Model Specification Testing in Nonparametric and Semiparametric Time Series Econometrics

JITI GAO

Department of Statistics
School of Mathematics and Statistics
The University of Western Australia
Crawley WA 6009, Australia
Email: jiti@maths.uwa.edu.au
Http://www.maths.uwa.edu.au/~jiti

This paper is recent joint work with Maxwell King.

Outline:

1. Motivation

2. Estimation and Testing

3. Example of Implementation

4. Discussion

1. Motivation and Examples

EXAMPLE 1.1 (Nonparametric time series regression):

$$Y_t = \mu(Y_{t-1}) + \sigma(Y_{t-1})e_t, \tag{1.1}$$

where

- $\mu(\cdot)$ and $\sigma(\cdot) > 0$–unknown; and
- $E[e_t|Y_{t-1}] = 0$ and $E[e_t^2|Y_{t-1}] = 1$.
- Estimation of $\mu(\cdot)$ and $\sigma(\cdot)$.
- Univariate testing:

$$H_0: \ \mu(y) = \beta_0(\alpha_0 - y) \ \text{ or } \ \sigma(y) = \sqrt{\sigma_0 + \sigma_1 y^2},$$

  with finite α_0 and β_0, $\sigma_0 > 0$ and $\sigma_1 \geq 0$.

- Simultaneous testing:

$$H_0: \ \mu(y) = \beta_0(\alpha_0 - y) \ \text{ and } \ \sigma(y) = \sqrt{\sigma_0 + \sigma_1 y^2}.$$

EXAMPLE 1.2 (Nonparametric diffusion model):

$$dr_t = \mu(r_t)dt + \sigma(r_t)dB_t, \tag{1.2}$$

where

- $\mu(\cdot)$–Drift function;
- $\sigma(\cdot)$–volatility function; and
- B_t–standard Brownian motion.

Simultaneous testing:

$$H_0 : \ \mu(r) = \mu(r; \theta_0) \ \text{ and } \ \sigma^2(r) = \sigma^2(r; \theta_0).$$

An approximate version of model (1.2) is

$$r_{(t+1)\Delta} - r_{t\Delta} = \mu(r_{t\Delta})\Delta + \sigma(r_{t\Delta}) \cdot (B_{(t+1)\Delta} - B_{t\Delta}), \tag{1.3}$$

where

- Δ–time between successive observations.
- In finance, deal with monthly, weekly, daily, or higher frequency observations.

Let $X_t = r_{(t-1)\Delta}$ and $Y_t = \frac{r_{t\Delta} - r_{(t-1)\Delta}}{\Delta}$.
Model (2.14) can then be written as

$$Y_t = \mu(X_t) + \sigma(X_t)e_t, \tag{1.4}$$

where

- e_t–Gaussian with $E[e_t] = 0$ and $\text{var}[e_t] = \Delta^{-1}$.

2. Estimation and Testing

Consider a general nonparametric time series model of the form

$$Y_t = \mu(X_t) + \sigma(X_t)e_t, \tag{2.1}$$

where

- $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are unknown functions defined over R^d;
- both X_t and e_t can be stationary time series; and
- $\{e_t\}$ satisfies $E[e_t|X_t] = 0$ and $E[e_t^2|X_t] = 1$.
- Let

$$\begin{aligned} m_1(x) &= E(Y_t|X_t = x) = \mu(x), \\ m_2(x) &= \text{var}(Y_t|X_t = x) = \sigma^2(x) \end{aligned} \tag{2.2}$$

  for $x \in S \subset R^d$.

- Define $m(x) = (m_1(x), m_2(x))^\tau$ be a bivariate vector; and $m_\theta(x) = (m_{1,\theta}(x), m_{2,\theta}(x))^\tau$ be a vector of some parametric alternatives, where $\theta \in \Theta \subset R^q$.

