Lasso

Shuofan Zhang

October 16, 2018

Detailed description

Given the data set we use (Stock & Watson 2016), when thresh=1E-16, the difference in sum of squared errors between OLS and Lasso ($\lambda = 0$) is 1.81e-06, while when thresh=1E-7, the difference is 0.005. But the difference when λ is nonzero is not very significant, so to reduce the computational burden, we choose to use 1E-10 with the maxit (maximum number of passes over the data for all lambda values) to be 10^9.

All the monthly data was aggregated into quarterly data.

Each series was standardized (centered, sd=1) before put into regression.

118 series were transformed by log().

GDP growth rate is used as the dependent variable.

$$y_t = log(GDP_t)$$

$$\Delta y_t = log(GDP_t) - log(GDP_{t-1})$$

Lasso 1

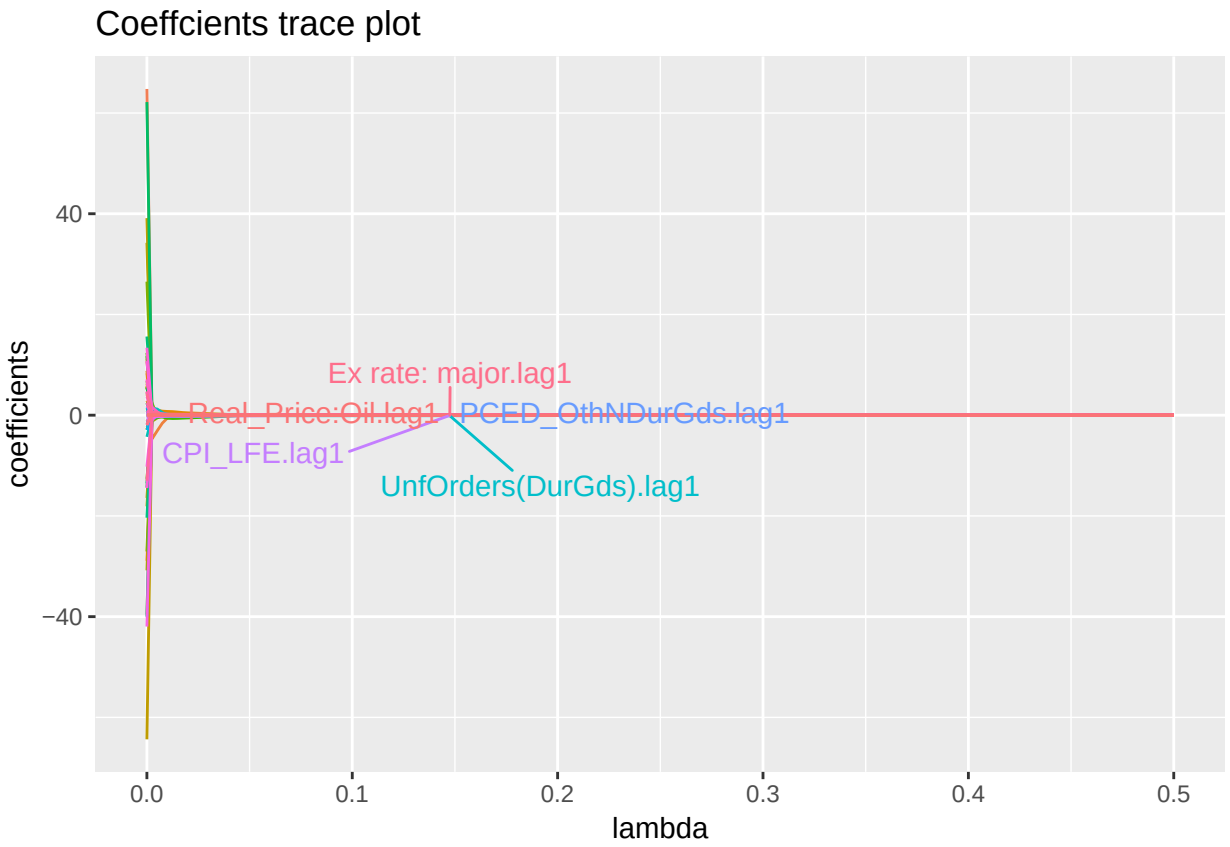


Table 1: Non-zero Coefficients with fixed lambda

variable	value
UnfOrders(DurGds).lag1	-0.0977381
PCED_OthNDurGds.lag1	-0.0487353
CPI_LFE.lag1	-0.0438177
Ex rate: major.lag1	0.0123589
Real_Price:Oil.lag1	-0.0242649

The regression being estimated is:

$$\Delta y_t = I(0)_{t-1} + I(1)_{t-1} + \Delta I(2)_{t-1}$$

1, 144 series were used as explanatory variables;

2, I(2) series were first differenced, no change to I(0) and I(1) series;

3, all explanatory variables are lagged by 1 quarter;

4, after first difference and one lag, we have 222 observations (lost 2).

Lasso 2

The regression being estimated is:

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

1, 149 series were used as explanatory variables, $log(GDP_{t-1})$ $\Delta log(GDP_{t-1})$ $\Delta log(GDP_{t-2})$ $\Delta log(GDP_{t-3})$ $\Delta log(GDP_{t-4})$ were added on the top of the Lasso 1;

2, I(2) series were first differenced;

3, all explanatory variables are lagged by 1 quarter;

4, after first difference and four lags, we have 219 observations (lost 5);

Lasso 3

Some notes: in this section, the number of parameters exceeds the number of observations, but glmnet still works when $\lambda = 0$ (why), and "lm" also works unless we set "singular.ok = FALSE".

The regression being estimated is:

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

1, 581 (1+4*145) series were used as explanatory variables;

I(1) series were first-differenced;

I(2) series were second-differenced;

2, all explanatory variables are now I(0) and lagged by 4 quarters;

3, after first difference, second difference and four lags, we have 218 observations (lost 6).

Lasso 4

The regression being estimated is:

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

1, 436 (1+3*145) series were used as explanatory variables;

I(1) series were first-differenced;

I(2) series were second-differenced;

2, all explanatory variables are now I(0) and lagged by 3 quarters;

3, after first difference, second difference and 3 lags, we have 219 observations (lost 5).

Lasso 5

The regression being estimated is:

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

1, 291 (1+2*145) series were used as explanatory variables;

I(1) series were first-differenced;

I(2) series were second-differenced;

2, all explanatory variables are now I(0) and lagged by 2 quarters;

3, after first difference, second difference and 2 lags, we have 219 observations (lost 5).

Lasso 6

The regression being estimated is:

$$\Delta y_t = y_{t-1} + \Delta y_{t-1} + I(0)_{t-1} + \Delta I(1)_{t-1} + \Delta^2 I(2)_{t-1}$$

1, 146 (1+145) series were used as explanatory variables;

I(1) series were first-differenced;

I(2) series were second-differenced;

2, all explanatory variables are now I(0) and lagged by 1 quarters;

3, after first difference, second difference and 1 lag, we have 221 observations (lost 3).

