Financial Econometrics

Return Predictability

Tim C.C. Hung 洪志清

April 11^{th}, 2022

Return

- Return processes are thought to be stationary, while prices are not.
- Simple return from $t-1$ to t:

$$R_t = \frac{P_t}{P_{t-1}} - 1$$

  where P_t is the price of the asset at time t.

- Simple gross return:

$$1 + R_t$$

- Compound return over k periods:

$$\begin{aligned} 1 + R_t(k) &\equiv (1+R_t)(1+R_{t-1})\cdots(1+R_{t-k+1}) \\ &= \frac{P_t}{P_{t-1}}\frac{P_{t-1}}{P_{t-2}}\cdots\frac{P_{-k+1}}{P_{t-k}} = \frac{P_t}{P_{t-k}} \end{aligned}$$

Return

- Continuous compounding:

$$r_t \equiv log(1+R_t) = log\frac{P_t}{P_{t-1}} = logP_t - logP_{t-1}$$

- Why use log?
  Multiperiod returns (time additivity):

$$\begin{aligned} 1 + R_t(k) &= (1+R_t)(1+R_{t-1})\cdots(1+R_{t-k+1}) \\ r_t(k) &= log(1+R_t(k)) = r_t + r_{t-1} + \cdots + r_{t-k+1} \end{aligned}$$

- Also: remember price is log-normally distributed.

Dividend

- When the asset in question pays periodic dividends D_t, the simple net return is:
$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1$$
- Supplementary Materials:
  - Shiller (1981,*AER*): Do stock prices move too much to be justified by subsequent changes in dividends?
  - Campbell, Shiller (1988,*RFS*): Campbell-Shiller decomposition
  - Cochrane (2008,*RFS*): The dog that did not bark: A defense of return predictability

Excess Return

- An **excess return** is the difference between the return on an asset of interest and the return of a reference asset, R_{0t}.
  - Quite often the reference asset is a riskfree asset or short-term T-bill, maybe a zero-beta asset.
  - The simple excess return on asset i is defined as:

$$Z_{it} = R_{it} - R_{0t}$$

- The log excess return is not the log of the excess return, but instead: $z_{it} = r_{it} - r_{0t}$, the difference between the log return on the asset and the log return on the reference asset.

- We will introduce the idea of **abnormal return** after tests of CAPM.

Efficient Market Hypothesis

National Taiwan University

- Fama (1970, *JF*):
  "A market in which prices always fully reflect available information is efficient."

- Malkiel (1992, *book chapter*):
  "A capital market is fully efficient if it correctly reflects all information in determining security prices. Formally, the market is said to be efficient with respect to some information set. . . if security prices would be unaffected by revealing that information to all market participants. Moreover, efficiency with respect to an information set. . . implies that it is impossible to make economic profits by trading on the basis of [the information in that set]."

EMH Discussion

- The second sentence of Malkiel's definition expands Fama's definition and suggests a test for efficiency useful in a laboratory.
- The third sentence suggests a way to judge efficiency that can be used in empirical work.
- The latter is what the finance literature has mainly focused on.
  - Examples: mutual fund managers profits: if they are true economic profits, then prices are not efficient with respect to their information.
  - Difficult to test for good reasons we will discuss.

Market Efficiency

National Taiwan University

- Price changes must be unforecastable.
- Prices always fully reflect available information.
- It is impossible to make economic profits by trading on the basis of an information set I.
- Forms of market efficiency:
  - Weak form: I includes **past prices**.
  - Semi-strong form: I includes **all public information**.
  - Strong form: I includes **all knowable information**.

What Does Profitable Mean?

- We're talking about economic profits, adjusting for risk and costs.
  $\rightarrow$ Need **models** of such things, particularly the risk adjustment.

- Difficulty: rely on joint tests of efficiency and certain asset pricing models, or benchmark at the same time!
  Cochrane (2006): positive and normative view of finance.

- Possible Models:
  Random Walk Model, CAPM, Fama-French 3-Factor Model, Carhart 4-Factor Model (Momentum), Fama-French 5-Factor Model, etc.