The interest of this paper is to test

$$\begin{aligned} \mathcal{H}_0: \ m_1(x) &= m_{1\theta}(x) \quad \text{and} \\ m_2(x) &= m_{2\theta}(x) \end{aligned} \tag{2.3}$$

for some $\theta \in \Theta$ against

$$\begin{aligned} \mathcal{H}_1: \ m_1(x) &= m_{1\theta}(x) + C_{1T}\Delta_{1T}(x) \quad \text{and} \\ m_2(x) &= m_{2\theta}(x) + C_{2T}\Delta_{2T}(x), \end{aligned} \tag{2.4}$$

where

- $\Delta_{1T}(x)$ and $\Delta_{2T}(x)$ are continuous and bounded functions over R^d; and
- $0 \le C_{iT} \le 1$ and $\lim_{T\to\infty} C_{iT} = 0$ for $i = 1, 2$.

Note that the above hypotheses are equivalent to

$$\begin{aligned} \mathcal{H}_0: \ m(x) &= m_\theta(x) \quad \text{versus} \\ \mathcal{H}_1: \ m(x) &= m_\theta(x) + C_T\Delta_T(x), \end{aligned}$$

where

- $C_T = (C_{1T}, C_{2T})^\tau$; and
- $\Delta_T(x) = (\Delta_{1T}(x), \Delta_{2T}(x))^\tau$.
- This contains the parametric case where $\Delta_T(\cdot) \equiv 0$.
- Let $\theta_0 \in \Theta$ denote the true value of θ if $\mathcal{H}_0$ is true. That is, $m(x) = m_{\theta_0}(x)$ for all $x \in S$ if $\mathcal{H}_0$ is true.

- Let K be a d-dimensional bounded probability density function with

$$\int uK(u)du = 0 \text{ and } \int uu^\tau K(u)du = \sigma_K^2 \mathcal{I}_d,$$

  where $\mathcal{I}_d$ is the d-dimension identity matrix and σ_K^2 is a positive constant.
- Let
$$R(K) = \int K^2(x)dx.$$
- Let h be a smoothing bandwidth satisfying $h \to 0$ and $Th^d \to \infty$ as $T \to \infty$.
- This paper considers only the case of $1 \le d \le 3$ due to the curse of dimensionality when $d \ge 4$.

- Define $K_h(u) = h^{-d}K(u/h)$. The Nadaraya-Watson (NW) estimators of $m_l(x)$ for $l = 1, 2$ are defined by

$$\widehat{m}_1(x) = \frac{\sum_{t=1}^T K_h(x - X_t)Y_t}{\sum_{t=1}^T K_h(x - X_t)}, \tag{2.5}$$
$$\widehat{m}_2(x) = \frac{\sum_{t=1}^T K_h(x - X_t)(Y_t - \widehat{m}_1(X_t))^2}{\sum_{t=1}^T K_h(x - X_t)}.$$

- This paper also considers using the only one smoothing parameter h. One can use two different bandwidth parameters h_1 and h_2 for $l = 1$ and $l = 2$ respectively. See Chen and Gao (2003).

- Similarly, for the parametric forms, one can estimate $m_{l,\theta}$ by

$$\widetilde{m}_{l,\tilde{\theta}}(x) = \frac{\sum_{t=1}^T K_h(x - X_t)m_{l,\tilde{\theta}}(X_t)}{\sum_{t=1}^T K_h(x - X_t)} \tag{2.6}$$

  for $l = 1, 2$, where $\tilde{\theta}$ is a consistent estimator of θ under $\mathcal{H}_0$.

- $\widehat{m}(x) = (\widehat{m}_1(x), \widehat{m}_2(x))^\tau$; and
- $\widetilde{m}_\theta(x) = (\widetilde{m}_{1,\theta}(x), \widetilde{m}_{2,\theta}(x))^\tau$.
- $\epsilon_t = Y_t - m_1(X_t)$ and $\eta_t = \epsilon_t^2 - m_2(X_t)$;
- $\sigma_{ij}(x) = E\left[\epsilon_t^i \eta_t^j | X_t = x\right]$ for $i, j = 0, 1, 2$;
- $$\Sigma_0(x) = \begin{pmatrix} \sigma_{20}(x) & \sigma_{11}(x) \\ \sigma_{11}(x) & \sigma_{02}(x) \end{pmatrix}.$$
- $s_0(x) = |\Sigma_0(x)|^{-1}$, where $|A|$ is the determinant of A.
- $f(x)$– the marginal density of $\{X_t\}$.
- $\Sigma(x) = f^{-1}(x)\Sigma_0(x)$.