Lasso 7

The regression being estimated is:

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

1, 697 series were used as explanatory variables;

I(1) series were first-differenced;

I(2) series were second-differenced;

2, all explanatory variables are now I(0) and lagged by 4 quarters;

3, after first difference, second difference and 4 lags, we have 218 observations (lost 6).

Lasso 8

The regression being estimated is:

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

The level of lag 1 of y (y_{t-1}) was added on the top of lasso 7.

1, 698 series were used as explanatory variables;

I(1) series were first-differenced;

I(2) series were second-differenced;

2, all explanatory variables are now I(0) and lagged by 4 quarters;

3, after first difference, second difference and 4 lags, we have 218 observations (lost 6).

Experiment Lasso

The regression being estimated is:

$$\Delta y_t = y_{t-1} + I(1)_{t-1} + \Delta I(2)_{t-1}$$

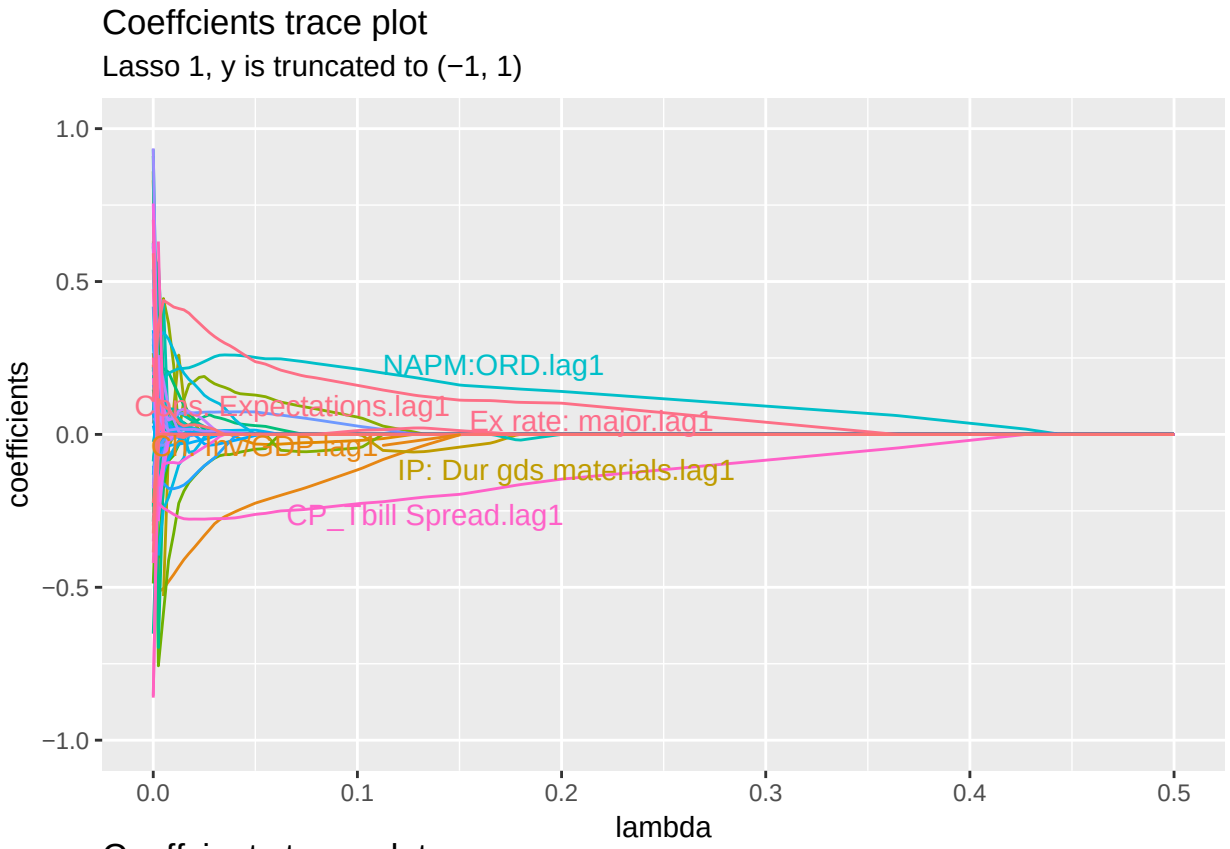
1, 118 series were used as explanatory variables;

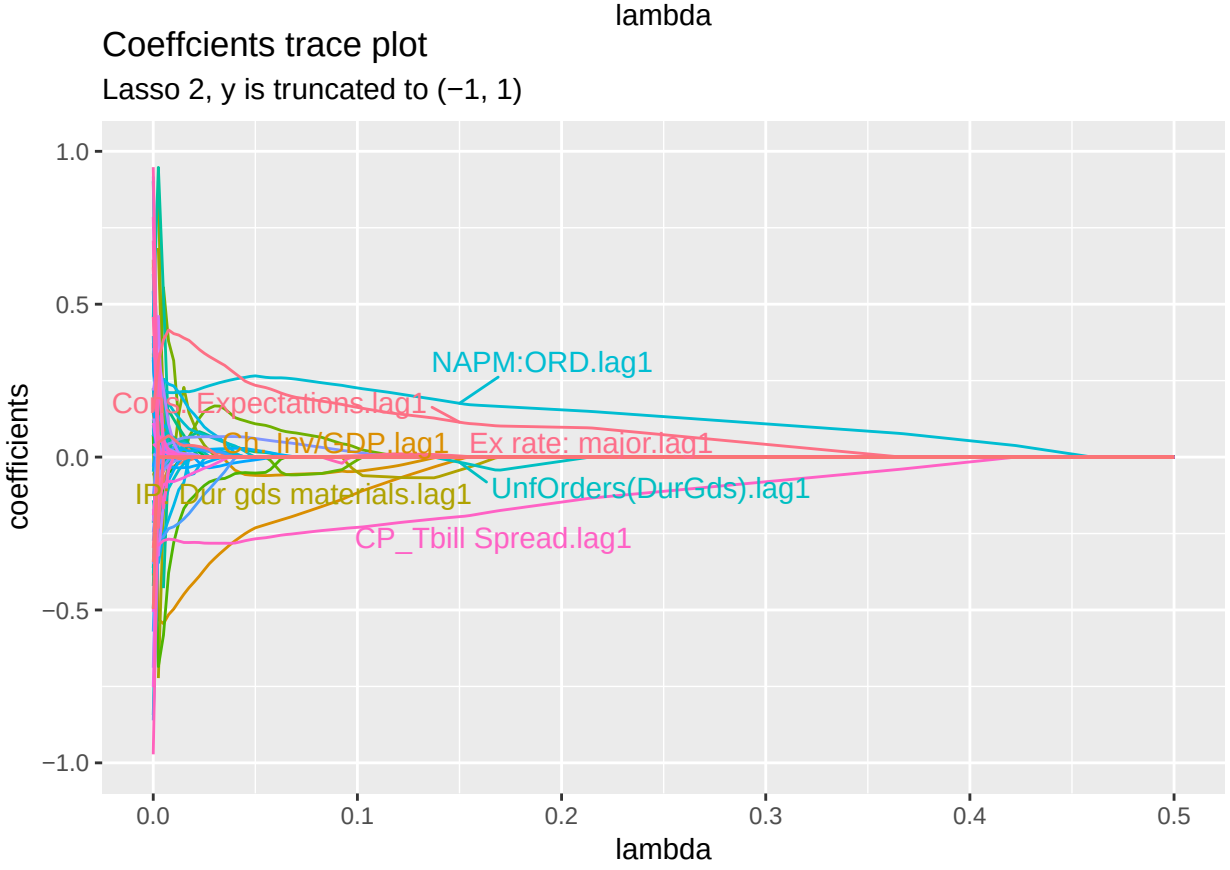
2, I(2) series were first differenced, no change to I(1) series;

