WISE WORKING PAPER SERIES

WISEWP0601

Serial Correlation and Serial Dependence

Yongmiao Hong

June 2006

厦门大学王亚南经济研究院
The Wang Yanan Institute for Studies in Economics, Xiamen University

COPYRIGHT© WISE, XIAMEN UNIVERSITY, CHINA

Serial Correlation and Serial Dependence

Yongmiao Hong

Department of Economics and Department of Statistical Science
Cornell University, U.S.A.
and
Wang Yanan Institute for Studies in Economics, Xiamen University, China

June 2006

I thank Steven Durlauf (editor) for suggesting this topic, and Linlin Duan and Jing Liu for excellent research assistance and references. All remaining errors are solely mine.

ABSTRACT

Serial correlation and serial dependence has been central to time series econometrics. The existence of serial correlation complicates statistical inference of econometric models, and in time series analysis, inference of serial correlation, or more generally, serial dependence, is crucial to model and capture the dynamics of time series processes. In this entry, we first discuss the impact of serial correlation on statistical inference of a linear regression model. In Section 2, we introduce various tests for serial correlation, for both estimated residuals and observed raw data, and discuss their relationship. Section 3 discusses serial dependence in nonlinear time series contexts, and related measures and tests for serial dependence.

1 Introduction

Serial correlation and serial dependence has been central to time series econometrics. The existence of serial correlation complicates statistical inference of econometric models, and in time series analysis, inference of serial correlation, or more generally, serial dependence, is crucial to model and capture the dynamics of time series processes. In this entry, we first discuss the impact of serial correlation on statistical inference of a linear regression model. In Section 2, we introduce various tests for serial correlation, for both estimated residuals and observed raw data, and discuss their relationship. Section 3 discusses serial dependence in a nonlinear time series context, introducing related measures and tests for serial dependence.

Consider a linear regression model

$$Y_t = X_t'\beta^0 + \varepsilon_t, \qquad t = 1, \cdots, n, \tag{1.1}$$

where Y_t is a dependent variable, X_t is a $k \times 1$ vector of explanatory variables, β^0 is an unknown $k \times 1$ parameter vector, and ε_t is an unobservable disturbance with $E(\varepsilon_t|X_t) = 0$. Suppose X_t is strictly exogenous such that $\text{cov}(X_t, \varepsilon_s) = 0$ for all t, s. Then the model (1.1) is called a static regression model.

In the classical linear regression, it is assumed that $\varepsilon_t|X \sim N(0, \sigma^2)$, where $X \equiv (X_1', X_2', \cdots, X_n')'$ is a $n \times k$ matrix, and I is an $n \times n$ identity matrix. Among other things, this implies conditional homoskedasticity (i.e., $\text{var}(\varepsilon_t|X) = \sigma^2$ a.s.) and conditionally serial uncorrelatedness (i.e., $\text{cov}(\varepsilon_t, \varepsilon_s|X) = 0$ a.s. for all $t \neq s$). Under this condition, the OLS estimator

$$\hat{\beta} \equiv (X'X)^{-1}X'Y$$

is the best linear unbiased estimator (BLUE), and conditional on X, $\hat{\beta}$ is normally distributed:

$$\hat{\beta}|X \sim N(\beta^0, \sigma^2(X'X)^{-1}).$$

The Student t-test statistic follows a Student t-distribution with $n-k$ degrees of freedom:

$$T \equiv \frac{R\hat{\beta} - r}{\sqrt{s^2 R(X'X)^{-1}R'}} \sim t_{n-k}$$

under the null hypothesis that $R\beta^0 = r$, where R is a $1 \times k$ vector, $s^2 = e'e/(n-k)$ is the residual sample variance, and $e = Y - X\hat{\beta}$ is the estimated OLS residual vector. The F-test statistic follows an F-distribution with J and $n-k$ degrees of freedom:

$$F \equiv \frac{(R\hat{\beta} - r)' \left[R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta} - r)/J}{s^2} \sim F_{J,n-k}$$

under the null hypothesis that $R\beta^0 = r$, where R is a $J \times k$ matrix, and r is a $J \times 1$ vector.

These appealing finite sample results do not hold once the normality assumption is abandoned, which is not realistic for many economic and financial data. However, suppose $\{Y_t, X_t'\}'$ is an ergodic stationary process with $\{X_t\varepsilon_t\}$ being a martingale difference sequence (*m.d.s.*), which includes independent and identically distributed (*i.i.d.*) processes as a special case. The classical results are approximately applicable for large samples when there exists conditional homoskedasticity (i.e., $\text{var}(\varepsilon_t|X_t) = \sigma^2$ a.s.). The OLS estimator $\hat{\beta}$ is asymptotically normal and asymptotically BLUE:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1}), \text{ as } n \to \infty,$$

where $Q \equiv E(X_t X_t')$, and $\xrightarrow{d}$ denotes convergence in distribution. Although the conventional t-test and F-test statistic have unknown finite sample distributions, they are asymptotically valid, with

$$T \xrightarrow{d} N(0, 1), \text{ as } n \to \infty,$$

and

$$J \cdot F \xrightarrow{d} \chi^2_{n-k}, \text{ as } n \to \infty,$$

under the null hypothesis of $R\beta^0 = r$.

When there exists conditional heteroskedasticity (i.e., $E(\varepsilon_t^2|X_t) \neq \sigma^2$), the OLS estimator $\hat{\beta}$ is asymptotically normal:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q^{-1}VQ^{-1}), \text{ as } n \to \infty,$$

where $V \equiv E(X_t X_t' \varepsilon_t^2)$. However, it is no longer asymptotically BLUE, and the t-test and F-test statistics cannot be used because they are based on an incorrect asymptotic variance estimator of $\hat{\beta}$. Instead, White's (1980) heteroskedasticity-consistent variance estimator should be used, and based on it, robust econometric procedures can be constructed.

One important implication of the *m.d.s.* assumption on $\{X_t\varepsilon_t\}$ is that $\{X_t\varepsilon_t\}$ is serially uncorrelated, which implies that $\{\varepsilon_t\}$ is serially uncorrelated when X_t contains an intercept, as is usually the case in practice. There exist various scenarios where $\{\varepsilon_t\}$ may be serially correlated. When $\{X_t\}$ is strictly exogenous, and $\{\varepsilon_t\}$ follows an AR(1) process

$$\varepsilon_t = \rho\varepsilon_{t-1} + z_t, \qquad \{z_t\} \sim i.i.d.(0, \sigma_0^2),$$

we can transform model (1.1) by taking a quasi-difference:

$$Y_t - \rho Y_{t-1} = (X_t - \rho X_{t-1})'\beta + z_t. \tag{1.2}$$

This suggests a two-step estimation procedure: (i) Obtain an OLS estimator $\hat{\beta}$ and save estimated

residuals $\{e_t\}$; (ii) Regress e_t on e_{t-1} to obtain an OLS estimator $\hat{\rho}$; (iii) Plug $\hat{\rho}$ in Eq. (1.2) to replace ρ and obtain a feasible OLS estimator $\tilde{\beta}$, which is asymptotically BLUE. This is the well-known Cochrane-Orcutt procedure. It can be easily extended to the case where $\{\varepsilon_t\}$ follows an AR(p) process for some fixed p.