2.1. Test Statistic I

To construct the first class of our test statistics, we have a look at the following null hypothesis:

$$\mathcal{H}_{01}: \ m_1(x) = m_{1\theta}(x) \quad \text{versus} \tag{2.7}$$
$$\mathcal{H}_{11}: \ m_1(x) = m_{1\theta}(x) + C_{1T}\Delta_{1T}(x).$$

To test (2.7), Härdle and Mammen (1993) suggested using the following test statistic

$$\text{HM}_T = (Th^d) \int \left(\widehat{m}_1(x) - \widetilde{m}_{1\tilde{\theta}}(x)\right)^2 \pi(x)dx, \tag{2.8}$$

where $\pi(x)$ is a non-negative weight function satisfying

$$\int \pi(x)dx = 1 \ \text{ and } \ \int \pi^2(x)dx < \infty.$$

It can be shown that under $\mathcal{H}_{01}$

$$\overline{\text{HM}}_T = \frac{\text{HM}_T - \mu_0}{\sigma_{0h}} \to_D N(0, 1) \tag{2.9}$$

as $T \to \infty$, where

- $$\mu_0 = R(K) \int \frac{\sigma^2(x)\pi(x)}{f(x)} dx, \quad \text{and}$$

- $$\sigma_{0h}^2 = 2h^d K^{(4)}(0) \int \left(\frac{\sigma^2(x)\pi(x)}{f(x)} \right)^2 dx,$$

- in which
$$K^{(4)}(0) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} K(y)K(x+y)dy \right)^2 dx.$$

To test (2.4), equation (2.9) thus motivates the use of a test statistic of the form

$$N_{1T}(h) = (Th^d) \int \{\widehat{m}(x) - \widetilde{m}_{\widehat{\theta}}(x)\}^{\tau} \hat{\Sigma}^{-1}(x) \{\widehat{m}(x) - \widetilde{m}_{\widehat{\theta}}(x)\} \pi(x) dx \tag{2.10}$$

provided that $\hat{\Sigma}^{-1}(x)$ exists, where

$$\begin{aligned}
\hat{\Sigma}^{-1}(x) &= \hat{f}(x)\hat{\Sigma}_0^{-1}(x), \qquad (2.11)\\
\hat{\Sigma}_0(x) &= \begin{pmatrix} \hat{\sigma}_{20}(x) & \hat{\sigma}_{11}(x) \\ \hat{\sigma}_{11}(x) & \hat{\sigma}_{02}(x) \end{pmatrix}, \\
\hat{f}(x) &= \frac{1}{Th^d} \sum_{t=1}^{T} K\left(\frac{x - X_t}{h} \right), \\
\hat{\sigma}_{ij}(x) &= \frac{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right) \hat{\epsilon}_t^i \hat{\eta}_t^j}{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right)}, \\
\hat{\epsilon}_t &= Y_t - \widehat{m}_1(X_t), \\
\hat{\eta}_t &= \hat{\epsilon}_t^2 - \widehat{m}_2(X_t).
\end{aligned}$$

Theorem 2.1. Under suitable conditions, we have under $\mathcal{H}_0$

$$L_{1T}(h) = \frac{N_{1T}(h) - 2}{\sigma_1(h)} \to_D N(0,1) \tag{2.12}$$

as $T \to \infty$, where

$$\sigma_1^2(h) = 4h^d K^{(4)}(0) R^{-2}(K) \int \pi^2(x) dx.$$

- As can be seen from the construction of L_{1T}, random denominators are involved in the form.
- Our experience in finite sample studies suggests that the involvement of such random denominators can reduce the power of the proposed test.
- This motivates the construction of the following test statistic.

2.2. Test Statistic II

To test (2.7), existing studies suggest using

$$N_{21T}(h) = \sum_{s=1}^{T} \sum_{t=1,\neq s}^{T} \hat{U}_t \, p_{st} \, \hat{U}_s, \text{ with } p_{st} = K\left(\frac{X_s - X_t}{h} \right), \tag{2.13}$$

which is a kernel–based sample analogue of the following distance function (see Li 1999):

$$\pi_p(\epsilon) = E\left[\epsilon E\left(\epsilon|X\right) f(X)\right] = E\left[E^2(\epsilon|X) f(X)\right], \tag{2.14}$$

in which $\epsilon = Y - m_{1\theta_0}(X)$ and $f(X)$ is the marginal density of X_t.

