FINANCIAL ECONOMICS AND MARTINGALE MODELS AS AN ALTERNATIVE TO RANDOM WALK MODELS.

Leroy (1989, JEL p. 1583) surveys this topic as follows. Capital markets are efficient if the publicly available knowledge is very large, and no profitable trading rules are available. An empirical test of efficiency is to see if an agent can gain a comparative advantage over his rivals by acquiring information. A French mathematician, Louis Bachalier formulated the random walk (RW) model in 1900 in his Ph.D. thesis. Many authors have found empirical and theoretical support for the RW or Brownian motion • where the Brownian motion used by Bachalier is simply a continuous time version of the RW. Does this mean that preferences and technology have no effect on stock market prices? Why do rational agents spend large amounts of money in gathering and analysis of information ? Why does investment advise fail to produce extra normal returns ? Can one forecast the stock market is a problem that has fascinated many thinkers for a long time. Forecast price changes ?p from past $(?p)_{t\text{-}j}$ only and no other information leads to RWalk and martingale, two most important ideas grew out of this exercise to ``beat the market".

The martingale model was advocated by Samuelson (1965) as an alternative for the RW model. Samuelson proved the proposition that properly anticipated prices fluctuate randomly. Thus, with simple assumptions randomness can be linked with the basic elements of economic equilibria including preferences and returns. It provides a precise way in which information is reflected in the asset prices. A stochastic process x_t is a martingale within an information set I_t if the best forecast of x_{t+1} based on current information I_t would be equal to x_t, or formally,

$$\text{Martingale:} \quad E(x_{t+1}| I_t) = x_t \qquad (1.1)$$

or conditionally expected returns are constant. (CONSTANT Conditionally Expected RETURNS)

A stochastic process y_t is a fair game if its forecast would be zero for any possible value of I_t.

$$\text{Fair Game:} \quad E(y_{t+1}| I_t) \ \text{œ}\ 0. \qquad (1.2)$$

Thus, one can easily infer that x_t is a martingale if and only if x_{t+1} • x_t is a fair game. The rates of return are a fair game only when stock returns (=prices + cumulative dividends discounted back to the present at rate 3) series is a martingale. Then stock price p_t is given by:

$$\text{(One-period) Fair Game Price:} \quad p_t \ \text{œ}\ (1+3)^{-1} E(p_{t+1} + d_{t+1}| I_t) \qquad (1.3)$$

where d denotes dividends. Now equation (1.3) tells us that the stock prices of today are equal to the sum of the expected future prices and dividends, discounted back to the present at the rate p. Equation (1.3) can be rewritten by removing I_t and introducing a subscript for E as:

$$\text{Fair Game Price:} \quad p_t \ \text{œ}\ (1+3)^{-1} E_t\,(p_{t+1} + d_{t+1}) \qquad (1.4)$$

Now the price or dividend variables are not martingales. The variable that is a martingale is the discounted value of an imaginary mutual fund that holds the particular stock whose price follows (1.4). To see this, let $u_t = (1+3)^{-t} p_t n_t$ be the value of the mutual fund discounted back to date zero. Here n_t is defined as the number of shares of stock that the mutual fund holds at t. At time t+1 the value includes the dividend as follows.

$$p_{t+1}n_{t+1} \ \text{œ}\ (p_{t+1} + d_{t+1})\, n_t \qquad (1.5)$$

$$E_t(u_{t+1}) \ \text{œ}\ E_t\,[(1+3)^{-(t+1)}p_{t+1}n_{t+1}]$$

$$\text{œ}\ E_t\,[(1+3)^{-(t+1)}(p_{t+1}+d_{t+1})\, n_t] \ \text{, using (1.5)}$$

$$\text{œ}\ (1+3)^{-t}p_t n_t \ \text{œ}\ u_t \qquad (1.6)$$

Hence u_t is a martingale by the definition (1.1).(Conditional Expectation of Discounted Present Value of mutual fund is constant)

In the realm of efficient markets literature, stock price should be understood to include reinvested dividends. To find a variable that predicts future returns could mean either of the following: (1) that the capital market is inefficient (i.e., it does not satisfy the martingale property), or (2) the predictive variable is not in the agent's information sets.