However, when X_t is not strictly exogenous, for example, when X_t contains lagged dependent variables, serial correlation in $\{\varepsilon_t\}$ will generally cause $E(X_t\varepsilon_t) \neq 0$, rendering inconsistent the OLS estimator and invalidating the Cochrane-Orcutt procedure. Even when X_t is strictly exogenous, the Cochrane-Orcutt procedure is not applicable if $\{\varepsilon_t\}$ has serial correlation of unknown form. In this case, one has to rely on the asymptotic distribution of the OLS estimator $\hat{\beta}$:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q^{-1}VQ^{-1}) \text{ as } n \to \infty,$$

where V is a so-called long-run variance-covariance matrix

$$V \equiv \sum_{j=-\infty}^{\infty} \gamma(j),$$

with $\gamma(j) = \text{cov}(X_t\varepsilon_t, X_{t-j}\varepsilon_{t-j})$.

Estimation of the long-run variance V has been challenging. A popular approach is based on the fact that

$$V = 2\pi h(0),$$

where $h(0)$ is the spectral density matrix at frequency 0 of the process $\{X_t\varepsilon_t\}$. The spectral density matrix

$$h(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j)e^{-ij\omega}, \qquad \omega \in [-\pi, \pi],$$

when $\gamma(j)$ is absolutely summable. The link of $V = 2\pi h(0)$ motivates a nonparametric estimation of V when serial correlation in $\{X_t\varepsilon_t\}$ has an unknown form. A popular nonparametric method has been the kernel method which usually takes the following general form

$$\hat{V} = \sum_{j=1-n}^{n-1} k(i/M)\hat{\gamma}(j),$$

where $\hat{\gamma}(j)$ is the sample autocovariance function of $\{X_te_t\}_{t=1}^n$, $k(\cdot)$ is a kernel function that assigns weightings to various lags, and M is a smoothing parameter that can be interpreted as a truncation lag order when $k(\cdot)$ has compact support (i.e., when $k(z) = 0$ if $|z| > 1$). Earlier works (e.g., Hansen 1982, White 1984) use the truncated kernel $k(z) = 1\,(|z| \leq 1)$ that assigns uniform weighting to all lags $|j| \leq M$, but the resulting variance estimator $\hat{V}$ may not be positive definite in finite samples. Newey and West (1987) use the Bartlett kernel $k(z) = (1 - |z|)1(|z| \leq 1)$ that

ensures positive semi-definiteness of $\hat{V}$, and the Parzen kernel used by Gallant and White (1988) also has this property. Andrews (1991) shows that the Quadratic-Spectral kernel is optimal in terms of an asymptotic mean squared error and it also delivers a positive semi-definite estimator $\hat{V}$. Andrews (1991), Andrews and Monahan (1992), and Newey and West (1994), among others, discuss data-driven methods to choose the smoothing parameter M, which depends on the degree of serial dependence in $\{X_t\varepsilon_t\}$ and is more important than the choice of the kernel function $k(\cdot)$.

Long-run variance estimators have found widespread applications in time series econometrics, not just in stationary linear regression models. Unfortunately, it has been well documented that associated procedures usually perform poorly in finite samples. Most tests tend to overreject the null hypothesis or confidence interval estimators tends to have undercovereges. Intuitively, macroeconomic data usually display persistent serial correlation, which generates a spectral peak at frequency 0. A kernel-based estimator for the long-run variance has a downward bias, thus underestimating the true variance V and causing overrejection for hypothesis testing and undercoverage for confidence interval estimation.

2 Testing for Serial Correlation

For a linear dynamic regression model where the regressor vector X_t contains lagged dependent variables, serial correlation in $\{\varepsilon_t\}$ will generally render inconsistent the OLS estimator. To see this, we consider a simple AR(1) model

$$Y_t = \beta_0^0 + \beta_1^0 Y_{t-1} + \varepsilon_t = X_t'\beta^0 + \varepsilon_t,$$

where $X_t = (1, Y_{t-1})'$. If ε_t also follows an AR(1) process, we will have $E\left(X_t\varepsilon_t\right) \neq 0$, rendering inconsistent the OLS estimator for β^0.

It is therefore important to check serial correlation for estimated residuals, which serves as a misspecification test for a linear dynamic regression model. Even for a static linear regression model, it is useful to check serial correlation. In particular, if there exists no serial correlation in $\{\varepsilon_t\}$ in a static regression model, then there is no need to use a long-run variance estimator.

Durbin–Watson Test

Testing for serial correlation has been a longstanding problem in time series econometrics. The most well known test for serial correlation in regression disturbances is the Durbin-Watson test, which is the first formal procedure developed for testing first order serial correlation

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, \qquad \{u_t\} \sim i.i.d.\left(0, \sigma^2\right)$$

using the OLS residuals in a static linear regression model. Specifically, Durbin and Watson

(1950,1951) propose a test statistic

$$d = \frac{\Sigma_{t=2}^{n} (e_t - e_{t-1})^2}{\Sigma_{t=1}^{n} e_t^2}$$

where $\{e_t\}$ is the OLS residual from a static linear regression model.

Durbin and Watson present tables of bounds at the 0.05, 0.025 and 0.01 significance levels of the d test statistic for static regressions with an intercept. Against the one-sided alternative that $\rho > 0$, if the observed value of d is less than the lower bound d_L, the null hypothesis that $\rho = 0$ is rejected, if the observed value of d is greater than the upper bound d_U, the null hypothesis is accepted. Otherwise, the test is equivocal. Against the one-sided alternative that $\rho < 0$, $4 - d$ can be used to replace d in the above procedure.

The Durbin-Watson test has been extended to test for lag 4 autocorrelation by Wallis (1972) and for autocorrelation at any lag by Vinod (1973). Several studies show that these tests are powerful against a variety of alternatives including first- and second-order autoregressive and moving average processes (Ali 1984; Blattberg 1973; Smith 1976).

Durbin's h Test.

The Durbin-Watson's d test is not applicable to dynamic linear regression models, because parameter estimation uncertainty in the OLS estimator $\hat{\beta}$ will have nontrivial impact on the distribution of the d statistic. Durbin (1970) developed the so-called h test for first-order autocorrelation in $\{\varepsilon_t\}$ that takes into account the impact of parameter estimation uncertainty in $\hat{\beta}$. Consider a simple dynamic linear regression model

$$Y_t = \beta_0^0 + \beta_1^0 Y_{t-1} + \beta_2^0 X_t + \varepsilon_t,$$

where X_t is strictly exogenous. Durbin's h statistic to test for first-order autocorrelation in $\{\varepsilon_t\}$ is defined as:

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n\hat{\text{var}}(\hat{\beta}_1)}},$$

where $\hat{\text{var}}(\hat{\beta}_1)$ is an estimator for the asymptotic variance of $\hat{\beta}_1$, $\hat{\rho}$ is the OLS estimator from regressing e_t on e_{t-1} (in fact, $\hat{\rho} \approx 1 - d/2$). Durbin (1970) shows that, under null hypothesis that $\rho = 0$,

$$h \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$

Breusch–Godfrey Test

A more convenient and generally applicable test for serial correlation is the Lagrange Multi-

plier test developed by Breusch (1978) and Godfrey (1978). Consider an auxiliary autoregression of order p :

$$\varepsilon_t = \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j} + z_t, \qquad t = p+1, \cdots, n. \tag{2.1}$$

The null hypothesis of no serial correlation implies $\alpha_j = 0$ for all $1 \leq j \leq p$. Under the null hypothesis of no serial correlation, we have $nR_{uc}^2 \xrightarrow{d} \chi_p^2$, where R_{uc}^2 is the uncentered R^2 of (2.1). However, the auxiliary autoregression (2.1) is infeasible because ε_t is unobservable. One can replace ε_t with the estimated OLS residual e_t :

$$e_t = \sum_{j=1}^{p} \alpha_j e_{t-j} + v_t, \qquad t = p+1, \cdots, n.$$

Such a replacement, however, may contaminate the asymptotic distribution of the test statistic because the estimated residual $e_t = \varepsilon_t - (\hat{\beta} - \beta)' X_t$ contains the estimation error $(\hat{\beta} - \beta)' X_t$ of which X_t may have nonzero correlation with the regressors e_{t-j} for $1 \leq j \leq p$ in dynamic regression models. This correlation affects the asymptotic distribution of nR_{uc}^2 so that it will not be χ_p^2. To purge this impact of the asymptotic distribution of the test statistic, one can consider the augmented auxiliary regression

$$e_t = X_t' \gamma + \sum_{j=1}^{p} \alpha_j e_{t-j} + v_t, \qquad t = p+1, \cdots, n. \tag{2.2}$$