To propose a similar test for testing (2.3), we first introduce the following notation:

$$\begin{aligned} Y &= m_{1\theta_0}(X) + \epsilon_t \ \text{ with } \ \epsilon = \sqrt{m_{2\theta_0}(X)}e, \\ \epsilon^2 &= m_{2\theta_0}(X) + \eta \ \text{ with } \ \eta = m_{2\theta_0}(X)(e^2 - 1). \end{aligned} \tag{2.15}$$

Obviously, $E[\epsilon|X] = E[\eta|X] = 0$ under $\mathcal{H}_0$. Let $\xi = \epsilon + \eta$. Then $E[\xi] = 0$ under $\mathcal{H}_0$. Equation (2.15) suggests using a distance function of the form:

$$\begin{aligned} \pi_v(\xi) &= E\left[\xi E\left(\xi|X\right) f(X)\right] \\ &= E\left[\left(E^2(\epsilon|X) + E^2(\eta|X)\right) f(X)\right] \\ &+ 2E\left[E(\epsilon\ \eta|X) f(X)\right]. \end{aligned} \tag{2.16}$$

This would suggest using a kernel–based sample analogue of (2.16) of the form

$$N_{2T}(h) = \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \hat{W}_s\ p_{st}\ \hat{W}_t, \tag{2.17}$$

where $\hat{W}_t = \hat{U}_t + \hat{V}_t$.

When $E[\epsilon\eta|X] = 0$, one may replace $N_{2T}(h)$ by

$$N_{3T}(h) = \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \hat{U}_s\ p_{st}\ \hat{U}_t + \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \hat{V}_s\ p_{st}\ \hat{V}_t, \tag{2.18}$$

(without involving the cross terms $\hat{U}_s\hat{V}_t$ and $\hat{U}_t\hat{V}_s$) which is a kernel–based sample analogue of the following distance:

$$\begin{aligned} \pi_w(\psi) &= E\left[\psi^\tau E\left(\psi|X\right) f(X)\right] \\ &= E\left[\left(E^2(\epsilon|X) + E^2(\eta|X)\right) f(X)\right], \end{aligned}$$

where $\psi = (\epsilon, \eta)^\tau$.

Theorem 2.2. Under suitable conditions, we have under $\mathcal{H}_0$

$$L_{2T} = \frac{N_{2T}(h)}{\sigma_2(h)} \to_D N(0,1) \tag{2.19}$$

as $T \to \infty$, where

$$\sigma_2^2(h) = 2T^2h^d \cdot \int K^2(x)dx \cdot \int \sigma_\xi^4(x) f^2(x)dx,$$

in which $\sigma_\xi^2(x) = E[\xi_t^2|X_t = x]$ and $\xi_t = \epsilon_t + \eta_t$ with

$$\epsilon_t = Y_t - m_{1\theta_0}(X_t) \text{ and } \eta_t = \epsilon_t^2 - m_{2\theta_0}(X_t).$$

This section then suggests using

$$L_2^* = \max_{h \in H_T} L_{2T}(h), \tag{2.20}$$

where H_T is a set of bandwidths satisfying certain conditions.

Simulation Scheme: We now discuss how to obtain a critical value for L_2^*. The exact α–level critical value, l_α^* $(0 < \alpha < 1)$ is the $1 - \alpha$ quantile of the exact finite-sample distribution of L_2^*. Because θ_0 is unknown, l_α^* cannot be evaluated in practice. We therefore suggest choosing a simulated α–level critical value, l_α, by using the following simulation procedure:

1. For each $t = 1, 2, \ldots, T$, generate $Y_t^* = m_{1\tilde{\theta}}(X_t) + \sqrt{m_{2\tilde{\theta}}(X_t)}\epsilon_t$, where $\{\epsilon_t\}$ is sampled independently from a specified distribution with $E\left[\epsilon_t\right] = 0$ and $E\left[\epsilon_t^2\right] = 1$. In Example 3.1 below, we consider the case where the specified distribution is either the standard Normal distribution–$N(0,1)$ or the normalized exponential distribution–Exp(1) $-$ 1.