Samuelson proved that if the stock market is efficient in the sense that the prices are a fair game as in (1.4), then stock prices should equal the expected present value of future dividends:

$$\text{Efficient market price} = p_t = \sum_{i=1}^{\infty} (1+\rho)^{-i} E_t(d_{t+i}) \tag{1.7}$$

Derivation of equation (1.7) from the Fair Game price model of (1.4):

First, we replace t by (t+1) in (1.4) $p_{t+1} = (1+\rho)^{-1} E_{t+1} (p_{t+2} + d_{t+2})$
and use the resulting equation to replace the price term p_{t+1} in (1.4: fair game price = discounted $(\text{price+dividend})_{t+1}$

$$p_t = (1+\rho)^{-1} E_t [(1+\rho)^{-1} E_{t+1} (p_{t+2} + d_{t+2}) + d_{t+1}] \tag{1.8}$$

Now the information set I_{t+1} is obviously more informative than I_t, as it is assumed that the agents never forget the past. Thus the LAW of Iterated Expectations (LIE) guarantees that $E_t [E_{t+1}(p_{t+2})] = E_t(p_{t+2})$. Therefore (1.8) becomes:

$$p_t = (1+\rho)^{-1} E_t (d_{t+1}) + (1+\rho)^{-2} E_t (p_{t+2} + d_{t+2}) \tag{1.9}$$

So far, we have 2 terms in

$$\text{Efficient market price} = p_t = \sum_{i=1}^{2} (1+\rho)^{-i} E_t(d_{t+i}) + (1+\rho)^{-2} E_t(p_{t+2})$$

We get the Efficient market price (1.7) after proceeding similarly n times, and assuming that the last extra term involving prices disappears as $(1+\rho)^{-n} E_t(p_{t+n})$ converges to zero due to heavy discounting (dividing by a large number) as n gets large. □

Samuelson was the first to use the LIE here which says that $E_t(p)=E_tE_{t+1}(p)$. If price change to t+1 is not forecastable with inf.ormation set at time t, then price change from t+1 to t+2 based on unavailable future information is also NOT forecastable at time t.

Stock Price Fundamentalists and the Fair Game Pricing by the Martingale

The stock market fundamentalists believe that the stock prices reflect the fundamental value of the companies they represent. The above derivation shows that fair game pricing model can lead to the prices being equal to the "expected present value of future dividends" from (1.7). Since the fundamental value of the firm may also be measured by (1.7), the fundamentalist position is consistent with the martingale. The only reason why the two seem diametrically opposite is that the fundamentalists focus attention on the stable part of the prices whereas the martingale model emphasizes the random part. Both are present, and it is merely a matter of emphasis. If enough traders seek and find the true fundamental value, the price will equal the fundamental value and trading profits will disappear leading to the fair game pricing.

The Assumptions Regarding Preferences Necessary for the Martingale

The martingale model is satisfied under several special assumptions regarding preferences noted below. First, let us assume that the agents have common and constant time preferences, have common probabilities, and are risk-neutral. If these conditions prevail, investors will prefer to hold that asset which generates the highest expected return, completely ignoring the risk factor. Ignoring risk is called risk neutrality, and it implies the martingale as in equation (1.1), but NOT the more restrictive random walk (RW) model:

$$p_{t+1} = p_t + \varepsilon_t \tag{1.1a}$$

where ε_t is STN$(0,\sigma^2_\varepsilon)$, a stationary series with mean zero and variance σ^2_ε. Although, $E(p_{t+1}) = p_t$ as in (1.1), the RW model also assumes that there is NO autocorrelation among the VARIANCES. For example, the stock prices may have successive periods when the variances are stable and other periods having sudden changes in the variances • sometimes called autoregressive conditional heteroscedasticity (ARCH) effects. Martingales permit the ARCH effects and imply risk neutrality.

Now since risk neutrality implies that NO ONE cares about variances, therefore, the very fact that future conditional variances are partly forecastable becomes totally irrelevant. Thus, Samuelson's results showed that the theoretical implications of efficient markets lead to the martingale and not to the random walk model. Furthermore, the empirical testing for the presence of serial correlation among stock returns do not test the market efficiency, but a weak form of the martingale model.

Fama's (1970) paper on this was devoted almost exclusively to empirical work. Fama identified market efficiency with the assumption that y_t is a fair game:

$$E(y_{t+1}| I_t) = 0 \tag{2.1}$$

where y_{t+1} is defined to equal the price of some security at t+1 less its conditional expectation (Exp of excess price is zero?):

$$y_{t+1} = p_{t+1} - E(p_{t+1}| I_t) \tag{2.2}$$

The problem with Fama's characterization of market efficiency is that (2.1) follows tautologously from the definition (2.2) of y_{t+1} − we just have to take conditional on I_t, on both sides of (2.2). By Fama's definition, any capital market is efficient and no empirical evidence can possibly bear on the question of market efficiency.

<u>Efficient</u> <u>Markets</u> <u>and</u> <u>Volatility</u> <u>of</u> <u>Stock</u> <u>Prices</u>

Leroy and Porter (1981) and Shiller (1979) show that the same models which implied that returns should be unforecastable also implied that asset prices should be less volatile than the dividends. However, the empirical studies show that the volatility of dividends is far less than that of stock prices. This leads to a <u>clean rejection</u> of efficient markets model because the measurement of volatility can be made without reference to the information sets of the agents. If the martingale model is tested directly, it involves the information set and leaves open the argument that the markets are inefficient because we have not included the correct variables in the information set.

