THE UNIVERSITY OF CHICAGO
Booth School of Business
Business 41914, Spring Quarter 2017, Mr. Ruey S. Tsay

Lecture 8: High-Dimensional Time Series (HDTS) Analysis

High-dimensional statistical analysis has attracted mucn interest in recent years. See Lasso regression and its variants. However, most studies in high-dimensional statistical analysis assume that the data are uncorrelated. For instance, many articles concerning Lasso regression assume that the data consist of independent observations. On the other hand, real applications often involve correlated data or observations that are dependent over time or in space. The goal of high-dimensional time series analysis is to extend the modern statistical analysis to dynamically dependent data. The dependence is either in time or in space or both. That is, HDTS analysis deals with *spatio-temperal* data. More specifically, HDTS analysis considers $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{kt})'$ for $t = 1, \ldots, T$ such that $\min\{k, T\} \to \infty$.
The research in HDTS is still in its infancy. There are many interesting research problems waiting to be explored. In this lecture, we provide some recent developments and discuss some open questions for further study. We start with a simple simulation demonstrating that the Lasso regression may fail in the presence of strong serial correlations in the data.

Example 1. Consider the AR(3) model

$$x_t = 1.9x_{t-1} - 0.8x_{t-2} - 0.1x_{t-3} + a_t, \quad a_t \sim_{iid} N(0, 1). \tag{8.1}$$

We generated 2,000 observations from the AR(3) model of Equation (8.1). Figure 8.1 shows the time plot of the data. The data generating model can be recovered by an AR(3) fit via the ordinary least squares method. See R output below.

```
> m4 <- lm(zt[4:2000]~-1+zt[3:1999]+zt[2:1998]+zt[1:1997])
> summary(m4)
Call: lm(formula = zt[4:2000] ~ -1 + zt[3:1999] + zt[2:1998] + zt[1:1997])

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
zt[3:1999]  1.94480    0.02237  86.925   <2e-16 ***
zt[2:1998] -0.88907    0.04477 -19.857   <2e-16 ***
zt[1:1997] -0.05574    0.02242  -2.486    0.013 *
---
Residual standard error: 0.9974 on 1994 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 3.626e+11 on 3 and 1994 DF,  p-value: < 2.2e-16

> Box.test(m4$residuals,lag=10,type="Ljung")
       Box-Ljung test
data:  m4$residuals
X-squared = 3.8571, df = 10, p-value = 0.9536
```

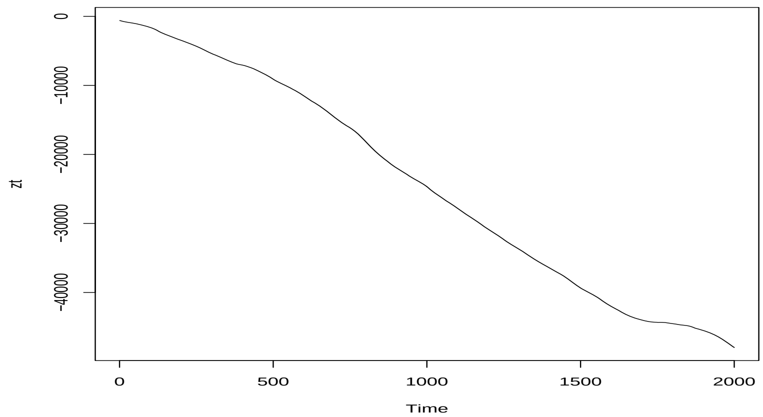

Figure 8.1: Time plot of a simulated time series with strong serial dependence

On the other hand, Figure 8.2 shows the result of the `lars` command of `lars` package. The result indicates that the lag-3 dependence is not recovered. This is not surprising because the AR(3) model has strong serial correlations. [The `glmnet` package also requires multiple iterations to see the data generating model.]
The goal of this simulation example is simply to demonstrate that strong serial correlations in the data may cause some difficulties for statistical methods developed for independent data. Therefore, there is a need to understand the impact of serial correlations on the Lasso regression and to develop methods that can deal with serial dependence in the data.

8.1 Testing for serial dependence

For multivariate time series, i.e. the dimension k is fixed, we can apply the multivariate Ljung-Box statistics QM(m) to check for serial dependence in the data. The performance of the QM(m) statistics, however, deteriorates quickly as k increases. New testing statistics are needed for the HDTS analysis. We shall mention three new methods in this lecture. Two of them were developed by Qiwei Yao of LSE and his associates. The other method is part of my recent work, Tsay (2017).