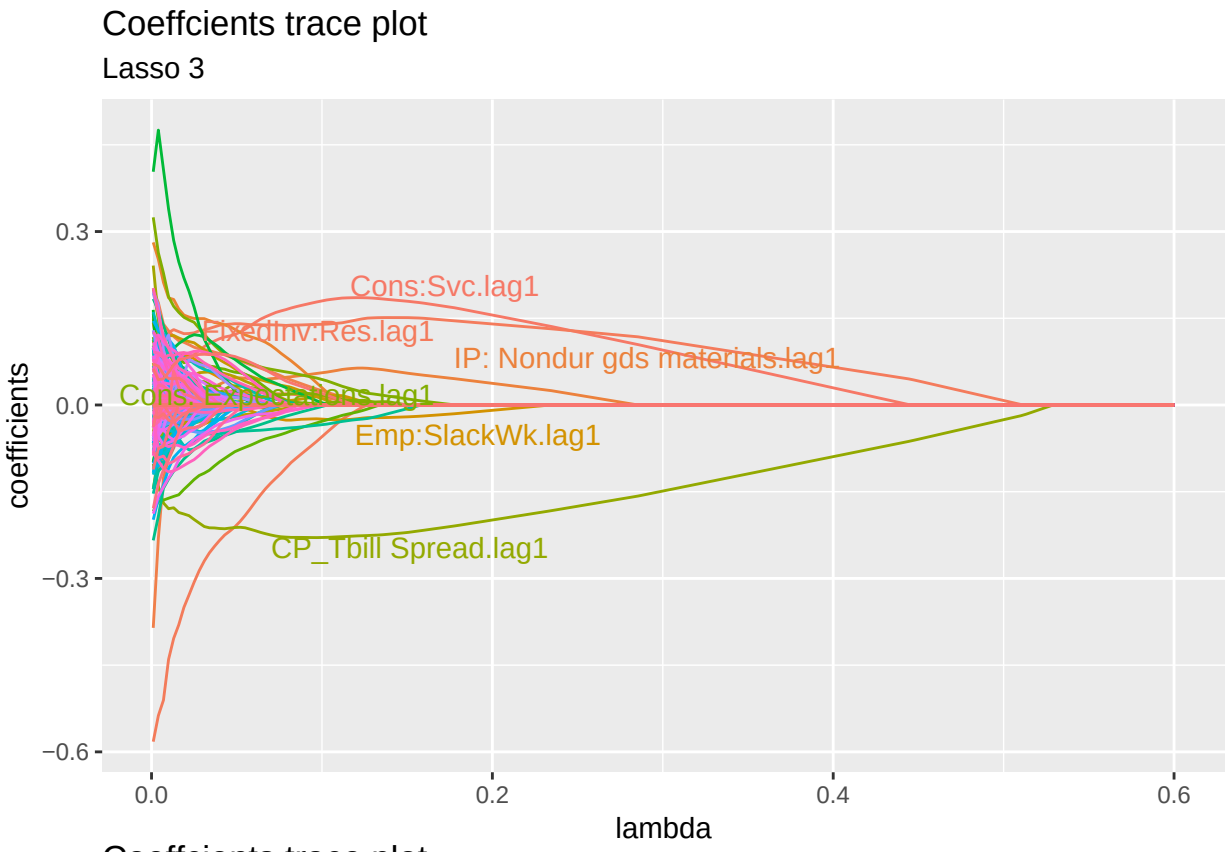
3, all explanatory variables are lagged by 1 quarter;

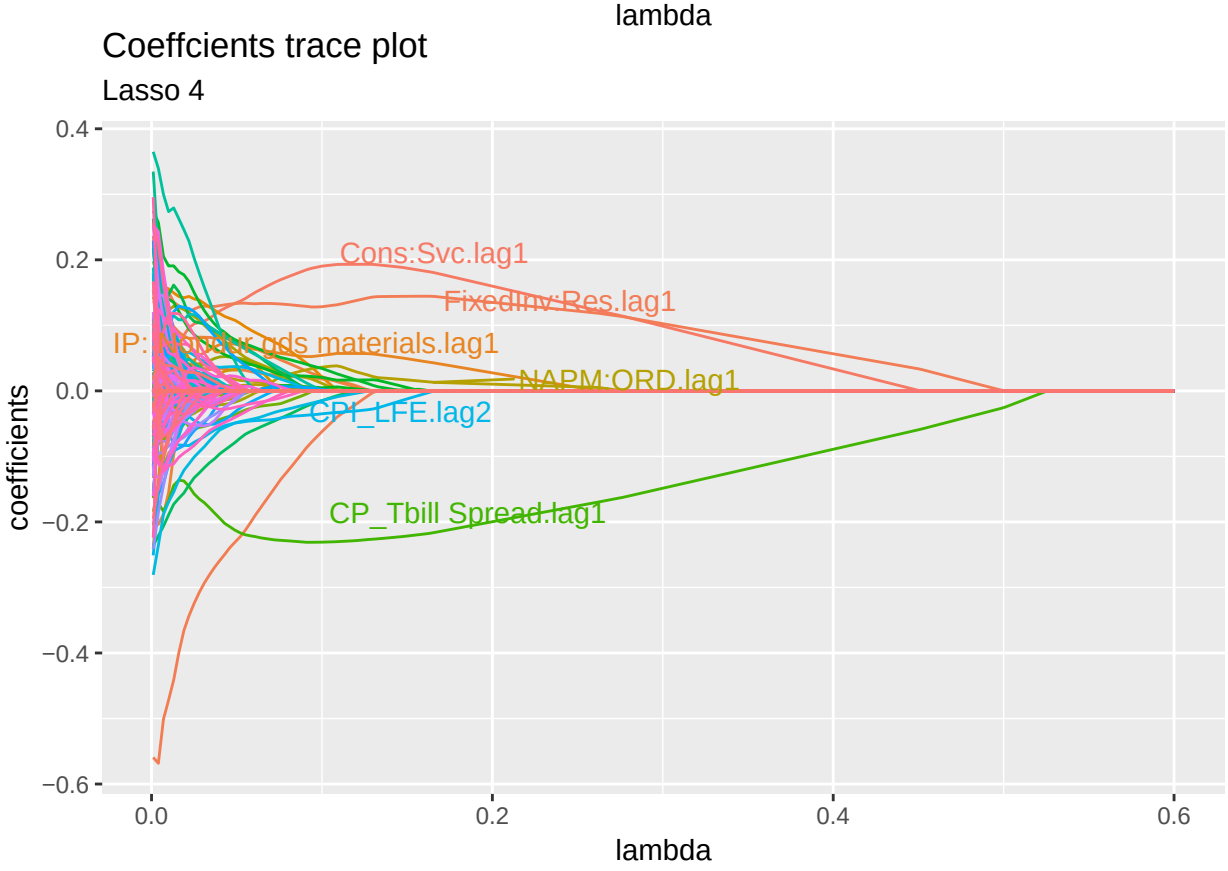
4, after first difference and one lag, we have 222 observations (lost 2).

