}_{MLE}}(Y_2,\ldots,Y_T|Y_1)}_{\text{Unrestricted}} - \underbrace{\log f_{\tilde{\theta}_{MLE}}(Y_2,\ldots,Y_T|Y_1)}_{\text{Restricted}}\right\}
$$

and under the null of restricted model:

$$
LR \to^d \chi_m^2
$$

and m is the number of restrictions

Others

- ECON0022 Review Questions
- ECON0022 Week 7 Questions

- F-Statistics:

$$F = \frac{SSR_R - SSR_U}{SSR_U} \times \frac{n-k}{q}$$

- Log-Normal Expectation:

$$\ln X \sim \mathcal{N}(\mu, \sigma^2) \implies \mathbb{E}[X] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$