Financial Econometrics

Eric Zivot

April 10, 2002

1 Lecture Outline

- Market Efficiency
- The Forms of the Random Walk Hypothesis
- Testing the Random Walk Hypothesis
- Long Horizon Returns
- Long Memory in Asset Returns

2 Martingales and Martingale Difference Sequences (MDS)

Let $\{Y_t\}$ be a sequence of random variables and let $\{I_t\}$ denote a sequence of information sets with $I_t \subset I$ for all t where $I =$ universal info set. (Y_t, I_t) is a *martingale* if

- $I_t \subset I_{t+1}$ (filtration)
- $Y_t \subset I_t$ (Y_t is adapted to I_t)
- $E[|Y_t|] < \infty$
- $E[Y_t | I_{t-1}] = Y_{t-1}$ (martingale property)

Example: Random Walk

$$\begin{aligned} Y_t &= Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim iid(0, \sigma^2) \\ I_t &= \{\varepsilon_t, \varepsilon_{t-1}, \ldots\} \\ E[Y_t | I_{t-1}] &= Y_{t-1} \end{aligned}$$

Example: Independent heteroskedastic $I(1)$ process

$$\begin{aligned} Y_t &= Y_{t-1} + \frac{\varepsilon_t}{t} = Y_{t-1} + u_t, \\ & \quad u_t \sim inid(0, \sigma_t^2) \\ I_t &= \{u_t, u_{t-1}, \ldots\} \\ E[Y_t | I_{t-1}] &= Y_{t-1} \end{aligned}$$

Law of Iterated Expectations. Let $\{Y_t, I_t\}$ be a martingale. Then

$$\begin{aligned} E[Y_t | I_{t-2}] &= E[E[Y_t | I_{t-1}] | I_{t-2}] \\ &= E[Y_{t-1} | I_{t-2}] = Y_{t-2} \end{aligned}$$

It follows that

$$E[Y_t | I_{t-k}] = Y_{t-k}$$

In general, for information sets I_t and J_t such that $I_t \subset J_t$ (J_t is the bigger info set). The Law of Iterated Expectations says

$$E[Y|I_t] = E[E[X|J_t]|I_t]$$

Martingale Difference Sequence (MDS). (X_t, I_t) is a MDS if (X_t, I_t) is an adapted sequence and

$$E[X_t|I_{t-1}] = 0$$

Remark 1: If Y_t is a martingale and $X_t = Y_t - E[Y_t|I_{t-1}]$ then X_t is a MDS by construction.

Remark 2: Let Z_t be any nonlinear function of the past history of Y_t so that $Z_t \subset I_t$. Then by the Law of Iterated Expectations

$$\begin{aligned} E[X_t Z_{t-1}] &= E[E[X_t Z_{t-1}|I_{t-1}] \\ &= E[Z_{t-1} E[X_t|I_{t-1}] \\ &= 0 \end{aligned}$$

so that X_t is uncorrelated with any nonlinear function of the history of Y_t

Example: ARCH Process

$$\begin{aligned} X_t &= \varepsilon_t \sigma_t,\ \varepsilon_t \text{ ˜ } iid\ (0,1) \\ \sigma_t^2 &= \alpha + \beta \sigma_{t-1}^2,\ \alpha > 0 \text{ and } |\beta| < 1 \\ E[X_t|I_{t-1}] &= E[\varepsilon_t \sigma_t | I_{t-1}] \\ &= \sigma_t E[\varepsilon_t | I_{t-1}] \\ &= 0 \end{aligned}$$

3 Market Efficiency

Unpredictable asset returns is the result of the Law of Iterated Expectations. *Samuelson's fameous result*: Let $V^* =$ fundamental value of asset and assume P_t is a rational forecast. Then

$$\begin{aligned} P_t &= E[V^*|I_t] \\ P_{t+1} &= E[V^*|I_{t+1}] \\ E[P_{t+1} - P_t|I_t] &= E[E[V^*|I_{t+1}] - E[V|I_t]|I_t] \\ &= E[V^*|I_t] - E[V^*|I_t] = 0 \end{aligned}$$

3.1 Types of Market Efficiency

- *Weak Form*: Information set includes only the history of prices or returns

- *Semistrong Form*: The information set includes all publicly available information

- *Strong Form*: The information set contains all public and private information

3.2 Testing Market Efficiency

- Any test of market efficiency must assume an equilibrium model that defines normal security returns (e.g. CAPM)

- Perfect efficiency is unrealistic. Grossman and Stiglitz (1980) argue that you need some inefficiency to promote information gathering activity.