- Suggest Reading: van Binsbergen, Berk (2016, *JFE*): Assessing asset pricing models using revealed preference

Testing Random Walk

- Random Walk Hypothesis of Stock Price:

$$P_t = \mu + P_{t-1} + \epsilon_t$$

  or in log prices:

$$p_t = \mu + p_{t-1} + \epsilon_t$$

- RW1: IID increments: $\epsilon_t \sim IID(0, \sigma^2)$
- RW2: Independent increments: $\epsilon_t \sim IND$
- RW3: Uncorrelated increments: $Cov(\epsilon_t, \epsilon_{t-k}) = 0$, for $k > 0$

Box-Pierce Q Test

- For a sample return series $\{r_t\}_{t=1}^{T}$,
  - $\bar{r_T} = \frac{1}{T}\sum_{t=1}^{T} r_t$
  - $\overline{\gamma}(k) = \frac{1}{T}\sum_{t=1}^{T-k}(r_t - \bar{r_T})(r_{t+k} - \bar{r_T})$
  - $\widehat{\rho}(k) = \frac{\overline{\gamma}(k)}{\overline{\gamma}(0)}$
- Random walk implies that all autocorrelations are zero.
- The Box-Pierce Q-statistic is a joint test for the first m autocorrelations. The test focuses on the more general alternative $\rho_j \neq 0$ for at least one $j \leq m$ given a known m.

$$Q_m = T\sum_{k=1}^{m}\rho^2(k)$$

- Under the random walk null hypothesis, asymptotically,

$$\hat{Q}_m = T\sum_{k=1}^{m}\hat{\rho}^2(k) \sim \chi^2(m)$$

Variance Ratio Test

- For all RW hypotheses, the variance of RW increments is linear in the time interval.

- For example, a 2-period continuously compounded return has twice the variance of a 1-period return.

$$
\begin{aligned}
VR(2) \equiv \frac{Var[r_t(2)]}{2Var[r_t]} &= \frac{Var[r_t + r_{t+1}]}{2Var[r_t]} \\
&= \frac{2Var[r_t] + 2cov[r_t, r_{t+1}]}{2Var[r_t]} \\
&= 1 + \rho(1)
\end{aligned}
$$

- Under the RW null, $\rho(1) = 0$, $VR(2) = 1$
- With positive (negative) first-order autocorrelation, $VR(2) > \ (<)\ 1$

Variance Ratio Test (Cont.)

- For a covariance stationary ARMA process:

$$VR(q) \equiv \frac{Var[r_t(q)]}{q\ Var[r_t]} = 1 + 2\sum_{k=1}^{q-1}(1-\frac{k}{q})\rho(k)$$

- For example, for an AR(1) process $r_t = \phi r_{t-1} + u_t$, $VR(q) = 1 + 2\sum_{k=1}^{q-1}(1-\frac{k}{q})\phi^k$

- Asymptotic Distribution:
- Under RW null,

$$\sqrt{Tq}(\widehat{VR}(q)-1) \sim N(0, 2(q-1))$$

CAPM

National Taiwan University

Capital Asset Pricing Model:

- Please refer to the supplementary material for derivation of CAPM.
- Consider the Security Market Line:

Theorem (Unconditional Form)

Sharpe-Lintner CAPM:

$$E[R_i] = R_f + \beta_i(E[R_m] - R_f)$$

Black CAPM (Zero-Beta CAPM):

$$E[R_i] = E[R_z] + \beta_i(E[R_m - R_z])$$

- Let's focus on $SL - CAPM$ first.

Testing CAPM

- Rewrite $SL-CAPM$ with excessive returns:

$$E[Z_i] = \beta_i E[Z_m]$$

where $Z_i = R_i - R_f$, $Z_m = R_m - R_f$

- Expected returns are not observable!
$\rightarrow$ tests are based on **realized returns**.

- Testable market model (with realized returns):

$$r^e_{i,t} = \alpha_i + \beta^e_i r_{m,t} + \epsilon_{i,t}$$

where $r^e_{i,t}$ is the realized return risk premium for stock i at time t; $r^e_{m,t}$ is the realized market risk premium for at time t.

- What are the differences from the original $CAPM$?

Testing CAPM (Cont.)

國立臺灣大學 National Taiwan University

- Three testable implications!

  1. $\alpha_i = 0$ for all i
  2. Beta completely captures cross-sectional variation of expected returns.
  3. Market risk premium $E[Z_m]$, or $r^e_{m,t}$, is positive.

Testing CAPM - Remarks

- For the CAPM to be true, the market portfolio must lie in the efficient set. This is in fact the economic implication of the CAPM analysis, the **identification** of an efficient portfolio.

