Economics

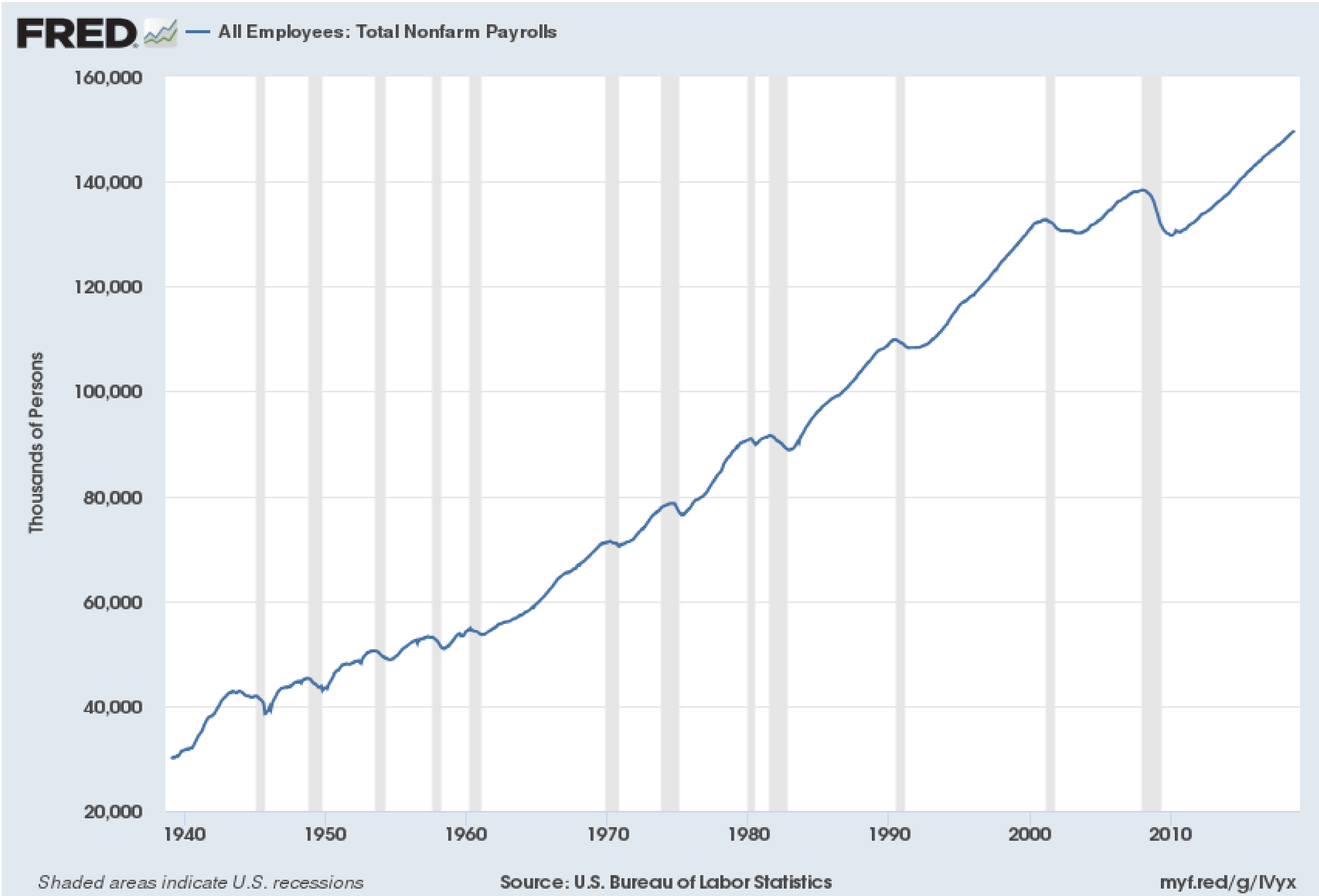
Lecture #11

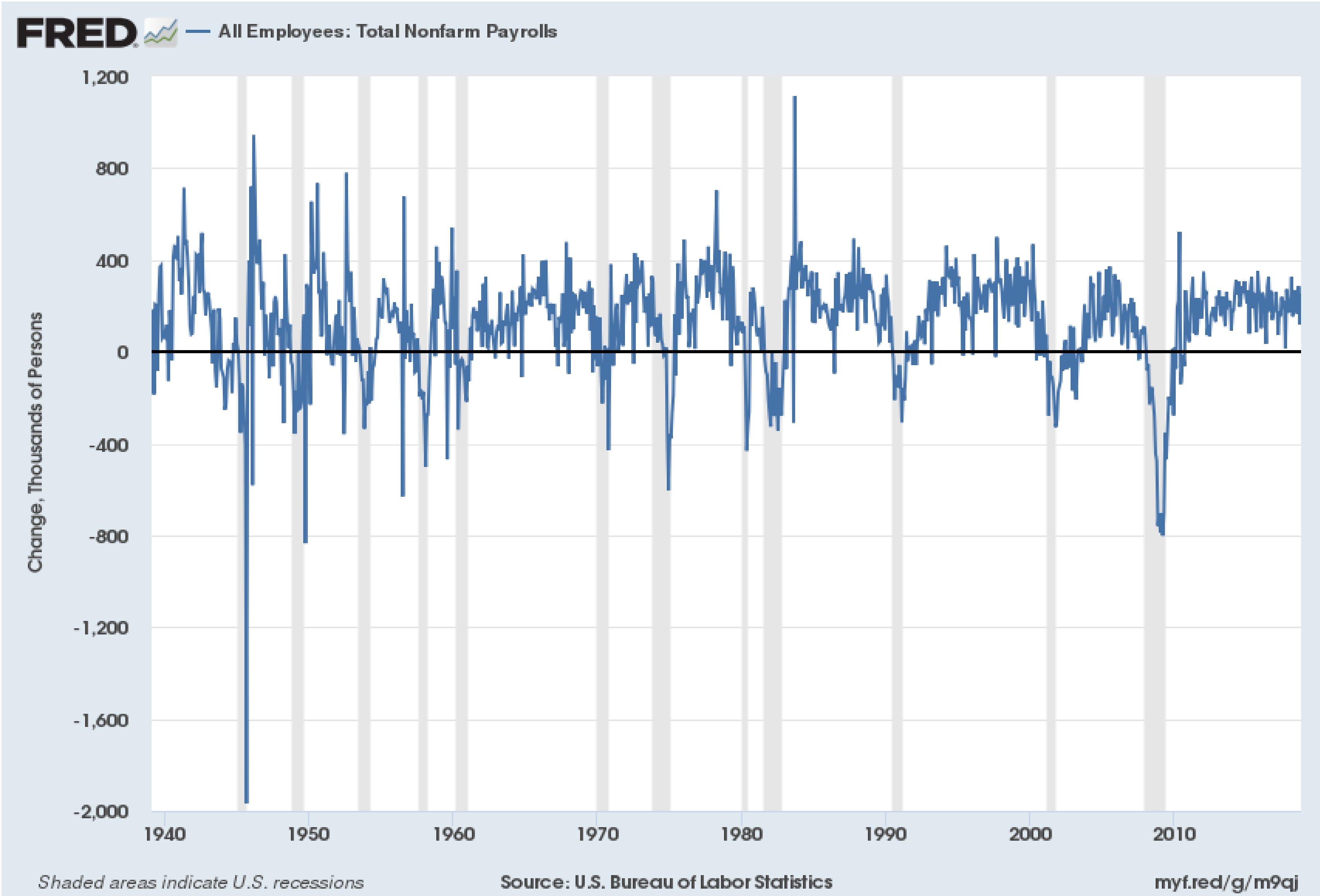
Forecasting I

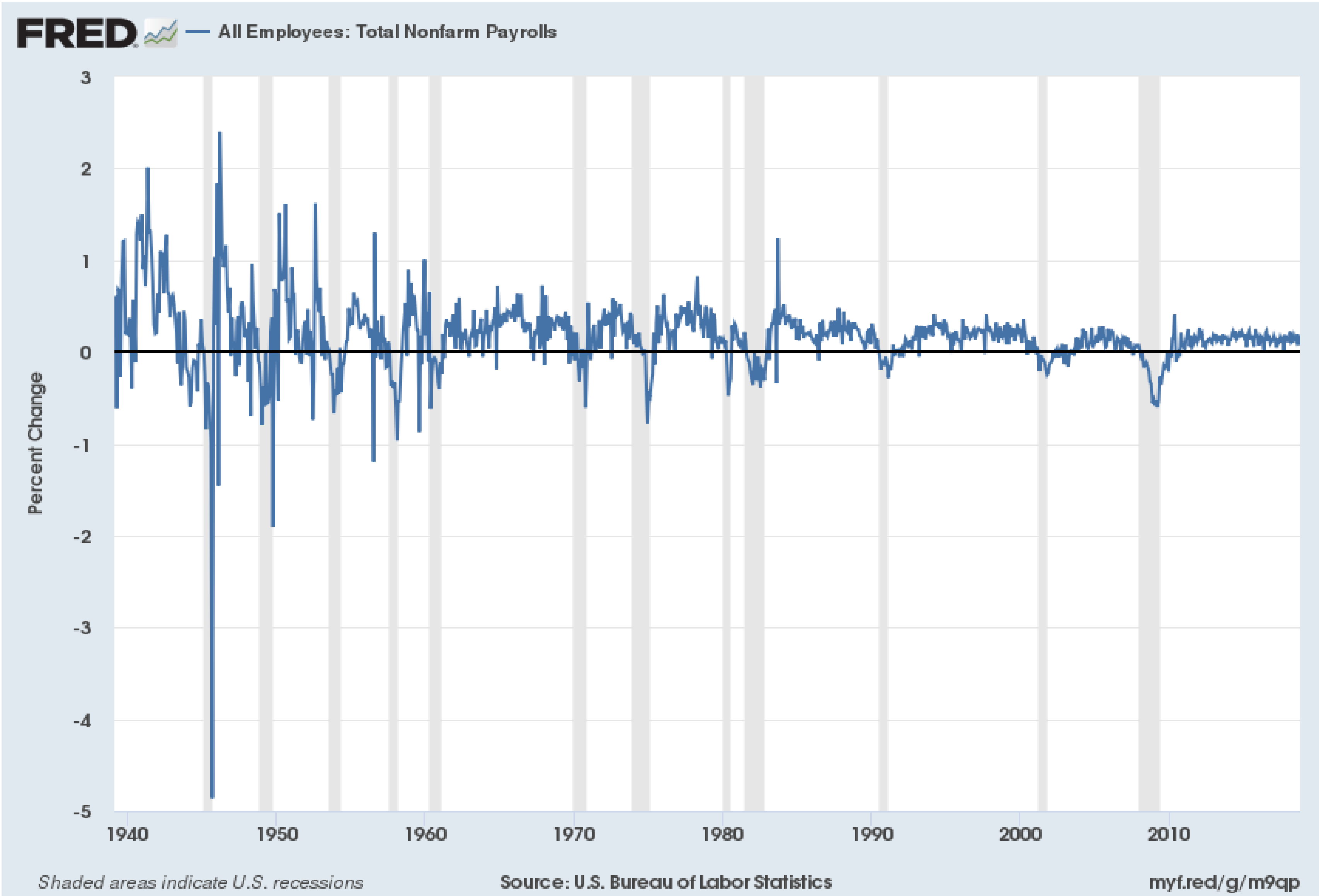