The inclusion of X_t will capture the impact of estimation error $(\hat{\beta} - \beta)' X_t$. As a result, the resulting test statistic $nR^2 \xrightarrow{d} \chi_p^2$ under the null hypothesis of no serial correlation, where, assuming that X_t contains an intercept, R^2 is the centered squared multi-correlation coefficient in (2.2). For a static linear regression model, it is not necessary to include X_t in the auxiliary regression, because $\{X_t\}$ and $\{\varepsilon_t\}$ are uncorrelated, but it does not harm the size of the test if X_t is included. Therefore, the nR^2 test is applicable to both static and dynamic regression models.

Box-Pierce-Ljung Test

In time series ARMA modelling, Box and Pierre (1970) propose a portmanteau test as a diagnostic check for the adequacy of an ARMA model

$$Y_t = \psi_0 + \sum_{j=1}^{r} \psi_j Y_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \qquad \{\varepsilon_t\} \sim i.i.d.(0, \sigma^2). \tag{(2.3)}$$

Suppose e_t is an estimated residual obtained from a maximum likelihood estimator. One can

define the residual sample autocorrelation function

$$\hat{\rho}(j) \equiv \frac{\hat{\gamma}(j)}{\hat{\gamma}(0)}, \qquad j = 0, \pm 1, \cdots, \pm(n-1),$$

where $\hat{\gamma}(j) = n^{-1}\Sigma_{t=|j|+1}^{n} e_t e_{t-|j|}$ is the residual sample autocovariance function.

Box and Pierce (1970) propose a portmanteau test

$$Q \equiv n \sum_{j=1}^{p} \hat{\rho}^2(j) \xrightarrow{d} \chi^2_{p-(r+q)},$$

where the asymptotic χ^2 distribution follows under the null hypothesis of no serial correlation, and the adjustment of degrees of freedom $r+q$ is due to the impact of parameter estimation uncertainty for the r autoregressive coefficients and q moving average coefficients in (2.3).

To improve small sample performance of the Q test, Ljung and Box (1978) propose a modified Q test statistic:

$$Q^* \equiv n(n+2) \sum_{j=1}^{p} (n-j)^{-1} \hat{\rho}^2(j) \xrightarrow{d} \chi^2_{p-(r+q)}.$$

The modification matches the first two moments of Q^* with those of the χ^2 distribution. This improves the size in small samples, although not the power of the test.

The Q test is applicable to test serial correlation in the OLS residuals $\{e_t\}$ of a linear static regression model, where

$$Q \xrightarrow{d} \chi^2_p$$

under the null hypothesis of no serial correlation. Unlike for ARMA models, there is no need to adjust the degrees of freedom for the χ^2 distribution because the estimation error $(\hat{\beta} - \beta)' X_t$ has no impact on it, due to the fact that $\text{cov}(X_t, \varepsilon_s) = 0$ for all t, s. In fact, it could be shown that the nR^2 statistic and the Q statistic are asymptotically equivalent under the null hypothesis. However, when applied to the estimated residual of a dynamic regression model which contains both endogenous and exogenous variables, the asymptotic distribution of the Q test is generally unknown (Breusch and Pagan 1980). One solution is to modify the Q test statistic as follows:

$$\hat{Q} \equiv n\hat{\rho}'(I - \hat{\Phi})^{-1}\hat{\rho} \xrightarrow{d} \chi^2_p \text{ as } n \to \infty,$$

where $\hat{\rho} = [\hat{\rho}(1), \cdots, \hat{\rho}(p)]'$, and $\hat{\Phi}$ captures the impact caused by nonzero correlation between $\{X_t\}$ and $\{\varepsilon_s\}$. See Hayashi (2000, Section 2.10) for more discussion.

Spectral Density-Based Test

Much criticism has been leveled at the possible low power of the Box-Pierce-Ljung portman-

teau tests, which also applies to the nR^2 test, due to the asymptotic equivalence between the Q test and the nR^2 test for a static regression. Moreover, there is no theoretical guidance on the choice of p for these tests. A fixed lag order p will render inconsistent any test for serial correlation of unknown form.

To test serial correlation of unknown form in the estimated residuals of a linear regression model, which can be static or dynamic, Hong (1996) uses a kernel spectral density estimator

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

and compares it with the flat spectrum implied by the null hypothesis of no serial correlation:

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

Under the null hypothesis, $\hat{h}(\omega)$ and $\hat{h}_0(\omega)$ are close. If $\hat{h}(\omega)$ is significantly different from $\hat{h}_0(\omega)$, then there is evidence of serial correlation. A global measure of the divergence between $\hat{h}(\omega)$ and $\hat{h}_0(\omega)$ is the quadratic form

$$L^2(\hat{h}, \hat{h}_0) \equiv \int_{-\pi}^{\pi} \left[\hat{h}(\omega) - \hat{h}_0(\omega)\right]^2 d\omega = \sum_{j=1}^{n-1} k^2(j/p)\hat{\gamma}^2(j).$$

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

$$M_o \equiv \left[n \sum_{j=1}^{n-1} k^2(j/p)\hat{\rho}^2(j) - \hat{C}_o(p)\right] / \sqrt{\hat{D}_o(p)} \xrightarrow{d} N(0,1)$$

where the centering and scaling factors

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

This test can be viewed as a generalized version of Box and Pierre's (1970) portmanteau test, the latter being equivalent to using the truncated kernel $k(z) = \mathbf{1}(|z| \leq 1)$, which gives equal weighting to each of the first p lags. In this case, M_o is asymptotically equivalent to

$$M_T \equiv \frac{n\sum_{j=1}^{p}\hat{\rho}^2(j) - p}{\sqrt{2p}} \xrightarrow{d} \frac{\chi_p^2 - p}{\sqrt{2p}} \sim N(0,1) \text{ as } p \to \infty$$

However, uniform weighting to different lags may not be expected to be powerful when a large number of lags is employed. For any weakly stationary process, the autocovariance function $\gamma(j)$ typically decays to 0 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. Hong (1996) shows that the Daniell kernel

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

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

It is important to note that the spectral density test M applies to both static and dynamic regression models, and no modification is needed when applied to a dynamic regression model. Intuitively, parameter estimation uncertainty causes some adjustment of degrees of freedom, which becomes asymptotically independent when the lag order $p \to \infty$ as $n \to \infty$. This differs from the case where p is fixed.

For similar spectral density-based tests for serial correlation, see also Paparoditis (2000), Chen and Deo (2004), and Fan and Zhang (2004).

Heteroskedasticity-Robust Tests

All tests for serial correlation discussed so far assume conditional homoskedasticity or even *i.i.d.* on $\{\varepsilon_t\}$, which rules out high frequency financial time series, which has been documented to have persistent volatility clustering. Some effort has been devoted to robustifying tests for serial correlation. Wooldridge (1990,1991) proposes a two-stage procedure to robustify the nR^2 test for serial correlation in estimated residuals $\{e_t\}$ of a linear regression model (1.1): (i) Regress $(e_{t-1}, \cdots, e_{t-p})$ on X_t and save the estimated $p \times 1$ residual vector $\hat{v}_t$; (ii) Regress 1 on $\hat{v}_t e_t$ and obtain SSR, the sum of squared residuals; (iii) Compare the $n - SSR$ statistic with the asymptotic χ^2_p distribution. The first auxiliary regression purges the impact of parameter estimation uncertainty in the OLS estimator $\hat{\beta}$ and the second auxiliary regression delivers a test statistic robust to conditional heteroskedasticity of unknown form.

