Inferences on Conditional Mean Dynamics and Tests of the Martingale Hypothesis

Yongmiao Hong
Cornell University & Xiamen University

April 28, 2009

1 Inference on the Martingale Hypothesis

The martingale difference sequence hypothesis is important in economics and finance, as illustrated by (e.g.) the efficient market hypothesis. Moreover, from a statistical point of view, before modeling the conditional mean dynamics, one may first like to know whether the dynamics in mean does exist and is statistically significant. Furthermore, if the dynamics in mean exists, one may like to know what is the possible pattern for the conditional mean dynamics. This information will be very useful for choosing a suitable time series model for conditional mean that can explain the stylized facts and predict future evolutions of the economy.

1.1 Hypotheses of Interest

Suppose $\{X_t\}$ is a weakly stationary process with $E(X_t) = \mu$.

The hypothesis of Interest are:

$$\mathbb{H}_0 : E(X_t|I_{t-1}) = \mu$$

versus

$$\mathbb{H}_A : E(X_t|I_{t-1}) \neq \mu.$$

Question: How to test $\mathbb{H}_0$ vs. $\mathbb{H}_A$ using a random sample $\{X_t\}_{t=1}^T$?

References: Campbell, Lo and MacKinlay (1997), Chapter 2.

Remarks:

(i) The hypothesis $\mathbb{H}_0$ is fundamental to the predictability of financial returns or financial variables using the historical information.

(ii) To test $\mathbb{H}_0$ is the first step to model the conditional mean dynamics of X_t. One would like to make sure that there exists some dependent structure before going further to model it.

(iii) It is possible that under $\mathbb{H}_0$, $h_t = \text{var}(X_t|I_{t-1})$ and other higher order conditional moments are time-varying. What is the impact of allowing conditional heteroksedatsicity on the test statistics for $\mathbb{H}_0$? (Some popular test statistics for $\mathbb{H}_0$ are derived under the conditional homoskedasticity assumption.)

(iv) No model parameter estimation is involved here. Thus, there is no need to consider the impact of parameter estimation uncertainty on test statistics proposed. Note that tests for asset pricing models are not covered here, because parameter estimation is involved for asset/derivative pricing models. Later, we will discuss the case when X_t depends on estimated parameters (e.g., X_t is an estimated residual from some regression model).

(v) I_{t-1} may contain the past history of X_t or the past history of both X_t and other variables.

1.2 Existing Conventional Methods

Remark: The MDS hypothesis $\mathbb{H}_0$ (after demeaning)

$$E(X_t|I_{t-1}) = \mu \text{ a.s.}$$

implies that (i) $\{X_t\}$ is serially uncorrelated:

$$\gamma(j) = \text{cov}(X_t, X_{t-j}) = 0 \text{ for all } j > 0$$

and equivalently, (ii), $\{X_t\}$ has a flat spectrum:

$$h(\omega) = \frac{1}{2\pi}\gamma(0) \text{ for all } \omega \in [-\pi, \pi].$$

Thus, we can test $\mathbb{H}_0$ by checking whether $\gamma(j) = 0$ for all $j > 0$. If $\mathbb{H}_0$ holds, then $\gamma(j) = 0$ for all $j \neq 0$. This provides a basis to construct tests with proper sizes (i.e., proper Type I errors). On the other hand, if $\gamma(j) \neq 0$, then $\mathbb{H}_0$ is false. A problem with tests based on $\gamma(j)$ is that they cannot detect non-m.d.s. alternatives that have zero autocorrelation.

Remark: The use of autocorrelation to test MDS can date back to at least Fama (1965).

We first introduce a popular autocorrelation test proposed by Box and Pierce (1970).

1.2.1 Autocorrelation Tests (Box-Pierce-Ljung Tests)

Define the sample autocovariance function

$$\hat{\gamma}(j) = T^{-1} \sum_{t=|j|+1}^{T} (X_t - \bar{X})(X_{t-|j|} - \bar{X}), \qquad j = 0, \pm 1, ..., \pm(T-1),$$

where $\bar{X}$ is the sample mean. Then the sample autocorrelation function

$$\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0), \qquad j = 0, \pm 1, ..., \pm(T-1).$$

Theorem: Suppose $\{X_t\}$ is IID(μ, σ^2). Then as $T \to \infty$, we have for any fixed positive integer j,

$$\begin{aligned} \sqrt{T}\hat{\rho}(j) &\to^d N(0,1), \\ \text{cov}[\sqrt{T}\hat{\rho}(i), \sqrt{T}\hat{\rho}(j)] &\to^p 0 \text{ for } i \neq j. \end{aligned}$$

Box and Pierce (1970, *Journal of the American Statistical Association*) propose a portmanteau test that is based on the sum of the first p squared sample autocorrelations. Under the i.i.d. assumption on $\{X_t\}$, we have

$$\begin{aligned} BP(p) &= T\sum_{j=1}^{p}\hat{\rho}^2(j) \\ &= \sum_{j=1}^{p}\left[\sqrt{T}\hat{\rho}(j)\right]^2 \\ &\to^d \chi_p^2. \end{aligned}$$

When T is small, $BP(p)$ is often found to be too conservative in practice. To improve the size of the Box-Pierce test in small and finite samples, Ljung and Box (1978, *Biometrika*) proposed a modified test

$$\begin{aligned} LB(p) &= T(T+2)\sum_{j=1}^{p}(T-j)^{-1}\hat{\rho}^2(j) \\ &\to {}^d\chi_p^2, \end{aligned}$$

where the convergence in distribution occurs under $\mathbb{H}_0$. The modification of the weighting for each lag improves the matching of the first two moments between LB(p) and the χ_p^2 distribution. This improves the size but not necessarily the power of the test.

Remarks:

(i) Much critism has been leveled at the possible low power of the Box-Pierre-Ljung tests. However, they remain a useful and important diagnostic tools in time series analysis. They are easy to understand conceptully and fall in line with the traditional approach to data analysis. In most cases, they are fairly easy to compute.

(ii) Under $\boldsymbol{H}_0$, the Box-Pierce test statistic is asymptotically equivalent to to the test statistic TR^2 in an auxiliary regression procedure

$$X_t = \alpha_0 + \sum_{j=1}^{p}\alpha_j X_{t-j} + \varepsilon_t, \qquad t = p+1, ..., T,$$

and test the joint hypothesis that $\alpha_1 = \cdots = \alpha_p = 0$.

(iii) IID is a very strong assumption. It implies $\mathbb{H}_0$ but not vice versa.

(iv) These asymptotic distribution results continue to hold under $\mathbb{H}_0$ with conditional homoskedasticity var$(X_t|I_{t-1}) = \sigma^2$ a.s. The reason is that $\hat{\gamma}^2(0)$ is a consistent estimator for the asymptotic variance of $\sqrt{T}\hat{\gamma}(j)$:

To see this, first, observe that for any fixed j, we have

$$\begin{aligned}\hat{\gamma}(j) &= T^{-1}\sum_{t=|j|+1}^{T}(X_t-\mu)(X_{t-|j|}-\mu)+O_P(T^{-1}) \\ &= \tilde{\gamma}(j)+O_P(T^{-1}),\end{aligned}$$

where the $O_P(T^{-1})$ term is the effect of replacing the sample mean $\bar{X}$ with μ, which is asymptotically negligible.

Second, under $\mathbb{H}_0$, for any given integer $j \neq 0$,

$$\begin{aligned}&\text{var}\left(\sqrt{T}\tilde{\gamma}(j)\right) \\ &= T^{-1}\sum_{t=|j|+1}^{T}E\left[(X_t-\mu)^2(X_{t-|j|}-\mu)^2\right] \\ &= T^{-1}\sum_{t=|j|+1}^{T}E\left\{E(X_t-\mu)^2|I_{t-1}](X_{t-|j|}-\mu)^2\right\} \\ &= (1-|j|/T)\gamma^2(0) \\ &\to \gamma^2(0).\end{aligned}$$

where the last equality follows by the law of iterated expectations and conditional homoskedasticity. Consequently, $\sqrt{T}\hat{\gamma}(j)/\hat{\gamma}(0)=\sqrt{T}\hat{\rho}(j)\to^d N(0,1)$ by the central limit theorem and the Slutsky theorem.

Finally, for $j,l>0, j\neq l$,

$$\begin{aligned}&\text{cov}\left[\sqrt{T}\tilde{\gamma}(j),\sqrt{T}\tilde{\gamma}(l)\right] \\ &= T^{-1}\sum_{t=j+1}^{T}E\left[(X_t-\mu)^2(X_{t-j}-\mu)(X_{t-l}-\mu)\right] \\ &= 0\end{aligned}$$

given $\text{var}(X_t|I_{t-1})=\sigma^2$ a.s.. In other words, $\{\sqrt{T}\hat{\gamma}(j)\}$ is asymptotically independent across different lags.

However, if there exists conditional heteroskedatsicity (e.g., ARCH effects), then these asymptotic distribution results are no longer valid.

Question: Why?

Answer: When there exists conditional heteroskedasticity, the asymptotic variance of

$\tilde{\gamma}(j)$ is

$$\begin{aligned} & \operatorname{var}\left(\sqrt{T}\tilde{\gamma}(j)\right) \\ = & T^{-1}\sum_{t=|j|+1}^{T} E\left[(X_t-\mu)^2(X_{t-|j|}-\mu)^2\right] \\ = & (1-|j|/T)\gamma_2(j) \\ = & (1-|j|/T)\left[\gamma^2(0)+\operatorname{cov}[(X_t-\mu)^2,(X_{t-j}-\mu)^2]\right] \\ \neq & (1-|j|/T)\gamma^2(0) > \gamma^2(0). \end{aligned}$$

Remark: When there exists volatility clustering, var$[\sqrt{T}\hat{\gamma}(j)]$ is large than $\gamma^2(0)$. This implies that the true critical value of $\sqrt{T}\hat{\rho}(j)$ at any given significance level is large than the counterpart of N(0,1).