LeRoy and Porter (1981) assume that all variables have finite means and variances. The fair game assumption of (1.4) plus the definition of rate of return imply that p_t can be written as:

$$p_t = (1+\rho)^{-1}(d_{t+1} + p_{t+1}) - (1+\rho)^{-1} e_{t+1} \tag{3.1}$$

where, e_{t+1} denotes the <u>unexpected</u> <u>component</u> of a one-period return on the stock. It too should be discounted. At time t+i, the unexpected component is:

$$e_{t+i} \equiv d_{t+i} + p_{t+i} - E_{t+i-1}(p_{t+i} + d_{t+i}) \tag{3.2}$$

Replacing the t by t+i in (3.1) and multiplying both sides by $(1+\rho)^{-i}$ we have:

$$(1+\rho)^{-i} p_{t+i} = (1+\rho)^{-(i+1)}(d_{t+i+1} + p_{t+i+1}) - (1+\rho)^{-(i+1)} e_{t+i+1} \tag{3.3}$$

Now summing (3.3) over i from zero to infinity and assuming convergence, we have the <u>ex</u> <u>post</u> rational stock price (Shiller's terminology):

$$p_t^* = p_t + x_t \quad (x_t \text{ is forecast error}) \tag{3.4}$$

where, $p_t^* = \sum_{i=1}^{\infty}(1+\rho)^{-i} d_{t+i}$ =price if divi. were perfectly forecast (3.5)

and $x_t = \sum_{i=1}^{\infty}(1+\rho)^{-i} e_{t+i}$, and efficient price p_t should be unbiased forecast of true

price p_t^* which is a discounted present value of future dividends. Hence the variance

$$V(x_t) = \sum_{i=1}^{\infty}(1+\rho)^{-2i}\sigma_\varepsilon^2 = \sigma_\varepsilon^2 \frac{1}{1-[1+\rho]^2} = \frac{V(e_t)}{2\rho+\rho^2} \tag{3.6}$$

p_t^* denotes the price of stock that would obtain if future realizations of dividends were perfectly forecastable. Taking conditional expectations, (3.4) yields:

p_t œ $E\,(p_t^*|\,I_t)$, (3.7)

so that p_t is a forecast of p_t^* given agents' information I_t. Given (3.7), (3.4) says that p_t^* can be expressed as the sum of a forecast (p_t) and a forecast error (x_t). Now, optimal forecasting of prices by agents implies that the forecasts and forecast errors are uncorrelated. The uncorrelatedness (orthogonality of errors) implies that:

$V(p_t^*)$ œ $V(p_t) + V(x_t)$ [where cov =0] (3.8)

Since the variances are always non-negative, $V(p_t^*)$ is an upper bound for $V(p_t)$. The implied (TESTABLE) variance inequality

Var. of prices $V(p_t)$ Ÿ $V\,(p_t^*)$ var of dividends and discount factors, (3.9)

is attractive because the upper bound depends only on the discount factor and dividends model (3.5) with perfect foresight (information), but not on agents' information sets. Hence upper bound it empirically measured cleanly and rejection of inequality rejects martingale model (i.e., Conditional Exp is not constant, prices are more volatile than dividends).

<u>Proposition:</u> Equation (3.8) also implies that the more the information with the agents: (i) the greater will be the variance of prices, and (ii) the lower will be the variance of discounted returns.
HDV: Over time the information is increasing, so V(p) should go up $V(p^*)$ should go down. It is!

To understand this proposition let H_t denote less information than I_t the actual information. Perfect foresight model assumes the other extreme where the agents are assumed to have more information than I_t. Let $\hat{p}_t$ denote the price of stock that would obtain under the lesser information set H_t: (Hat is ignorant price?)

$\hat{p}_t$ œ $E(p_t^*|\,H_t)$ where H_t § I_t (3.10)

Now $\hat{p}_t$, like p_t^*, is a fictional stock price series that would obtain if investors had different information than they actually do. $\hat{p}_t$ and p_t^* are on opposite sides of p_t: the former price would prevail if agents have lesser information than now, while the latter would prevail under perfect information. Now $V(p_t^*)$ is an upper bound for $V(p_t)$ and it turns out that $V(p_t)$ is an upper bound for $V(\hat{p}_t)$.

The two basic facts about martingale models:
(1) Variance of stock price V(p) and variance of returns add up to $V(p^*)$ the variance of ex-post rational price [3.7 and 3.8] or
(2) Variance of ex-post rational price $V(p^*)$ does not depend on how much information agents have.

These martingale model implications also imply that hypothetical variations in agents' information set should induce a negative relation between the variance of price and the variance of returns (i.e. mean reversion)

If we can place bounds on agents' information, these will induce bounds on the variances of price and returns. Now, perfect information will be the obvious choice for the upper bound on agents' information implying that $V(p_t^*)$ is an upper bound for $V(p_t)$. Thus Fama's definition of weak-form efficiency is the obvious choice of a lower bound on agents' information implying that $V(\hat{p}_t)$ is a lower bound for $V(p_t)$.