Whang (1998) also proposes a semiparametric test for serial correlation in estimated residuals of a possibly nonlinear regression model. Assuming that $\varepsilon_t = \sigma[Z_t(\alpha)]z_t$, where $\{z_t\} \sim$ i.i.d.$(0,1)$, and $\text{var}(\varepsilon_t|I_{t-1}) = \sigma^2[Z_t(\alpha)]$ depends on a random vector with fixed dimension (e.g., $Z_t(\alpha) = (\varepsilon^2_{t-1}, ..., \varepsilon^2_{t-K})'$ for a fixed K), but the functional form $\sigma^2(\cdot)$ is unknown. This covers a variety of conditionally heteroskedastic processes, although it rules out non-Markovian processes such as Bollerslev's (1986) GARCH model. Whang (1998) first estimates $\sigma^2[Z_t(\alpha)]$ nonparametrically using a kernel method, and then constructs a Box-Pierce type test for serial correlation in the

estimated regression residuals standardized by the square root of the nonparametric variance estimator.

The assumption imposed on $\text{var}(\varepsilon_t|I_{t-1})$ in Whang (1998) rules out popular GARCH models, and both Wooldridge (1991) and Whang (1998) test serial correlation up to a fixed lag order only. Hong and Lee (2006) have recently robustified Hong's (1996) spectral density-based consistent test for serial correlation of unknown form:

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

where the centering and scaling factors

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

where

$$\hat{\gamma}_{22}(j) \equiv n^{-1} \sum_{t=j+1}^{n-1} [e_t^2 - \hat{\gamma}(0)][e_{t-j}^2 - \hat{\gamma}(0)]$$

is the sample autocovariance function of $\{\varepsilon_t^2\}$, and

$$\hat{C}(0,j,l) \equiv n^{-1} \sum_{t=\max(j,l)+1}^{n} [\varepsilon_t^2 - \hat{\gamma}(0)] e_{t-j} e_{t-l}$$

is the sample fourth order moment. Intuitively, the centering and scaling factors have taken into account possible volatility clustering and asymmetric features of volatility dynamics, so the $\hat{M}$ test is robust to these effects. It allows for various volatility processes, including GARCH models, Nelson's (1991) EGARCH, and Glosten *et al.*'s (1993) Threshold GARCH models.

Martingale Tests

Several tests for serial correlation are motivated for testing the *m.d.s.* property of an observed time series $\{Y_t\}$, say asset returns, rather than estimated residuals of a regression model. We now present a unified framework to view some martingale tests for observed data.

Extending an idea of Cochrance (1988), Lo and MacKinlay (1988) first rigorously present an asymptotic theory for a variance ratio test for the *m.d.s.* hypothesis of asset returns $\{Y_t\}$. Recall

that $\sum_{j=1}^{p} Y_{t-j}$ is the cumulative asset return over a total of p periods. Then under the *m.d.s.* hypothesis, which implies $\gamma(j) = 0$ for all $j > 0$, one has

$$\frac{\text{var}\left(\sum_{j=1}^{p} Y_{t-j}\right)}{p \cdot \text{var}(Y_t)} = \frac{p\gamma(0) + 2p\sum_{j=1}^{p}(1 - j/p)\gamma(j)}{p\gamma(0)} = 1.$$

This unity property of the variance ratio can be used to test the *m.d.s.* hypothesis because any departure from unity is evidence against the *m.d.s.* hypothesis.

The variance ratio test is essentially based on the statistic

$$\text{VR}_o \equiv \sqrt{n/p}\sum_{j=1}^{p}(1 - j/p)\hat{\rho}(j) = \frac{\pi}{2}\sqrt{n/p}\left[\hat{f}(0) - \frac{1}{2\pi}\right],$$

where $\hat{f}(0)$ is a kernel-based normalized spectral density estimator at frequency 0, with the Bartlett kernel $K(z) = (1 - |z|)\mathbf{1}(|z| \leq 1)$ and a lag order equal to p. In other words, VR_o is based on a spectral density estimator of frequency 0, and because of this, it is particularly powerful against long memory processes, whose spectral density at frequency 0 is infinity (see Robinson 1994, for discussion on long memory processes).

Under the *m.d.s.* hypothesis with conditional homoskedasticity, Lo and MacKinlay (1988) show that for any fixed p,

$$\text{VR}_o \xrightarrow{d} N[0, 2(2p-1)(p-1)/3p] \text{ as } n \to \infty.$$

Lo and MacKinlay (1988) also consider a heteroskedasticity-consistent variance ratio test:

$$\text{VR} \equiv \sqrt{n/p}\sum_{j=1}^{p}(1 - j/p)\hat{\gamma}(j)/\sqrt{\hat{\gamma}_2(j)},$$

where $\hat{\gamma}_2(j)$ is a consistent estimator for the asymptotic variance of $\hat{\gamma}(j)$ under conditional heteroskedasticity. Lo and MacKinlay (1988) assume a fourth order cumulant condition that

$$E\left[(Y_t - \mu)^2(Y_{t-j} - \mu)(Y_{t-l} - \mu)\right] = 0, \qquad j, l > 0, j \neq l. \tag{(2.4)}$$

Intuitively, this condition ensures that the sample autocovariances at different lags are asymptotically uncorrelated; that is, $\text{cov}[\sqrt{n}\hat{\gamma}(j), \sqrt{n}\hat{\gamma}(l)] \to 0$ for all $j \neq l$. As a result, the heroskedasticity-consistent VR has the same asymptotic distribution as VR_o. However, the condition in (2.4) rules out many important volatility processes, such as EGARCH and Threshold GARCH models. Moreover, the variance ratio test only exploits the implication of the *m.d.s.* hypothesis on the spectral density at frequency 0; it does not check the spectral density at nonzero frequencies. As a result, it is not consistent against serial correlation of unknown form. See

Durlauf (1991) for more discussion.

Durlauf (1991) considers testing the *m.d.s.* hypothesis for observed raw data $\{Y_t\}$, using the spectral distribution function

$$H(\lambda) \equiv 2\int_0^{\pi\lambda} h(\omega)d\omega = \gamma(0)\lambda + \sqrt{2}\sum_{j=1}^{\infty}\gamma(j)\frac{\sqrt{2}\sin(j\pi\lambda)}{j\pi}, \qquad \lambda \in [0,1],$$

where $h(\omega)$ is the spectral density of $\{Y_t\}$:

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

Under the *m.d.s.* hypothesis, $H(\lambda)$ becomes a straight line:

$$H_0(\lambda) = \gamma(0)\lambda,\ \ \lambda \in [0,1].$$

An *m.d.s.* test can be obtained by comparing a consistent estimator for $H(\lambda)$ and $\hat{H}_0(\lambda) = \hat{\gamma}(0)\lambda$. Although the periodogram (or sample spectral density function)

$$\hat{I}(\omega) \equiv \frac{1}{2\pi n}\left|\sum_{t=1}^{n}(Y_t - \bar{Y})e^{it\omega}\right|^2 = \frac{1}{2\pi}\sum_{j=1-n}^{n-1}\hat{\gamma}(j)e^{-ij\omega}$$

is not consistent for the spectral density $h(\omega)$, the integrated periodogram

$$\hat{H}(\lambda) \equiv 2\int_0^{\lambda\pi}\hat{I}(\omega)d\omega = \hat{\gamma}(0)\lambda + \sqrt{2}\sum_{j=1}^{n-1}\hat{\gamma}(j)\frac{\sqrt{2}\sin(j\pi\lambda)}{j\pi}$$

is consistent for $H(\lambda)$, thanks to the additional smoothing provided by the integration. Among other things, Durlauf (1991) proposes a Cramer-von Mises type statistic

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

Under the *m.d.s.* hypothesis with conditional homoskedasticity (i.e., $\mathrm{var}(Y_t|I_{t-1}) = \sigma^2$ a.s.), Durlauf (1991) shows

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

where $\{\chi_j^2(1)\}_{j=1}^{\infty}$ is a sequence of *i.i.d.* χ^2 random variables with one degree of freedom. This asymptotic distribution is nonstandard, but it is distribution-free and can be easily tabulated

or simulated. An appealing property of Durlauf's (1991) test is its consistency against serial correlation of unknown form, and there is no need to choose a lag order p.