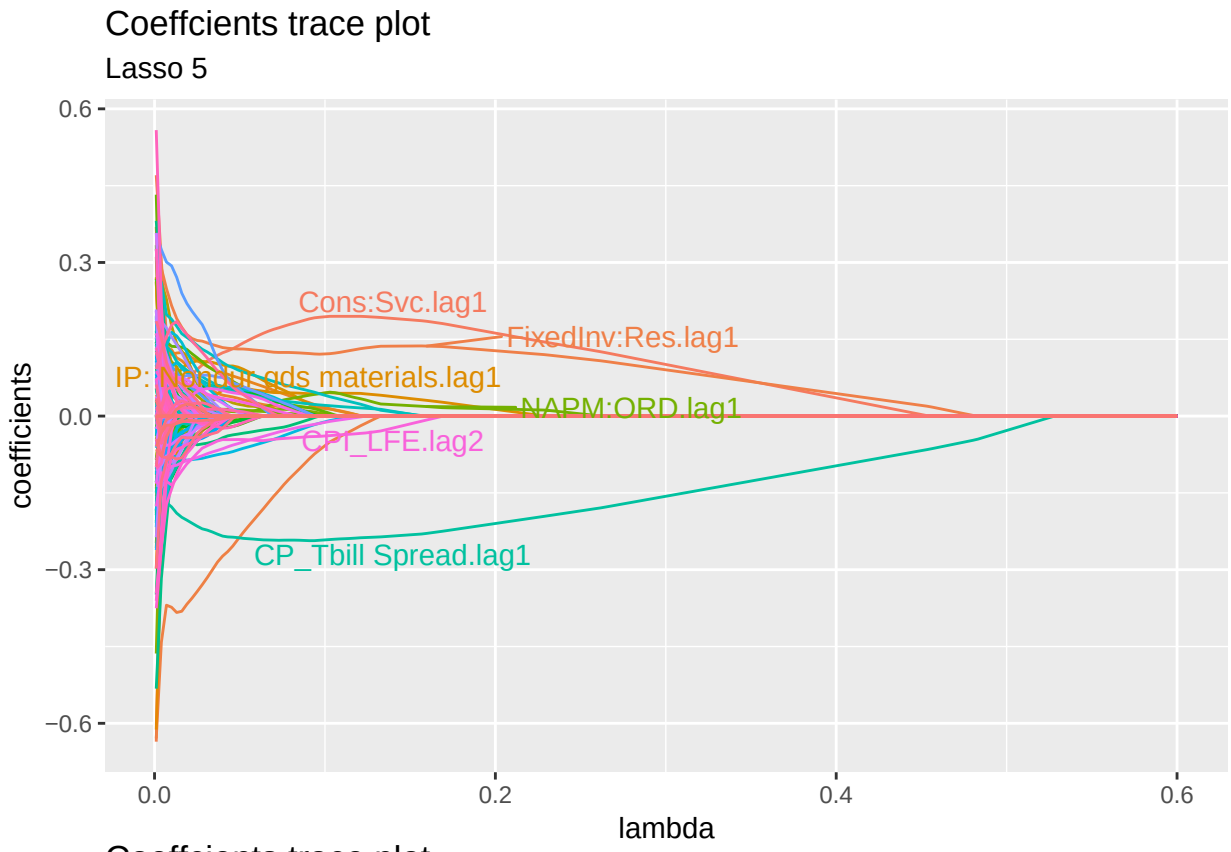
Graphs

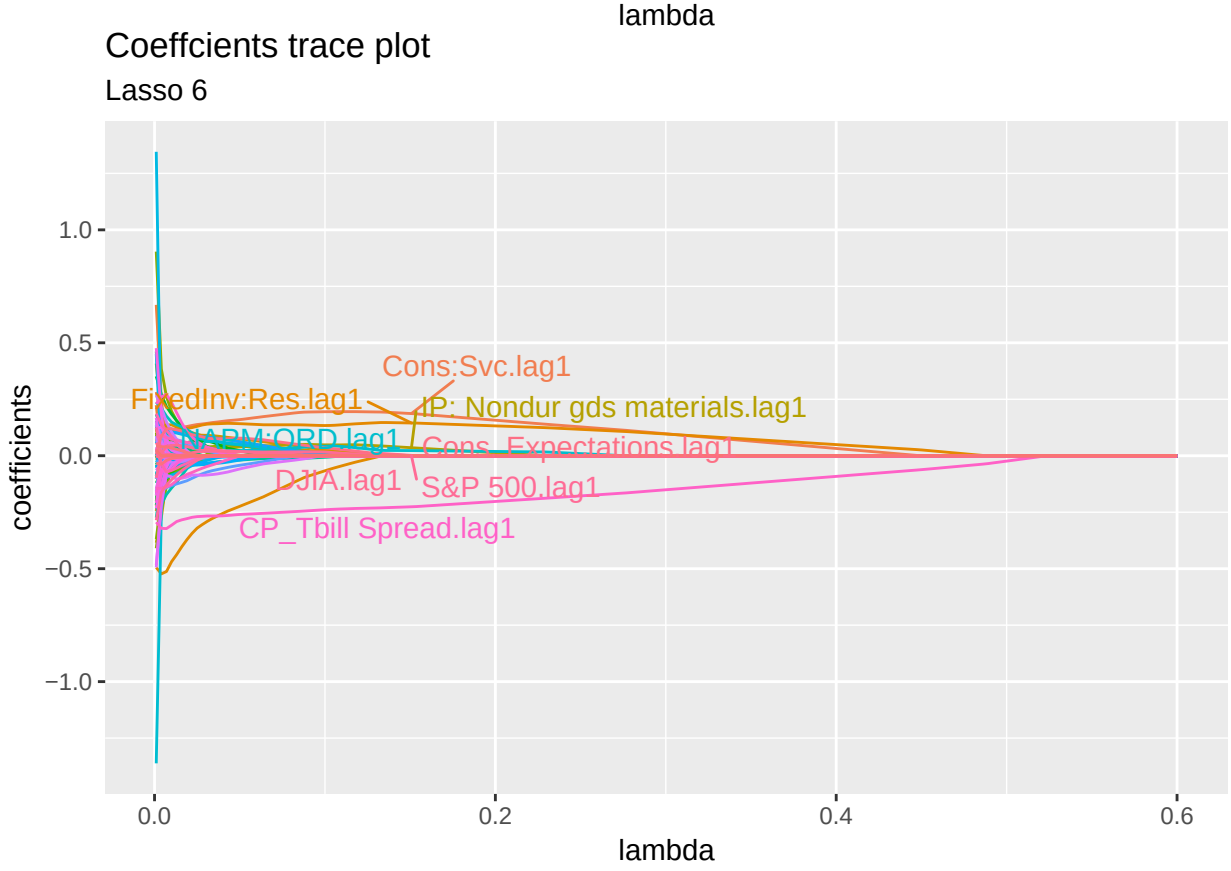

Coeffcients trace plot
Lasso 1, y is truncated to (−1, 1)
coefficients
lambda
NAPM:ORD.lag1
Ex rate: major.lag1
IP: Dur gds materials.lag1