There are two types of tests:

l. BOUNDS TESTS Ä The null hypothesis in a bounds test is satisfied if the variance of price (or returns) is less than its theoretical upper bound.

2. ORTHOGONALITY TEST Ä This is a test of the implications for variances of the equality restrictions on parameters implied by the orthogonality of forecasts and forecast errors.

Some of the variance-bounds tests were subject to severe econometric problems. M. Flavin in 1983 demonstrated that small-sample problems led to bias against acceptance of efficiency. The estimated variances of both p* and p were biased downwards. The reason is that sample means of both p* and p must be estimated and the usual procedure of reducing the degrees of freedom by one gives an inadequate correction for the induced downward bias. Now since p* is more highly autocorrelated than p, the downward bias is greater in estimating the variance of p* than of p.

Kleidon in 1986 focused on the econometric consequences of a stationary assumption. He showed that if the dividends have unit roots, problems similar to that of Flavin's could persist even in arbitrarily large samples.

If p_t is an optimal estimator of p_t^*, the difference between the two will be uncorrelated with investors' information variables, and therefore, also with $p_t - p_t^0$, where p_t^0 denotes any naive forecast * represents what we want to forecast. Thus:
$(p_t^* - p_t^0) \equiv (p_t^* - p_t) + (p_t - p_t^0)$ is an algebraic identity. Squaring both sides and taking expectations:

$$E(p_t^* - p_t^0)^2 \equiv E(p_t^* - p_t)^2 + E(p_t - p_t^0)^2 \quad (3.11)$$

implying in turn:

$$\text{Var of naive } E(p_t^* - p_t^0)^2 \qquad E(p_t^* - p_t)^2 \text{ or Var of optimal} \quad (3.12)$$

and

$$\text{Var diff(true-naive)}= E(p_t^* - p_t^0)^2 \qquad E(p_t - p_t^0)^2 \text{ Var of (optimal} - \text{naive)} \quad (3.13)$$

Mankiw, Romer and Shapiro constructed the sample counterparts of the population parameters in (3.12) and (3.13) and then checked the associated inequalities empirically. They found that both were reversed suggesting excess volatility of p_i. They characterized this exercise as an unbiased test of the variance bounds inequality.

The crucial difference between conventional efficiency tests and variance bounds is that while the former tests the orthogonality of returns over short intervals, (Return $\perp$ short) the latter tests the orthogonality of a smooth average of past returns over a period of years and a similar smooth average of future returns. (past $\perp$ future).

Fama and French U-Shaped pattern in return autocorrelations (zero, negative, zero)

Now to evaluate for the differing results of variance-bound tests and the conventional return autocorrelation tests is to estimate directly the correlation between average returns over a specified maximum lag of T. For one year lag the correlation was essentially zero. For t on the order of 3 to 5 years the correlation was negative. For T of 10 years, the correlation reverts to approximately zero. Hence they found a U-shaped pattern.

Question is what sort of model would generate the U-shaped pattern in the return autocorrelation. Shiller and Summers proposed that instead of modeling stock price as a martingale, analysts should consider assuming that prices comprise a random walk plus a fad variable where the latter is modeled as a slowly mean-reverting stationary series. This generates the (U-shape) forecastability pattern required.

For small T the dispersion dominates the drift, implying that the return autocorrelations for any stationary stochastic process look like those of a fair game (zero). Similarly, return autocorrelations over long horizons approach zero because the random walk term dominates the mean reverting component of price. In between there is a negative correlation (or mean reversion).

A different test – a hybrid of variance bounds and return autocorrelation tests – determines directly whether price predicts future returns. This usually leads to strong rejection of the martingale model.

Another way to test for mean reversion is to use variance ratios. Variance of k-period returns is k for a random walk model. $(1/k)\text{Var}(y_t - y_{t-k})$ This should be one if divided by variance of one period returns. Under a random walk the variance ratio should equal unity for any value of k. However, Poterba and Summers (1988) showed that the variance ratios declined with k, indicating presence of a mean-reverting component. y_t is not really N(0,t) as the random walk model implies. (RW is NOT mean reverting, it is nonstationary)

The presence of a mean reverting component in stock prices implies substantial forecastability of intermediate term returns and, therefore, also substantial differences between price and fundamentals [rational expectations of ex-post rational price.] Thus there is a surprisingly small degree of forecastability of short term returns which is consistent with a surprisingly large discrepancy between price and fundamental value.

Lo and MacKinlay (1988) found that weekly and monthly stock returns had positive autocorrelation coefficient of 30%, contradicting the findings of almost zero autocorrelation reported in early efficient markets literature and the prediction of approximately zero autocorrelation from mean-reversion model. Kim, Nelson and Startz (1988) found evidence of mean reversion only in the 1930s data sets.