Deo (2000) shows that under the *m.d.s.* hypothesis with conditional heteroskedasticity, Durlauf's (1991) test statistic can be robustified as follows:

$$CVM = \sum_{j=1}^{n-1} \frac{\left[\hat{\gamma}^2(j)/\hat{\gamma}_2(j)\right]}{(j\pi)^2} \xrightarrow{d} \sum_{j=1}^{\infty} \chi_j^2(1)/(j\pi)^2$$

where $\hat{\gamma}_2(j)$ is a consistent estimator for the asymptotic variance of $\hat{\gamma}(j)$ and the asymptotic distribution remains unchanged. Like Lo and MacKinlay (1988), Deo (2000) also assume the crucial fourth order joint cumulant condition in (2.4).

3 Serial Dependence in Nonlinear Models

The autocorrelation function $\gamma(j)$, or equivalently, the power spectrum $h(\omega)$, of a time series $\{Y_t\}$, is a measure for linear association. When $\{Y_t\}$ is stationary Gaussian (i.e., for any admissible t and k , the set of random variables $\{Y_t, Y_{t+1}, \cdots, Y_{t+k}\}$ follows a multivariate normal distribution with constant mean and constant variance-covariance matrix), $\gamma(j)$ or $h(\omega)$ can completely determine the full dynamics of $\{Y_t\}$.

It has been well documented, however, that most economic and financial time series, particularly high-frequency economic and financial time series, are not Gaussian. For non-Gaussian processes, $\gamma(j)$ and $h(\omega)$ may not capture the full dynamics of $\{Y_t\}$. Consider two nonlinear process examples:

- Bilinear (BL) autoregressive process:

$$Y_t = \alpha\varepsilon_{t-1}Y_{t-2} + \varepsilon_t, \ \{\varepsilon_t\} \sim i.i.d.(0, \sigma^2), \tag{3.1}$$

- Nonlinear moving average (NMA) process

$$Y_t = \alpha\varepsilon_{t-1}\varepsilon_{t-2} + \varepsilon_t, \ \{\varepsilon_t\} \sim i.i.d.(0, \sigma^2). \tag{3.2}$$

For these two processes, there exists nonlinearity in conditional mean. For the BL process in (3.1),

$$E(Y_t|I_{t-1}) = \alpha\varepsilon_{t-1}Y_{t-2};$$

for the NMA process in (3.2),

$$E(Y_t|I_{t-1}) = \alpha\varepsilon_{t-1}\varepsilon_{t-2}.$$

However, both processes are serially uncorrelated. If asset returns follow either BL in (3.1) or NMA in (3.2), the market is not efficient but $\gamma(j)$ and $h(\omega)$ will miss them. Hong and Lee (2003a) document that indeed, for foreign currency markets, most foreign exchange changes are serially uncorrelated, but they are all not *m.d.s.* There may exist predictable nonlinear components in the conditional mean dynamics of foreign exchange markets.

It is also possible that serial dependence exists only in higher order conditional moments. An example is Engle's (1982) first order autoregressive conditional heteroskedatic (ARCH (1)) process:

$$\begin{cases} Y_t = \sigma_t \varepsilon_t, \\ \sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2, \\ \{\varepsilon_t\} \sim i.i.d.\,(0,1)\,. \end{cases} \tag{3.3}$$

For this process, the conditional mean

$$E(Y_t|I_{t-1}) = 0,$$

which implies $\gamma(j) = 0$ for all $j > 0$. However, the conditional variance,

$$var(Y_t|I_{t-1}) = \alpha_0 + \alpha_1 Y_{t-1}^2,$$

depends on the previous volatility. Both $\gamma(j)$ and $h(\omega)$ will miss such higher order dependence.

In nonlinear time series modeling, it is important to measure serial dependence, i.e., any departure from *i.i.d.*, rather than serial correlation. As Priestley (1988) points out, the main purpose of nonlinear time series analysis is to find a filter $h(\cdot)$ such that

$$h(Y_t, Y_{t-1}, \cdots .) = \varepsilon_t \sim i.i.d.(0, \sigma^2).$$

In other words, the filter $h(\cdot)$ can capture all serial dependence in $\{Y_t\}$ so that the "residual" $\{\varepsilon_t\}$ becomes an *i.i.d.* sequence. One example of $h(\cdot)$ in modeling the conditional probability distribution of Y_t given I_{t-1} is the probability integral transform

$$Z_t(\beta) = \int_{-\infty}^{Y_t} f(y|I_{t-1}, \beta)dy,$$

where $f(y|I_{t-1}, \beta)$ is a conditional density model for Y_t given I_{t-1}, and β is an unknown parameter. When $f(y|I_{t-1}, \beta)$ is correctly specified for the conditional probability density of Y_t given I_{t-1}, that is, when the true conditional density coincides with $f(y|I_{t-1}, \beta^0)$ for some β^0, the probability integral transforms becomes

$$\{Z_t(\beta^0)\} \sim i.i.d.U[0,1]. \tag{3.4}$$

Thus, one can test whether $f(y|I_{t-1}, \beta)$ is correctly specified by checking the *i.i.d.*U[0,1] for the

probability integral transform series.

Bispectrum and Higher Order Spectra

Due to the very nature of measuring linear association, the autocorrelation function $\gamma(j)$ and the spectral density $h(\omega)$ are rather limited in nonlinear time series analysis. Various alternative tools have been proposed to capture nonlinear serial dependence (e.g., Granger and Terasvirta 1993, Tjøstheim 1996). In nonlinear time series analysis, one often uses the third order cumulant function

$$C(j,k) \equiv E[(Y_t - \mu)(Y_{t-j} - \mu)(Y_{t-k} - \mu)], j, k = 0, \pm 1, \cdots .$$

This is also called the biautocovariance function of time series $\{Y_t\}$.. It can capture certain nonlinear time series, particularly those displaying asymmetric behaviors such as skewness. Hsieh (1989) proposes a test based on $C(j,k)$ for a given pair of (j,k) which can detect certain predictable nonlinear components in asset returns.