Thus, we have

$$\sqrt{T}\hat{\gamma}(j)/\sqrt{\hat{\gamma}_2(j)} \to^d N(0,1)$$

under $\mathbb{H}_0$, where

$$\hat{\gamma}_2(j) = T^{-1}\sum_{t=|j|+1}^{T}(X_t-\bar{X})^2(X_{t-|j|}-\bar{X})^2$$

is a consistent estimator for the variance of $\tilde{\gamma}(j)$. This is a White's (1980) version of heteroskedatsicity-consistent asymptotic variance estimator for $\sqrt{T}\hat{\gamma}(j)$.

However, we still do not have

$$\sum_{j=1}^{p}\left[\sqrt{T}\hat{\gamma}(j)/\sqrt{\hat{\gamma}_2(j)}\right]^2 \to^d \chi_p^2$$

unless the folllowing condition holds:

$$E\left[(X_t-\mu)^2(X_{t-j}-\mu)(X_{t-l}-\mu)\right]=0, \qquad j,l>0, j\neq l.$$

This condition ensures that cov$[\sqrt{T}\hat{\gamma}(j),\sqrt{T}\hat{\gamma}(l)] \to 0$.

Question: What is the implication of the above condition?

Answer: It rules out asymmetric volatility models. For example, it rules out EGARCH and TGARCH processes. See Lecture 09 for discussions on volatility models.

1.2.2 Spectral Distribution Tests

Durlauf (1991, *Journal of Econometrics*) considers testing the martingale hypothesis using the spectral distribution function

$$\begin{aligned} H(\lambda) &= 2\int_0^{\pi\lambda} h(\omega)d\omega \\ &= \gamma(0)\lambda + \sqrt{2}\sum_{j=1}^{\infty}\gamma(j)\frac{\sqrt{2}\sin(j\pi\lambda)}{j\pi}, \qquad 0\leq\lambda\leq 1, \end{aligned}$$

where $h(\omega)$ is the power spectral density:

$$h(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) \cos(j\omega), \qquad -\pi \le \omega \le \pi.$$

Under $\mathbb{H}_0, H(\lambda)$ becomes a straight line:

$$H_0(\lambda) = \gamma(0)\lambda \text{ for } \lambda \in [0,1].$$

A test for $\mathbf{H}_0$ can be obtained by comparing a consistent estimator for $H(\lambda)$ and $\hat{H}_0(\lambda) = \hat{\gamma}(0)\lambda$.

Question: How to estimate $H(\lambda)$ consistently?

Although the periodogram

$$\begin{aligned}
\hat{I}(\omega) &= \frac{1}{2\pi T} |\sum_{t=1}^{T} (X_t - \bar{X}) e^{it\omega}|^2 \\
&= \frac{1}{2\pi} \sum_{j=1-T}^{T-1} \hat{\gamma}(j) e^{-ij\omega}
\end{aligned}$$

is not a consistent estimator for the spectral density $h(\omega)$, as was shown earlier, the integrated periodogram

$$\begin{aligned}
\hat{H}(\lambda) &= 2\int_0^{\lambda\pi} \hat{I}(\omega) d\omega \\
&= \hat{\gamma}(0)\lambda + \sqrt{2} \sum_{j=1}^{T-1} \hat{\gamma}(j) \frac{\sqrt{2}\sin(j\pi\lambda)}{j\pi}
\end{aligned}$$

is consistent for $H(\lambda)$.

Question: Why is $\hat{H}(\lambda)$ consistent for $H(\lambda)$?

Answer: For simplicity, we assume $E(X_t) = 0$ and we know it. Then $\hat{\gamma}(j) = T^{-1}\Sigma_{t=|j|+1}^{T} X_t X_{t-|j|}$. We then write

$$\begin{aligned}
& \hat{H}(\lambda) - H(\lambda) \\
= & \left[\hat{\gamma}(0) - \gamma(0)\right] \lambda \\
& + \sqrt{2} \sum_{j=1}^{T-1} \left[\hat{\gamma}(j) - \gamma(j)\right] \frac{\sqrt{2}\sin(j\pi\lambda)}{j\pi} \\
& + \sqrt{2} \sum_{j=T}^{\infty} \gamma(j) \frac{\sqrt{2}\sin(j\pi\lambda)}{j\pi}.
\end{aligned}$$

Because $\text{var}[\sqrt{T}\hat{\gamma}(\lambda)] \leq C$ for some bounded constant, we can show that $\hat{H}(\lambda) - H(\lambda) \to^p 0$.

Durlauf (1991) uses the following Cramer-von Mises type statistic

$$\begin{aligned} CVM &= \frac{1}{2}T\int_0^1 \left[\hat{H}(\lambda)/\hat{\gamma}(0) - \lambda\right]^2 d\lambda \\ &= T\sum_{j=1}^{T-1}\hat{\rho}^2(j)/(j\pi)^2. \end{aligned}$$

Under $\mathbb{H}_0$ with conditional homoskedasticity ($\text{var}(X_t|I_{t-1}) = \sigma^2$), we have

$$CVM \to^d \sum_{j=1}^{\infty}\chi_j^2(1)/(j\pi)^2,$$

where $\{\chi_j^2(1)\}$ is a sequence of i.i.d. chi-square random variables with one degree of freedom. This asymptotic distribution is nonstandard, but it is distribution-free and can be easily tabulated or simulated. For example, the critical values at the 10%, 5% and 1% significance levels are respectively.

Remark: Under $\mathbb{H}_0$ with conditional heteroskedasticity, the statistic CVM should be modified as follows:

$$\begin{aligned} CVM &= \sum_{j=1}^{T-1}\frac{\left[\hat{\gamma}^2(j)/\hat{\gamma}_2(j)\right]}{(j\pi)^2} \\ &\to {}^d\sum_{j=1}^{\infty}\chi_j^2(1)/(j\pi)^2 \end{aligned}$$

if the following condition holds:

$$E\left[(X_t-\mu)^2(X_{t-j}-\mu)(X_{t-l}-\mu)\right] = 0, \qquad j,l>0, j\neq l.$$

See Deo (2001, *Journal of Econometrics*) for more discussion.

1.2.3 Variance Ratio Tests

Extending an idea of Cochrance (1988), Lo and MacKinlay (1988, *Review of Financial Study*) consider a variance ratio test. The basic idea is as follows. Recall that $\sum_{j=1}^p X_{t-j}$ is the cumulative return over a period of p days. Then under $\mathbb{H}_0$,

$$\begin{aligned} &\frac{\text{var}\left(\sum_{j=1}^p X_{t-j}\right)}{p\cdot\text{var}(X_t)} \\ &= \frac{p\gamma(0) + 2p\sum_{j=1}^p(1-j/p)\gamma(j)}{p\gamma(0)} \\ &= 1 \end{aligned}$$

given $\gamma(j) = 0$ for all $j \neq 0$. This unity property of the variance ratio can be used to test $\mathbb{H}_0$ because any departure from 1 is evidence against $\mathbb{H}_0$.

When $p \to \infty$ as $T \to \infty$, the variance ratio test statistic is asymptotically equivalent to

$$\begin{aligned} \mathrm{VR}(p) &= \sqrt{T/p}\sum_{j=1}^{p}(1 - j/p)\hat{\rho}(j) \\ &= \frac{\pi}{2}\sqrt{T/p}\left[\hat{f}(0) - \frac{1}{2\pi}\right], \end{aligned}$$

where $\hat{f}(0)$ is a kernel-based normalized spectral density estimator at frequency zero, with the Bartlett kernel

$$K(z) = (1 - |z|)\mathbf{1}(|z| \leq 1)$$

and a lag order equal to p.

Under $\mathbb{H}_0$ with conditional homoskedasticity, Lo and MacKinlay (1988) show that

$$\mathrm{VR}(p) \to^d N[0, 2(2p-1)(p-1)/3p].$$

Remark: Under $\mathbf{H}_0$ with conditional heteroskedasticity, $VR(p)$ should be modified as follows:

$$\mathrm{VR}(p) = \sqrt{T/p}\sum_{j=1}^{p}(1 - j/p)\hat{\gamma}(j)/\sqrt{\hat{\gamma}_2(j)}$$

to take into account the impact of ARCH effects, provided the condition that

$$\begin{aligned} E\left[(X_t - \mu)^2(X_{t-j} - \mu)(X_{t-l} - \mu)\right] &= 0, \\ j, l &> 0, \\ j &\neq l \end{aligned}$$

is satisfied. See Campbell, Lo and MacKinlay (1997, Section 2.4) for more discussion.

1.2.4 Spectral Density Tests

Hong (1996, *Econometrica*) considers a kernel spectral density estimator

$$\begin{aligned} \hat{h}(\omega) &= \frac{1}{2\pi}\sum_{j=1-T}^{T-1} k(j/p)\hat{\gamma}(j)e^{-ij\omega}, \\ \omega &\in [-\pi, \pi], \end{aligned}$$

and compares it with the flat spectrum implied by $\mathbf{H}_0$:

$$\hat{h}_0(\omega) = \frac{1}{2\pi}\hat{\gamma}(0), \qquad \omega \in [-\pi, \pi].$$

Under $\mathbb{H}_0$, $\hat{h}(\omega)$ and $\hat{h}_0(\omega)$ are close. If $\hat{h}(\omega)$ is significantly different from $h_0(\omega)$, then $\mathbb{H}_0$ must be false. A global measure of the divergence between $\hat{h}(\omega)$ and $\hat{h}_0(\omega)$ is the quadratic form

$$\begin{aligned}\hat{Q}(\hat{h},\hat{h}_0) &= \int_{-\pi}^{\pi}\left[\hat{h}(\omega)-\hat{h}_0(\omega)\right]^2 d\omega \\ &= \sum_{j=1}^{T-1} k^2(j/p)\hat{\gamma}^2(j).\end{aligned}$$

Under the i.i.d. assumption for $\{X_t\}$, the test statistic

$$\begin{aligned}&\hat{M}_o(p) \\ \equiv\ &\left[T\sum_{j=1}^{T-1} k^2(j/p)\hat{\rho}^2(j)-\hat{C}_o(p)\right]/\sqrt{\hat{D}_o(p)} \\ \to\ &^{d}N(0,1)\end{aligned}$$

where the centering and scaling factors

$$\begin{aligned}\hat{C}_o(p) &= \sum_{j=1}^{T-1} k^2(j/p) \\ \hat{D}_o(p) &= 2\sum_{j=1}^{T-2} k^4(j/p).\end{aligned}$$

Remarks:

(i) This test can be viewed as a generalized version of Box and Pierre's (1970) test, which is equivalent to using the truncated kernel

$$k(z) = \mathbf{1}(|z| \le 1).$$

This kernel puts equal weighting to each of the first p lags.