These findings further raise questions about whether the variance bounds violations are empirically the same thing as mean reversion. To date this question remains unresolved.

NON-MARTINGALE MODELS

Martingale model assumes risk neutrality but people in general are risk averters. If agents are risk averse they will hold only risky assets if expected returns vary so as to compensate them for these changes in risk. Therefore, one would expect the returns to be partly forecastable (High risk Ê High return). Thus risk-aversion will lead to a departure from the martingale model.

In production economics where corner solutions are possible, prices represent both the technology (at the corner) as well as preferences whenever corner solutions occur, so risk neutrality by itself is insufficient to generate the martingale.

Lucas model presumes that the more risk averse agents there are, the more volatile asset prices will be.

There was a possibility that the variance-bounds violations reflected departures from the martingale model induced by risk aversion. There has been an attempt to determine whether asset price fluctuations could be interpreted as reflecting risk averse agents' attempts to smooth consumption over time. Now results to date have been disappointing as consumption based models of asset pricing imply that stock returns will be positively and strongly correlated with consumption growth, and this turns out to be not true empirically.

The introduction of cheap computing and large financial data bases infused new anomalies which posed a serious problem. The best known is the "January effect" Ä stock returns averaged 3.5% in January and 0.5% over the other months. This is inconsistent with a martingale. Banz (1981) found that small firms have higher returns than is consistent with their riskiness. Keim (1983) showed that "small-firm" effect and "January effect" are the same thing. Tinic and West (1984) found that risk return trade-off occurs entirely in January.

Another anomaly-weekend effect-stock returns are on average negative from the close of trading on Fridays to the opening of trading on Mondays.

R. Ariel (1987) showed that returns are positive on average only in the first half of the calendar month.

High volume of trade on organized securities exchanges is a very striking piece of evidence conflicting with market efficiency.

There has been a traditional hostility towards irrationality [e.g. towards Shiller's fad variables.] But Fischer Black, by renaming irrational trading as noise trading, sanitized irrationality and made it palatable to analysts.

Efficient market theory implies that returns should be explainable ex- post by fundamentals. Roll's recent studies are an exception. He showed that weather information could explain empirically only a small fraction of the variation in the prices of orange crops in Florida. He could not identify any variable that explained the remainder of the variation.

The recent wave of leveraged buyouts seem to provide strong evidence against market efficiency.

The majority of traders will value the firm correctly; only the acquirer is led by "hubris" to overpay. Following this Roll argued that the market which he identified with the majority of traders is efficient. However, this argument does not hold. Now no systematic pattern of price decline should occur in the wake of a publicly known event like a successful takeover.

On October 19, 1987, stock values dropped by about half a trillion dollars on a single day in complete absence of news that can plausibly be related to market fundamentals. Thus nonfundamental factors exist even when the stock market is functioning normally.

CONCLUSION

The transition between market efficiency and the martingale model is far from direct. The variance bounds violations were interpreted as constituting evidence against the assumed stationarity of dividends rather than as conflicting with market efficiency.

It was suggested that traders who act irrationally will lose wealth on average. But Bradford DeLong (1988) questions this hypothesis stating that in a population of risk-averse agents the average rewards to risk takers exceed those to risk avoiders and the law of large number implies that the risk takers as a whole do better then risk averters. Thus irrationality may actually be rewarded in the aggregate.

Recent literature on cognitive psychology has documented systematic biases in the manner in which people use information to make decisions. Some of the biases can be connected to securities market behavior. Example Ä The presence of points of reference, though irrelevant, distorts the agents' decisions. They further systematically overweight current information and underweight background information relative to what Bayes' theorem implies. Economists assumed that these biases would disappear where stakes were high as in securities markets, but this line is wearing thin.

Flood and Hodrick (1990) survey various tests for bubbles and related issues.

Finance References

Samuelson, P.A. (1965) “Proof that Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review* 6, 41-49.

Flood, R.P. and R. J. Hodrick (1990) “On Testing for Speculative Bubbles," *Journal of Economic Perspectives* 4(2), 85-102.

Martingales

The assumption of independence in probability is somewhat restrictive in time series applications, and is relaxed with the use of Markov and martingale processes. A sequence X_t for $t=l,2,\ldots,T$ with $|E(X_t)|<\infty$ is called a martingale if the conditional expectation given the past values does not depend on t, that is

$$E[X_{t+l}|\, X_l, X_2, \ldots, X_t] = X_t \text{ almost everywhere (ae), } t \quad 1$$

It is useful to think of X_t as the size of a gambler's cash position at time t. Games of chance, including gambling, are regarded as “fair" if the gambler's expected winnings are unchanged after playing the game one more time. Imagine that the gambler starts with two different endowments X_l or X_l plus a constant. If it turns out that the change in endowment does not materially change the outcome of our analysis, the martingale property has “translation invariance." Then, one can simply assume $E(X_t) = 0$ without loss of generality.