The Fourier transform of $C(j,k)$,

$$b(\omega_1, \omega_2) \equiv \frac{1}{(2\pi)^2} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C(j,k) e^{-ij\omega_1 - ik\omega_2}, \qquad \omega_1, \omega_2 \in [-\pi, \pi],$$

is called the bispectrum. When $\{Y_t\}$ is *i.i.d.*, $b(\omega_1, \omega_2)$ becomes a flat bispectral surface:

$$b_0(\omega_1, \omega_2) \equiv E(Y_t^3), \quad \omega_1, \omega_2 \in [-\pi, \pi].$$

Any deviation from a flat bispectral surface will indicate the existence of serial dependence in $\{Y_t\}$. Moreover, the bispectrum $b(\omega_1, \omega_2)$ can be used to distinguish some linear time series processes from nonlinear time series processes. When $\{Y_t\}$ is a linear process with *i.i.d.* innovations, i.e., when

$$Y_t = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j \varepsilon_{t-j} + \varepsilon_t, \quad \{\varepsilon_t\} \sim i.i.d. \left(0, \sigma^2\right),$$

the normalized bispectrum

$$\widetilde{b}(\omega_1, \omega_2) \equiv \frac{b(\omega_1, \omega_2)}{h(\omega_1)\, h(\omega_2)\, h(\omega_1 + \omega_2)} = E(\varepsilon_t^3)$$

is a flat surface. Any departure from a flat normalized bispectral surface will indicate that $\{Y_t\}$ is not a linear time series with *i.i.d.* innovations.

The bispectrum $b(\omega_1, \omega_2)$ can capture the BL and NMA processes in (3.1) and (3.2), because the third order cumulant $C(j,k)$ can distinguish them from an *i.i.d.* process. However, it may still miss some important alternatives. For example, it will easily miss ARCH (1) with *i.i.d.*N(0,1) innovation $\{\varepsilon_t\}$. In this case, $b(\omega_1, \omega_2)$ becomes a flat bispectrum and cannot distinguish ARCH

(1) from an *i.i.d.* sequence. One could use higher order spectra or polyspectra (Brillinger and Rosenblatt 1967a,1967b), which are the Fourier transforms of higher order cumulants. However, higher order spectra have met with some difficulty in practice: Their spectral shapes are difficult to interpret, and their estimation is not stable in finite samples, due to the assumption of the existence of higher order moments. Indeed, it is often a question whether economic and financial data, particularly high-frequency data, have finite higher order moments.

Nonparametric Measures of Serial Dependence

In the recent literature, nonparametric measures for serial dependence have been proposed, which avoid assuming the existence of moments. Granger and Lin (1994) propose a nonparametric entropy measure for serial dependence to identify significant lags in nonlinear time series. Define the Kullback-Leibler information criterion

$$I(j) = \int \ln \left[\frac{f_j(x,y)}{g(x)g(y)} \right] f_j(x,y)dxdy, \qquad j = 1, 2, \ldots .$$

where $f_j(x,y)$ is the joint probability density function of Y_t and Y_{t-j}, and $g(x)$ is the marginal probability density of $\{Y_t\}$. The Granger-Lin normalized entropy measure is defined as follows:

$$e^2(j) = 1 - \exp[-2I(j)],$$

which enjoys a number of appealing features such as invariance to monotonic continuous transformation. Because $f_j(x,y)$ and $g(x)$ are unknown, Granger and Lin (1994) use nonparametric kernel density estimators. They establish the consistency of their normalized entropy estimator (say $\hat{I}(j)$) but do not derive its asymptotic distribution, which is important for confidence interval estimation and hypothesis testing.

In fact, Robinson (1991) has elegantly explained the difficulty of obtaining the asymptotic distribution of a nonparametric entropy estimator for serial dependence, namely it is a degenerate statistic so that the usual root-n normalization does not deliver a well-defined asymptotic distribution. As a sensible solution, Robinson (1991) uses a weighting device and considers a modified entropy estimator

$$\hat{I}_\gamma(j) = n^{-1} \sum_{t=j+1}^{n} C_t(\gamma) \ln \left[\frac{\hat{f}_j(Y_t, Y_{t-j})}{\hat{g}(Y_t)\hat{g}(Y_t)} \right],$$

where $\hat{f}_j(\cdot,\cdot)$ and $\hat{g}(\cdot)$ are nonparametric kernel density estimators, $C_t(\gamma) = 1 - \gamma$ if t is odd, $C_t(\gamma) = 1 + \gamma$ if t is even, and γ is a prespecified parameter. The weighting device does not affect the consistency of $\hat{I}(j)$ to $I(j)$ and affords a well-defined asymptotic N(0,1) distribution under the *i.i.d.* hypothesis.

Skaug and Tjøsthem (1993a,1996) use a different weighting function to avoid the degeneracy of the entropy estimator:

$$\hat{I}_w(j) = n^{-1} \sum_{t=1}^{n} W(Y_t, Y_{t-j}) \ln \left[\frac{\hat{f}_j(Y_t, Y_{t-j})}{\hat{g}(Y_t)\hat{g}(Y_{t-j})} \right]$$

where $W(Y_t, Y_{t-j})$ is a weighting function of observations X_t and X_{t-j}. Unlike Robinson's (1991) weighting device, this modified entropy estimator is not consistent for the population entropy $I(j)$, but it also delivers a well-defined asymptotic N(0,1) distribution after a root-n normalization.

Intuitively, the use of weighting devices slow down the convergence rate of the entropy estimator, giving a well-defined asymptotic N(0,1) distribution after the usual root-n normalization. However, this is achieved at the cost of an efficiency loss, due to the slower convergence rate of the weighted entropy estimator. Moreover, this approach breaks down when $\{Y_t\}$ is uniformly distributed, as in the case of the probability integral transforms of the conditional density in (3.4). Hong and White (2005) take a different approach. Instead of using a weighting device, Hong and White (2005) exploit the degeneracy of the entropy estimator and use a degenerate U-statistic theory to establish the asymptotic normality. Specifically, Hong and White (2005) show

$$nh\hat{I}(j) + hd_n^0 \xrightarrow{d} N(0, V),$$

where $h = h(n)$ is the bandwidth, and d_n^0 and V are nonstochastic factors. A payoff of this approach is that it preserves the convergence rate of the unweighted entropy estimator, giving sharper confidence interval estimation and more powerful hypothesis tests. It also avoids choosing a weighting device and is applicable when $\{Y_t\}$ is uniformly distributed.

Skaug and Tjøstheim (1993b) also use an empirical Hoeffding measure to test serial dependence. See also Delgado (1996) and Hong (1998, 2000). The empirical Hoeffding measures are based on the empirical distribution functions.

Generalized Spectrum

Without assuming the existence of higher order moments, Hong (1999) proposes a generalized spectrum as an alternative analytic tool to power spectrum and higher order spectra. The basic idea is to transform $\{Y_t\}$ via a complex-valued exponential function

$$Y_t \to \exp(iuY_t), \quad u \in (-\infty, \infty),$$

and then consider the spectrum of the transformed series. Let

$$\psi(u) \equiv E(e^{iuY_t}),$$

be the marginal characteristic function of $\{Y_t\}$ and let

$$\psi_j(u,v) \equiv E[e^{i\left(uY_t+vY_{t-|j|}\right)}], \qquad j = 0, \pm 1, \cdots,$$

be the pairwise joint characteristic function of $(Y_t, Y_{t-|j|})$. Define the covariance function between transformed variables e^{iuY_t} and $e^{ivY_{t-|j|}}$:

$$\sigma_j(u,v) \equiv \mathrm{cov}(e^{iuY_t}, e^{ivY_{t-|j|}})$$

Straightforward algebra yields

$$\sigma_j(u,v) = \psi_j(u,v) - \psi(u)\,\psi(v),$$

which is zero for all u, v if and only if Y_t and $Y_{t-|j|}$ are independent. Thus $\sigma_j(u,v)$ can capture any type of pairwise serial dependence over various lags, including those with zero autocorrelation. For example, $\sigma_j(u,v)$ can capture the BL, NMA and ARCH (1) processes in (3.1)–(3.3), all of which are serially uncorrelated.