- The CAPM is a single-period model. In econometric analysis, the model must be estimated over time. Must add assumption about time-series behavior of returns.

- Assumption: returns are independent and identically distributed over time, and jointly multivariate normal.

- Implicit assumption: Probability distributions for returns on stocks and bonds are stationary.

Fama-MacBeth Regression

- Presents a cross-sectional test of CAPM.
- Test of linearity and pricing of non-beta factors.
- Methodology to account for cross-sectional correlation.

Fama-MacBeth Procedure

1. Estimate beta with OLS time-series regressions for each stock/portfolio.

$$r_{i,t}^e = \alpha_i + \beta_i r_{m,t}^e + \epsilon_{i,t}$$

2. Estimate γ_{0t} and γ_{1t} for each time period (cross-section).

$$r_{i,t}^e = \gamma_{0t} + \gamma_{1t}\hat{\beta_{it}} + u_{it}$$

3. Analyze the time-series of estimated γ_{0t} and γ_{1t}.

Fama-MacBeth Procedure (Cont.)

- Define $\gamma_0 = E[\gamma_{0t}]$ and $\gamma_1 = E[\gamma_{1t}]$

- Implications of the CAPM are
  $\gamma_0 = 0$ and $\gamma_1 > 0$ (γ_1 is the average market risk premium)

- Use t-test:

$$t_j = \frac{\hat{\gamma_j}}{\sigma_{\hat{\gamma_j}}} \sim t(T-1)$$

  where

$$\hat{\gamma_j} = \frac{1}{T}\sum_{i=1}^{T}\hat{\gamma_{jt}} \quad \text{and} \quad {\sigma_{\hat{\gamma_j}}}^2 = \frac{1}{T(T-1)}\sum_{i=1}^{T}(\hat{\gamma_{jt}} - \hat{\gamma_j})^2, \quad j = 0, 1$$

Fama-MacBeth Recipe

National Taiwan University

- Assign stocks to one of the 20 portfolios. (Why portfolio?)
- Estimate Betas over a five-year window (e.g. 1980-1984).
- Use betas estimated in the last step as the portfolio beta estimate for the following four years (e.g. 1985-1988).
- Fit the cross-sectional regression for each month (and adjust beta monthly for security delistings) in years 1985 through 1988.
  - Their specification: $\tilde{r_{i,t}} = \tilde{\gamma_{0t}} + \tilde{\gamma_{1t}}\beta_i + \tilde{\gamma_{2t}}\beta_i^2 + \tilde{\gamma_{3t}}S_i + \eta_{it}$ (Why γ_{2t}?)
- Repeat for the next year.

Fama-MacBeth - Remarks

- Questions:
  1. Does the standard error estimator account for cross-sectional correlations?
  2. What is the implicit assumption here about the serial correlation of the monthly regression estimates?

- Answer: assume that $\tilde{\gamma_{jt}}$ are IID.
  - What's the rationale behind this assumption?
  - People also use Newey-West procedures to take care of possible serial correlations and heteroskedasticity in γ_{jt}.
  - We will discuss about Newey-West shortly.

- We can also include characteristics, such as size, book-to-market, past returns in the FMB regressions. $\rightarrow$ The S_i stuff in their specification.

- Nowadays, FMB has been regarded as a more general method to study panel data: estimate cross-sectional regressions first and then test the time-series of coefficient estimates. $\rightarrow$ Similar to time fixed effects.

Fama-MacBeth Empirical Results

- Fama, MacBeth (1973, JPE)

TABLE 3

Summary Results for the Regression

$$R_p = \hat{\gamma}_{0t} + \hat{\gamma}_{1t}\hat{\beta}_p + \hat{\gamma}_{2t}\hat{\beta}^2{}_p + \hat{\gamma}_{3t}\bar{s}_p(\hat{\epsilon}_i) + \hat{\eta}_{pt}$$