CP_Tbill Spread.lag1

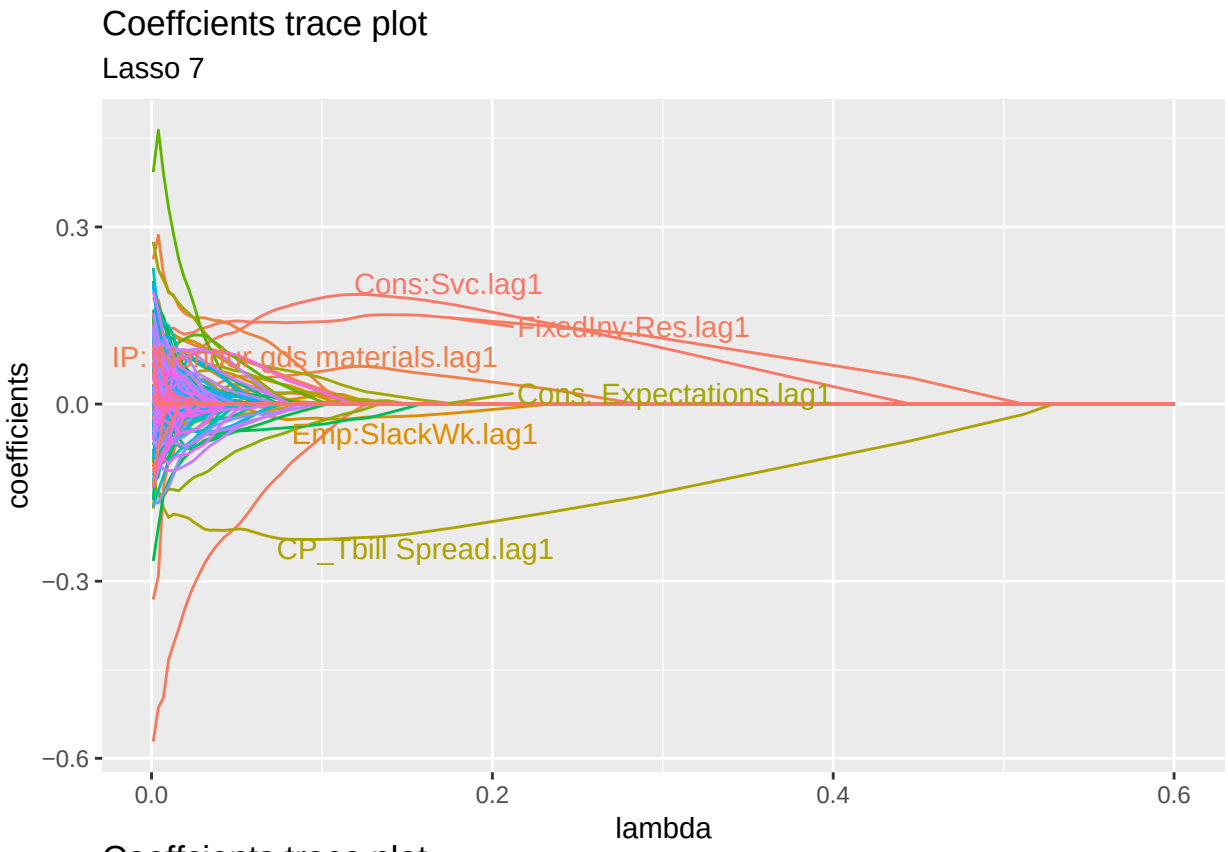

Coeffcients trace plot
Lasso 2, y is truncated to (−1, 1)
coefficients
lambda
NAPM:ORD.lag1
Ex rate: major.lag1
UnfOrders(DurGds).lag1
CP_Tbill Spread.lag1

