SW and ADF test (GDP)

Shuofan Zhang

November 23, 2018

Data

Data: 145+1-5=141 series in total.

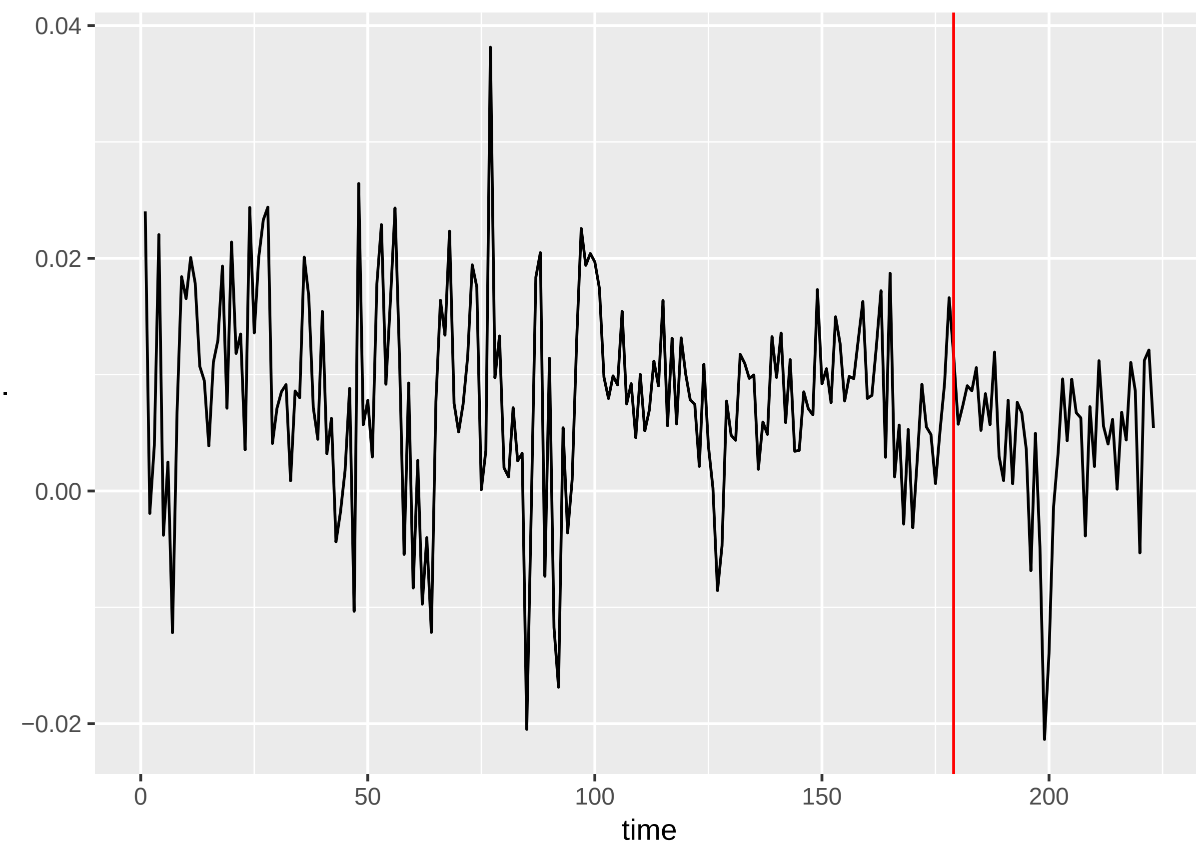
The "spread" series (difference between two I(1) series) are removed.

One of the I(1), "CP3FM" (short name: "Com Paper") was omitted in the original data set, now is added.

log() is done.

y is GDP.

The earliest 80% of data were used to estimate the coefficients, the rest 20% were used to calculate the out-of-sample MSE.



ADF test

Step 1, ADF test to the 146 original series.

Step 2, mark "I(0)" variables as "I(0)".

Step 3, ADF test to the first-differenced 146 series.

Step 4, check for contradictions, found "PCED_RecServices" in AIC.

Step 5, mark "I(1)" variable as "I(2)" (including "PCED_RecServices").

Step 6, mark the rest as "I(1)".

Step 7, repeat the above 6 steps for both "AIC" and "BIC".

AR(p) modeL

- The estimation method is OLS; lags are chosen by AIC.

Lasso 1

$$\begin{aligned}\Delta y_t = y_{t-1}\\ &+ \Delta y_{t-1} + \Delta y_{t-2} + \Delta y_{t-3} + \Delta y_{t-4}\\ &+ I(0)_{t-1} + I(0)_{t-2} + I(0)_{t-3} + I(0)_{t-4}\\ &+ I(1)_{t-1} + I(1)_{t-2} + I(1)_{t-3} + I(1)_{t-4}\\ &+ \Delta I(2)_{t-1} + \Delta I(2)_{t-2} + \Delta I(2)_{t-3} + \Delta I(2)_{t-4}\end{aligned}$$

Lasso 2

$$\begin{aligned}\Delta y_t = y_{t-1}\\ &+ \Delta y_{t-1} + \Delta y_{t-2} + \Delta y_{t-3} + \Delta y_{t-4}\\ &+ I(0)_{t-1} + I(0)_{t-2} + I(0)_{t-3} + I(0)_{t-4}\\ &+ \Delta I(1)_{t-1} + \Delta I(1)_{t-2} + \Delta I(1)_{t-3} + \Delta I(1)_{t-4}\\ &+ \Delta^2 I(2)_{t-1} + \Delta^2 I(2)_{t-2} + \Delta^2 I(2)_{t-3} + \Delta^2 I(2)_{t-4}\end{aligned}$$