1. *On testing a high-dimensional white noise* by Li, Yao, Lam and Yao (2016). The paper is available from Prof. Yao's web at

   `http://stats.lse.ac.uk/q.yao/yao.links/publicationsHighDTS.html`

2. *Testing for high-dimensional white noise using maximum cross correlations* by Chang and Yao (2016, to appear in Biometrika). Also available on Prof. Yao's web page.

3. *Testing for serial correlations in high-dimensional time series via extreme value theory* by Tsay (2017).

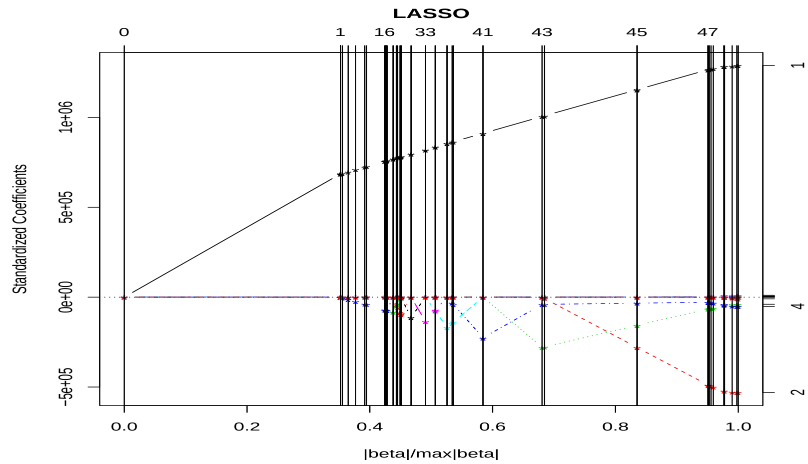

Figure 8.2: Results of `lars` command in R for serially correlated data.

In what follows, I briefly describe each method. You should read the papers more carefully for details. The basic series considered is

$$\boldsymbol{z}_t = \boldsymbol{\Sigma}^{1/2} \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{a}_{t-i}, \tag{8.2}$$

where $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{kt})'$ is a k-dimensional stationary linear time series. The $\boldsymbol{a}_t$ series are assumed to be a sequence of independent k-dimensional random vectors with mean zero and the coefficient matrices satisfy that $\sum_{i=0}^{\infty} \|\boldsymbol{\psi}_i\| < \infty$. This basically is the linear time series we discussed in the lectures. Different methods put additional assumptions on $\boldsymbol{a}_t$.
Let $\boldsymbol{\Gamma}_\ell$ be the lag-ℓ cross-covariance matrix of $\boldsymbol{z}_t$, i.e.

$$\boldsymbol{\Gamma}_\ell = E(\boldsymbol{z}_t \boldsymbol{z}_{t-\ell}), \quad \ell = 0, 1, \ldots$$

Two types of null hypotheses are often considered. The first null hypothesis is that $\boldsymbol{z}_t$ has no lag-ℓ serial correlations for a fixed ℓ. The second null hypothesis is that $\boldsymbol{z}_t$ has no serial correlations for $\ell = 1, \ldots, m$ for some fixed m. We shall refer the test statistics for these two null hypotheses as the single and joint test, respectively.
Single test: In Paper 1, the authors further assume that $E(a_{it}^2) = 1$ and $E(a_{it}^4) < \infty$. The sample cross-covariance matrix is

$$\widehat{\boldsymbol{\Gamma}}_\ell = \frac{1}{T} \sum_{t=\ell+1}^{T} \boldsymbol{z}_t \boldsymbol{z}_{t-\ell}.$$

Since $\widehat{\boldsymbol{\Gamma}}_\ell$ is no symmetric in general, the authors employ

$$\widehat{\boldsymbol{M}}_\ell = \frac{1}{2}(\widehat{\boldsymbol{\Gamma}}_\ell + \widehat{\boldsymbol{\Gamma}}'_\ell). \tag{8.3}$$

Under the null hypothesis, $E(\widehat{\boldsymbol{M}}_\ell) = \boldsymbol{0}$. The test statistic is

$$\phi_\ell = \frac{T}{k} L_\ell - \frac{k}{2}, \quad \text{with} \quad L_\ell = \sum_{j=1}^{k} \lambda_{j,\ell}^2 = tr(\widehat{\boldsymbol{M}}'_\ell \widehat{\boldsymbol{M}}_\ell), \tag{8.4}$$

where $\lambda_{j,\ell}$ are the eigenvalues of $\widehat{\boldsymbol{M}}_\ell$. Ideally, L_ℓ should be zero under the null hypothesis. Therefore, the null hypothesis is rejected for large values of ϕ_ℓ.
To study the limiting properties of ϕ_ℓ, the authors employ the so-called *Marcenko-Pastur regime* in which $c_k = \frac{k}{T} \to c > 0$ when $k, T \to \infty$.
Theorem 2.1 of Paper 1: Let $\ell \geq 1$ be a fixed integer, and assume that

1. $\{\boldsymbol{a}_t\}$ are independently distributed satisfying $E(a_{it}) = 0$, $E(a_{it}^2) = 1$ and $E(a_{it}^2) = v_4 < \infty$.
2. $c_k = \frac{k}{T} \to c > 0$ when $k, T \to \infty$.