	Statistic																			
Period	$\bar{\hat{\gamma}}_0$	$\bar{\hat{\gamma}}_1$	$\bar{\hat{\gamma}}_2$	$\bar{\hat{\gamma}}_3$	$\overline{\hat{\gamma}_0 - R_f}$	$s(\hat{\gamma}_0)$	$s(\hat{\gamma}_1)$	$s(\hat{\gamma}_2)$	$s(\hat{\gamma}_3)$	$\rho_0(\hat{\gamma}_0 - R_f)$	$\rho_M(\hat{\gamma}_1)$	$\rho_0(\hat{\gamma}_2)$	$\rho_0(\hat{\gamma}_3)$	$t(\bar{\hat{\gamma}}_0)$	$t(\bar{\hat{\gamma}}_1)$	$t(\bar{\hat{\gamma}}_2)$	$t(\bar{\hat{\gamma}}_3)$	$t(\overline{\hat{\gamma}_0 - R_f})$	$\bar{r}^2$	$s(r^2)$
Panel A:																				
1935–6/68 ..	.0061	.0085	...	...	.0048	.038	.066	...	...	.15	.02	...	...	3.24	2.57	...	...	2.55	.29	.30
1935–45	.0039	.0163	...	...	.0037	.052	.098	...	...	.10	−.03	...	...	.86	1.92	...	...	.82	.29	.29
1946–55	.0087	.0027	...	...	.0078	.026	.041	...	...	.18	.07	...	...	3.71	.70	...	...	3.31	.31	.32
1956–6/68 ..	.0060	.0062	...	...	.0034	.030	.044	...	...	.27	.15	...	...	2.45	1.73	...	...	1.39	.28	.29
1935–40	.0024	.0109	...	...	.0023	.064	.116	...	...	.07	−.09	...	...	.32	.79	...	...	.31	.23	.30
1941–45	.0056	.0229	...	...	.0054	.034	.069	...	...	.23	.15	...	...	1.27	2.55	...	...	1.22	.37	.28
1946–50	.0050	.0029	...	...	.0044	.031	.047	...	...	.20	.04	...	...	1.27	.48	...	...	1.10	.39	.33
1951–55	.0123	.0024	...	...	.0111	.019	.035	...	...	.20	.08	...	...	5.06	.53	...	...	4.56	.24	.29
1956–60	.0148	−.0059	...	...	.0128	.020	.034	...	...	.37	.18	...	...	5.68	−1.37	...	...	4.89	.22	.31
1961–6/68 ..	.0001	.0143	...	...	−.0029	.034	.048	...	...	.22	.09	...	...	.03	2.81	...	...	−.80	.32	.27
Panel B:																				
1935–6/68 ..	.0049	.0105	−.0008	...	.0036	.052	.118	.056	...	.03	−.11	−.11	...	1.92	1.79	−.29	...	1.42	.32	.31
1935–45	.0074	.0079	.0040	...	.0073	.061	.139	.074	...	−.10	−.31	−.21	...	1.39	.65	.61	...	1.36	.32	.30
1946–55	−.0002	.0217	−.0087	...	−.0012	.036	.095	.034	...	.04	.00	.00	...	−.07	2.51	−2.83	...	−.38	.36	.32
1956–6/68 ..	.0069	.0040	.0013	...	.0043	.054	.116	.053	...	.17	.07	.03	...	1.56	.42	.29	...	.97	.30	.30
1935–40	.0013	.0141	−.0017	...	.0012	.069	.160	.075	...	−.13	−.36	−.35	...	.16	.75	−.19	...	.14	.24	.30
1941–45	.0148	.0004	.0108	...	.0146	.050	.111	.073	...	−.04	−.19	−.04	...	2.28	.03	1.15	...	2.24	.39	.29
1946–50	−.0008	.0152	−.0051	...	−.0015	.037	.104	.032	...	.14	.04	.00	...	−.18	1.14	−1.24	...	−.32	.44	.32
1951–55	.0004	.0281	−.0122	...	−.0008	.030	.085	.035	...	−.17	−.14	−.01	...	.10	2.55	−2.72	...	−.20	.28	.29
1956–60	.0128	−.0015	−.0020	...	.0108	.030	.072	.029	...	.35	.11	.26	...	3.38	−.16	−.54	...	2.84	.25	.31
1961–6/68 ..	.0029	.0077	.0034	...	−.0000	.066	.138	.064	...	.14	.06	−.01	...	.42	.53	.51	...	−.01	.34	.29

Fama-MacBeth Conclusion

"We cannot reject the hypothesis that average returns on New York Stock Exchange common stocks reflect the attempts of risk-averse investors to hold efficient portfolios. Specifically, on average there seems to be a positive tradeoff between return and risk, with risk measured from the portfolio viewpoint. In addition, although there are "stochastic nonlinearities" from period to period, we cannot reject the hypothesis that on average their effects are zero and unpredictably different from zero from one period to the next."