(ii) Intuition: why is $\hat{M}_0(p)$ asymptotically N(0,1)? Consider the special case of $k(z) = \mathbf{1}(|z| \le 1)$. In this case, $\hat{M}_o(p)$ is asymptotically equivalent to

$$\begin{aligned}\hat{M}_o(p) &\equiv \left[T\sum_{j=1}^{p}\hat{\rho}^2(j)-p\right]/\sqrt{2p} \\ &\to {}^{d}\frac{\chi_p^2-p}{\sqrt{2p}} \\ &\to {}^{d}N(0,1) \text{ as } p\to\infty.\end{aligned}$$

(iii) However, uniform weighting to different lags is expected to be not efficient when a large number of lags are employed. For a stationary process, the autocorrelation typically decays to zero as lag order j increases. Thus, it is more efficient to discount higher order lags. This can be achieved by using non-uniform kernels. Most commonly used kernels, such as the Bartlett, Pazren and Quadratic-Spectral kernels, discount higher order lags.

(iv) Hong (1996) shows that the Daniell kernel

$$k(z) = \frac{\sin(\pi z)}{\pi z}, \qquad -\infty < z < \infty,$$

maximizes the power over a class of the kernel functions when $p \to \infty$. The optimal kernel for hypothesis testing is different from the optimal kernel for spectral density estimation.

(v) Somewhat surprisingly, the test statistic $\hat{M}_o(p)$ is also asymptotically valid under $\mathbb{H}_0$ with conditional heteroskedasticity. However, simulation studies show that it overrejects $\mathbb{H}_0$ at a given significantly level. Therefore, we do not recommend using $M_o(p)$ in the presence of conditional heteroskedasticity. Instead, we suggest using a robust version of the spectral density test described below.

A Robust Version of the Spectral Density Test:

A modified version of the spectral density-based test statistic that is robust to volatility clustering and improves the finite sample performance can be given below:

$$\begin{aligned} &\hat{M}(p) \\ \equiv\ &\left[T^{-1}\sum_{j=1}^{T-1} k^2(j/p)\hat{\gamma}^2(j) - \hat{C}(p)\right] / \sqrt{\hat{D}(p)}, \end{aligned}$$

where the centering and scaling factors

$$\begin{aligned} \hat{C}(p) \equiv\ & \hat{\gamma}^2(0)\sum_{j=1}^{T-1} k^2(j/p) \\ & + \sum_{j=1}^{T-1} k^2(j/p)\hat{\gamma}_{22}(j), \end{aligned}$$

$$\begin{aligned} \hat{D}(p) \equiv\ & 2\hat{\gamma}^4(0)\sum_{j=1}^{T-2} k^4(j/p) \\ & + 4\hat{R}^2(0)\sum_{j=1}^{T-2} k^4(j/p)\hat{\gamma}_{22}(j) \\ & + 2\sum_{j=1}^{T-2}\sum_{l=1}^{T-2} k^2(j/p)k^2(l/p)\left[\hat{C}(0,j,l)\right]^2, \end{aligned}$$

where

$$\begin{aligned} & \hat{\gamma}_{22}(j) \\ \equiv\ & T^{-1} \sum_{t=j+1}^{T-1} [(X_t - \bar{X})^2 - \hat{\gamma}(0)][(X_{t-j} - \bar{X})^2 - \hat{\gamma}(0)] \end{aligned}$$

is the sample autocovariance function of $\{(X_t - \mu)^2\}$, and

$$\begin{aligned} & \hat{C}(0, j, l) \\ \equiv\ & T^{-1} \sum_{t=\max(j,l)+1}^{T} [(X_t - \bar{X})^2 - \hat{\gamma}(0)] \\ & \times (X_{t-j} - \bar{X})(X_{t-l} - \bar{X}) \end{aligned}$$

is the fourth order moment. Intuitively, the centering and scaling constants have taken into account possible volatility clustering and asymmetric features of volatility dynamics, so the test $\hat{M}(p)$ is robust to these effects.

Remarks:

(i) The condition

$$E\left[(X_t - \mu)^2 (X_{t-j} - \mu)(X_{t-l} - \mu)\right] = 0, \qquad j, l > 0, j \neq l,$$

which is assumed in Lo and MacKinlay (1988) and Deo (2000), is not required. This condition will rule out the "leverage effect" or other asymmetric volatility clustering, among other things.

(ii) The $\hat{M}(p)$ test is also applicable to the estimated residuals from a time series regression model. No modification is needed. In other words, parameter estimation uncertainty does not affect the asymptotic distribution of the test statistic. This is different from other tests for serial correlation (e.g., the Durlauf-Deo test, the Box-Pierce-Ljung test).

(iii) Some simulation evidence:

Data Generating Processes

We now examine the finite sample performance of the proposed tests $\hat{M}(p)$ and $\hat{M}_o(p)$ in comparison with a number of popular tests for serial correlation. We compare them with the popular tests of Lo and MacKinlay (1988) and Deo (2000).

We consider a variety of data generating processes (DGPs):

DGP S.1 [*i.i.d.*(0,1)]:

$$X_t = z_t;$$

DGP S.2 [Stochastic Volatility]

$$\begin{aligned} X_t &= \exp(h_t) z_t, \\ h_t &= 0.936 h_{t-1} + 0.5 v_t; \end{aligned}$$

DGP S.3 [EGARCH(1,1)],

$$\begin{cases} X_t = \exp(\frac{1}{2}h_t)z_t, \\ h_t = -5.496(1-0.856) + 0.856h_{t-1} + 0.265(|z_{t-1}| - \sqrt{2/\pi}) - 0.08z_{t-1}, \end{cases}$$

DGP S.4 [GARCH(1,1)],

$$\begin{aligned} X_t &= h_t^{1/2} z_t, \\ h_t &= 0.1 + 0.7h_{t-1} + 0.2X_{t-1}^2. \end{aligned}$$

where $\{z_t\}$ is $i.i.d.$N(0,1), $\{v_t\}$ is $i.i.d.$N(0,1), and both $\{z_t\}$ and $\{v_t\}$ are mutually independent. The coefficients of DGP S.2–S.4 are empirically relevant. Except for DGP S.1, all other DGPs display conditional heteroskedasticity. Both DGP S.2 and DGP S.3 have a finite eighth moment condition on X_t. GARCH(1,1) does not satisfy this moment condition, but we include this DGP to see how the tests perform in this case.

To compute our tests $\hat{M}(p)$ and $\hat{M}_o(p)$, we use the Parzen kernel

$$k(z) = \begin{cases} 1 - 6z^2 + 6|z|^3, & \text{if } |z| \le \frac{1}{2}, \\ 2(1-|z|)^3, & \text{if } \frac{1}{2} \le |z| \le 1, \\ 0, & \text{otherwise}, \end{cases}$$

which has bounded support and is computationally efficient. Our simulation experience suggests that the choice of $k(\cdot)$ has little impact on both the size and power of our tests, provided a non-uniform kernel is used. To examine the impact of the choice of lag order p on the size and power of $\hat{M}(p)$, we choose $p = cT^{\frac{1}{5}}$, where $c > 0$. The rate $p \propto T^{\frac{1}{5}}$ is optimal for the Parzen kernel-based spectral density estimation. To examine the impact of the choice of p, we consider a variety of values for the tuning parameter $c = 2, 4, 6, 8, 10$. These values of c give a rather wide range for p : $[5, 20]$ for $T = 100$, $[6, 24]$ for $T = 250$, $[6, 27]$ for $T = 500$, and $[7, 31]$ for $T = 1000$.

Tables 1 and 2 report the empirical rejection rates of the tests under $\mathbb{H}_0$ for observed raw data at the 10% and 5% significance levels, using asymptotic theory. All results are based on 1000 iterations, using GAUSS Windows Version 5.0 random number generator. When $\{X_t\}$ is $i.i.d.$ (DGP S.1, Table 1), both $\hat{M}(p)$ and $\hat{M}_o(p)$ have similar and reasonable sizes in most cases, particularly at the 10% level. At the 5% level, there exist some but not excessive overrejections. The sizes of $\hat{M}(p)$ and $\hat{M}_o(p)$ are relatively robust to the choice of lag order p, though there is a tendency that larger lag orders p give slightly better sizes. On the other hand, the VR(p) test also has better sizes when small lag orders p are used. It performs reasonably well for most lag orders p when $T \ge 500$. Among all tests, DEO has the best sizes at both the 10% and 5% levels in most cases.

Next, we consider the case when $\{X_t\}$ follows a stochastic volatility process (DGP S.2, Table 1). The $\hat{M}(p)$ test shows some underrejection for $T \le 500$ but its sizes are reasonable for all p and all $T \ge 250$. As is expected, the homoskedasticity-specific test $\hat{M}_o(p)$ shows strong overrejection, which becomes worse as T increases. The V(p) test also shows underrejection and smaller lag orders give better sizes. Again, DEO performs very well at both the 10% and 5% levels, particularly for $T \ge 500$.

When $\{X_t\}$ follows DGP S.3 (EGARCH(1,1), Table 2), $\hat{M}(p)$ performs better than under DGP S.2 and has reasonable sizes. The $\hat{M}_o(p)$ test also has better sizes than under DGP S.2, but still displays some overrejection. The sizes of $\hat{M}(p)$ and $\hat{M}_o(p)$ are relatively robust to the choice of the lag order p. The VR(p) test underrejects $\mathbb{H}_0$ when $T \leq 250$ but performs well when $T \geq 500$. Smaller lag orders p give better sizes for VR(p). Again, DEO performs rather well for all samples.