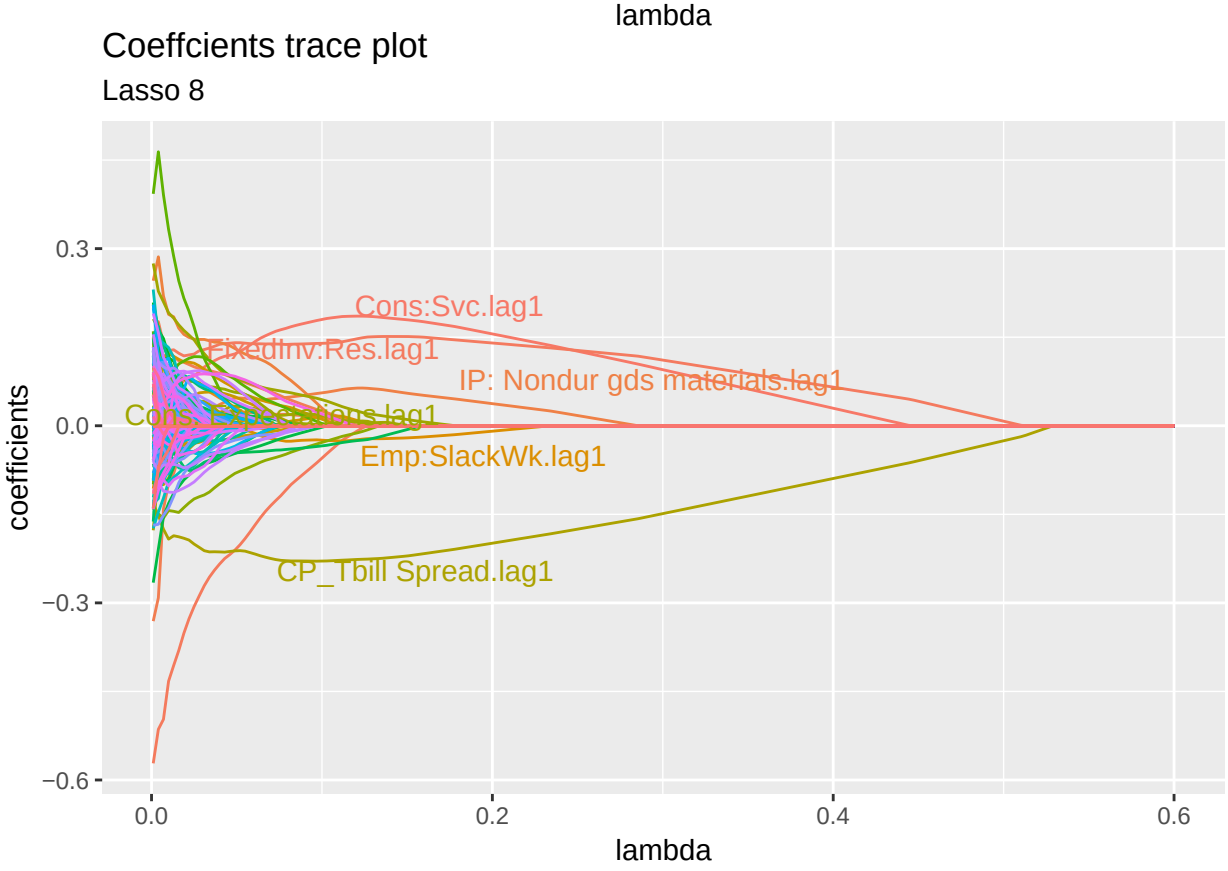

Coeffcients trace plot
Lasso 3
coefficients
lambda
Cons:Svc.lag1
FixedInv:Res.lag1
IP: Nondur gds materials.lag1
Cons: Expectations.lag1
Emp:SlackWk.lag1
CP_Tbill Spread.lag1
0.3
0.0
−0.3
−0.6
0.0
0.2
0.4
0.6

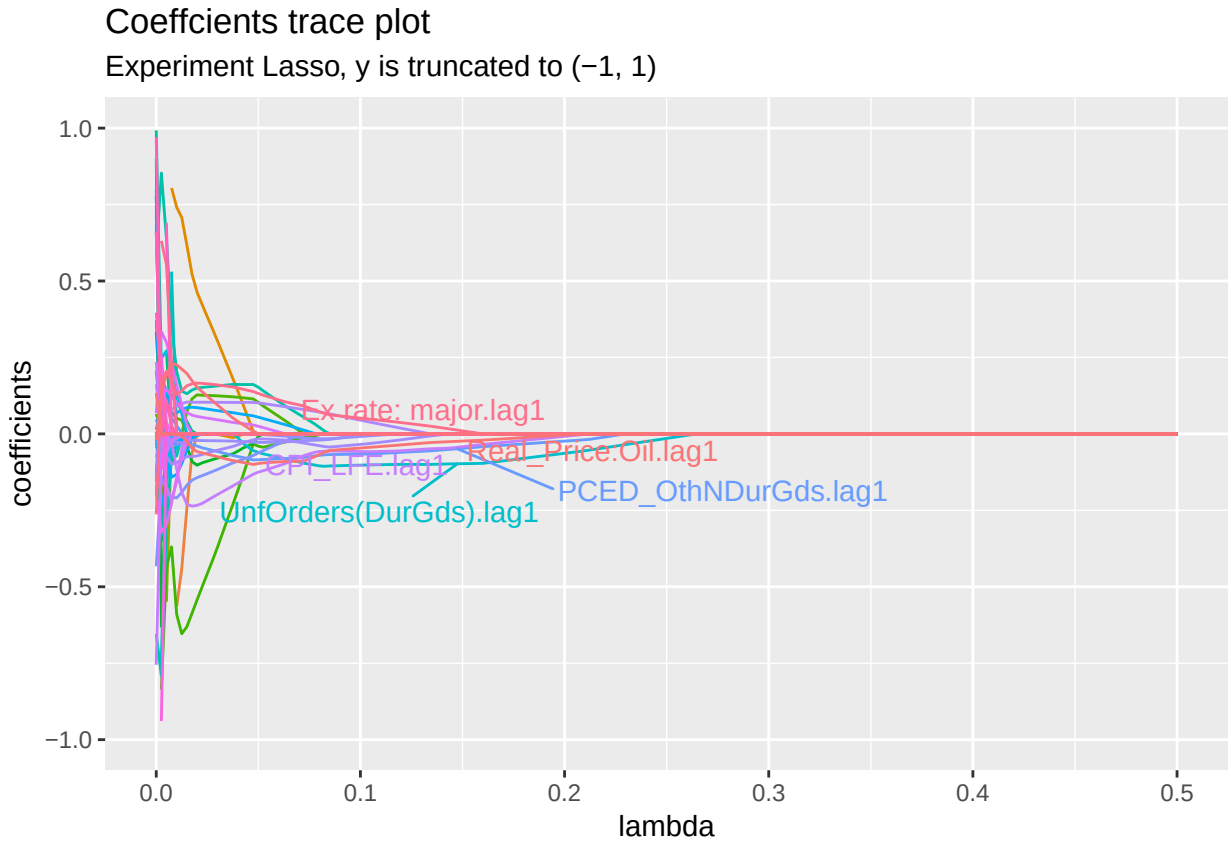

Coeffcients trace plot
Lasso 4
coefficients
lambda
Cons:Svc.lag1
FixedInv:Res.lag1
IP: Nondur gds materials.lag1
NAPM:ORD.lag1
CPI_LFE.lag2
CP_Tbill Spread.lag1
0.4
0.2
0.0
−0.2
−0.4
−0.6
0.0
0.2
0.4
0.6


Coeffcients trace plot
Lasso 5
coefficients
lambda
Cons:Svc.lag1
FixedInv:Res.lag1
IP: Nondur gds materials.lag1
NAPM:ORD.lag1
CPI_LFE.lag2
CP_Tbill Spread.lag1