Roll's Critique

- Background: CAPM does not work well in the late 70's.
- Roll (1977, JFE): A critique of the asset pricing theorys tests Part I: On past and potential testability of the theory
- The only empirically testable hypothesis is that the market portfolio is mean-variance efficient.
  $\rightarrow$ We don't know the true composition of the market portfolio!
  - If the proxy used as market portfolio is mean-variance efficient, CAPM will be accepted regardless of whether the **true** market portfolio is mean-variance efficient.
  - If the market proxy is not mean-variance efficient, CAPM will be rejected regardless of whether the **true** market portfolio is mean-variance efficient.
- A test of CAPM not using the market portfolio is only a test of whether the chosen index portfolio was ex-post efficient or not!
- Roll's conclusion: CAPM is not testable. Any market index is not complete: human capital, real estate, foreign assets, etc. are omitted.

- Standing at the time period t:

Theorem (Unconditional and Conditional CAPM)

Sharpe-Lintner CAPM (Unconditional):

$$E[R_i] = R_f + \beta_i(E[R_m] - R_f)$$

Sharpe-Lintner CAPM (Conditional):

$$E_t[R_{i,t+1}] = R_{f,t+1} + \beta_{i,t+1}(E_t[R_{m,t+1}] - R_{f,t+1})$$

Conditional CAPM (Cont.)

- Replace expected return with realized return, subtract risk-free rate, and rewrite the Security Market Line:

$$r^e_{i,t+1} = \alpha_{i,t+1} + \beta_{i,t+1} r^e_{m,t+1} + \epsilon_{i,t+1}$$

- Take expectation at time t with information only up to t:

$$\begin{aligned}
E_t[r^e_{i,t+1}] &= E_t[\alpha_{i,t+1} + \beta_{i,t+1} r^e_{m,t+1} + \epsilon_{i,t+1}] \\
&= E_t[\alpha_{i,t+1}] + E_t[\beta_{i,t+1} r^e_{m,t+1}] + E_t[\epsilon_{i,t+1}] \\
&= E_t[\alpha_{i,t+1}] + E_t[\beta_{i,t+1} r^e_{m,t+1}] \\
&= E_t[\alpha_{i,t+1}] + E_t[\beta_{i,t+1}] E_t[r^e_{m,t+1}] + Cov_t[\beta_{i,t+1}, r^e_{m,t+1}]
\end{aligned}$$

- If we have $Cov_t[\beta_{i,t+1}, r^e_{m,t+1}] = 0$, we have the static CAPM back.

Conditional CAPM (Cont.)

- In general, $Cov_t[\beta_{i,t+1}, r^e_{m,t+1}] \neq 0$.
- During bad economic times, the expected market risk premium is relatively high (Why is this the case?), more leveraged firms are likely to face more financial difficulties and have higher conditional betas.
- Given I_t, $Cov_t[\beta_{i,t}, r^e_{m,t}] = 0$ is testable.
  $\rightarrow$ This is the base of conditional CAPM.
- For more details, check Lewellen, Nagel (2006, JFE).

Empirical Findings in the 80s/90s: Anomalies

- Under-reactions and positive corrections:
  - French (1980, JFE): Monday effect
  - Reinganum (1983, JFE): January effect
  - Bernard, Thomas (1989,JAR): post-earnings announcement drift (PEAD)
- Over-reactions and negative reversals:
  - Jegadeesh, Titman (1993, JF): mid-horizon momentum (3-12 months)
  - DeBondt, Thaler (1985, JF): long-horizon reversals (3-5 years)
- Pricing Factors:
  - Banz (1981, JFE): firm size is negatively related to returns.
  - Fama, French (1992, JF): size and book-to-market effect

Arbitrage Pricing Theory

- Ross (1977, JET): Arbitrage Pricing Theory (APT).
  $\rightarrow$ theoretical base for multi-factor models.

- For some stock i:

$$r^e_{i,t} = \alpha_i + \beta_i r^e_{m,t} + \epsilon_{i,t}$$

- As long as $\epsilon_{i,t}$ is not correlated with $\epsilon_{j,t}$ for all $j \neq i$, we can form portfolios (for simplicity, equal-weighting) using N stocks with $r^e_{p,t} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i + (\frac{1}{N}\sum_{i=1}^{N}\beta_i) r^e_{m,t}$

- We could then conduct a $beta - hedge$ with one long position in $r^e_{p,t}$ and $(\frac{1}{N}\sum_{i=1}^{N}\beta_i) \equiv \beta_p$ short position in the market premium.
  We end up with a riskless $\frac{1}{N}\sum_{i=1}^{N}\alpha_i$ return!