4 The Random Walk Hypotheses

$$
\begin{aligned}
p_t &= \mu + p_{t-1} + \varepsilon_t,\ p_t = \ln(P_t) \\
&\Longrightarrow\ r_t = \mu + \varepsilon_t,\ r_t = \Delta p_t
\end{aligned}
$$

- RW1: ε_t is independent and identically distributed (iid) $(0, \sigma^2)$. Not realistic

- RW2: ε_t is independent (allows for heteroskedasticity). Test using filter rules, technical analysis

- RW3: ε_t is uncorrelated (allows for dependence in higher moments). Test using autocorrelations, variance ratios, long horizon regressions

4.1 Autocorrelation Tests

Assume that r_t is covariance stationary and ergodic. Then

$$\begin{aligned} \gamma_k &= cov(r_t, r_{t-k}) \\ \rho_k &= \gamma_k / \gamma_0 \end{aligned}$$

and sample estimates are

$$\begin{aligned} \hat{\gamma}_k &= \frac{1}{T} \sum_{t=1}^{T-k} (r_t - \bar{r})(r_{t+k} - \bar{r}), \; \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \\ \bar{r} &= \frac{1}{T} \sum_{t=1}^{T} r_t \end{aligned}$$

Result: Under RW1

$$\begin{aligned} E[\hat{\rho}_k] &= -\frac{T-k}{T(T-1)} + O(T^2) \\ & \sqrt{T}\hat{\rho}_k \overset{A}{\sim} N(0,1) \end{aligned}$$

Box-Pierce Q-statistic: Consider testing $H_0 : \rho_1 = \cdots = \rho_m = 0$. Under RW1

$$MQ = T(T+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{T-k} \sim \chi^2(m)$$

4.2 Variance Ratios

Intuition. Under RW1

$$VR(2) = \frac{var(r_t(2))}{2 \cdot var(r_t)} = \frac{var(r_t + r_{t-1})}{2 \cdot var(r_t)} = \frac{2\sigma^2}{2\sigma^2} = 1$$

If r_t is a covariance stationary process then

$$\begin{aligned} VR(2) &= \frac{var(r_t) + var(r_{t-1}) + 2cov(r_t, r_{t-1})}{2 \cdot var(r_t)} \\ &= \frac{2\sigma^2 + 2\gamma_1}{2\sigma^2} = 1 + \rho_1 \end{aligned}$$

Three cases:

- $\rho_1 = 0 \Longrightarrow VR(2) = 1$
- $\rho_1 > 0 \Longrightarrow VR(2) > 1$ (mean aversion)
- $\rho_1 < 0 \Longrightarrow VR(2) < 1$ (mean reversion)

General $q-period$ variance ratio under stationarity

$$
\begin{aligned}
VR(q) &= \frac{var(r_t(q))}{q \cdot var(r_t)} = 1 + 2\sum_{k=1}^{q-1}\left(1-\frac{k}{q}\right)\rho_k \\
r_t(q) &= r_t + r_{t-1} + \cdots + r_{t-q+1}
\end{aligned}
$$

Remark 1: Under RW1, $VR(q) = 1$.

Remark 2: For stationary and ergodic returns with a 1-summable Wold representation

$$
\begin{aligned}
r_t &= \mu + \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j},\ \varepsilon_t \sim iid(0,\sigma^2) \\
\psi_0 &= 1, \sum j|\psi_j| < \infty
\end{aligned}
$$

it can be shown that

$$
\begin{aligned}
\lim_{q\to\infty} VR(q) &= \frac{\sigma^2\psi(1)^2}{\gamma_0} \\
&= \frac{lrv(r_t)}{var(r_t)} = \frac{\text{long-run variance}}{\text{short-run variance}}
\end{aligned}
$$