   It should be noted that under $\mathcal{H}_0$, $m_1(X_t)$ and $m_2(X_t)$ can be specified simultaneously. As a result, $m_2(X_t)$ can be specified parametrically as $m_{2\theta}(X_t)$ under $\mathcal{H}_0$. When only $m_1(X_t)$ is specified, one usually needs to replace $m_2(X_t)$ by a nonparametric estimate.

2. Use the data set $\{Y_t^*, X_t : t = 1, 2, \ldots, T\}$ to estimate θ. Denote the resulting estimate by $\hat{\theta}$. Compute the statistic $\hat{L}_2^*$ that is obtained by replacing Y_t and $\tilde{\theta}$ with Y_t^* and $\hat{\theta}$ on the right–hand side of (2.20).

3. Repeat the above steps M times and produce M versions of $\hat{L}_2^*$ denoted by $\hat{L}_{2m}^*$ for $m = 1, 2, \ldots, M$. Use the M values of $\hat{L}_{2m}^*$ to construct their empirical bootstrap distribution function: $F^*(u) = \frac{1}{M}\sum_{m=1}^{M} I(\hat{L}_{2m}^* \leq u)$. Use the empirical bootstrap distribution function to estimate the asymptotic critical value, l_α.

We now state the following result.

Theorem 2.3. Under suitable conditions, we have under $\mathcal{H}_0$

$$\lim_{T\to\infty} P(L_2^* > l_\alpha) = \alpha.$$

This result shows that l_α is an asymptotically correct α–level critical value under any model in $\mathcal{H}_0$.

Let $\rho(\mathcal{H}_0, \mathcal{H}_1)$ define the distance between $\mathcal{H}_0$ and $\mathcal{H}_1$. The following result shows that a consistent test will reject a false $\mathcal{H}_0$ with probability approaching one as $T \to \infty$.

Theorem 2.4. Assume that the conditions of Theorem 2.3 hold. In addition, if there is some $C_0 > 0$ such that $\lim_{T\to\infty} P\left(\rho(\mathcal{H}_0, \mathcal{H}_1) \geq C_0\right) = 1$ holds, then

$$\lim_{T\to\infty} P(L_2^* > l_\alpha) = 1.$$

3. An example of implementation

Example 3.1. Consider a nonlinear time series model of the form

$$\begin{aligned} Y_t &= \alpha + \beta X_t + \sigma \cdot \sqrt{1 + 0.5X_t^2} \cdot e_t, \\ X_t &= 0.5X_{t-1} + \epsilon_t,\ 1 \leq t \leq T, \end{aligned} \tag{3.1}$$

where

- α, β and $\sigma > 0$ are unknown parameters;
- $\{\epsilon_t : t \geq 1\}$ and $\{e_t : t \geq 1\}$ are i.i.d. and independent of X_0;
- $\epsilon_t \sim U(-0.5, 0.5)$, $X_0 \sim U(-1, 1)$; and
- $\{e_t\}$ is either the standard $N(0, 1)$ or the normalized exponential Exp(1) − 1 error, which has mean zero and variance one.

In the following detailed simulation, we consider using a class of alternatives of the form

$$\begin{aligned} Y_t &= \alpha + \beta X_t + \frac{1}{\psi}\phi(X_t/\psi) \\ &+ \left(\sigma \cdot \sqrt{1 + 0.5X_t^2} + \frac{1}{\psi}\phi(X_t/\psi)\right) e_t, \end{aligned} \tag{3.2}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $\psi \neq 0$ is to be chosen.