Arbitrage Pricing Theory (Cont.)

- By the no-arbitrage argument,

$$\frac{1}{N}\sum_{i=1}^{N}\alpha_i \text{ must} = 0$$

- At the end, we have

$$E[r^e_{p,t}] = E[\beta_p r^e_{m,t}] \text{ and } E[r^e_{i,t}] = E[\beta_i r^e_{i,t}]$$

- The critical assumption behind is:
  No correlation between $\epsilon_{i,t}$ and $\epsilon_{j,t}$ for all $i \neq j$!!!

- In practice, no single factor accounts for all the correlations between asset returns. There are many other sources of common variation: HML, SMB, industry effects, etc. $\rightarrow$ **Call for more pricing factors!!**

Fama French 3/5 Factor Model

- Fama, French (1992, JF)

$$r_{i,t}^{e} = \alpha_i + \beta_i^{m} r_{m,t}^{e} + \beta_i^{smb} r_{SMB,t} + \beta_i^{hml} r_{HML,t} + \epsilon_{i,t}$$

- Fama, French (2015, JFE)

$$\begin{aligned} r_{i,t}^{e} = \alpha_i + \beta_i^{m} r_{m,t}^{e} + \beta_i^{smb} r_{SMB,t} + \beta_i^{hml} r_{HML,t} \\ + \beta_i^{rmw} r_{RMW,t} + \beta_i^{cma} r_{CMA,t} + \epsilon_{i,t} \end{aligned}$$

- Check K. French's data library for details about construction methods.

Multi-Factor Models - Remarks

- Applications:
  - Portfolio selection and risk management
  - Performance evaluation
  - Measuring abnormal returns in event studies
  - Estimating the cost of capital
- Rationale: a manager who has a high expected return, but also has a strong loading on the SMB or HML factors, is not necessarily a good investor; he or she is simply exploiting the known empirical fact that there are high expected returns associated with SMB and HML.
- Problem: Is it risk or anomaly? Still up for debates.

Benchmarks for Abnormal Returns

- Fama-French 3 factor model:
  - MKTRF (market), SMB (size), HML (value).

- Carhart 4 factor model:
  - FF 3 factors, Momentum (UMD).

- DGTW (Daniel, Grinblatt, Titman and Wermers (1997, JF) Style-Adjusted Model
  - 125 portfolio sorted based on market cap, book-to-market, momentum ($5 \times 5 \times 5$).
  - Use the Size/BM/Momentum bucket portfolio which a stock falls into as the comparing benchmark.

Time-Series Regression - Caveats

國立臺灣大學 National Taiwan University

- After learning all the knowledge about asset pricing models, now it's the time to conduct return predictability tests.

- You may run time-series regressions for stock or portfolios to test how 1.past returns; 2.factors; 3.firm characteristics (accounting-related or trading/investment-related) affect future returns.

- Before we move on, there are two issues we need to deal with:
  1. Potential serial correlation in the error terms.
     → detect using Box-Pierce, variance ratio, or Durbin-Watson tests.
     → How to fix it if we find serial correlation? Newey-West SE!

  2. Problems of over-fitting.
     → How to avoid in-sample over-fitting? Out-of-sample tests!
     → Check Welch, Goyal (2008, RFS)!

- Newey, West (1987, ECMA): **HAC** estimator for standard errors $\rightarrow$ **Heteroskedasticity and Autocorrelation Consistent**.

- Let's begin with a standard regression setting:

$$\mathbf{Y}_{T\times 1} = \mathbf{X}_{T\times k}\boldsymbol{\beta}_{k\times 1} + \mathbf{u}_{T\times 1}$$

or at individual-level

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t, \quad \text{where} \quad \mathbf{x}_t \text{ is } (k \times 1)$$

$$u_t \sim (0, \sigma_t^2), \quad t = 1, 2, \cdots, T$$

Newey-West Standard Errors (Cont.)