Coeffcients trace plot
Lasso 6
coefficients
lambda
Cons:Svc.lag1
FixedInv:Res.lag1
IP: Nondur gds materials.lag1
NAPM:ORD.lag1
Cons_Expectations.lag1
DJIA.lag1
S&P 500.lag1
CP_Tbill Spread.lag1


Coeffcients trace plot
Lasso 7
coefficients
0.3
0.0
−0.3
−0.6
Cons:Svc.lag1
FixedInv:Res.lag1
IP: Nondur gds materials.lag1
Cons. Expectations.lag1
Emp:SlackWk.lag1
CP_Tbill Spread.lag1
0.0
0.2
0.4
0.6
lambda


Coeffcients trace plot
Lasso 8
coefficients
0.3
0.0
−0.3
−0.6
Cons:Svc.lag1
FixedInv:Res.lag1
IP: Nondur gds materials.lag1
Cons. Expectations.lag1
Emp:SlackWk.lag1
CP_Tbill Spread.lag1
0.0
0.2
0.4
0.6
lambda



Data transformation table

Table 2: Number of series with 'No-transformation' is 12

tcode	short	long
I(0)	Ch. Inv/GDP	Ch. Inv/GDP
I(0)	AWH Man	Average Weekly Hours: Manufacturing
I(0)	VendPerf	ISM Manufacturing: Supplier Deliveries Index©
I(0)	NAPM:INV	ISM Manufacturing: Inventories Index©
I(0)	NAPM:ORD	ISM Manufacturing: New Orders Index©; Index;
I(0)	NAPM com price	ISM Manufacturing: Prices Paid Index©
I(0)	BAA_GS10	BAA-GS10 Spread
I(0)	tb6m_tb3m	tb6m-tb3m
I(0)	GS1_tb3m	GS1_Tb3m
I(0)	GS10_tb3m	GS10_Tb3m
I(0)	CP_Tbill Spread	CP3FM-TB3MS
I(0)	Cons. Expectations	Consumer expectations NSA (Copyright University of Michigan)

Table 3: Number of 'First-differenced' series is 15

tcode	short	long
I(1)	LF Part Rate	LaborForce Participation Rate (16 Over) SA
I(1)	Unemp Rate	Urate

tcode	short	long
I(1)	Urate_ST	Urate Short Term (< 27 weeks)
I(1)	Urate_LT	Urate Long Term (>= 27 weeks)
I(1)	Urate: Age16-19	Unemployment Rate - 16-19 yrs
I(1)	Urate:Age>20 Men	Unemployment Rate - 20 yrs. & over Men
I(1)	Urate: Age>20 Women	Unemployment Rate - 20 yrs. & over Women
I(1)	AWH Overtime	Average Weekly Hours: Overtime: Manufacturing
I(1)	FedFunds	Effective Federal Funds Rate
I(1)	TB-3Mth	3-Month Treasury Bill: Secondary Market Rate
I(1)	TM-6MTH	6-Month Treasury Bill: Secondary Market Rate
I(1)	TB-1YR	1-Year Treasury Constant Maturity Rate
I(1)	TB-10YR	10-Year Treasury Constant Maturity Rate
I(1)	AAA Bond	Moody's Seasoned Aaa Corporate Bond Yield
I(1)	BAA Bond	Moody's Seasoned Baa Corporate Bond Yield

Table 4: Number of 'First-differenced in logs' series is 86

tcode	short	long
log, I(1)	GDP	Real Gross Domestic Product 3 Decimal
log, I(1)	Consumption	Real Personal Consumption Expenditures
log, I(1)	Cons:Dur	Real Personal Consumption Expenditures: Durable Goods Quantity Index
log, I(1)	Cons:Svc	Real Personal Consumption Expenditures: Services Quantity Index
log, I(1)	Cons:NonDur	Real Personal Consumption Expenditures: Nondurable Goods Quantity Index
log, I(1)	Investment	Real Gross Private Domestic Investment 3 Decimal
log, I(1)	FixedInv	Real Private Fixed Investment Quantity Index
log, I(1)	Inv:Equip	Real Nonresidential Investment: Equipment Quantity Idenx
log, I(1)	FixInv:NonRes	Real Private Nonresidential Fixed Investment Quantity Index
log, I(1)	FixedInv:Res	Real Private Residential Fixed Investment Quantity Index
log, I(1)	Gov.Spending	Real Government Consumption Expenditures & Gross Investment 3 Decimal
log, I(1)	Gov:Fed	Real Federal Consumption Expenditures Quantity Index
log, I(1)	Gov:State&Local	Real State & Local Consumption Expenditures Quantity Index
log, I(1)	Exports	Real Exports of Goods & Services 3 Decimal
log, I(1)	Imports	Real Imports of Goods & Services 3 Decimal
log, I(1)	Disp-Income	Real Disposable Personal Income
log, I(1)	Ouput:NFB	Nonfarm Business Sector: Output
log, I(1)	Output:Bus	Business Sector: Output
log, I(1)	IP: Total index	IP: Total index
log, I(1)	IP: Final products	Industrial Production: Final Products (Market Group)
log, I(1)	IP: Consumer goods	IP: Consumer goods
log, I(1)	IP: Materials	Industrial Production: Materials
log, I(1)	IP: Dur gds materials	Industrial Production: Durable Materials