The vector of unknown parameters, $\theta = (\alpha, \beta, \sigma)$, involved in (3.1) was then estimated using the pseudo–maximum likelihood method, which is quite common in the estimation of parametric ARCH models. We choose the following weight function and the kernel function given by

$$\pi(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

and

$$K(x) = \begin{cases} \frac{15}{16}(1 - x^2)^2 & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

With the choice of $\pi(\cdot)$ and $K(\cdot)$ in (3.3) and (3.4), the constant $C(K, \pi)$ involved in $L_{1T}(h)$ is $\frac{93}{10}$. In order to calculate both $L_{1T}(h)$ and $L_{2T}(h)$, one needs to find suitable h values, which are chosen by the following simulation procedure:

- For the simulation, we start with some initial values for θ_0 and X_0.
- For each $t = 1, 2, \ldots, T$, generate the data (X_t, Y_t) from (3.2).
- Use the data set $\{(Y_t, X_t) : t = 1, 2, \ldots, T\}$ to estimate θ. Denote the resulting estimate by $\tilde{\theta}$. For each fixed h, compute the resulting function of h given by

  $$\widehat{L}_1(h) = \frac{\widehat{N}_{1T}(h) - 2}{\sqrt{\frac{186h}{5}}}.$$

- Repeat the above steps $M = 1000$ times and produce M versions of $\hat{L}_1(h)$ denoted by $\hat{L}_{1m}(h)$ for $m = 1, 2, \ldots, M$. Use the M functions of h, $\hat{L}_{1m}(h)$ for $m = 1, 2, \ldots, M$, to construct their empirical bootstrap distribution function, that is,

  $$F_{1h}(u) = \frac{1}{M} \sum_{m=1}^{M} I(\hat{L}_{1m}(h) \leq u),$$

  where $I(U \leq u)$ is the usual indicator function.

- For the given empirical value $l_{0.05} = 1.65$, one can calculate the following power function

$$\phi_1(h) = 1 - F_{1h}(l_{0.05}).$$

- Find approximately at which h value the power function $\phi_1(h)$ is maximized. Denote the maximizer by h^*. Similarly, one can find the maximizer, h_*, of the corresponding power function $\phi_2(h)$ for

$$\widehat{L}_2(h) = \frac{\sum_{t=1}^T \left(\sum_{s=1,\neq t}^T p_{st}\hat{W}_s\right)\hat{W}_t}{\hat{\sigma}_2(h)},$$

where $\hat{\sigma}_2^2(h) = 2\sum_{t=1}^T \sum_{s=1}^T p_{st}^2 \hat{W}_t^2 \hat{W}_s^2$.

We now can calculate the following test statistic

$$L_1^* = \widehat{L}_1(h^*). \tag{3.5}$$

Similarly, we can compute

$$L_2^* = \widehat{L}_2(h_*). \tag{3.6}$$

We suggest choosing two simulated 5%–level critical values, $l_{1,0.05}$ and $l_{2,0.05}$, by using the following simulation procedure:

- For the simulation, we start with some initial values θ_0 and X_0.
- For each $t = 1, 2, \ldots, T$, generate the data (X_t, Y_t) from model (3.1).
- Use the data set $\{(Y_t, X_t) : t = 1, 2, \ldots, T\}$ to estimate θ. Denote the resulting estimate by $\tilde{\theta}$. For the chosen H_T, compute the statistics L_1^* and L_2^* given by (3.5) and (3.6).
- Repeat steps 2–3 $M = 1000$ times and produce M versions of L_1^* and L_2^* denoted by L_{1m}^* and L_{2m}^* for $m = 1, 2, \ldots, M$. Use the M values of L_{1m}^* and L_{2m}^* to construct their empirical bootstrap distribution functions, that is, $F_i^*(u) = \frac{1}{M}\sum_{m=1}^M I(L_{im}^* \leq u)$ for $i = 1, 2$. Use the empirical bootstrap distribution functions to calculate the two bootstrap simulated critical values, $l_{1,0.05}$ and $l_{2,0.05}$.

For each case where both ψ and T are chosen, we can compute the rejection rates. The number of simulations in producing Tables 3.1 and 3.2 below was 1000. The detailed results are given in Tables 3.1 and 3.2 below.

Table 3.1. Rejection Rates for the Simultaneous Tests at the 5% level

Normal Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_1^*	L_2^*
	250	0.054	0.060
	500	0.063	0.056
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_1^*	L_2^*
10	250	0.551	0.723
10	500	0.776	1.000
25	250	0.357	0.533
25	500	0.691	0.866
Normalized Exponential Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_1^*	L_2^*
	250	0.049	0.062
	500	0.053	0.058
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_1^*	L_2^*
10	250	0.679	0.887
10	500	0.847	1.000
25	250	0.462	0.667
25	500	0.717	0.933

Remark 3.1.