Lasso 3

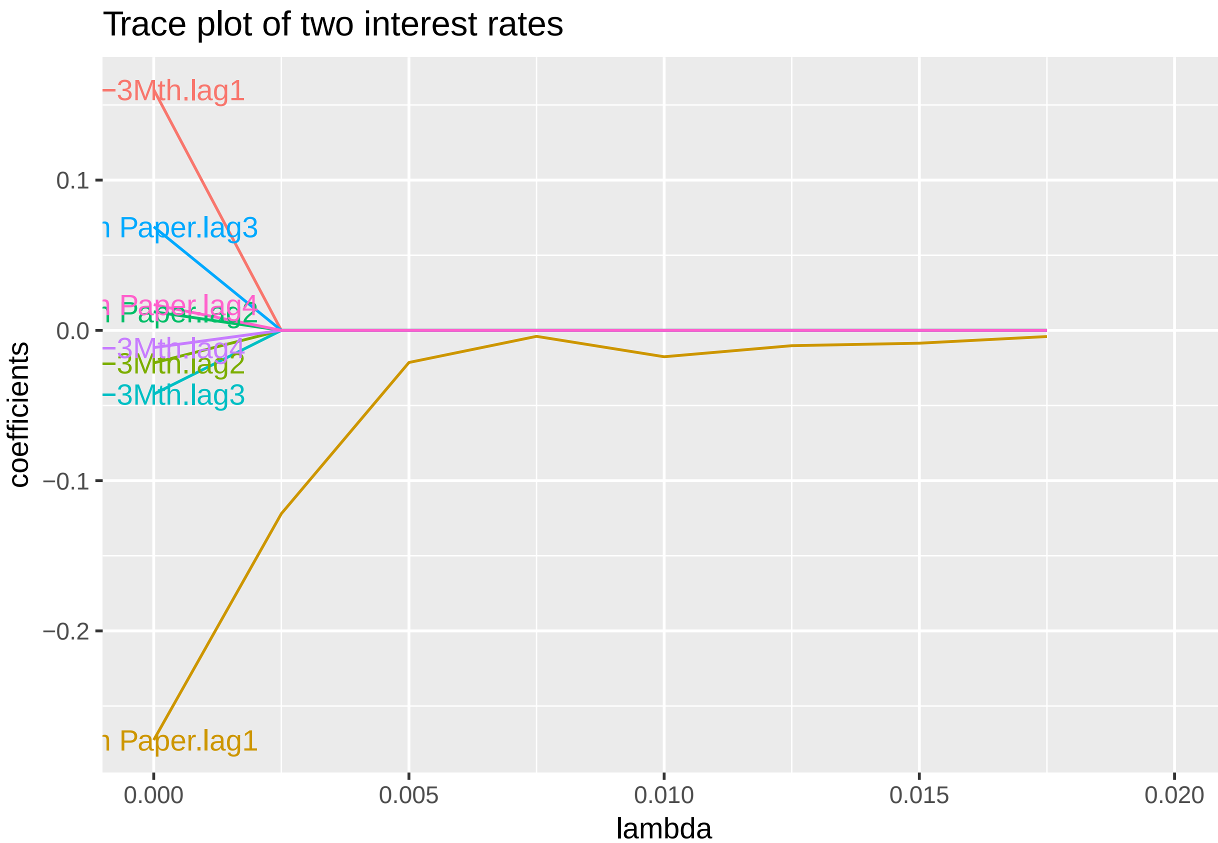
$$\begin{aligned}\Delta y_t = y_{t-1}\\ &+ \Delta y_{t-1} + \Delta y_{t-2} + \Delta y_{t-3} + \Delta y_{t-4}\\ &+ I(0)_{t-1} + I(0)_{t-2} + I(0)_{t-3} + I(0)_{t-4}\\ &+ \Delta I(1)_{t-1} + \Delta I(1)_{t-2} + \Delta I(1)_{t-3} + \Delta I(1)_{t-4}\\ &+ \Delta^2 I(2)_{t-1} + \Delta^2 I(2)_{t-2} + \Delta^2 I(2)_{t-3} + \Delta^2 I(2)_{t-4}\\ &+ I(1)_{t-1} + \Delta I(2)_{t-1}\end{aligned}$$

Table 1: out-of-sample MSE, when the dependent variable is inflation

	SW	AIC	BIC
LASSO 1	0.7607	0.726	0.7059
LASSO 2	0.4971	0.4971	0.4993
LASSO 3	0.4981	0.4981	0.5002

- AIC chooses 2 lags for the GDP growth.
- The out-of-sample MSE of AR(2) is 0.89.

- The estimated AR(1) coefficient is 0.3078.



The trace plot shows the paths of the four lags of the two interest rates that are I(1) variables: "3-month treasury bill (TB-3Mth)" and "3-Month AA Financial Commercial Paper Rate post 1997 ... linked to XLI CP90 before 1997 (Com Paper)".

Previously, when we had the difference of these two variables "CP_Tbill Spread" in the data, it was selected by the LASSO and the absolute value of its coefficient was the biggest among other non-zero coefficients. Since we have removed all the spreads in the data, I try to figure out what is happening to the two "parents" I(1) variables.

From the graph, we see:

1, when $\lambda = 0$, which is the OLS, the magnitude and signs of the estimated coefficients all make sense, for example, lag-1 of TB-3Mth is positive and has a big coefficient, while lag-1 of Com Paper is negative and also has a big absolute coefficient. I think this may indicate that OLS works OK in terms of estimating co-integrating vectors.

2, Once λ becomes non-zero, (in our case, the smallest non-zero $\lambda = 0$ is 0.0025) seven out of eight of these variables are dropped immediately, which is a bit disappointing.

Note: Although this graph only shows the trace of eight variables (two series with four lags), the regression was conducted with all 146 variables.