Note the white noise process mentioned before is related to the martingale as follows.

A formal mathematical discussion of martingales is often in terms of sums of independent random variables. The dependence is introduced by the fact that one cumulates the independent random variables. Consider a summation process, S_n say $S_n = a_l + a_2 + \ldots a_n$ where a_i are white noise processes. Now S_n is a martingale as shown below:

$$E[S_{n+l} \mid S_l, \ldots, S_n] = E[a_{n+l} + S_n \mid a_l, \ldots, a_n] = S_n$$

which says that the (conditionally) expected value of the sequence at n+1 is the same as the value at n, satisfying the definition given above. See M.M Rao(1984, p.168). Any martingale can be characterized in terms of a sum of independent r.v.'s, and under mild conditions the difference between successive martingales is an independent r.v.'s. For example, $S_{n+1} - S_n = a_{n+1}$, which is white noise.

7) Martingale Differences

Starting with $S_l = a_l$, define $Y_2 = S_2 - S_l, \ldots, Y_t = S_t - S_{t-l}$ for $t > l$ the martingale property implies that for the martingale differences Y_t we have the white noise,

$$E[Y_{t+l} \mid Y_l \ldots Y_t] = 0.$$

which may be readily verified for the white noise process.

Hamilton (1994, p. 189) notes the importance of martingale difference sequences (MDS). $\{Y_t\}$ is an MDS wrt $\mathcal{H}_t$ (information from $Y_1 \ldots Y_t$ and $X_1 \ldots X_t$). If $E(Y_t|\mathcal{H}_{t-1})=0$ for $t=2, 3, \ldots$

MDS is stronger than absence of serial correlation. It allows nonlinear forecasts. The notion of independence refers to all moments, here we have only the expectation zero which does not depend on past values. The variance is allowed to depend on past values. This concept is stronger than mere absence of serial correlation, which says that no linear forecasting formula can forecast Y_t. What about a nonlinear forecasting formula? zero serial correlation does not rule out nonlinear forecasting formula for Y_t, but MDS does rule it out. The central limit theorem (CLT) for MDS says that sqrt(T) times a sample mean of MDS's is asymptotically normal with zero mean and variance σ^2 estimated by (1/T) times sum of squares of Y_t from t=1 to t=T. Before the theory of MDS was developed, it appeared that CLT cannot proved unless one assumes independence: f(u,v)=f(u)f(v), where f denotes a generic density. A weaker requirement is factoring of expectations E(u,v)=E(u)E(v). The higher moments need not satisfy similar factoring, yet CLT can be proved.

Hendry (1995, p.58 p.733) discusses the innovation process as one where best predictor of X_t is its past value (random walk) $X_t=X_{t-1}+\varepsilon_t$. The word innovation means that it is truly new, i.e., information set based on past history I_{t-1} cannot predict it, or its distribution does not depend on I_{t-1}. Thus innovation must be white noise if the information set contains past history, but not conversely. (Banerjee et al 1993, p.12). Innovation must be a MDS. Economic examples are stock market prices and interest rates. It is unpredictable in mean. Then ΔX_t is martingale difference sequences (MDS). In general, economic data are rarely martingales, but defined as deviations: $\varepsilon_t= y_t-\Sigma_{j=1}^{s}\beta_j\, y_{t-j}$ from an autoregressive model is a derived process. It can be regarded as an MDS, though it cannot be regarded as independent. We can assume successive ε_t to be uncorrelated. Among properties of MDS, besides CLT, it satisfies the Chebyshev inequality and strong law of large numbers (LLN) [sample mean converges to population mean]

8) <u>Submartingale</u> <u>(Favorable game)</u> <u>in</u> <u>Asymptotic</u> <u>Theory</u>

If the gambler's cash position is becoming favorable to the gambler, it is called a submartingale defined by the following property. $\{X_t\}$ is a submartingale if

$$E[X_{t+1} \mid X_1, \ldots, X_t] > X_t.$$

A transformation of a martingale by a convex transformation having a finite mean yields a submartingale. By a theorem of Doob [1953, p. 321] under appropriate measurability conditions a submartingale can be uniquely decomposed into a martingale and an increasing sequence of non-negative random variables. Doob's martingale convergence theorem is useful in proving that certain limiting distributions are standard normal. See Hall and Heyde (1980, p. 64, p. 115).

9) <u>Weak</u> <u>Law</u> <u>of</u> <u>Large</u> <u>Numbers</u> (Chebyshev)

Let the sample mean $\bar{x}_n = (1/n) \sum_{t=1}^{n} x_t$, where x_t is a realization from a sequence of random variables such that $E(X_n) = \mu$ and cov $(X_t, X_s) = \delta_{ts}\, \sigma^2$ for all t and s, where δ_{ts} is the Kronecker delta, we have $\bar{x}_n \xrightarrow{p} \mu$. In other words the sample mean converges in probability to the population mean.