- As can be seen from Table 3.1, for the standard Normal error the power can be close to one when $T = 500$ and the value of ψ^{-1} is between 4% and 10%.
- This may show that L_2^* is not only asymptotically optimal but also practically applicable to both the small and medium sample cases, since the differences between $\mathcal{H}_0$ and $\mathcal{H}_1$ were made deliberately close.
- We also computed the power of the tests for the case where $\psi = 1$ or 0.25, our small sample results showed that the power of L_2^* was already 100% even when $T = 250$.
- In Table 3.2, we have provided some small sample results for the case where the error is the normalized exponential random variable. The results show that the power of L_2^* is uniformly higher than that for the standard $N(0,1)$ case.
- Table 3.1 also shows that L_2^* is more powerful than L_1^*.

To test $\mathcal{H}_{01} : m_1(x) = m_{1\theta_0}(x)$, existing studies suggest using

$$L_{21T}(h) = \frac{\sum_{s=1}^{T} \sum_{t=1,\neq s}^{T} p_{st} \hat{U}_t \hat{U}_s}{S_{21T}}, \tag{3.7}$$

where $S^2_{21T} = 2\sum_{t=1}^T \sum_{s=1}^T p^2_{st}\hat{U}^2_s\hat{U}^2_t$.

We also propose testing $\mathcal{H}_{02} : m_2(x) = m_{2\theta_0}(x)$ by

$$L_{22T}(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T p_{st}\hat{V}_s\hat{V}_t}{S_{22T}}, \tag{3.8}$$

where $S^2_{22T} = 2\sum_{t=1}^T \sum_{s=1}^T p^2_{st}\hat{V}^2_s\hat{V}^2_t$.

- When there is a model specification problem with only either the conditional mean or the conditional variance,
- but it is difficult to determine which one may have caused the model specification problem,
- it would be interesting to know whether there would be any significant reduction of the power when using L^*_2.

Table 3.2. Rejection Rates for Testing the Conditional Mean at the 5% level

Normal Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L^*_2	L^*_{21}
	250	0.052	0.059
	500	0.047	0.054
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L^*_2	L^*_{21}
40	250	0.198	0.267
40	500	0.401	0.478
25	250	0.602	0.667
25	500	0.827	0.866
Normalized Exponential Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L^*_2	L^*_{21}
	250	0.047	0.053
	500	0.057	0.049
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L^*_2	L^*_{21}
40	250	0.362	0.404
40	500	0.617	0.733
25	250	0.643	0.679
25	500	0.933	1.000

Table 3.3. Rejection Rates for Testing the Conditional Variance at the 5% level

Normal Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_2^*	L_{22}^*
	250	0.052	0.046
	500	0.061	0.058
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_2^*	L_{22}^*
40	250	0.193	0.264
40	500	0.467	0.591
25	250	0.278	0.376
25	500	0.593	0.632
Normalized Exponential Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_2^*	L_{22}^*
	250	0.051	0.055
	500	0.047	0.059
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_2^*	L_{22}^*
40	250	0.267	0.404
40	500	0.523	0.732
25	250	0.309	0.443
25	500	0.764	0.898

4. Future Work

Extension to nonstationarity.

Nonstationary cases are much more difficult but important. For example, a nonlinear random walk of the form

$$Y_t = Y_{t-1} + g(Y_{t-1}) + \sigma(Y_{t-1})\epsilon_t \tag{4.1}$$

is nonstationary. See Granger, *et al* (1997): *J. Econometrics.*

It would be interesting to test whether

$$H_0: \; g(\cdot) = 0 \;\text{ and }\; \sigma(\cdot) = \sigma_0.$$

In other words, before using model (4.1) one should test whether a random walk model of the form $Y_t = Y_{t-1} + e_t$ would still be useful in modelling some nonlinear and nonstationary time series.

Consider a nonparametric cointegration of the form

$$\begin{aligned} Y_t &= m(X_t) + \sigma(X_t)e_t, \\ X_t &= X_{t-1} + u_t, \end{aligned} \tag{4.2}$$

where $m(\cdot)$ and $\sigma(\cdot) > 0$ are both unknown to be specified.