Outline (Forecasting I and II)

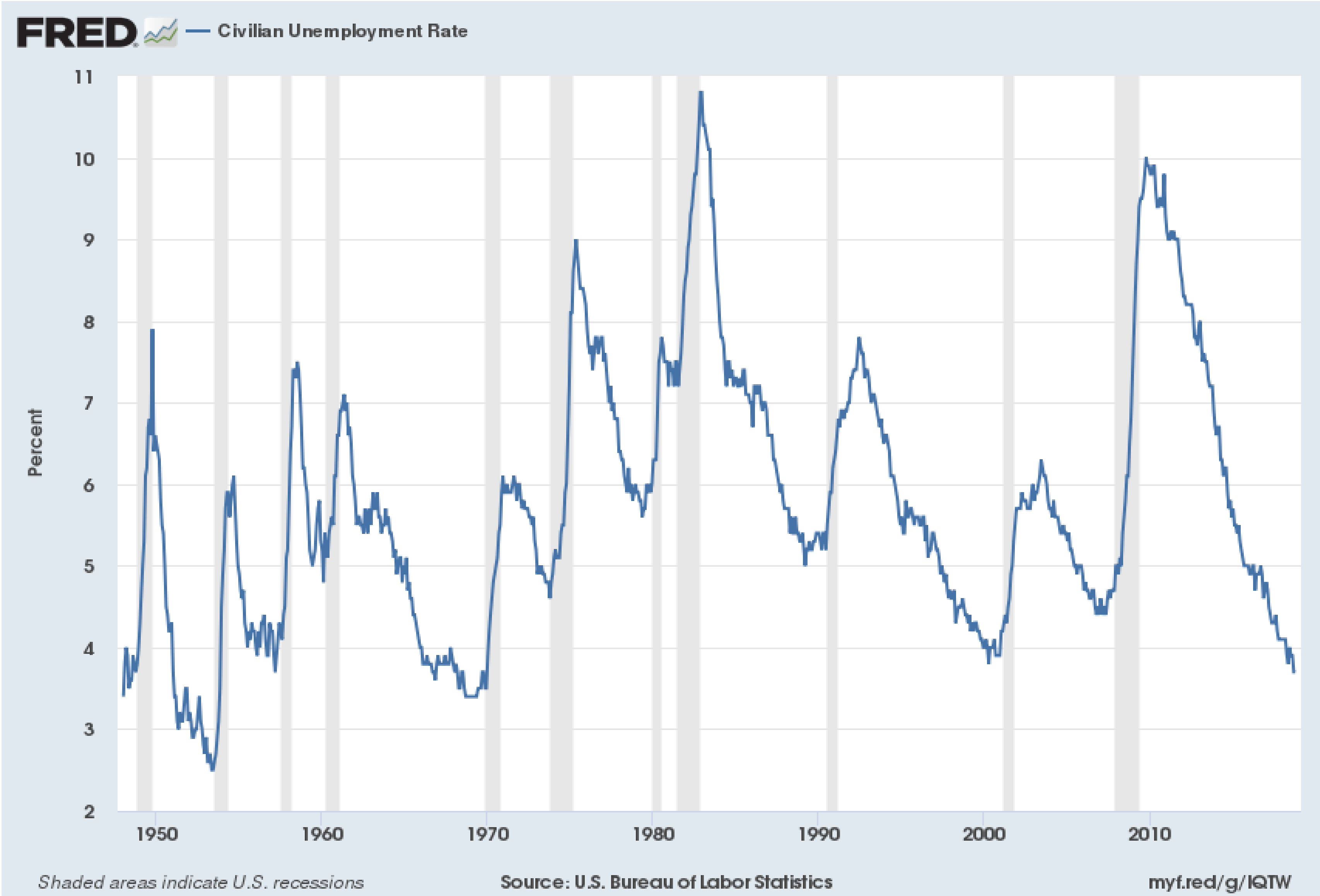
1. Intro: Economic time series data & econometric issues
2. Lags, differences, autocorrelation, and stationarity