Then, if $\boldsymbol{z}_t = \boldsymbol{a}_t$ (white noise), then

$$\phi_\ell \to_d N\left(\frac{1}{2}, 1 + \frac{3(v_4 - 1)}{2} c\right).$$

Theorem 2.2 of Paper 1: Let $\ell \geq 1$ be a fixed integer, and assume that

1. $\{\boldsymbol{a}_t\}$ are independently distributed satisfying $E(a_{it}) = 0$, $E(a_{it}^2) = 1$ and $E(a_{it}^2) = v_4 < \infty$.
2. $k/T \to 0, k^3/T = O(1)$ when $k, T \to \infty$.

Then, if $\boldsymbol{z}_t = \boldsymbol{a}_t$ (white noise), then

$$\phi_\ell \to_d N\left(\frac{1}{2}, 1\right).$$

Obviously, both Theorems can be combined as Theorem 2 can be regarded as a special case of Theorem 1 with $c = 0$. In practice, one can estimate c by k/T. For a given type-I error α, one rejects the null hypothesis if $\phi_\ell \geq \frac{1}{2} + z_{1-\alpha/2}\sqrt{1 + \frac{3(\hat{v}_4 - 1)}{2}\frac{k}{T}}$.

Remark: In practice, $E(a_{it}^2) = \sigma_i^2$. In this case, one can standardize each component a_{it} by its sample standard deviation. In particular, if $\boldsymbol{z}_t = \boldsymbol{a}_t$, then the standardized the series by sample standard errors.

<u>Joint test</u>: Consider the joint test of $H_0 : \boldsymbol{\Gamma}_1 = \cdots = \boldsymbol{\Gamma}_m = \boldsymbol{0}$. Paper 1 employs a non-overlapping $k(m+1)$-dimensional time series

$$\boldsymbol{y}_j = \begin{bmatrix} \boldsymbol{z}_{j(m+1)-m} \\ \boldsymbol{z}_{j(m+1)-m+1} \\ \vdots \\ \boldsymbol{z}_{j(m+1)} \end{bmatrix}, \quad j = 1, \ldots, N,$$

where $N = \lfloor \frac{T}{m+1} \rfloor$ is the largest integer less than or equal to $T/(m+1)$. The covariance matrix of $\boldsymbol{y}_j$ is

$$\mathrm{cov}(\boldsymbol{y}_j) = \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1' & \cdots & \boldsymbol{\Gamma}_m' \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_{m-1}' \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_m & \boldsymbol{\Gamma}_{m-1} & \cdots & \boldsymbol{\Gamma}_0 \end{bmatrix}_{(m+1)k \times (m+1)k}.$$

Under the null hypothesis, $\mathrm{cov}(\boldsymbol{y}_j)$ is a block-diagonal matrix. The paper proposes to use the test statistic (which is referred to as John's test)

$$U_m = \frac{\frac{1}{k(m+1)}\sum_{i=1}^{k(m+1)}(\lambda_{i,m} - \bar{\lambda}_m)^2}{\bar{\lambda}_m^2}, \tag{8.5}$$

where $\lambda_{i,m}$ are the eigenvalues of the sample covariance matrix of $\boldsymbol{y}_j$ and $\bar{\lambda}_m$ is the mean of all the eigenvalues. The sample covariance matrix of $\boldsymbol{y}_j$ is defined as

$$\boldsymbol{S}_m = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{y}_j \boldsymbol{y}_j'.$$

Theorem 3.1. Case of $\boldsymbol{\Gamma}_0 = \sigma^2 \boldsymbol{I}_k$. Let $m \geq 1$ be a fixed integer. Assume that

1. a_{it} are all independently distributed satisfying $E(a_{it}) = 0$, $E(a_{it}^2) = 1$ and $E(a_{it}^2) = v_4 < \infty$.

2. $k, T \to \infty$, $c_k = \frac{k}{T} \to c \geq 0$.