Finally, we consider the case when $\{X_t\}$ follows DGP S.4 (GARCH(1,1), Table 2). Although the eighth moment of X_t does not exist, $\hat{M}(p)$ has reasonable levels, which are robust to the choice of p. In contrast, $\hat{M}_o(p)$ displays very strong overrejection, which increases as T increases. VR(p) underreject $\mathbb{H}_0$ in most cases, but its sizes improve as T increases, and small lag orders p give better sizes. DEO performs well, although its sizes at the 5% level become smaller as $T \to \infty$.

TABLE 1. Empirical levels of Tests Under i.i.d. errors

	$M(p)$			$M_o(p)$			W(p)			VR(p)			DEO		
p	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
							$T = 100$								
5	10.4	7.0	3.3	10.3	6.9	2.9	5.4	2.7	0.4	9.3	3.7	0.7	9.3	5.4	1.1
10	8.8	6.3	2.6	9.9	6.4	3.2	4.3	1.4	0.1	8.2	2.2	0.9	9.3	5.4	1.1
15	8.3	5.6	2.5	9.3	6.6	3.3	3.2	0.7	0.0	5.7	2.3	0.9	9.3	5.4	1.1
20	8.1	5.0	2.6	8.8	6.2	3.5	2.2	0.5	0.1	4.0	2.3	0.9	9.3	5.4	1.1
25	7.8	5.1	2.5	8.8	6.0	3.2	1.2	0.2	0.0	3.1	2.4	0.9	9.3	5.4	1.1
							$T = 250$								
6	11.0	7.0	3.9	10.9	7.0	4.0	5.5	2.9	0.7	9.7	4.5	0.5	11.8	5.4	1.3
12	10.3	7.0	3.1	10.4	6.8	3.5	6.1	2.5	0.4	9.2	4.0	0.6	11.8	5.4	1.3
18	10.8	6.6	3.2	11.1	7.0	3.4	4.7	2.2	1.0	8.7	3.3	0.8	11.8	5.4	1.3
24	10.8	6.9	3.0	11.4	7.4	3.7	4.7	1.7	0.0	8.7	3.2	0.8	11.8	5.4	1.3
30	11.2	7.5	3.0	11.6	7.8	3.7	3.4	0.9	0.0	6.9	2.8	1.0	11.8	5.4	1.3
							$T = 500$								
6	10.1	6.8	3.0	10.2	7.1	3.1	6.0	2.5	0.4	10.5	5.2	0.9	10.4	5.1	0.8
13	10.3	6.6	2.5	10.4	6.9	2.4	6.3	2.4	0.1	10.4	5.2	0.9	10.4	5.1	0.8
20	10.1	6.1	2.2	10.5	6.6	2.2	5.8	2.1	0.1	10.6	0.1	1.0	10.4	5.1	0.8
27	10.2	6.4	2.4	10.9	6.9	2.5	4.3	1.9	0.2	9.2	4.5	1.4	10.4	5.1	0.8
34	9.8	6.3	2.6	10.7	7.0	2.8	5.0	1.2	0.2	8.6	3.9	1.1	10.4	5.1	0.8
							$T = 1000$								
7	12.2	7.6	3.4	12.2	7.5	3.3	7.0	3.3	0.8	10.4	4.8	0.9	10.0	5.5	1.0
15	11.9	8.1	3.7	12.2	8.1	3.8	8.3	4.1	0.5	10.2	4.4	0.5	10.0	5.5	1.0
23	11.0	6.7	3.8	11.7	7.1	3.7	5.9	2.9	0.4	11.1	5.3	0.7	10.0	5.5	1.0
31	10.3	6.9	2.8	10.6	7.1	3.1	6.1	2.5	0.7	10.7	5.4	0.8	10.0	5.5	1.0
39	10.3	6.5	2.4	10.4	6.6	2.9	6.3	2.9	0.8	9.1	4.9	1.2	10.0	5.5	1.0

Notes : (i) Y_t=ε_t, ε_t ∼i.i.d. N(0,1);

(ii) 1000 iterations;

(iii) $M(p)$, robust spectral density test, $M_o(p)$, Hong's (1996) spectral density test; W(p), Wooldridge's (1991) test;

VR(p), Lo and McKinlay's (1988) variance ratio test; DEO, Deo's (2000) Cramer-von Mises test.

(iv) The rule to select p : $p = cT^{1/5}$, $c = 2, 4, 6, 8, 10$. The Parzen kernel is used for both $M(p)$ and $M_o(p)$.

TABLE 2. Empirical levels of Tests Under SV errors

	$M(p)$			$M_o(p)$			W(p)			VR(p)			DEO		
p	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
							$T = 100$								
5	7.5	4.1	2.1	28.5	21.4	14.2	3.4	1.2	0.0	6.8	3.4	1.3	8.0	3.4	0.5
10	6.5	4.0	2.1	31.3	24.6	14.9	2.2	0.5	0.0	5.7	3.5	0.7	8.0	3.4	0.5
15	6.3	4.3	2.1	30.7	24.5	15.9	1.4	0.3	0.0	4.3	2.2	1.1	8.0	3.4	0.5
20	5.8	4.2	1.7	30.8	23.5	15.2	1.7	0.5	0.0	3.2	2.0	0.9	8.0	3.4	0.5
25	6.0	4.2	1.3	29.2	22.5	14.4	1.1	0.0	0.0	3.1	1.7	0.6	8.0	3.4	0.5
							$T = 250$								
6	8.2	4.5	1.9	42.6	35.7	25.9	3.5	1.7	0.1	7.3	2.9	1.0	8.6	3.9	0.6
12	7.7	4.0	1.7	50.5	42.9	31.8	3.6	1.5	0.1	6.3	2.6	1.1	8.6	3.9	0.6
18	7.5	4.7	1.8	51.3	44.3	32.8	3.2	1.4	0.1	6.0	2.6	0.9	8.6	3.9	0.6
24	8.4	4.7	1.6	51.0	43.4	32.6	2.8	0.7	0.1	5.1	2.6	1.3	8.6	3.9	0.6
30	8.1	4.5	1.6	49.7	42.8	31.2	2.2	0.6	0.0	4.9	2.1	1.4	8.6	3.9	0.6
							$T = 500$								
6	8.2	5.1	1.6	47.6	41.7	32.2	4.8	2.4	0.0	8.1	3.7	0.6	10.1	4.1	0.7
13	7.2	4.6	1.9	58.8	51.8	39.2	4.8	1.5	0.1	6.7	3.6	1.0	10.1	4.1	0.7
20	6.9	4.5	1.8	62.0	53.9	42.2	3.0	1.0	0.0	6.9	3.4	1.3	10.1	4.1	0.7
27	6.9	4.5	1.8	61.0	55.3	41.5	2.8	1.2	0.1	6.0	2.8	1.4	10.1	4.1	0.7
34	7.1	4.6	1.6	60.5	54.2	40.8	3.1	1.1	0.2	5.4	2.9	1.5	10.1	4.1	0.7
							$T = 1000$								
7	10.5	6.5	3.1	56.6	50.1	39.8	6.1	2.8	0.0	10.7	5.2	2.0	11.1	5.9	1.2
15	9.9	6.2	2.7	68.8	63.3	50.0	4.8	1.9	0.0	8.8	4.4	1.8	11.1	5.9	1.2
23	10.0	6.2	2.3	71.7	65.2	52.6	4.8	1.7	0.1	8.6	4.2	1.2	11.1	5.9	1.2
31	9.4	5.9	1.9	71.8	64.6	52.4	4.4	1.2	0.0	7.6	3.9	1.2	11.1	5.9	1.2
39	9.2	5.7	1.7	71.1	63.2	51.1	4.7	1.2	0.3	7.0	3.8	1.2	11.1	5.9	1.2

Notes : (i) $Y_t = e^{h_t}\varepsilon_t$, $h_t = 0.936h_{t-1} + 0.5Y_t$, $\varepsilon_t \sim$i.i.d. N(0,1);

(ii) 1000 iterations;

(iii) $M(p)$, robust spectral density test, $M_o(p)$, Hong's (1996) spectral density test; W(p), Wooldridge's (1991) test;

VR(p), Lo and McKinlay's (1988) variance ratio test; DEO, Deo's (2000) Cramer-von Mises test.

(iv) The rule to select p : $p = cT^{1/5}$, $c = 2, 4, 6, 8, 10$. The Parzen kernel is used for both $M(p)$ and $M_o(p)$.