3. Autoregressive (AR) models
4. Autoregressive distributed lag (ADL) models & Granger Causality tests
5. Lag length selection using information criteria
6. Stationarity, nonstationarity, breaks, and model stability
7. Forecast intervals and fan charts

1. Intro to time series data: some U.S. macro time series

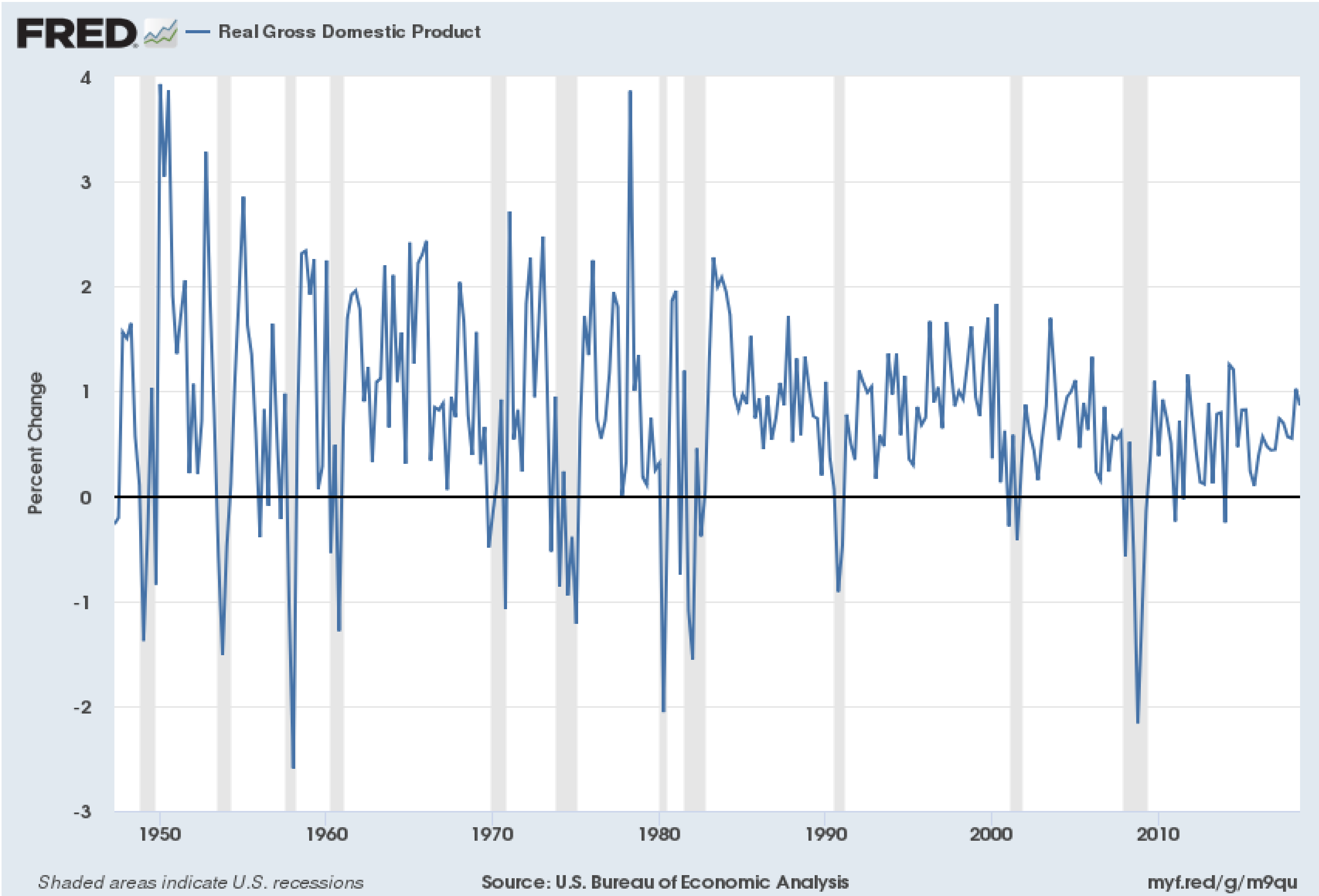


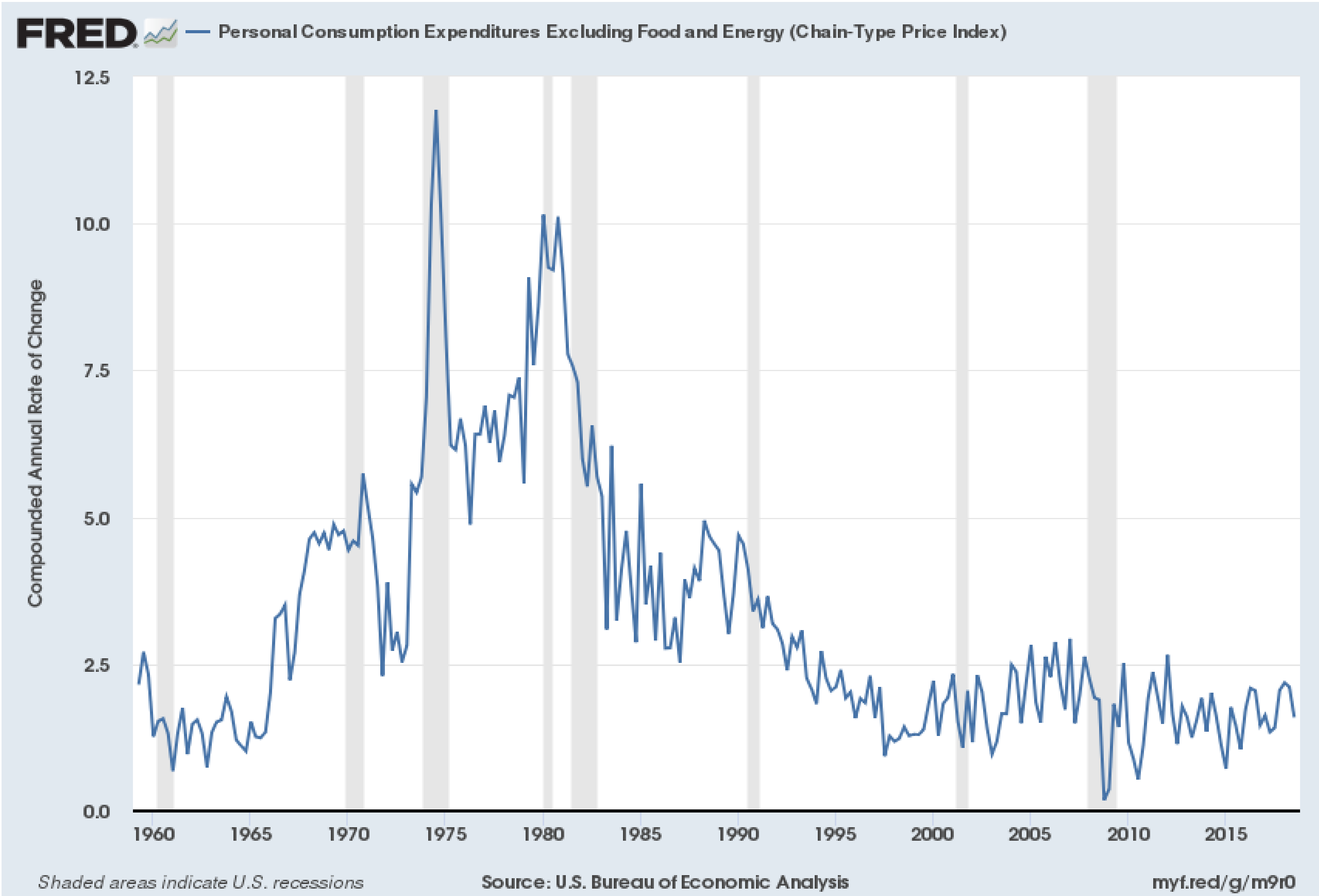

FRED
All Employees: Total Nonfarm Payrolls
1,200
800
400
0
-400