It would then be interesting to test

$$H_0: \ m(X_t) = m(X_t, \theta_0) \ \text{ and } \ \sigma(X_t) = \sigma(X_t, \theta_0).$$

5. By–products

Lemma A.1. Let ξ_t be a r-dimensional strictly stationary and α-mixing stochastic process. Let $\phi(\cdot,\cdot)$ be a symmetric Borel function defined on $R^r \times R^r$. Assume that for any fixed $x \in R^r$, $E[\phi(\xi_1, x)] = 0$ and $E[\phi(\xi_i, \xi_j)|\Omega_0^{j-1}] = 0$ for any $i < j$, where Ω_i^j denotes the σ–field generated by $\{\xi_s : i \le s \le j\}$. Let $\phi_{st} = \phi(\xi_s, \xi_t)$,

$$\sigma_{st}^2 = \mathrm{var}(\phi_{st}) \text{ and } \sigma_T^2 = \sum_{1 \le s < t \le T} \sigma_{st}^2.$$

For some small constant $0 < \delta < 1$, let

$$M_{T1} = \max_{1 \le i < j < k \le T} \max \left\{ E|\phi_{ik}\phi_{jk}|^{1+\delta} \right\},$$

$$M_{T2} = \max_{1 \le i < j < k \le T} \max \left\{ E|\phi_{ik}\phi_{jk}|^{2(1+\delta)} \right\}.$$

Assume that all the $M_T's$ are finite. Let

$$M_T = \max \left\{ T^2 M_{T1}^{\frac{1}{1+\delta}} \right\}$$

and

$$N_T = \max \left\{ T^{\frac{3}{2}} M_{T2}^{\frac{1}{2(1+\delta)}} \right\}.$$

If as $T \to \infty$

$$\frac{\sigma_T^2}{T} \to 0 \ \text{ and } \ \frac{\max\{M_T, N_T\}}{\sigma_T^2} \to 0,$$

then as $T \to \infty$

$$\frac{1}{\sigma_T} \sum_{1 \le s < t \le T} \phi(\xi_s, \xi_t) \to_D N(0, 1).$$

Case study: As an application, one can show that as $T \to \infty$

$$L_T = \frac{e^\tau P e - E[e^\tau P e]}{\sqrt{\mathrm{var}(e^\tau P e)}} \to N(0, 1),$$

where e is a random vector and P is a random matrix.

When $P = X(X^\tau X)^{-1} X^\tau$ is a non-random matrix with $\mathrm{tr}(P) = q \to \infty$ as $T \to \infty$ and $e \sim N(0, I)$, one can have

$$L_T = \frac{\chi_q^2 - q}{\sqrt{2q}} \to N(0, 1)$$

as $T \to \infty$.

Lemma B.1. (i) Let $\psi(\cdot,\cdot,\cdot)$ be a symmetric Borel function defined on $R^r \times R^r \times R^r$. Let the process ξ_i be defined as in Lemma A.1. Assume that for any fixed $x, y \in R^r$, $E[\psi(\xi_1, x, y)] = 0$. Then

$$E \left\{ \sum_{1 \le i < j < k \le T} \psi(\xi_i, \xi_j, \xi_k) \right\}^2 \le C T^3 M_1^{\frac{1}{1+\delta}},$$

where $0 < \delta < 1$ is a small constant, $C > 0$ is a constant independent of T and the function ψ, and

$$M_1 = \max_{1<i<j\leq T} \max \left\{ E|\psi(\xi_1, \xi_i, \xi_j)|^{2(1+\delta)} \right\}.$$

(ii) Let $\phi(\cdot, \cdot)$ be a symmetric Borel function defined on $R^r \times R^r$. Let the process ξ_i be defined as in Lemma B.1. Assume that for any fixed $x \in R^r$, $E[\phi(\xi_1, x)] = 0$. Then

$$E \left\{ \sum_{1\leq i<j\leq T} \phi(\xi_i, \xi_j) \right\}^2 \leq CT^2 M_2^{\frac{1}{1+\delta}},$$

where $\delta > 0$ is a constant, $C > 0$ is a constant independent of T and the function ϕ, and

$$M_2 = \max_{1<i<j\leq T} \max \left\{ E|\phi(\xi_1, \xi_i)|^{2(1+\delta)} \right\}.$$