TABLE 3. Empirical levels of Tests Under GARCH(1,1) errors

	$M(p)$			$M_o(p)$			W(p)			VR(p)			DEO		
p	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
							$T = 100$								
5	10.1	7.3	2.9	17.2	13.8	9.4	4.6	2.2	0.2	9.0	3.7	1.0	10.1	5.3	1.2
10	9.5	6.2	2.8	19.5	14.9	8.9	3.1	1.4	0.0	6.3	2.5	1.2	10.1	5.3	1.2
15	9.3	6.0	2.2	20.4	14.7	9.1	2.0	0.9	0.0	4.5	2.8	1.1	10.1	5.3	1.2
20	8.9	5.4	2.1	20.9	15.1	8.4	1.5	0.5	0.0	3.5	2.4	0.9	10.1	5.3	1.2
25	7.8	5.1	2.0	20.2	7.8	7.9	1.1	0.1	0.0	2.9	2.2	1.0	10.1	5.3	1.2
							$T = 250$								
6	10.5	6.6	3.0	25.5	20.9	14.2	4.6	1.8	0.4	9.2	4.6	0.8	9.4	5.0	0.8
12	10.8	7.4	2.9	32.8	26.8	18.4	4.5	1.3	0.1	7.2	3.7	1.5	9.4	5.0	0.8
18	11.0	7.0	2.7	36.4	29.5	21.1	3.6	1.3	0.0	6.8	3.5	1.9	9.4	5.0	0.8
24	10.4	6.7	2.7	38.2	32.0	22.2	2.7	0.8	0.0	5.8	3.2	1.9	9.4	5.0	0.8
30	10.0	6.4	2.3	39.2	32.7	23.6	1.7	0.4	0.0	5.1	3.3	1.8	9.4	5.0	0.8
							$T = 500$								
6	10.1	6.2	2.8	38.1	31.3	21.2	5.4	1.2	0.0	9.5	5.3	1.9	9.5	4.6	0.8
13	10.0	5.6	2.0	45.7	39.5	28.1	3.6	1.3	0.1	8.9	4.9	2.1	9.5	4.6	0.8
20	9.5	5.2	1.2	50.4	43.8	32.4	3.4	0.9	0.1	7.7	4.4	1.8	9.5	4.6	0.8
27	9.8	5.4	1.6	53.4	45.4	33.8	3.1	0.9	0.0	6.8	3.8	1.7	9.5	4.6	0.8
34	9.4	5.5	1.4	55.6	47.6	34.7	2.1	0.7	0.1	6.6	3.8	1.6	9.5	4.6	0.8
							$T = 1000$								
7	10.6	6.0	2.6	47.9	42.1	32.5	6.4	3.5	0.2	8.4	4.6	1.0	9.7	3.9	0.8
15	10.7	6.7	2.9	60.4	53.5	43.6	5.4	2.6	0.5	7.1	3.7	1.9	9.7	3.9	0.8
23	10.2	6.5	3.0	65.2	60.5	48.8	4.7	1.8	0.2	6.7	3.7	2.0	9.7	3.9	0.8
31	9.8	6.4	2.7	68.5	63.3	52.2	3.8	1.6	0.0	6.6	4.2	1.8	9.7	3.9	0.8
39	9.8	6.1	2.8	70.9	65.1	54.4	4.2	1.0	0.0	5.9	4.1	1.9	9.7	3.9	0.8

Notes : (i) $Y_t = h_t^{1/2}\varepsilon_t$, $h_t = 0.936 + 0.5Y_{t-1}^2$, $\varepsilon_t \sim$i.i.d. N(0,1);

(ii) 1000 iterations;

(iii) $M(p)$, robust spectral density test, $M_o(p)$, Hong's (1996) spectral density test; W(p), Wooldridge's (1991) test;

VR(p), Lo and McKinlay's (1988) variance ratio test; DEO, Deo's (2000) Cramer-von Mises test.

(iv) The rule to select p : $p = cT^{1/5}$, $c = 2, 4, 6, 8, 10$. The Parzen kernel is used for both $M(p)$ and $M_o(p)$.

TABLE 4. Empirical levels of Tests Under EGARCH(1,1) errors

	$M(p)$			$M_o(p)$			W(p)			VR(p)			DEO		
p	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
							$T = 100$								
5	9.2	6.3	2.5	15.6	11.4	6.5	4.1	1.1	0.0	7.3	3.7	1.5	9.5	4.6	0.8
10	9.9	6.5	2.8	17.4	12.1	7.2	3.9	1.2	0.0	6.5	3.1	1.8	9.5	4.6	0.8
15	10.0	6.0	3.0	16.7	12.7	7.5	1.2	0.3	0.0	4.7	2.5	1.2	9.5	4.6	0.8
20	10.0	6.0	2.7	16.9	12.3	6.8	1.3	0.3	0.0	3.2	2.1	1.1	9.5	4.6	0.8
25	10.0	6.3	2.8	15.9	12.1	6.5	0.7	0.4	0.0	2.5	1.8	0.9	9.5	4.6	0.8
							$T = 250$								
6	9.2	5.8	2.6	18.0	14.0	7.2	4.4	1.6	0.3	7.6	2.9	0.8	9.5	4.5	1.3
12	8.0	5.2	2.2	17.9	11.9	7.3	3.5	1.4	0.2	6.9	3.0	0.8	9.5	4.5	1.3
18	8.8	4.8	1.7	17.3	12.4	6.6	3.4	1.2	0.2	5.0	2.2	0.8	9.5	4.5	1.3
24	8.3	5.3	1.6	17.3	12.2	6.8	3.4	1.4	0.2	5.4	1.9	1.0	9.5	4.5	1.3
30	8.4	5.4	1.8	17.0	11.7	6.9	2.8	1.0	0.3	4.9	1.7	1.0	9.5	4.5	1.3
							$T = 500$								
6	10.0	6.1	2.6	19.2	15.2	8.0	5.3	2.9	0.5	10.1	5.1	0.9	10.4	5.4	0.6
13	10.5	6.8	3.0	20.5	15.8	8.5	5.0	2.6	0.4	9.6	5.5	1.3	10.4	5.4	0.6
20	11.0	6.6	3.2	19.7	15.3	8.5	6.2	3.1	0.3	8.8	3.6	1.3	10.4	5.4	0.6
27	11.0	6.9	2.7	20.2	15.1	7.7	5.8	2.6	0.5	7.9	3.3	1.2	10.4	5.4	0.6
34	11.0	7.0	3.0	19.7	15.1	7.2	4.7	1.9	0.4	7.1	3.2	1.1	10.4	5.4	0.6
							$T = 1000$								
7	9.3	6.3	2.9	21.3	14.7	8.8	4.4	1.9	0.3	11.0	6.3	1.7	10.9	5.5	0.8
15	9.3	5.8	2.8	21.0	15.5	8.2	5.4	2.0	0.3	10.0	4.7	1.7	10.9	5.5	0.8
23	9.0	6.1	2.1	22.2	14.7	7.4	6.2	2.7	0.3	8.2	4.2	1.5	10.9	5.5	0.8
31	9.6	5.7	1.9	21.8	14.7	6.6	5.6	2.6	0.6	7.7	3.3	0.9	10.9	5.5	0.8
39	9.8	5.6	2.0	20.9	14.0	5.8	6.9	2.3	0.3	8.1	3.0	0.7	10.9	5.5	0.8

Notes: (i) $Y_t = 0.5Y_{t-1} + e^{0.5h_t}\varepsilon_t$, $h_t = -5.496(1-0.856) + 0.856h_{t-1} + 0.265\left(|\varepsilon_{t-1}| - \sqrt{2/\pi}\right) - 0.08\varepsilon_{t-1}$

$\varepsilon_t \sim$i.i.d. N(0,1);

(ii) 1000 iterations;

(iii) $M(p)$, robust spectral density test, $M_o(p)$, Hong's (1996) spectral density test; W(p), Wooldridge's (1991) test;

VR(p), Lo and McKinlay's (1988) variance ratio test; DEO, Deo's (2000) Cramer-von Mises test.

(iv) The rule to select p : $p = cT^{1/5}$, $c = 2, 4, 6, 8, 10$. The Parzen kernel is used for both $M(p)$ and $M_o(p)$.

1.3 New Approach: Generalized Spectral Derivative Tests

Motivation:

The autocorrelation function or the power spectrum only checks the white noise property. As is well-known, there is a gap between a white noise and an MDS. Although an MDS is a white noise, a white noise may not be an MDS. An example is the nonlinear moving average process:

$$X_t = \alpha z_{t-1} z_{t-2} + z_t,$$

where $\{z_t\} \sim$ i.i.d.$(0, \sigma^2)$. For this process, $\text{cov}(X_t, X_{t-j}) = 0$ for all $j > 0$ but $E(X_t|I_{t-1}) = \alpha z_{t-1} z_{t-2} \neq 0$. Any autocorrelation-based or spectral-based test may easily miss this process. This will forfeit some profitable opportunities.

Question: Why do there exist nonlinearities in asset returns?

Answer: Many aspects of economic behavior may not be linear. Experimental evidence and casual introspection suggest that the investor' attitudes toward risk and expected return are nonlinear. The terms of many financial contracts such as options and other derivative securities are nonlinear. And the strategic interactions among market participants, the process by which information is incorporated into security prices, and the dynamics of economic-wide fluctuations are all inherently nonlinear. Therefore, a natural frontier for financial econometrics is the modeling of nonlinear phenomena, firstly in conditional mean.

1.3.1 Generalized Spectral Tests

We now apply the generalized spectral density $f(\omega, u, v)$, introduced in Lecture 05, to construct a new test for the MDS hypothesis $\mathbf{H}_0$. Recall the generalized spectral density of $\{X_t\}$:

$$\begin{aligned} f(\omega, u, v) &\equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v) e^{-\mathbf{i}j\omega}, \\ \omega &\in [-\pi, \pi], \end{aligned}$$

where the generalized covariance function

$$\begin{aligned} \sigma_j(u, v) &\equiv \text{cov}(e^{\mathbf{i}uX_t}, e^{\mathbf{i}vX_{t-|j|}}) \\ &= \varphi_j(u, v) - \varphi(u)\varphi(v). \end{aligned}$$

Here, $\varphi_j(u, v) \equiv Ee^{\mathbf{i}(uX_t + vX_{t-|j|})}$ and $\varphi(u) \equiv Ee^{\mathbf{i}uX_t}$ are the pairwise joint characteristic function and marginal characteristic function of $\{X_t\}$ respectively.