-800
-1,200
-1,600
-2,000
Change, Thousands of Persons
1940
1950
1960
1970
1980
1990
2000
2010
Shaded areas indicate U.S. recessions
Source: U.S. Bureau of Labor Statistics
myf.red/g/m9qj

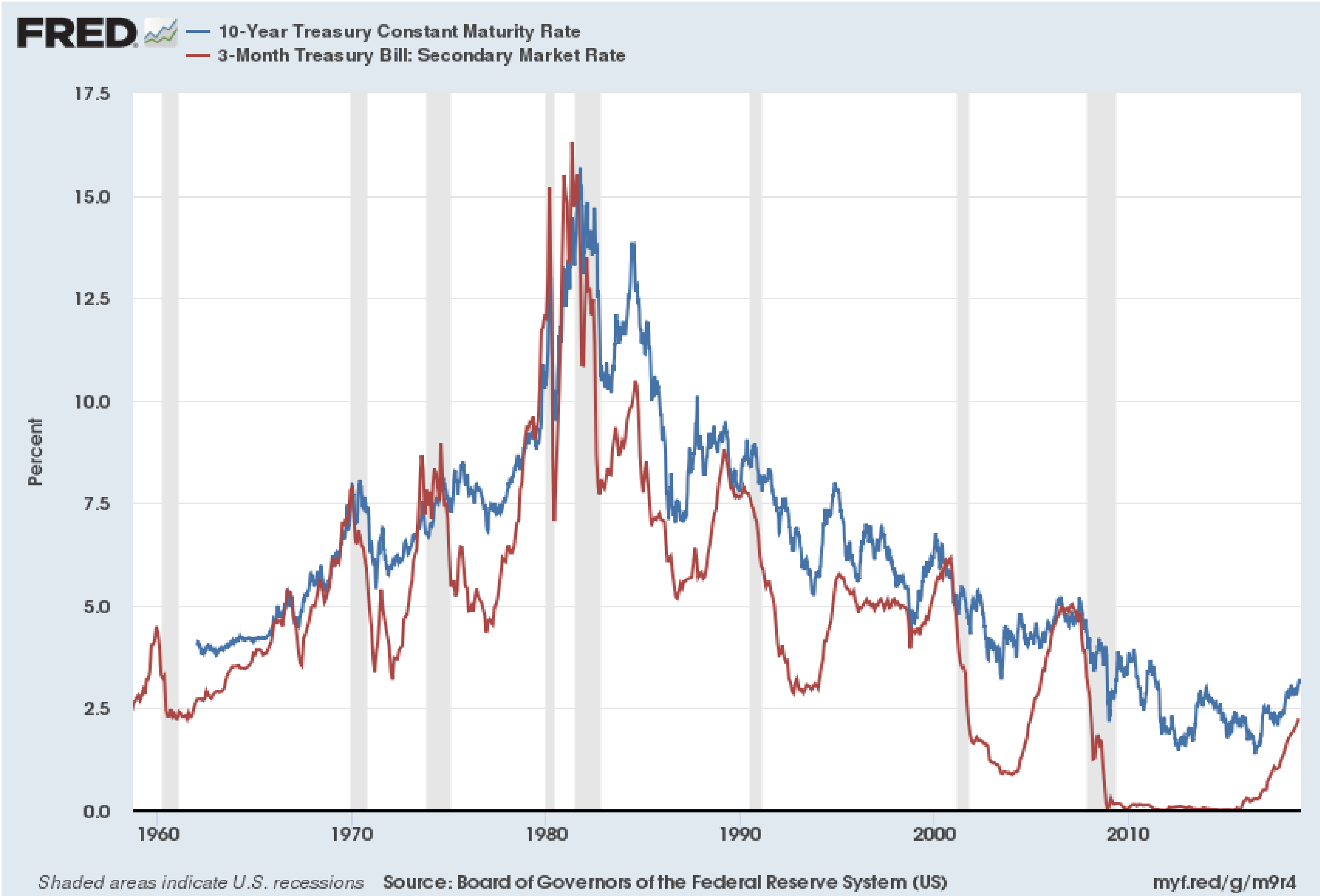

FRED
All Employees: Total Nonfarm Payrolls
3
2
1
0
-1
-2
-3
-4
-5
Percent Change
1940
1950
1960