Remark 3: Under RW2 and RW3, $VR(q) = 1$ provided

$$\frac{1}{T}\sum_{t=1}^{T} var(r_t) \to \bar{\sigma}^2 > 0$$

4.2.1 Lo and MacKinlay's Test Statistics

Under RW1, the standardized variance ratio

$$\begin{aligned}\psi(q) &= (VR(q) - 1) \cdot \left(\frac{2(2q-1)(q-1)}{3Tq}\right)^{-1/2} \\ &\overset{A}{\sim} N(0,1)\end{aligned}$$

Under RW2 and RW3 the heteroskedasticty-robust standardized variance ratio

$$
\begin{aligned}
\psi^*(q) &= (VR(q) - 1) \cdot \Omega(q)^{-1/2} \\
\Omega(q) &= \sum_{j=1}^{q-1} \left(\frac{2(q-j)}{j} \right)^2 \delta_j \\
\delta_j &= \frac{\sum_{t=j+1}^{T} \alpha_{0t}\alpha_{jt}}{\left(\sum_{t=1}^{T} \alpha_{0t} \right)^2} \\
\alpha_{jt} &= \left(r_{t-j} - r_{t-j-1} - \frac{1}{T}(r_T - r_0) \right)^2
\end{aligned}
$$

is asymptotically standard normal.

4.3 Empirical Results

CML chapter 2, section 8. CRSP value-weighted and equal weighted indices, individual securities from 1962 - 1994

- Daily, weekly and monthly cc returns from VW and EW indices show significant 1st order autocorrelation

- $VR(q) > 1$ and $\psi^*(q)$ statistics reject RW3 for EW index but not VW index.

  – Market capitalization or size may be playing a role. In fact, $VR(q) > 1$ and $\psi^*(q)$ are largest for portfolios of small firms.

- For individual securities, typically $VR(q) < 1$ (negative autocorrelation) and $\psi^*(q)$ is not significant!!! How can portfolio $VR(q) > 1$ when individual security $VR(q) < 1$?

4.3.1 Cross lag autocorrelations and lead-lag relations

Result: Portfolio returns can be positively correlated and securities returns can be negatively correlated if there are positive cross lag autocorrelations between the securities in the portfolio.

Let $\mathbf{R}_t$ denote an $N \times 1$ vector of security returns. Define

$$
\begin{aligned}
\gamma_{ij}^k &= cov(r_{it}, r_{jt-k}) = \text{ cross lag autocorrelation} \\
\mathbf{\Gamma}_k &= cov(\mathbf{R}_t, \mathbf{R}_{t-k}) = \begin{pmatrix} \gamma_{11}^k & \gamma_{12}^k & \cdots & \gamma_{1N}^k \\ \gamma_{21}^k & \gamma_{22}^k & \cdots & \gamma_{2N}^k \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N1}^k & \gamma_{N2}^k & \cdots & \gamma_{NN}^k \end{pmatrix}
\end{aligned}
$$

Let $R_{mt} = \mathbf{1}'\mathbf{R}_t/N =$ equally weighted portfolio. Then

$$
\begin{aligned}
cov(R_{m,t}, R_{m,t-1}) &= \frac{1}{N^2}\mathbf{1}'\mathbf{\Gamma}_1\mathbf{1} \\
corr(R_{m,t}, R_{m,t-1}) &= \frac{\mathbf{1}'\mathbf{\Gamma}_1\mathbf{1} - tr(\mathbf{\Gamma}_1)}{\mathbf{1}'\mathbf{\Gamma}_0\mathbf{1}} + \frac{tr(\mathbf{\Gamma}_1)}{\mathbf{1}'\mathbf{\Gamma}_0\mathbf{1}}
\end{aligned}
$$

5 Long Horizon Returns

Define

$$
\begin{aligned}
r_t &= \ln(P_t/P_{t-1}) = \text{monthly cc return} \\
r_t(12) &= \ln(P_t/P_{t-12}) = \sum_{j=0}^{11} R_{t-j} = \text{annual cc return}
\end{aligned}
$$

Suppose r_t ~$iid(\mu, \sigma^2)$ and consider a sample $\{r_1, r_2, \ldots, r_T\}$ of size T of monthly returns. A sample of annual returns may be created in two ways:

- *Overlapping sample*:

  $$\{r_{12}(12), r_{13}(12), \ldots, r_T(12)\}$$

  is a sample of $T - 11$ monthly overlapping annual returns

- *Non-overlapping sample*:

  $$\{r_{12}(12), r_{24}(12), \ldots, r_T(12)\}$$

  is a sample of $T/2$ non-overlapping annual returns

Result: If r_t ~$iid(\mu, \sigma^2)$ then $r_t(12)$ in overlapping sample follows an MA(11) process since:

$$\begin{aligned}\gamma_j &= cov(r_t(12), r_{t-j}(12)) = (12 - j)\sigma^2 \text{ for } j < 12 \\ \gamma_j &= 0 \text{ for } j \geq 12\end{aligned}$$

Implication 1:

$$\begin{aligned}\bar{r}(12) &= \frac{1}{T}\sum_{t=1}^{T-11} r_t(12) \overset{A}{\sim} N\left(\mu_{12}, \frac{lrv}{T}\right) \\ lrv &= \gamma_0 + 2\sum_{j=1}^{11}\gamma_j = \text{ long-run variance}\end{aligned}$$

Implication 2: Newey-West HAC standard errors should always be computed in regressions where the dependent variable is multi-period returns from overlapping data! Note: Monte Carlo studies have shown that Newey-West HAC standard errors are not very good in small samples.

6 Long Memory

A *fractionally integrated* white noise process y_t has the form

$$(1-L)^d p_t = \varepsilon_t,\ \varepsilon_t \,\tilde{}\, WN(0,\sigma^2)$$

where $(1-L)^d$ has the binomial series expansion representation (valid for any $d > -1$)

$$\begin{aligned}(1-L)^d &= \sum_{k=0}^{\infty} \binom{d}{k}(-L)^k \\ &= 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 +\end{aligned}$$

Hence, p_t has an AR(∞) representation

$$\begin{aligned}\sum_{k=0}^{\infty} \phi_k p_{t-k} &= \varepsilon_t \\ \phi_k &= (-1)^k \binom{d}{k} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}\end{aligned}$$

as well as a MA(∞) representation

$$\begin{aligned} p_t &= (1-L)^{-d}\varepsilon_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \\ \psi_k &= \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} \end{aligned}$$

Special cases:

- $d = 1$ then p_t is a random walk

- $d = 0$ then p_t is white noise.

- For $0 < d < 1$ it can be shown that
$$\rho_k \propto k^{2d-1}$$
so that the ACF for p_t declines hyperbolically to zero at a speed that depends on d.

  – p_t is stationary and ergodic for $0 < d < 0.5$

  – The variance of p_t is infinite for $0.5 \leq d < 1$.

7 Long Horizon Regressions of Returns on Valuation Ratios

Cochrane (2001) gives the following summary

- Dividend/Price ratios forecast excess returns on stocks. Regression coefficients and R^2 rise with the forecast horizon. This is a result of the fact that the forecasting variable is persistent

Consider the stylized model relating returns to a persistent valuation ratio like dividend/price

$$\begin{aligned} r_{t+1} &= \beta x_t + \varepsilon_{t+1},\ \beta > 0 \\ x_{t+1} &= \rho x_t + \eta_{t+1},\ 0 < \rho < 1 \end{aligned}$$

The relationship between $r_{t+2}(2) = r_{t+2} + r_{t+1}$ and x_t is

$$\begin{aligned} r_{t+1}(2) &= \beta x_{t+1} + \beta x_t + \varepsilon_{t+2} + \varepsilon_{t+1} \\ &= \beta(\rho x_t + \eta_{t+1}) + \beta x_t + \varepsilon_{t+2} + \varepsilon_{t+1} \\ &= \beta(1+\rho)x_t + w_{2t} \\ &= \beta_2 x_t + w_t,\ \beta_2 > \beta. \end{aligned}$$

In general,

$$
\begin{aligned}
r_{t+1}(k) &= \beta(1+\rho+\cdots+\rho^{k-1})x_t + w_{kt} \\
&= \beta_k x_t + w_{kt},\ \beta_k > \beta_{k-1}
\end{aligned}
$$

Note: The population value of the numerator in the long-horizon regression is

$$E[(r_{t+k} + r_{t+k-1} + \cdots + r_{t+1})x_t]$$

which, under stationarity is the same as

$$E[r_{t+1}(x_t + x_{t-1} + \cdots + x_{t-k+1})]$$

so that the regression of $r_{t+1}(k)$ on x_t behaves like the regression of r_{t+1} on k lags of x_t