10) <u>Ensemble</u> <u>and</u> <u>Ergodicity</u>

An observed time series X_t is thought to be only one realization of a random process from the set of all possible realizations. An ensemble is a collection of all possible records. To understand a time series process we may be interested in knowing the ensemble average μ of a process. To estimate μ we naturally consider the time average, the sample mean $\bar{X}$ defined over n points. Is this valid ? Yes, if the observed process is ergodic, a term commonly used in statistical mechanics. It can be shown that a time average converges in mean square to the corresponding ensemble average as $n \to \infty$, (e.g., $\bar{X} \xrightarrow{ms} \mu$).

Consider the set of all integers, I mentioned earlier and a random variable $\{X_t\!: t \in I\}$. Let S denote a shift transformation of t, e.g., S(t) = t + h for all h > 0. Recall that if X_t is stationary, its probability structure does not change under the (shift) transformation S. Such a S is sometimes called measure preserving since it does not change probability measures. The transformation S is characterized as being ergodic by the so-called ergodic theorem, whose statement involves advanced mathematics beyond the scope of this book. The word ergodic comes from a Greek term meaning wanderer, and the notion of returning to the original set after several transformations by S is involved in the ergodic theorem. It is relevant for proving the law of large numbers for time series, Doob (1953, p. 565), and for time series regressions.

Banerjee, et al (1993, p. 16) note that ergodicity, uniform mixing and strong mixing are three types of asymptotic independence concepts. Events A and B are independent if P(A,B)=P(A)P(B). Here P is replaced by distribution function F, A is $y_1, y_2, \ldots, y_n$ and B is $y_{h+1}, y_{h+2}, \ldots, y_{h+n}$. We write independence as $|F(A,B) - F(A)F(B)| \to 0$ as $h \to \infty$. As the distance between sub sequences increases ($h \to \infty$) the joint distribution approaches the product of distributions of sub sequences. Ergodic process is stationary, and if for any t, the limit of $\Sigma_{\tau=1}^{T}$ cov$(y_t, y_{t+\tau})/T = 0$ as

$T \to \infty$. Note this has summation of all lagged covariances. A suff. condition for ergodicity (not necessary) is that the cov $\to 0$ as $\tau \to \infty$.

Uniform mixing, also called ϕ-mixing and strong mixing is also called α-mixing and is based on the P(A|B)=P(A) and P(A,B)=P(A)P(B) properties of independence being satisfied asymptotically.

<u>The LAW of Iterated Expectations (LIE)</u>

Campbell, Lo and MacKinley (CamLM97, Etrics of Fin. markets, Princeton U. 1997) p.23. Discounted present value of stock prices does permit randomness in security returns. The key to understanding this is the LIE. Define two information sets $I_t \subset J_t$. where J has superior or extra information. Consider r.v. X and its expectations under the information set I or J is denoted by $E_I(X)$ and $E_J(X)$. The LIE states that:

$E_I(X)=E_I\, E_J(X)$

or E_I of forecast error $[X - E_J(X)]=0$, that is: LIE states that one cannot use limited information I to predict the forecast error one would make if one had superior information set J.

CamLM97p24 state that market efficiency is not testable very well. One can make a joint test of equilibrium model and market efficiency only. Some abnormal returns compensate the gatherers for the cost of information. However relative efficiency of one market compared to another is empirically measurable.

<u>Returns:</u>

CamLM97p9 state: simple return $R_t=(P_t/P_{t-1}) - 1$.

Gross Return=1+ $R_t=(P_t/P_{t-1})$.

Two-period gross return $(1+R_t(2))=(1+R_t)(1+R_{t-1})$, a product of two one-period returns and goes back two time periods.

The k-period gross return is $(1+ R_t(k))= (P_t/P_{t-1})(P_{t-1}/P_{t-2}) \ldots =(P_t/P_{t-k})$

because of consecutive cancellations. Returns always refer to a time period and are flow variables. Continuously compounded return is simply log of gross return denoted in lower case for logs notation as $r_t=\log(1+R_t)$. And continuously compounded multi-period return is simply a sum of r_t values: $r_t(k)=r_t+r_{t-1}+r_{t-2}+ \ldots +r_{t-k+1}$

<u>Excess return</u> on i-th asset $Z_{it}=R_{it} - R_{0t}$, where R_{0t} is a reference return (e.g. on market or on risk free asset). If this is not zero arbitrage traders can profit without any net investment. At short time horizons the historical returns show some skewness and strong excess kurtosis. Stable distributions have variance $\to \infty$, which is counterfactual. If return is conditionally normal, conditional on a variance parameter which is itself random (random volatility?) then the unconditional distribution of returns is a MIXTURE of normal distributions bringing about fat tails. The resulting unconditional distribution has finite variance and finite higher moments, yet it is convenient, because CLT still applies. Long horizon returns are close to normally distributed and short horizon return distributions have fat tails.