Then, if $\boldsymbol{z}_t = \boldsymbol{a}_t$, we have

$$NU_m - k(m+1) \to_d N(v_4 - 2, 4),$$

where $N = \lfloor \frac{T}{m+1} \rfloor$.
Remark: To make use the full sample, the paper also considers sub-sampling in constructing $\boldsymbol{y}_j$. For instance, starting with $t = 2$ in defining $\boldsymbol{y}_j$. This results in $m + 1$ possible series of $\boldsymbol{y}_j$ to perform the test.
Turn to the general covariance matrix $\boldsymbol{\Gamma}_0$. Paper 1 uses a result of Srivastava (2005) that depends on the normality assumption. Let $b_i = \frac{1}{k(m+1)} tr(\boldsymbol{\Gamma}_y^i)$ and

$$\widehat{U}_m = \frac{\frac{1}{k(m+1)} tr(\widehat{\boldsymbol{S}}_m^2)}{[\frac{1}{k(m+1)} tr(\widehat{\boldsymbol{S}}_m)]^2},$$

where $\widehat{\boldsymbol{S}}_m = \frac{1}{N-1}\sum_{j=1}^{N}(\boldsymbol{y}_j - \bar{\boldsymbol{y}})(\boldsymbol{y}_j - \bar{\boldsymbol{y}})'$ with $\bar{\boldsymbol{y}} = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{y}_j$.

Proposition 3.1. Let $m \geq 1$ be a fixed integer. Assume that

1. As $k \to \infty$, $b_i \to b_i^o$, $0 < b_i^o < \infty$ for $i = 1, \ldots, 8$.

2. $k, T \to \infty$, $N = O(p^\delta)$, $\quad 0 < \delta \leq 1$. Then,

$$\frac{(N-1)^3}{(N-2)(N+1)}\widehat{U}_m - \frac{k(m+1)(N-1)^2}{(N-2)(N+1)} - \frac{b_2}{b_1^2}(N-1) \to_d N(0, 4\tau_1^2),$$

where $\tau^2 = \frac{2N(b_4 b_1^2 - 2b_1 b_2 b_3 + b_2^3)}{k(m+1)b_1^6} + \frac{b_2^2}{b_1^4}$.

Under H_0, $\text{cov}(\boldsymbol{y}_j) = \text{diag}\{\boldsymbol{\Gamma}_0, \ldots, \boldsymbol{\Gamma}_0\}$ so that $b_i = \frac{1}{k}Tr(\boldsymbol{\Gamma}_0^i)$. Paper 1 defines $\hat{d}_i = \frac{1}{k}tr(\widehat{\boldsymbol{\Gamma}}_0^i)$ and $c_k = \frac{k}{T}$, and considers the estimates

$$
\begin{aligned}
\hat{b}_1 &= d_1 \\
\hat{b}_2 &= d_2 - c_p\hat{d}_1^2 \\
\hat{b}_3 &= d_3 - 3c_kd_1d_2 + 2c_k^2d_1^3 \\
\hat{b}_4 &= d_4 - 4c_kd_1d_3 - 2c_kd_2^2 + 10c_kd_1^2d_2 - 5c_kd_1^4 \\
\hat{\tau}_1^2 &= \frac{2N(\hat{b}_4\hat{b}_1^2 - 2\hat{b}_a\hat{b}_2\hat{b}_3 + \hat{b}_2^3)}{k(m+1)\hat{b}_1^6} + \frac{\hat{b}_2^2}{\hat{b}_1^4}.
\end{aligned}
$$

The null hypothesis is rejected if

$$
\frac{(N-1)^3}{(N-2)(N+1)}\widehat{U}_m > \frac{k(m+1)(N-1)^2}{(N-2)(N+1)} + \frac{\hat{b}_2}{\hat{b}_1^2}(N-1) + 2\hat{\tau}_1 Z_{1-\alpha/2}.
$$

Paper 1 also provides some simulation results demonstrating the performance of the proposed test statistics in finite samples.