The function $f(\omega, u, v)$ can capture all pairwise serial dependencies in $\{X_t\}$ over various lags. Therefore, it is not suitable to test the martingale difference sequence (MDS) hypothesis that $E(X_t|I_{t-1}) = \mu$. For example, $f^{(0,1,0)}(\omega, 0, v)$ can capture a m.d.s. process that is not serially independent, such as an ARCH(1) process. However, the first order partial derivative of $f(\omega, u, v)$,

$$
\begin{aligned}
f^{(0,1,0)}(\omega, 0, v) &\equiv \frac{\partial}{\partial u} f(\omega, u, v)|_{u=0} \\
&= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j^{(1,0)}(0, v) e^{-ij\omega},
\end{aligned}
$$

is suitable to test the martingale hypothesis, because

$$
\begin{aligned}
\sigma_j^{(1,0)}(0, v) &\equiv \frac{\partial}{\partial u} \sigma_j(u, v)|_{u=0} \\
&= \text{cov}\left(iX_t, e^{ivX_{t-|j|}}\right) \\
&= 0
\end{aligned}
$$

if and only if

$$E(X_t|X_{t-|j|}) = \mu \text{ a.s..}$$

Under $\mathbb{H}_0$, the partial generalized derivative becomes

$$f_0^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sigma_0^{(1,0)}(0, v) \text{ for all } \omega \in [-\pi, \pi].$$

Remarks:

(i) $f^{(0,1,0)}(\omega, 0, v)$ checks whether there exists serial dependence in mean. Specially, it checks whether $E(X_t|X_{t-j}) = E(X_t)$ for all $j > 0$, and so it is a more suitable test for MDS than those based on the autocorrelation, the variance ratio and the power spectrum. It can detect a wide range of deviations from MDS, including those with zero autocorrelations.

(ii) Because $f^{(0,1,0)}(\omega, 0, v)$ is always flat no matter whether higher order conditional moments are changing over time, one could construct a test for MDS that is robust to time-varying conditional higher order conditional moments. See Hong and Lee (2005) for more discussions.

(iii) To test $\mathbb{H}_0$, we can compare two consistent estimators for $f^{(0,1,0)}(\omega, 0, v)$ and $f_0^{(0,1,0)}(\omega, 0, v)$.

Question: Suppose a random sample $\{X_t\}_{t=1}^T$ is given. How to estimate $f_0(\omega, u, v) = \frac{1}{2\pi}\sigma_0(u, v)$?

Answer: Use

$$\hat{f}_0(\omega, u, v) = \frac{1}{2\pi} \hat{\sigma}_0(u, v),$$

where

$$
\begin{aligned}
\hat{\sigma}_j(u, v) &= \hat{\varphi}_j(u, v) - \hat{\varphi}_j(u, 0)\hat{\varphi}_j(0, v), \\
\hat{\varphi}_j(u, v) &= \frac{1}{T - |j|} \sum_{t=|j|+1}^{T} e^{iuX_t + ivX_{t-|j|}}, \\
j &= 0, \pm 1, ..., \pm (T - 1).
\end{aligned}
$$

Question: How to estimate $f(\omega, u, v)$?

Hong (1999, JASA, Theorem 1) shows that $f(\omega, u, v)$ can be consistently estimated by the smoothed kernel estimator

$$\begin{aligned} & \hat{f}(\omega, u, v) \\ \equiv \ & \frac{1}{2\pi} \sum_{j=1-T}^{T-1} (1 - |j|/T)^{\frac{1}{2}} k\left(j/p\right) \hat{\sigma}_j(u, v) e^{-\mathbf{i}j\omega}, \end{aligned}$$

where $p \equiv p(T)$ is a bandwidth or lag order, and $k(\cdot)$ is a kernel function or "lag window".

Remarks:

(i) Commonly used kernels include the Bartlett, Daniell, Parzen and Quadratic-Spectral kernels.

(ii) The factor $(1 - |j|/T)^{\frac{1}{2}}$ modifies the variance of $\hat{\sigma}_j(u, v)$. It could be replaced by 1, but it gives better finite sample performance for the tests based on $\hat{f}(\omega, u, v)$.

Theorem [Hong 1999, Theorem 1]: Suppose a set of regularity conditions hold, and let $p = cT^{\delta}$ for $\delta \in (0, 1)$. Then the IMSE

$$\begin{aligned} & E \int \int |\hat{f}(\omega, u, v) - f(\omega, u, v)|^2 d\omega dW_1(u) dW_2(v) \\ = \ & O(p/T + p^{-2q}) \to 0, \end{aligned}$$

where $W_1(\cdot)$ and $W_2(\cdot)$ are two nonnegative nondecreasing weighting functions with bounded total variation.

Remark: Examples of $W_1(\cdot)$ and $W_2(\cdot)$:

Let m, l be nonnegative integers. Define the generalized spectral derivative estimators

$$\begin{aligned} \hat{f}^{(0,m,l)}(\omega, u, v) & \equiv \frac{1}{2\pi} \sum_{j=1-T}^{T-1} (1 - |j|/T)^{\frac{1}{2}} k(j/p) \hat{\sigma}_j^{(m,l)}(u, v) e^{-\mathbf{i}j\omega}, \\ \hat{f}_0^{(0,m,l)}(\omega, u, v) & \equiv \frac{1}{2\pi} \hat{\sigma}_0^{(m,l)}(u, v), \end{aligned}$$

where

$$\hat{\sigma}_j^{(m,l)}(u, v) \equiv \frac{\partial^{m+l}}{\partial^m u \partial^l v} \hat{\sigma}_j(u, v).$$

Hong (1999) proposes a general class of tests based on the quadratic norm:

$$\begin{aligned} & Q\left(\hat{f}^{(0,m,l)}, \hat{f}_0^{(0,m,l)}\right) \\ \equiv \ & \int \int_{-\pi}^{\pi} |\hat{f}^{(0,m,l)}(\omega, u, v) - \hat{f}_0^{(0,m,l)}(\omega, u, v)|^2 d\omega dW_1(u) dW_2(v) \\ = \ & \frac{2}{\pi} \int \sum_{j=1}^{T-1} k^2(j/p)(1 - j/T) |\hat{\sigma}_j^{(m,l)}(u, v)|^2 dW_1(u) dW_2(v), \end{aligned}$$

where the second equality follows by Parseval's identity, the unspecified integrals are taken over the support of $W_1(\cdot)$ and $W_2(\cdot)$. As we will see below, proper choices of $W_1(\cdot)$ and $W_2(\cdot)$ as well as (m,l) allow us to test various specific aspects of serial dependence.

The test statistic is a standardized version of the quadratic form:

$$\begin{aligned} & M_0(m,l) \\ \equiv & \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j)\int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 dW_1(u)dW_2(v) - \hat{C}_0^{(m,l)}\sum_{j=1}^{T-1} k^2(j/p)\right] \\ & \div \left[\hat{D}_0^{(m,l)}\sum_{j=1}^{T-2} k^4(j/p)\right]^{1/2}, \end{aligned}$$

where the centering and standardization factors

$$\begin{aligned} \hat{C}_0^{(m,l)} &\equiv \int \hat{\sigma}_0^{(m,m)}(u,-u)dW_1(u)\int \hat{\sigma}_0^{(l,l)}(v,-v)dW_2(v), \\ \hat{D}_0^{(m,l)} &\equiv 2\int |\hat{\sigma}_0^{(m,m)}(u,u')|^2 dW_1(u)dW_1(u')\int |\hat{\sigma}_0^{(l,l)}(v,v')|^2 dW_2(v)dW_2(v'). \end{aligned}$$

Interpretation of $M_0(m,l)$: What does $M_0(m,l)$ test for?

$M(m,l)$	$W_1(\cdot)$	$W_2(\cdot)$	Test Function	Alternatives
$M(0,0)$	N(0,1)-CDF	N(0,1)-CDF	$\sigma_j(u,v)$	dependence
$M(1,0)$	$\delta(\cdot)$	N(0,1)-CDF	$E(X_t\vert X_{t-j})$	MDS
$M(1,1)$	$\delta(\cdot)$	$\delta(\cdot)$	$\text{cov}(X_t,X_{t-j})$	WN
$M(1,2)$	$\delta(\cdot)$	$\delta(\cdot)$	$\text{cov}(X_t,X_{t-j}^2)$	arch-in-mean
$M(1,3)$	$\delta(\cdot)$	$\delta(\cdot)$	$\text{cov}(X_t,X_{t-j}^3)$	skewness-in-mean
$M(1,4)$	$\delta(\cdot)$	$\delta(\cdot)$	$\text{cov}(X_t,X_{t-j}^4)$	kurtosis-in-mean
$M(2,0)$	$\delta(\cdot)$	$\delta(\cdot)$	$E(X_t^2\vert X_{t-j})$	homoskedasticity
$M(2,1)$	$\delta(\cdot)$	$\delta(\cdot)$	$\text{cov}(X_t^2,X_{t-j})$	leverage effect
$M(2,2)$	$\delta(\cdot)$	$\delta(\cdot)$	$\text{cov}(X_t^2,X_{t-j}^2)$	linear ARCH
$M(3,0)$	$\delta(\cdot)$	$\delta(\cdot)$	$E(X_t^3\vert X_{t-j})$	skewness
$M(3,3)$	$\delta(\cdot)$	$\delta(\cdot)$	$\text{cov}(X_t^3,X_{t-j}^3)$	skewness
$M(4,0)$	$\delta(\cdot)$	$\delta(\cdot)$	$E(X_t^4\vert X_{t-j})$	kurtosis
$M(4,4)$	$\delta(\cdot)$	$\delta(\cdot)$	$\text{cov}(X_t^4,X_{t-j}^4)$	kurtosis

Notes: $\delta(\cdot)$ is the Dirac delta function, i.e., $\delta(u)=0$ if $u\neq 0$, and $\int_{-\infty}^{\infty}\delta(u)du=1$.

Remark: The class of test statistics $M_0(m,l)$ is derived under the i.i.d. assumption.

Theorem [Hong 1999, Theorem 3]: Suppose a set of regularity conditions hold. Let $p=cT^{\delta}$ for $\delta\in(0,1)$. Then when $\{X_t\}\sim$ IID,

$$M_0(m,l)\to^d N(0,1)$$

for any given pair of nonnegative integers (m, l).

Remarks:

(i) For a kernel $k(\cdot)$ with unbounded support, $M_0(m, l)$ employs all $T-1$ lags in the sample. This is desirable when the alternative has persistent serial dependence.