<u>RWlk 1,2,3 and Mgl</u>

CamLM97p28: r_t is log gross return or $\log(1+R_t)=\log(P_t/P_{t-1})$. Let f and g be some arbitrary functions of $f(r_t)$ and $g(r_{t+k})$. Note that g is for time t+k. Now assume that:

Orthogonality Condition (OrthC): $Cov[\, f(r_t),\ g(r_{t+k})] = 0$ for all t and $k \neq 0$.

If f and g restricted to be linear functions, zero cov implies uncorrelated (RW3)

If f can be nonlinear but g is linear, then martingale or fair game $E(r_{t+k}|r_t)= \mu$

If both f and g can be nonlinear, we have independent increments and RW1(iid)/RW2(inid).

<u>Mgl Hyp.</u> states that Exp of tomorrow's price= today's price, given the entire price history of the asset. Best forecast of tomorrow's price is today's price.

Linear forecasting rules are ineffective. Most efficient market should completely random and unpredictable? Martingale Hyp. places restrictions on expected returns, but ignores risk. Hence, Mgl is neither necessary nor sufficient for rationally determined prices. However, if we adjust for risk, it works.

<u>RWlk-1:</u> iid normal increments $p_t=\mu+p_{t-1}+\varepsilon_t$, where p_t is log of price. The log has the advantage that it recognizes limited liability. It avoids negative prices.

$E(p_t)=\mu t+p_0$, $Var(p_t)=\sigma^2 t$, both depend on time so it is nonstationary.

Now continuously compounded returns are iid normal $N(\mu, \sigma^2)$

<u>RWlk-2:</u> inid increments (indep. but not identically distributed, allows unconditional heterosc.) this is more realistic than RWlk-1.

RWlk-3: uncorrelated increments, nonlinear dependence is allowed. $Cov(\epsilon_t \epsilon_{t-k})=0$ but cov among squares of ϵ's is nonlinear and it can be nonzero.

TESTS of RW1: Traditional tests include rank, runs, reversals, Portmanteau (T times Sum of sq. of $\hat{\rho}(k)$, Ljung-Box has T(T+2) times SSq $\hat{\rho}^2$/(T-k)

For RW3,

test of uncorrelated increments using autocorrelation coeff.'s. $E\hat{\rho}(k)=\frac{-(T-k)}{T(T-1)}+O(T^{-2})$

Bias corrected $\tilde{\rho}(k)=\hat{\rho}(k)+\frac{(T-k)}{(T-1)^2}+(1-\hat{\rho}^2(k))$

Note that $(\sqrt{T})\hat{\rho}(k) \sim N(0,1)$ and $(T/\sqrt{(T-k)})\tilde{\rho}(k) \sim N(0,1)$, asymptotically.

Variance ratio $VR(q)=\frac{V(r_t(q))}{qV(r_t)} - 1$, where $r_t(q)=r_t+r_{t-1}+\ldots r_{t-q+1}$, the q-period compound return.

$VR(2)=V(r_t+r_{t-1})/2V(r_t)$. Note the var(sum) > sum of variances if positively autocorrelated returns exist. hence the VR(2) >1, this is not good. Some correction is needed.

$VR(q)=1+2\,\Sigma(1-\frac{k}{q})\rho(k)$. Note that VR is a linear combination, with declining weights: (1-(k/q)), Bartlett's window? Variance difference estimator uses Hausman's statistic.

For higher order returns higher order autocorrelations come into play

Stochastic Volatility Model (different from ARCH better fit to finance theory)

High freq. financial series (daily stock prices or exchange rates) often have heterosc. errors.

Esther Ruitz (J of Etr 63, 1994, p289) uses quasi Maximum likelihood rather than GMM.

$y_t=\epsilon_t \exp[h_t/2]$, $h_t=\gamma+\phi h_{t-1}+\eta_t$, η_t is $NID(0,\sigma^2_\eta)$

where $h_t=\ln(\sigma_t^2)$, ϵ_t is white noise with unit variance indep of η and $|\phi|<1$.

Take squares

$y_t^2=\epsilon_t^2 \exp[h_t]$, take logs $\Rightarrow$ linear state space model:

$\ln(y_t^2)=E(\ln \epsilon_t^2)+h_t+\xi_t$ non normal error, mean $E(\ln \epsilon_t^2) \approx -1.27$, Var is $\pi^2/2$

$h_t=\gamma+\phi h_{t-1}+\eta_t$

QML estimation under stationary case, RW+noise and Fat Tails cases are different.

They have good finite sample properties, easy to implement and better than GMM.