25 Spring 439/639 TSA: Lecture 22

Dr Sergey Kushnarev

Table of contents

Spurious correlation (continued)

Last time we mentioned the issue of spurious correlation for vector time series (X_t, Y_t). When sample size n is large, the sampling distribution of the sample CCF $r_m(X,Y)$ is approximately

$$r_m(X,Y) \sim \mathcal{N}\left(\rho_m(X,Y),\ \frac{1}{n}\left(1+2\sum_{k=1}^{\infty}\rho_k(X)\,\rho_k(Y)\right)\right).$$

The variance in this sampling distribution can be non-negligibly larger than $\frac{1}{n}$. If this happens, then the results reported by the default method in software are not not reliable.

We notice that, if one of the series (X_t) (or (Y_t)) is a white noise, then $\rho_k(X) = 0$ for any $k \geq 1$. This makes the variance

$$\frac{1}{n}\left(1+2\sum_{k=1}^{\infty}\rho_k(X)\,\rho_k(Y)\right) = \frac{1}{n}.$$

Idea: if we can transform (X_t) or (Y_t) into a white noise, then we may get rid of spurious correlation. This idea formally leads to the method prewhitening.

Prewhitening

Suppose (X_t) follows an ARMA(p,q)

$$\Phi(B)\ X_t = \Theta(B)\ e_t.$$

Assume the MA part $\Theta(B)$ is invertible, then we have the invertible representation

$$e_t = \Theta(B)^{-1}\ \Phi(B)\ X_t = \Pi(B)\ X_t = \sum_{j=0}^{\infty}\pi_j X_{t-j},$$

where $\Pi(B)$ is called a **prewhitening filter**.

This is the basic idea of Prewhitening: if we apply the prewhitening filter $\Pi(B)$ to (X_t), we can theoretically get a white noise (e_t). Then we can apply it simultaneously to both X_t and Y_t,

$$\Pi(B)\ X_t = e_t, \quad \Pi(B)\ Y_t = \widetilde{Y}_t.$$

(X_t) is transformed in to a white noise (e_t), and (Y_t) is transformed in to a new time series $(\widetilde{Y}_t)$. The dependence between (X_t, Y_t) is preserved in $(e_t, \widetilde{Y}_t)$. Then we can look at the CCF between $(e_t, \widetilde{Y}_t)$, and the previous spurious correlation issue is solved since one of the series is white noise.

In practice, the procedure can be briefly summarized as:

1. Make $(X_t), (Y_t)$ both stationary (by taking difference $\nabla^{d_1}\nabla_{s_1}^{D_1} X_t$, $\nabla^{d_2}\nabla_{s_2}^{D_2} Y_t$).
2. Fit an AR(p) model to (X_t) (choose a large p). Then the fitted AR filter $\Phi(B)$ can be approximately seen as the prewhitening filter $\Pi(B)$.
3. Apply $\Phi(B)$ to (Y_t), to get $\Phi(B)\ Y_t = \widetilde{Y}_t$.
4. Estimate the CCF between $(\widetilde{X}_t, \widetilde{Y}_t)$.

Note: Theoretically, we have $e_t \approx \Phi(B)\ X_t$ in this framework. But in practice, we can only get $\widetilde{X}_t = \Phi(B)\ X_t$, which is the resulted samples of X_t after filtering.

Example. Consider the specific regression model

$$Y_t = \sum_{h=-\infty}^{+\infty} \beta_h X_{t-h} + Z_t \approx \sum_{h=-m_1}^{m_2} \beta_h X_{t-h} + Z_t,$$

where the model assumes (Z_t) is a white noise and (Z_t) is independent of (X_t). The series (X_t) is not necessarily a white noise. The range from $-m_1$ to m_2 may be large, so we want to reduce it to a more "accurate" regression model given the observed data from $(X_t), (Y_t)$.

We use the previous idea of prewhitening. Fit an AR model for (X_t) to get a prewhitening filter $\Pi(B)$. Suppose this filter makes

$$\Phi(B)\ X_t = \widetilde{X}_t, \quad \Phi(B)\ Y_t = \widetilde{Y}_t, \quad \Phi(B)\ Z_t = \widetilde{Z}_t.$$

Then under this model, we should have

$$\widetilde{Y}_t \approx \sum_{h=-m_1}^{m_2} \beta_h \widetilde{X}_{t-h} + \widetilde{Z}_t.$$

Since $(\widetilde{X}_t)$ is approximately a white noise, and $(\widetilde{X}_t)$ is still independent of $(\widetilde{Z}_t)$, the theoretical CCF between $(\widetilde{X}_t)$ and $(\widetilde{Y}_t)$ is

$$\rho_k(\widetilde{X}, \widetilde{Y}) = \mathrm{corr}(\widetilde{X}_t, \widetilde{Y}_{t-k}) \approx \beta_{-k} \frac{\sigma_{\widetilde{X}}}{\sigma_{\widetilde{Y}}}.$$

If β_{-k} is zero, then approximately we have $\rho_k(\widetilde{X}, \widetilde{Y}) = 0$, so the sample CCF $r_k(\widetilde{X}, \widetilde{Y}) \in \left[\pm \frac{2}{\sqrt{n}}\right]$ with 95% probability. (By our earlier analysis, spurious correlation is no longer a issue after prewhitening.)

So we can look at the sample CCF between $(\widetilde{X}_t)$ and $(\widetilde{Y}_t)$. If $r_k(\widetilde{X}, \widetilde{Y}) \notin \left[\pm \frac{2}{\sqrt{n}}\right]$, then β_{-k} is probably nonzero. In practice, we detect all the k such that $r_k(\widetilde{X}, \widetilde{Y})$ is significantly nonzero (outside the interval $\left[\pm \frac{2}{\sqrt{n}}\right]$). If these k's are $k_1, ..., k_l$, then we keep the corresponding lags (be careful of the $-k$) in the original regression model:

$$Y_t \sim \beta_{-k_1} X_{t+k_1} + \beta_{-k_2} X_{t+k_2} + \cdots + \beta_{-k_l} X_{t+k_l}.$$