National Taiwan University

- With homoskedasticity assumption:

$$\widehat{\boldsymbol{\beta}}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \text{and} \quad Var(\widehat{\boldsymbol{\beta}}^{OLS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

or

$$\widehat{\boldsymbol{\beta}}^{OLS} = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t y_t\right)$$

and

$$Var(\widehat{\boldsymbol{\beta}}^{OLS}) = \sigma^2\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}$$

Newey-West Standard Errors (Cont.)

- Without homoskedasticity assumption:

$$Var(\widehat{\boldsymbol{\beta}}^{OLS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$$

  $\boldsymbol{\Omega}$ is the var-cov matrix of innovation terms.

- The White robust standard error is:

$$Var^R(\widehat{\boldsymbol{\beta}}^{OLS}) = \frac{T}{T-k}\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\hat{u}_t{}^2\right)\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}$$

- The White robust SE deals with heteroskedastic innovation terms but still assume for 0 covariance.

Newey-West Standard Errors (Cont.)

- With heteroskedastic and autocorrelated innovation terms, Newey-West (1987) provides an intuitive way to estimate the $(\mathbf{X}'\mathbf{\Omega}\mathbf{X})$ component:

$$\left[\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \hat{u}_t{}^2 + \sum_{j=1}^{q} \left(1 - \frac{j}{q+1}\right) \sum_{t=j+1}^{T} \left(\mathbf{x}_t \hat{u}_t \hat{u_{t-j}} \mathbf{x}_{t-j}' + \mathbf{x}_{t-j} \hat{u_{t-j}} \hat{u}_t \mathbf{x}_t'\right)\right]$$

- For detail derivation, please check 陳旭昇 (2013), Chapter 15.

- The number of q is up to your own choice/judgement.

- When $q = 1$, The red part is:

$$\frac{1}{2} \quad \hat{\gamma}(1) \quad \sum_{t=2}^{T} (\mathbf{x}_t \mathbf{x}'_{t-1} + \mathbf{x}_{t-1} \mathbf{x}'_t)$$

- When $q = 2$, The red part is:

$$\frac{2}{3} \quad \hat{\gamma}(1) \quad \sum_{t=2}^{T} (\mathbf{x}_t \mathbf{x}'_{t-1} + \mathbf{x}_{t-1} \mathbf{x}'_t) \quad + \quad \frac{1}{3} \quad \hat{\gamma}(2) \quad \sum_{t=3}^{T} (\mathbf{x}_t \mathbf{x}'_{t-2} + \mathbf{x}_{t-2} \mathbf{x}'_t)$$

- In STATA, use the *newey* command for Newey-West estimator.

Out-Of-Sample test

- Forecasting error:
  difference between realized value and expected (predicted) value

$$e_{t,t+k} = y_{t+k} - E[y_{t+k}]$$

- How do we evaluate a forecasting?
  $\rightarrow$ how long of forecasting error can we attain.

- Most common evaluation: comparing Mean Square Error (MSE), the quadratic loss function.

$$\widehat{MSE} = \frac{1}{T} \sum_{t=1}^{T} \hat{e}^2_{t,t+k}$$

Diebold-Mariano test

- For two time-series models A and B, if we have $\widehat{MSE}^A < \widehat{MSE}^B$, we can say that model A is a better prediction model.

- How large should the difference between $\widehat{MSE}^A$ and $\widehat{MSE}^B$ for us to make a statistical inference?

- Define:

$$d_t = (\hat{e}^A_{t,t+k})^2 - (\hat{e}^B_{t,t+k})^2 \quad \text{and} \quad \bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$$

- Diebold, Mariano (1995):

$$DM = \frac{\bar{d}}{\sqrt{\frac{\hat{G}}{T-1}}} \sim t(T-1), \quad \text{where} \quad \hat{G} = \hat{\gamma}(0) + 2\sum_{j=1}^{m} \hat{\gamma}(j)$$

Diebold-Mariano test (Cont.)

National Taiwan University

- DM suggest $m = T^{1/3}$ (round it to the nearest integer).
- With large T, $DM \xrightarrow{d} N(0,1)$
- Diebold-Mariano test is flexible in the loss function. We could also use Mean Absolute Error (MAE) or other loss function instead of MSE.
- General OOS test:
  - In most cases, the benchmark model we test against is random walk!
- Four ways to separate in-sample and out-of-sample observations:
  1. Half the data. $\leftarrow$ Least favorable.
  2. Rolling (fixed) windows.
  3. Rolling recursively.
  4. Real-time OOS test. $\leftarrow$ Most favorable but least feasible.