(ii) Non-uniform kernels, such as the Daniell kernel $k(z) = \sin(\pi z)/\pi z$, $z \in (-\infty, \infty)$, usually weight down higher order lags. This is expected to enhance good power of the tests in practice, because economic agents normally discount past information. In fact, the Daniell kernel maximizes the power of $M_0(m, l)$ over a class of kernels that include Parzen and Quadratic-Spectral kernels. The latter is optimal for spectral density estimation, but not necessarily for hypothesis testing (cf. Hong 1999, Theorem 6).

Question: The test statistic $M_0(1, 0)$ cannot be directly used to test the martingale hypothesis when there exists volatility clustering. Why?

Answer: The form of $M_0(1, 0)$ as well as its limit distribution are derived under i.i.d., which implies the martingale hypothesis. However, it is possible that the martingale hypothesis holds but $\{X_t\}$ is not i.i.d. In this case, $M_0(1, 0)$ will generally not converge to $N(0, 1)$ in distribution. Thus, there exists the Type I error when $\{X_t\}$ is MDS but not IID.

Question: How to make the test robust to volatility clustering and other higher order dependence (e.g., time-varying skewness and kurtosis)?

There are two solutions to this difficulty.

One solution: *Bootstrapping the Generalized Spectral Tests*

Remark: The bootstrap is a popular resampling technique that can approximate the finite sample distribution of a statistic up to an accurate degree.

References: Hall (1992, *Bootstrap and Edgeworth Expansion*), Horowitz (2001, *Handbook of Econometrics*), Shao and Tu (2000, *The Kackknife and Bootstrap*).

We use the wild bootstrap procedure (Wu 1986, Liu 1988).

Question: What is the wild bootstrap?

The wild Bootstrap Procedure:

- Step 1: Suppose $\{X_t\}_{t=1}^T$ is an observed real data. Compute $M_0(m, l)$ using $\{X_t\}_{t=1}^T$
- Step 2: For the wild bootstrap procedure, let F_t be the CDF of X_t, $t = 1, \ldots, T$. We generate a bootstrap sample $\{X_t^b\}_{t=1}^T$ according to the formula that

$$X_t^b = \begin{cases} aX_t & \text{with prob } p = a/\sqrt{5} \\ (1-a)X_t & \text{with prob } 1-p, \end{cases}$$

  where $a = (1+\sqrt{5})/2$. It follows that $E(Y_t^b|F_t) = 0$, $E[(Y_t^b)^2|F_t] = Y_t^2$, and $E[(Y_t^b)^3|F_t] = Y_t^3$. Thus, the wild bootstrap procedure can preserve the first three sample moments of the original data.

- Step 3: Use $\{X_t^b\}_{t=1}^T$ to compute $M_0^b(m,l)$.
- Step 4: Repeat Steps 2 and 3 B times when B is large (say $B = 1000$). We obtain a total of B test statistics $\{M_{01}^b(m,l), M_{02}^b(m,l), ..., M_{0B}^b(m,l)\}$.
- Step 5: Compute the bootstrap P-value

$$p_b^* = \frac{1}{B}\sum_{\tau=1}^{B} \mathbf{1}[M_{0\tau}^b(m,l) > M_0(m,l)].$$

- Step 6: If $p_b^* < 0.05$, then reject the null hypothesis of MDS at the 5% significance level.

Remarks:

(i) Under certain regularity conditions, the wild bootstrap procedure is expected to approximate the finite sample distribution of $M_0(m,l)$ rather well. However, this is not justified in theory. We check it only by simulation.

(ii) **Simulation Evidence:**

We consider a GARCH(1,1)-N(0,1) data generation process,

$$\begin{cases} X_t = h_t^{1/2} z_t, \\ h_t = 0.1 + 0.7h_{t-1} + 0.2X_{t-1}^2, \\ \{z_t\} \sim i.i.d.N(0,1), \end{cases}$$

which is an MDS but displays volatility clustering, as is typical for high frequency financial time series. Since the data is conditionally heteroskedastic, we use the wild bootstrap. Table 3 reports the bootstrap sizes of these tests for MDS, with $B = 300$. The results show that the wild bootstrap yields adequate sizes for the $M_0(1,l)$ tests of MDS.

TABLE. Sizes of Generalized Spectral Tests of MDS Using Wild Bootstrap

DGP: $Y_t = \sigma_t e_t$ where $\sigma_t^2 = 0.1 + 0.2e_{t-1}^2 + 0.7\sigma_{t-1}^2$ and e_t is IID N(0, 1)

	$T = 100$								
α	0.10			0.05			0.01		
$\bar{p}$	6	10	15	6	10	15	6	10	15
$M(1,0)$	.088	.096	.096	.044	.048	.046	.012	.012	.012
$M(1,1)$	.084	.090	.090	.044	.044	.042	.012	.014	.014
$M(1,2)$	.130	.134	.138	.070	.074	.074	.014	.014	.014
$M(1,3)$	.110	.104	.098	.056	.052	.052	.006	.006	.004
$M(1,4)$	.108	.106	.100	.036	.042	.042	.004	.006	.008

Notes: We compute bootstrap p-values using 300 bootstrap replications, and the empirical rejection rates are based on 500 Monte Carlo replications. $\bar{p}$ is a preliminary lag order.

2 Application I: Do Foreign Exchange Rates follow a Martingale?

References: Hong and Lee (2003, *Review of Economics and Statistics*)

Conventional Wisdom: Exchange rate changes approximately follow an martingale difference sequence (MDS) so that future changes are unpredictable using publicly available information (e.g., Meese and Rogoff 1983, *Journal of International Economics*).

Remark: This hypothesis has been tested using the autocorrelation function (Box-Pierce-Ljung's portmanteau tests), the variance-ratio (Lo and MacKinlay 1988) and the power spectrum (Durlauf 1991), based on data of various frequencies and sample periods.

Meese and Rogoff (1983) tested various models based on economic theory against the simple "random walk" model. The surprising result was the non-ability of the economic models to outperform the naive prediction (i.e., the random walk prediction) in out-of-sample forecasts. Some researchers (e.g., Ruelle 1991, p.82) even think that in general economies are too complex to allow for reliable forecasts.

Purpose: We now use the generalized spectral derivative test to explore whether there exists a gap between a white noise and an MDS for foreign exchange rate changes.

Specially, we use the generalized spectral tests to explore serial dependence of five major foreign exchange rates—the nominal exchange rates of

- Canada (CD),
- Germany (DM),
- UK (BP),
- Japan (JY), and
- France (FF)

per US dollar. The data is a weekly series from 1/1/1975 to 12/31/1998. The daily noon buying rates in New York City certified by the Federal Reserve Bank of New York for customs and cable transfers purposes are obtained from the Chicago Federal Reserve Board (www.frbchi.org). The weekly series is generated by selecting Wednesdays series (if a Wednesday is a holiday then the following Thursday is used), which has 1253 observations. The use of weekly data avoids the so-called weekend effect, as well as other biases associated with nontrading, bid-ask spread, asynchronous rates and so on, which are often present in higher frequency data. We use the scaled logarithmic difference $X_t = 100\ln(P_t/P_{t-1})$, where ξ_t is an exchange rate level.

The statistic $M_0(m,l)$ involves the choice of a bandwidth p, which may have significant impact on power. Hong (1999) proposes a data-driven method to choose p, which, to some extent, lets data themselves speak for an appropriate bandwidth. This method still involves the choice of a preliminary bandwidth $\bar{p}$. Simulations in Hong (1999) and Section 3 show that the choice of $\bar{p}$ is less important than the choice of p. We consider $\bar{p}$ in the range $6-15$ to examine the robustness of $M_0(m,l)$ with respect to the choice of $\bar{p}$.

Table 4 reports the values of $M_0(m,l)$ together with their bootstrap p-values, for CD, DM, BP, JY and FF, using the medium preliminary lag order $\bar{p}=10$. We use the 5% level here. For comparison, note that $M_0(m,l)$ has an asymptotic one-sided N(0,1) distribution, so the asymptotic critical value at the 5% level is 1.65.

TABLE 4. Generalized Spectral Tests: Bootstrap p-Values for Five Currencies

	CD			DM			BP		
	statistic	P_B	P_W	statistic	P_B	P_W	statistic	P_B	P_W
$M(0,0)$	12.283	.000		10.717	.000		6.121	.002	
$M(1,0)$	3.187	.012	.026	4.798	.000	.008	5.720	.002	.002
$M(1,1)$	2.396	.028	.090	1.235	.146	.180	0.769	.158	.382
$M(1,2)$	-0.099	.410	.674	-0.356	.444	.704	-0.496	.608	.872
$M(1,3)$	-0.012	.340	.686	-0.187	.420	.616	2.463	.050	.414
$M(1,4)$	1.730	.060	.294	0.829	.158	.346	2.340	.048	.466
$M(2,0)$	10.283	.000		7.488	.000		5.023	.000	
$M(2,1)$	5.264	.004		0.539	.302		1.964	.058	
$M(2,2)$	15.369	.002		18.942	.000		35.280	.000	
$M(3,0)$	2.311	.034		1.748	.098		4.086	.006	
$M(3,3)$	-0.651	.332		5.585	.028		14.932	.012	
$M(4,0)$	2.800	.020		2.691	.050		1.210	.104	
$M(4,4)$	-0.710	.250		4.467	.030		17.971	.014	

Notes: P_B denotes the naive bootstrap p-values and P_W denotes the wild bootstrap p-values, both are based on 500 bootstrap replications.

TABLE 4. Generalized Spectral Tests: Bootstrap p-Values for Five Currencies (continued.)

	JY			FF		
	statistic	P_B	P_W	statistic	P_B	P_W
$M(0,0)$	20.243	.000		10.601	.000	
$M(1,0)$	10.175	.000	.000	4.572	.000	.008
$M(1,1)$	8.152	.000	.002	1.306	.116	.178
$M(1,2)$	11.351	.000	.016	0.978	.132	.302
$M(1,3)$	5.695	.004	.100	-0.170	.424	.610
$M(1,4)$	9.076	.000	.068	1.352	.094	.282
$M(2,0)$	4.074	.010		7.685	.000	
$M(2,1)$	3.582	.020		3.201	.028	
$M(2,2)$	13.752	.000		14.148	.002	
$M(3,0)$	-0.482	.664		2.041	.062	
$M(3,3)$	4.150	.020		7.006	.018	
$M(4,0)$	-0.699	.760		2.414	.030	
$M(4,4)$	-0.050	.080		7.370	.022	

Notes: P_B denotes the naive bootstrap p-values and P_W denotes the wild bootstrap p-values, both are based on 500 bootstrap replications.