tcode	short	long
log, I(1)	IP: Nondur gds materials	Industrial Production: nondurable Materials
log, I(1)	IP: Dur Cons. Goods	Industrial Production: Durable Consumer Goods
log, I(1)	IP: Auto	IP: Automotive products
log, I(1)	IP:NonDur Cons God	Industrial Production: Nondurable Consumer Goods
log, I(1)	IP: Bus Equip	Industrial Production: Business Equipment
log, I(1)	Emp:Nonfarm	Total Nonfarm Payrolls: All Employees
log, I(1)	Emp: Private	All Employees: Total Private Industries
log, I(1)	Emp: mfg	All Employees: Manufacturing
log, I(1)	Emp:Services	All Employees: Service-Providing Industries
log, I(1)	Emp:Goods	All Employees: Goods-Producing Industries
log, I(1)	Emp: DurGoods	All Employees: Durable Goods Manufacturing
log, I(1)	Emp: Nondur Goods	All Employees: Nondurable Goods Manufacturing
log, I(1)	Emp: Const	All Employees: Construction
log, I(1)	Emp: Edu&Health	All Employees: Education & Health Services
log, I(1)	Emp: Finance	All Employees: Financial Activities
log, I(1)	Emp: Infor	All Employees: Information Services
log, I(1)	Emp: Bus Serv	All Employees: Professional & Business Services
log, I(1)	Emp:Leisure	All Employees: Leisure & Hospitality
log, I(1)	Emp:OtherSvcs	All Employees: Other Services
log, I(1)	Emp: Mining/NatRes	All Employees: Natural Resources & Mining
log, I(1)	Emp:Trade&Trans	All Employees: Trade Transportation & Utilities
log, I(1)	Emp: Gov	All Employees: Government
log, I(1)	Emp:Retail	All Employees: Retail Trade
log, I(1)	Emp:Wholesal	All Employees: Wholesale Trade
log, I(1)	Emp: Gov(Fed)	Employment Federal Government
log, I(1)	Emp: Gov (State)	Employment State government
log, I(1)	Emp: Gov (Local)	Employment Local government
log, I(1)	Emp: Total (HHSurve)	Emp Total (Household Survey)
log, I(1)	U: Dur<5wks	Number Unemployed for Less than 5 Weeks
log, I(1)	U:Dur5-14wks	Number Unemployed for 5-14 Weeks
log, I(1)	U:dur>15-26wks	Civilians Unemployed for 15-26 Weeks
log, I(1)	U: Dur>27wks	Number Unemployed for 27 Weeks & over
log, I(1)	Emp:SlackWk	Employment Level - Part-Time for Economic Reasons All Industries
log, I(1)	EmpHrs:Bus Sec	Business Sector: Hours of All Persons
log, I(1)	EmpHrs:nfb	Nonfarm Business Sector: Hours of All Persons
log, I(1)	Orders (DurMfg)	Mfrs' new orders durable goods industries (bil. chain 2000 $)
log, I(1)	Orders(ConsumerGoods/Mat.)	Mfrs' new orders consumer goods and materials (mil. 1982 $)
log, I(1)	UnfOrders(DurGds)	Mfrs' unfilled orders durable goods indus. (bil. chain 2000 $)
log, I(1)	Orders(NonDefCap)	Mfrs' new orders nondefense capital goods (mil. 1982 $)
log, I(1)	Real_AHE:Const	Average Hourly Earnings: Construction Defl by PCE(LFE) Def
log, I(1)	Real_AHE:MFG	Average Hourly Earnings: Manufacturing Defl by PCE(LFE) Def
log, I(1)	CPH:NFB	Nonfarm Business Sector: Real Compensation Per Hour
log, I(1)	CPH:Bus	Business Sector: Real Compensation Per Hour

tcode	short	long
log, I(1)	OPH:nfb	Nonfarm Business Sector: Output Per Hour of All Persons
log, I(1)	OPH:Bus	Business Sector: Output Per Hour of All Persons
log, I(1)	ULC:Bus	Business Sector: Unit Labor Cost
log, I(1)	ULC:NFB	Nonfarm Business Sector: Unit Labor Cost
log, I(1)	UNLPay:nfb	Nonfarm Business Sector: Unit Nonlabor Payments
log, I(1)	Real_mbase	St. Louis Adjusted Monetary Base; Bil. of $; M; SA; Defl by PCE(LFE) Def
log, I(1)	Real_m1	M1 Money Stock Defl by PCE(LFE) Def
log, I(1)	Real_m2	M2SL Defl by PCE(LFE) Def
log, I(1)	Real_mzm	MZM Money Stock Defl by PCE(LFE) Def
log, I(1)	Real_C&Lloand	Commercial and Industrial Loans at All Commercial Banks Defl by PCE(LFE) Def
log, I(1)	Real_ConsLoans	Consumer (Individual) Loans at All Commercial Banks - Outlier Code because of change in data in April 2010 ... see FRB H8 Release Defl by PCE(LFE) Def
log, I(1)	Real_NonRevCredit	Total Nonrevolving Credit Outstanding Defl by PCE(LFE) Def
log, I(1)	Real_LoansRealEst	Real Estate Loans at All Commercial Banks Defl by PCE(LFE) Def
log, I(1)	Real_ConsuCred	Total Consumer Credit Outstanding Defl by PCE(LFE) Def
log, I(1)	S&P 500	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)
log, I(1)	DJIA	COMMON STOCK PRICES: DOW JONES INDUSTRIAL AVERAGE
log, I(1)	Ex rate: major	FRB Nominal Major Currencies Dollar Index (Linked to EXRUS in 1973:1)
log, I(1)	IP: Energy Prds	IP: Consumer Energy Products
log, I(1)	Petroleum Stocks	U.S. Ending Stocks excluding SPR of Crude Oil and Petroleum Products (Thousand Barrels); SA using X11 in RATS
log, I(1)	Real_Price:Oil	PPI: Crude Petroleum Defl by PCE(LFE) Def

Table 5: Number of 'Second-differenced in logs' series is 32

tcode	short	long
log, I(2)	PCED	Personal Consumption Expenditures: Chain-type Price Index
log, I(2)	PCED_LFE	Personal Consumption Expenditures: Chain-type Price Index Less Food and Energy
log, I(2)	GDP Defl	Gross Domestic Product: Chain-type Price Index
log, I(2)	GPDI Defl	Gross Private Domestic Investment: Chain-type Price Index
log, I(2)	BusSec Defl	Business Sector: Implicit Price Deflator
log, I(2)	PCED_Goods	Goods
log, I(2)	PCED_DurGoods	Durable goods
log, I(2)	PCED_NDurGoods	Nondurable goods
log, I(2)	PCED_Serv	Services
log, I(2)	PCED_HouseholdServices	Household consumption expenditures (for services)

tcode	short	long
log, I(2)	PCED_MotorVec	Motor vehicles and parts
log, I(2)	PCED_DurHousehold	Furnishings and durable household equipment
log, I(2)	PCED_Recreation	Recreational goods and vehicles
log, I(2)	PCED_OthDurGds	Other durable goods
log, I(2)	PCED_Food_Bev	Food and beverages purchased for off-premises consumption
log, I(2)	PCED_Clothing	Clothing and footwear
log, I(2)	PCED_Gas_Enrgy	Gasoline and other energy goods
log, I(2)	PCED_OthNDurGds	Other nondurable goods
log, I(2)	PCED_Housing-Utilities	Housing and utilities
log, I(2)	PCED_HealthCare	Health care
log, I(2)	PCED_TransSvg	Transportation services
log, I(2)	PCED_RecServices	Recreation services
log, I(2)	PCED_FoodServ_Acc.	Food services and accommodations
log, I(2)	PCED_FIRE	Financial services and insurance
log, I(2)	PCED_OtherServices	Other services
log, I(2)	CPI	Consumer Price Index For All Urban Consumers: All Items
log, I(2)	CPI_LFE	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy
log, I(2)	PPI:FinGds	Producer Price Index: Finished Goods
log, I(2)	PPI:FinConsGds	Producer Price Index: Finished Consumer Goods
log, I(2)	PPI:FinConsGds(Food)	Producer Price Index: Finished Consumer Foods
log, I(2)	PPI:IndCom	Producer Price Index: Industrial Commodities
log, I(2)	PPI:IntMat	Producer Price Index: Intermediate Materials: Supplies & Components