The Fourier transform of the generalized covariance $\sigma_j(u,v)$:

$$f(\omega,u,v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j(u,v)e^{-ij\omega}, \qquad \omega \in [-\pi,\pi]$$

is called the "generalized spectral density" of $\{Y_t\}$. Like $\sigma_j(u,v)$, $f(\omega,u,v)$ can capture any type of pairwise serial dependencies in $\{Y_t\}$ over various lags. Unlike the power spectrum and higher order spectra, the generalized spectrum $f(\omega,u,v)$ does not require any moment condition on $\{Y_t\}$. When $\mathrm{var}(Y_t)$ exists, the power spectrum of $\{Y_t\}$ can be obtained by differentiating $f(\omega,u,v)$ with respect to (u,v) at $(0,0)$:

$$h(\omega) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma(j)e^{-ij\omega} = -\frac{\partial^2}{\partial u\partial v}f(\omega,u,v)|_{(u,v)=(0,0)}, \qquad \omega \in [-\pi,\pi].$$

This is the reason why $f(\omega,u,v)$ is called the "generalized spectral density" of $\{Y_t\}$.

When $\{Y_t\}$ is *i.i.d.*, $f(\omega,u,v)$ becomes a flat generalized spectrum as a function of ω:

$$f_0(\omega,u,v) = \frac{1}{2\pi}\sigma_0(u,v), \qquad \omega \in [-\pi,\pi]$$

Any deviation of $f(\omega, u, v)$ from the flat generalized spectrum $f_0(\omega, u, v)$ is evidence of serial dependence. Thus, the generalized spectrum is suitable to capture any departures from *i.i.d.*. Hong and Lee (2003b) use the generalized spectrum to develop a test for the adequacy of nonlinear time series models by checking whether the standardized model residuals are *i.i.d.*. Tests for *i.i.d.* are more suitable than tests for serial correlation in nonlinear contexts. Indeed, Hong and Lee (2003b), in an empirical application, show that some popular EGARCH models are inadequate in capturing full dynamics of asset returns, although the standardized model residuals do not display serial correlation.

Insight into the ability of $f(\omega, u, v)$ can be gained by considering a Taylor series expansion

$$f(\omega, u, v) = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \frac{(iu)^m (iv)^l}{m!l!} \left[\frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \operatorname{cov}(X_t^l, X_{t-|j|}^m) e^{-ij\omega} \right].$$

Although $f(\omega, u, v)$ has no physical interpretation, it can be used to characterize cyclical movements caused by linear and nonlinear serial dependence. Examples of nonlinear cyclical movements include cyclical volatility clustering, and cyclical tail clustering (e.g., Engle and Manganlli's (2004) CAVaR model). Intuitively, the supremum function

$$s(\omega) = \sup_{-\infty < u, v < \infty} |f(\omega, u, v)|, \qquad \omega \in [-\pi, \pi],$$

can measure the maximum dependence at frequency ω of time series $\{Y_t\}$. This function can be viewed as an operational frequency domain analog of Granger and Terasvirta's (1993) maximum correlation measure

$$mm\rho(j) = \max_{g(\cdot), h(\cdot)} |\operatorname{corr}[g(Y_t), h(X_{t-j})]|.$$

Once generic serial dependence is detected using $f(\omega, u, v)$ or any other measure, one may like to explore the nature and pattern of serial dependence. For example, one may be interested in the following questions:

- Is serial dependence operative primarily through the conditional mean or through conditional higher order moments?
- If serial dependence exists in conditional mean, is it linear or nonlinear?
- If serial dependence exists in conditional variance, does there exist linear or nonlinear and asymmetric ARCH?

Different types of serial dependence have different economic implications. For example, the efficient market hypothesis fails if and only if there is no serial dependence in conditional mean.

Just as the characteristic function can be differentiated to generate various moments, generalized spectral derivatives, when exists, can capture various specific aspects of serial dependence, thus providing information on possible types of serial dependence. Suppose $E[(Y_t)^{2\max(m,l)}] < \infty$ for some nonnegative integers m, l. Then the following generalized spectral derivative exists:

$$f^{(0,m,l)}(\omega, u, v) \equiv \frac{\partial^{m+l}}{\partial u^m \partial v^l} f(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j^{(m,l)}(u, v) e^{-ij\omega}$$

where

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

As an illustrative example, consider the generalized spectral derivative of order $(m, l) = (1, 0)$:

$$f^{(0,1,0)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j^{(1,0)}(u, v) e^{-ij\omega},$$

Observe

$$\sigma_j^{(1,0)}(0, v) = \text{cov}(iY_t, e^{-ivY_{t-|j|}}) = 0, \text{ for all } v \in (-\infty, \infty)$$

if and only if

$$E(Y_t | Y_{t-|j|}) = E(Y_t) \text{ a.s..}$$

The function $E(Y_t|Y_{t-|j|})$ is called the autoregression function of $\{Y_t\}$ at lag j. It can capture a variety of linear and nonlinear dependencies in conditional mean, including the BL and NMA processes in (3.1) and (3.2). The use of $\sigma_j^{(1,0)}(0, v)$, which can be easily estimated by a sample average, avoids smoothed nonparametric estimation of $E(Y_t|Y_{t-j})$. Thus, the generalized spectral derivative, $f^{(0,1,0)}(\omega, u, v)$, can be used to capture a wide range of serial dependence in conditional mean. In particular, the function

$$s(\omega) = \sup_{-\infty < v < +\infty} |f^{(0,1,0)}(\omega, 0, v)|$$

can be viewed as an operational frequency domain analog of Granger and Terasvirta's (1993) maximum mean correlation measure

$$mm(j) = \max_{h(\cdot)} |\text{corr}(Y_t, h(Y_{t-j})|.$$

Suppose one has found evidence of serial dependence in conditional mean using $f^{(0,1,0)}(\omega, u, v)$ or any other measure, one can go further to explore whether there exists linear serial dependence

in mean. This can be done by using the $(1,1)$-th order generalized derivative

$$f^{(0,1,1)}(\omega, 0, 0) = -h(\omega),$$

which checks serial correlation. Moreover, one can further use $f^{(0,1,l)}(\omega, u, v)$ for $l \geq 2$ to reveal nonlinear serial dependence in mean. In particular, these higher order derivatives can suggest that there exists

- ARCH-in-mean effect (e.g., Engle, Lilian and Robin 1988) if $\text{cov}(Y_t, Y_{t-j}^2) \neq 0$;
- Skewness-in-mean effect (e.g., Harvey and Siddique 2000) if $\text{cov}(Y_t, Y_{t-j}^3) \neq 0$;
- Kurtosis-in-mean effect (e.g., Brock *et al* 2005) if $\text{cov}(Y_t, Y_{t-j}^4) \neq 0$.

These effects may arise from the existence of time-varying risk premium, asymmetry of market behaviors, and improper account of the concern on large losses, respectively.

4 Conclusion

In this entry, we discuss serial correlation in a linear time series regression context and serial dependence in a nonlinear time series context. We first discuss the impact of serial dependence on the statistical inference of linear time series regression models, either static or dynamic. We then discuss various tests for serial correlation for both estimated regression residuals and observed raw data. Particular attention has been paid to the impact of parameter estimation uncertainty on the asymptotic distribution of test statistics. Finally, we discuss the drawback of serial correlation in nonlinear time series models and introduce a number of measures that can capture nonlinear serial dependence and reveal useful information about serial dependence.

References

Ali, M. M. (1984), "An Approximation to the Null Distribution and Power of the Durbin-Watson Statistic", *Biometrika* 71, 253-261.

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

Andrews, D. W. K. and J. C. Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator", *Econometrica* 60, 953-966.

Blattberg, R. C. (1973), "Evaluation of the Power of the Durbin-Watson Statistic for Non-First Order Serial Correlation Alternatives", *Review of Economics and Statistics* 55, 508-515.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedastcity", *Journal of Econometrics* 31, 307-327.