Summary: Stylized facts of Foreign Currecies

1. There exists strong serial dependence for the changes of all the five exchange rates. The geometric random walk hypothesis (possibly with drift) is strongly rejected for all the five currencies.

2. Although the changes of exchange rates are often serially uncorrelated (as is the case for CD, DM, BP, FF), they are clearly not an MDS for all the five currencies. There exists strong nonlinearity in mean for the changes of all the five exchange rates.

3. For CD, DM, BP and FF, the nonlinearity in mean cannot be explained by the polynomials of lagged exchange rate changes. It is of a complicated and unknown form.

4. There exist strong ARCH effects for all the five currencies. While the leverage effect is significant for CD, JY and FF, the ARCH-in-mean effect is significant only for JY.

5. There are significant conditional skewness and/or conditional kurtosis.

Remarks:

(i) While it is now documented that stock returns or foreign exchange rate changes do not follow a martingale difference sequence, it is rather difficult to find significant evidence on the out-of-sample predictive ability of most time series models relative to the martingale difference sequence model, particularly in terms of the MSE criterion.

Question: What are possible reasons?

- Difference between the predictability of data and the predictive ability of models.
- Regime shifts or structural changes in the out-of-sample period.
- The failure of the stationarity assumption and other important auxiliary regularity conditions for in-sample observations.

(ii) The predictability of financial returns and the profitability of models/trading rules.

It is possible that the conditional mean of financial returns is predictable using the historical information, but this may not guarantee the profitability of a forecast model/trading rule, due to the existence of transaction costs.

(iii) The predictability of financial returns and market efficiency

Question: What is the definition of market efficiency?

In a standard exchange economy in the sense of Lucas (1978, *Econometrica*), a representative economic agent maximizes his intertemporal utility over the entire lifetime subject to a sequence of budget constraints, and perhaps market equilibrium conditions. In this situation, the optimal behavior of the economic agent is characterized by the Euler equation

$$e^{-r_t} \int R_{it} f_t^*(C_t) dC_t = 1$$

or

$$e^{-r_t} E_{t-1}^*(R_{it}|I_{t-1}) = 1,$$

where R_{it} is the gross return on asset i from time $t-1$ to t, and r_t is the riskfree interest rate, and $f_t^*(C_t)$ is the risk-neural probability density of C_t conditional on I_{t-1}. In this setup, the financial return is generally predictable using I_{t-1} although the economic agent behaves optimally and markets are in equilibrium. However, the risk adjusted return is not predictable using the past information.

Question: Can we test the Euler condition? And if so, how?

(iv) The predictability of data and the predictive ability of a model

Suppose we have found evidence of predictability of financial returns $\{X_t\}$. How to construct a model for the conditional mean of X_t given I_{t-1}?

It will be informative if one can get some information on:

- Is the serial dependence in mean linear or nonlinear?
- If it is linear, is it a short memory or a long memory?
- If it is nonlinear, what is the nature of nonlinearity?

In the next lecture, we will introduce different models for conditional mean of X_t.

2.1 Application II: Predictability and Nonlinearity in U.S. Stock Returns

Hong and Lee (2005, *Review of Economic Studies*)

- Lo and MacKinlay (1988, *Review of Financial Studies*) find significant positive serial correlation for weekly and monthly holding-period returns. The statistical significance of their results is robust to conditional heteroskedasticity of unknown form.

- **Nonsynchronous trading:** It has been argued in the literature that the significant autocorrelation in stock returns may be superficial: it may be due to the fact that small capitalization stocks trade less frequently than large stocks. As a consequence, new information is impounded first into large capitalization stock prices and then into small capitalization stock prices with a time lag. This lag will induce a positive serial correlation in a weighted index of stock returns. Thus, the rejections of the martingale hypothesis for stock prices using autocorrelation may very well be the result of this nonsynchronous phenomenon.

- Using a hypothesized nonsynchronous trading model, Lo and MacKinlay (1988) point out that nonsynchronous trading can at most account for part of the documented positive autocorrelation coefficient and thus stock returns are predictable. However, Lo and Mackinlay's (1988) nonsynchronous trading model has not been empirically tested. It appears to be a difficult task to quantify how much of the significant autocorrelation could be attributed to nonsynchronous trading.

- We further contribute to this literature by taking a new approach. Instead of arguing how much of the significant autocorrelation can be explained by nonsynchronous trading, we check whether the stock returns are still predictable even when we remove all linear serial correlation in stock returns. In other words, we check whether there exists predictable nonlinear components in mean in addition to linear dependence. Note that Lo and MacKinlay's (1989) variance ratio test cannot be used for this purpose, because it is only able to capture linear dependence.

- We consider two daily stock price indices:

  1. S&P 500 index and
  2. NASDAQ index, from 12/01/1972 to 12/31/2001, obtained from CRSP.

- Define stock return $X_t = 100 \ln(P_t/P_{t-1})$, where P_t is the stock price index at time t. A graphical examination of $\{X_t\}$ shows that there exists strong volatility clustering for both S&P 500 and NASDAQ. Thus heteroskedasticity-robust tests should be used in checking the predictability of asset returns.

- We first apply the $\hat{M}(1,0)$ test to the raw stock returns $\{X_t\}$ to test whether $\{X_t\}$ is a *m.d.s.* in the presence of conditional heteroskedasticity of unknown form. Table 3, under subtitle "martingale testing", reports the $\hat{M}(1,0)$ statistics that are based on a data-driven lag order $\hat{p}_0$ (see Hong and Lee 2005 for more detailed discussion), with the preliminary bandwidth $\bar{p} = c(10T)^{1/5}$, for $c = 1, ..., 10$. These heteroskedasticity-robust statistics are quite robust to the choice of the preliminary lag order $\bar{p}$ and have essentially zero asymptotic p-values (the upper-tailed N(0,1) critical values at the 1% level is 2.28), suggesting rather strong evidence against the martingale hypothesis for both S&P 500 and NASDAQ daily returns.

TABLE 3 (a). Predictability of Stock Returns

	Martingale Testing			
$\hat{M}^{(1,0)}(\hat{p}_0)$	S&P 500		Nasdaq	
c	Statistics	p-value	Statistics	p-value
1	13.85	0.000	30.04	0.000
2	13.85	0.000	27.19	0.000
3	12.98	0.000	25.11	0.000
4	12.27	0.000	23.51	0.000
5	11.70	0.000	22.19	0.000
6	11.25	0.000	21.10	0.000
7	10.90	0.000	20.23	0.000
8	10.61	0.000	19.51	0.000
9	14.65	0.000	32.43	0.000
10	14.65	0.000	32.43	0.000

- Next, we examine whether the stock returns still contain predictable nonlinear components after removing linear dependence. This is a test of whether stock returns are a linear or nonlinear process in mean. For this purpose, we specify an AR(d) model

$$X_t = \alpha_0 + \sum_{j=1}^{d} \alpha_j X_{t-j} + u_t,$$

and estimate it with the lag order d selected using the BIC criterion,

$$\begin{aligned} d &= \arg\min_d \left[\ln \hat{s}^2 + \frac{(d+1)\ln(T)}{T} \right] \\ &= \arg\min_d [\text{goodness of fit} + \text{model complexity}] \end{aligned}$$

which delivers a consistent order selection for weakly stationary linear processes (Hannan 1980).

- The selected model is an AR(2) for S&P 500, and an AR(1) for NASDAQ, with significant but small AR coefficients.

- Table 3, under the subtitle "linearity testing", reports the $\hat{M}^{(1,0)}(\hat{p}_0)$ statistics applied to the OLS AR(d) residuals. These statistic values are much smaller than the $\hat{M}^{(1,0)}(\hat{p}_0)$ statistics applied to raw return data, but they are still quite significant. The asymptotic p-values are below 2% level for S&P 500 and are essentially zero for NASDAQ. These results strongly suggest that stock returns contain a predictable nonlinear component in mean, and the evidence is stronger for NASDAQ than for S&P 500.

TABLE 3 (a). Nonlinearity of Stock Returns

	Linearity Testing			
$\hat{M}^{(1,0)}(\hat{p}_0)$	S&P 500		Nasdaq	
c	Statistics	p-value	Statistics	p-value
1	2.57	0.005	7.78	0.000
2	2.57	0.005	7.04	0.000
3	2.47	0.007	6.50	0.000
4	2.37	0.009	6.11	0.000
5	2.28	0.011	5.80	0.000
6	2.23	0.013	5.55	0.000
7	2.18	0.014	5.36	0.000
8	2.15	0.016	5.21	0.000
9	2.69	0.004	8.40	0.000
10	2.69	0.004	8.40	0.000

Some Discussions

Question: Does the predictability of financial markets imply the existence of the inefficiency of financial markets?

Answer: Not necessary! References: Lucas (1978, *Econometrica*).

A new (modern) version of EMH that says that the risk-adjusted asset returns are m.d.s. is more suitable. In other words, one may define EMH by the Euler condition

$$E\left[M_t R_t | I_{t-1}\right] = 0,$$

where M_t is the stochastic discount factor, and R_t is the excess return on the asset.

Question: How to test the EMH as defined above? The key is how to have a model-free consistent estimate of M_t. (If M_t is obtained by specifying some model on the economic agent's preference among other things (e.g., the utility function of a representative agent), then when EMH is rejected, one does not know whether it is due to the misspecfication of M_t or due to the rejection of the Euler condition.

Question: Can a nonlinear time series model make money by exploiting the nonlinearity in conditional mean of stock returns?

Answer: It depends on the transaction costs.