Paper 2: Maximum cross-correlation in absolute values
Let $\widehat{\boldsymbol{\Gamma}}_\ell$ be the ℓ-th lag sample cross-covariance matrix of a realization $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T\}$ of a weakly stationary k-dimensional time series $\boldsymbol{z}_t$, and $\widehat{\boldsymbol{\rho}}_\ell$ be the corresponding cross-correlation matrix.
In Paper 2, Chang, Yao and Zhou (2016) define

$$
T_{T,\ell} = \sqrt{T} \times \max_{1 \le i,j \le k} |\hat{\rho}_{\ell,ij}|,
$$

which is the maximum sample cross-correlation, in absolute value, at lag ℓ, and

$$
T_T = \max_{1 \le \ell \le m} T_{T,\ell}, \tag{8.6}
$$

where m is a given fixed positive integer, representing the maximum lag specified for testing. The null hypothesis of no serial and cross correlations is rejected for large values of T_T. The authors provide a bootstrap method to obtain critical values of the test statistics. Specifically, the null hypothesis is rejected if $T_T > v_\alpha$, where the critical value v_α is given by

$$
Pr(T_T > v_\alpha) = \alpha.
$$

To evaluate v_α, Paper 2 considers a multivariate normal random vector, i.e. $N(\boldsymbol{0}, \boldsymbol{\Xi}_T)$, where

$$
\boldsymbol{\Xi}_T = (\boldsymbol{I}_m \otimes \boldsymbol{W})E(\boldsymbol{\xi}_T\boldsymbol{\xi}_T')(\boldsymbol{I}_m \otimes \boldsymbol{W}),
$$

where $\boldsymbol{\xi}_T = \sqrt{T}(\text{vec}(\widehat{\boldsymbol{\Gamma}}_1)', \cdots, \text{vec}(\widehat{\boldsymbol{\Gamma}}_m)')'$, $\boldsymbol{W} = \text{diag}(\widehat{\boldsymbol{\Gamma}}_0)^{-1/2} \otimes \text{diag}(\widehat{\boldsymbol{\Gamma}}_0)^{-1/2}$

Conditions:

(1) There exists a constant $C_1 > 0$ independent of k such that $\text{Var}(z_{it}) \ge C_1$ uniformly holds for any $i = 1, \ldots, k$.

(2) There exist three constants $C_2, C_3 > 0$ and $r_1 \in (0,2]$ independnet of k such that $\sup_t \sum_{1\leq i\leq k} Pr(|z_{it}| > x) \leq C_2 \exp(-C_3 x^{r_1})$ for any $x > 0$.

(3) Assume that $\{\boldsymbol{z}_t\}$ is β-mixing in the sense that $\beta_\ell \equiv \sup_t E[\sup_{B\in F_{t+\ell}^{\infty}} |Pr(B|F_{-\infty}^t) - Pr(B)|] \to 0$ as $\ell \to \infty$, where $F_{-\infty}^u$ and $F_{u+\ell}^{\infty}$ are the σ-fields generated, espectively, by $\{\boldsymbol{z}_t\}_{t\leq u}$ and $\{\boldsymbol{z}_t\}_{t\geq u+\ell}$. Furthermore there exist two constants $C_4 > 0$ and $r_2 \in (0,1]$ independent of k such that $\beta_\ell \leq \exp(-C_4 \ell^{r_2})$ for all $\ell \geq 1$.

(4) There exist constant $C_5 > 0$ and $v > 0$ independent of k such that

$$\begin{aligned} C_5^{-1} &< \lim\inf_{q\to\infty} \inf_{n\geq 0} E\left(\left|\frac{1}{q^{1/2}}\sum_{t=n+1}^{n+q} z_{i,t+\ell} z_{j,t}\right|^{2+v}\right) \\ &\leq \lim\sup_{q\to\infty} \sup_{n\geq 0} E\left(\left|\frac{1}{q^{1/2}}\sum_{t=n+1}^{t=n+q} z_{i,t+\ell} z_{jt}\right|^{2+v}\right) < C_5, \quad i,j = 1,\ldots,k; \ell = 1,\ldots,m. \end{aligned}$$

Proposition 1. Assume Conditions 1-4 hold and $\boldsymbol{G} \sim N(\boldsymbol{0}, \boldsymbol{\Xi}_T)$. There exists a positive constant C depending only on the constants appeared in Conditions 1-4 for which $\ln(p) \leq CT^{\delta_1}$ for some constant $C > 0$. Then it holds under H_0 that

$$\sup_{s\geq 0} |Pr(T_T > s) - Pr(|\boldsymbol{G}|_\infty > s)| \to 0, \quad T \to \infty,$$

where $|\boldsymbol{A}|_\infty$ denotes the maximum elements of the matrix $\boldsymbol{A}$ in modulus.
This proposition enables us to generate many random draws from multivariate normal distribution $N(\boldsymbol{0}, \boldsymbol{\Xi}_T)$. Each random draw provides a value $|\boldsymbol{G}|_\infty$. Therefore, critical values for the test statistic can be obtained.

Paper 3: A relatively simple approach. It is robust and simple. Finite-sample adjustments deserve further study.