Box, G.E.P. and D.A. Pierce (1970), "Distribution of Residual Autorrelations in Autoregressive Moving Average Time Series Models," *Journal of the American Statistical Association* 65, 1509-1526.

Breusch, T. S. (1978), "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers* 17,334-355.

Breusch, T. S. and A. Pagan (1980), "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics," *Review of Economic Studies* 47, 239–253.

Brillinger, D. R. and M. Rosenblatt (1967a), "Computation and Interpretation of the kth Order Spectra," in *Spectral Analysis of Time Series*, ed. B.Harris, New York: Wiley, pp.153-188.

Brillinger, D. R. and M. Rosenblatt (1967b**), "**Computation and Interpretation of the kth Order Spectra**,"** in *Spectral Analysis of Time Series*, ed. B.Harris, New York: Wiley, pp.189-232.

Brooks, C., S. Burke and G. Persand (2005), "Autoregressive Conditional Kurtosis," *Journal of Financial Econometrics* 3, 399-421.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

Chen, W. and R. Deo (2004), "A Generalized Portmanteau Goodness-of-fit Test for Time Series Models," *Econometric Theory* 20, 382-416.

Cochrane, J. H. (1988), "How Big is the Random Walk in GNP?" *Journal of Political Economy* 96, 893–920.

Delgado, M. A. (1996), "Testing Serial Independence Using the Sample Distribution Function", *Journal of Time Series Analysis* 17, 271-285.

Deo, R. S. (2000), "Spectral Tests of the Martingale Hypothesis under Conditional Heteroscedasticity", *Journal of Econometrics* 99, 291–315.

Durbin, J. (1970), "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica* 38, 422-421.

Durbin,J. and G. S. Watson (1950), "Testing for Serial Correlation in Least Squares Regression: I," *Biometrika* 37, 409-428.

Durbin, J. and G. S. Watson (1951), "Testing for Serial Correlation in Least Squares Regression: II," *Biometrika* 37, 159-178.

Durlauf, S. N. (1991), "Spectral Based Testing of the Martingale Hypothesis," *Journal of Econometrics* 50, 355–376.

Engle, R. (1982), "Autoregressive Conditional Hetersokedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50, 987-1008.

Engle, R., D. Lilien and R. P., Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model," *Econometrica* 55, 391-407.

Engle, R. and S. Manganelli (2004), "CARViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business and Economic Statistics* 22, 367-391.

Fan J. and W. Zhang (2004), "Generalized likelihood ratio tests for spectral density," *Biometrika* 91, 195-209.

Gallant, A.R. and H. White (1988), A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. New York: Basil Blackwell.

Glosten, R., R. Jagannathan and D. Runkle (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance* 48, 1779-1801.

Godfrey, L. G. (1978), "Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables," *Econometrica* 46, 1293-1301.

Granger, C. W. J. and J. L. Lin (1994), "Using the Mutual Information Coefficient to Identify Lags in Nonlinear Models", *Journal of Time Series Analysis* 15, 371-384.

Granger, C. J. W. and T. Terasvirta (1993), *Modeling Nonlinear Economic Relationships*, Oxford University Press: Oxford.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.

Harvey, C.R. and A.Siddique (2000), "Conditional Skewness in Asset Pricing Tests," *The Journal Of Finance* 51, 1263-1295.

Hayashi, F. (2000), "Econometrics", Princeton University Press: Princeton.

Hong, Y. (1996), "Consistent Testing for Serial Correlation of Unknown Form," *Econometrica* 64, 837-864.

Hong, Y. (1998), "Testing for Pairwise Serial Independence via the Empirical Distribution Function", *Journal of the Royal Statistical Society, Series B*, 60, 429-453.

Hong, Y. (1999), "Hypothesis Testing in Time Series via the Empirical Characteristic Function: a Generalized Spectral Density Approach," *Journal of the American Statistical Association* 94, 1201–1220.

Hong, Y. (2000), "Generalized Spectral Tests for Serial Dependence", *Journal of the Royal Statistical Society, Series B*, 62, 557-574.

Hong, Y. and T. H. Lee (2003a), "Inference on Predictability of Foreign Exchange Rates via Generalized Spectrum and Nonlinear Time Series Models", *Review of Economics and Statistics* 85, 1048-1062.

Hong, Y. and T. H. Lee (2003b), "Diagnostic Checking for the Adequacy of Nonlinear Time Series Models," *Econometric Theory* 19, 1065-1121.

Hong, Y. and Y. J. Lee (2005), "Generalized Spectral Testing for Conditional Mean Models in Time Series with Conditional Heteroskedasticity of Unknown Form," *Review of Economic Studies* 72, 499-451.

Hong, Y. and Y. J. Lee (2006), "Consistent Testing for Serial Correlation of Unknown Form Under General Conditional Heteroskedasticity," Submitted to *Econometric Theory*.

Hong, Y. and H. White (2005), "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence", *Econometrica* 73, 837–901.

Hsieh, D. A. (1989), "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates", *Journal of Business* 62, 339–368.

Ljung, G. M. and G. E. P. Box (1978), "On a Measure of Lack of Fit in Time Series models," *Biometrika* 65, 297–303.

Lo, A. W. and A. C. MacKinlay (1988), "Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies* 1, 41-66.

Nelson, D. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica* 59, 347-370.

Newey, W. K. and West, K. D. (1987), "A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

Newey, W. K. and West, K. D. (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, Blackwell Publishing 61, 631-653.

Paparodistis, E. (2000), "Spectral Density Based Goodness-of-fit Tests for Time Series Models," *Scandinavian Journal of Statistics* 27, 143-176.

Priestley, M. B. (1988), *Non-Linear and Non-Stationary Time Series Analysis*, London: Academic Press.

Robinson, P. M. (1991), "Consistent Nonparametric Entropy-Based Testing", *The Review of Economic Studies* 58, 437-453.

Robinson, P. M. (1994), "Time Series with Strong Dependence", in *Advances in Econometrics, Sixth World Congress*, Vol. 1, ed. by C. Sims. Cambridge: Cambridge University Press, 47-95.

Skaug, H. J. and D. Tjøstheim (1993a), "Nonparametric tests of serial independence", in *Developments in Time Series Analysis*, S. Rao, ed., Chapman and Hall, 207-229.

Skaug, H. J. and D. Tjøstheim (1993b), "A Nonparametric Test of Serial Independence

Based on the Empirical Distribution Function", *Biometrika* 80, 591-602.

Skaug, H. J. and D. Tjøstheim (1996), "Measures of Distance between Densities with Application to Testing for Serial Independence", In *Time Series Analysis in Memory of E. J. Hannan*, 363-377, New York: Springer.

Smith, V. K. (1976), "The Estimated Power of Several Tests for Autocorrelation with Non-First-Order Alternatives", *Journal of the American Statistical Association* 71, 879-883.

Tjøstheim, D. (1996), "Measures and Tests of Independence: A Survey," *Statistics* 28, 249-284.

Vinod, H. D. (1973), "Generalization of the Durbin-Watson Statistic for Higher Order Autoregressive Processes", *Communications in Statistics* 2, 115-144.

Wallis, K. F. (1972), "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations", *Econometrica* 40, 617-636.

Whang, Y. J. (1998), "A Test of Autocorrelation in the Presence of Heteroskedasticity of Unknown Form", *Econometric Theory* 14, 87-122.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance matrix Estimator and a Direct Test for Heterokedasticity," *Econometrica* 48, 817-838.

Wooldridge, J. M. (1990), "An encompassing approach to conditional mean tests with applications to testing nonnested hypotheses," *Journal of Econometrics* 45, 331-350.

Wooldridge, J. M. (1991), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics* 47, 5-46.