1970
1980
1990
2000
2010
Shaded areas indicate U.S. recessions
Source: U.S. Bureau of Labor Statistics
myf.red/g/m9qp

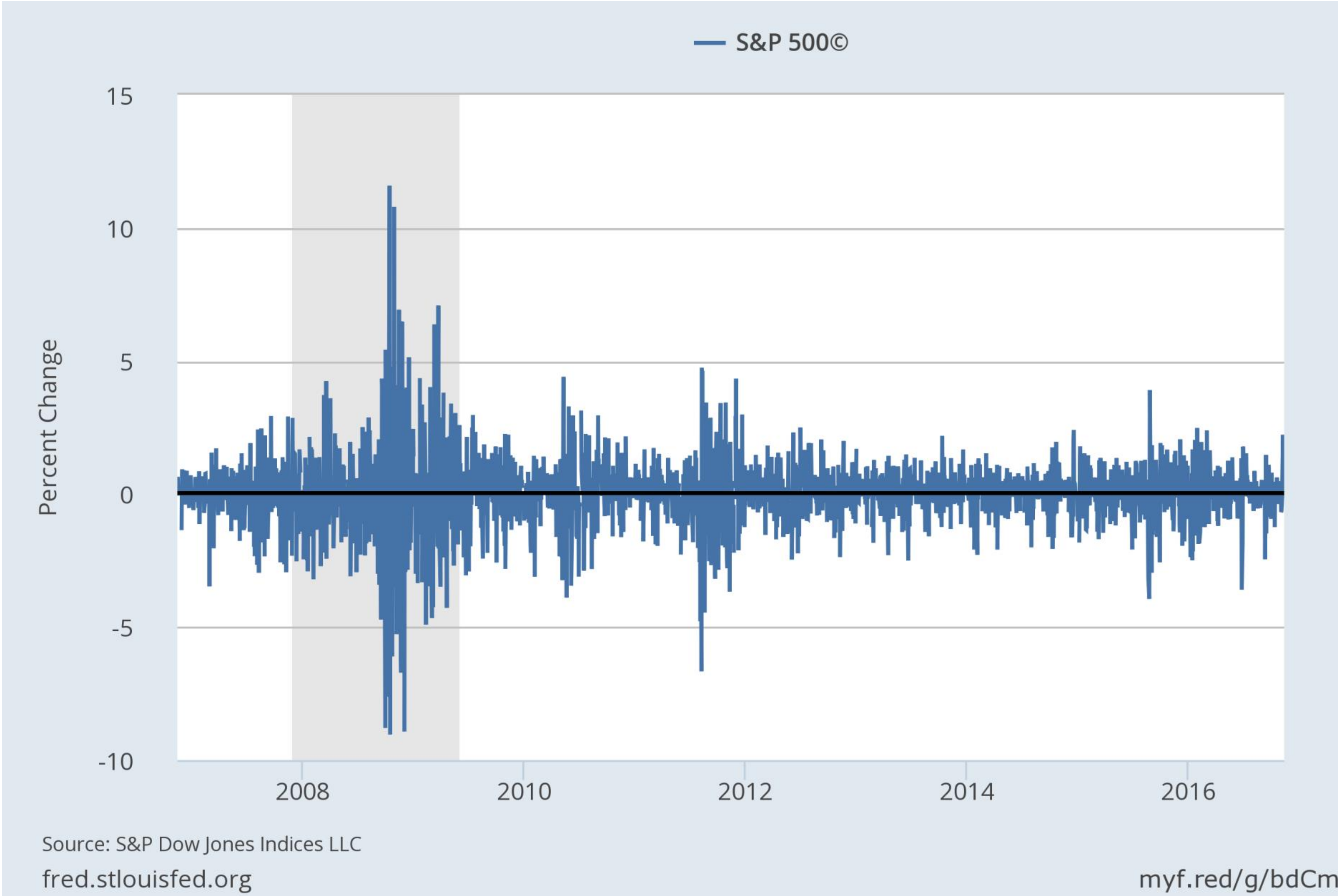


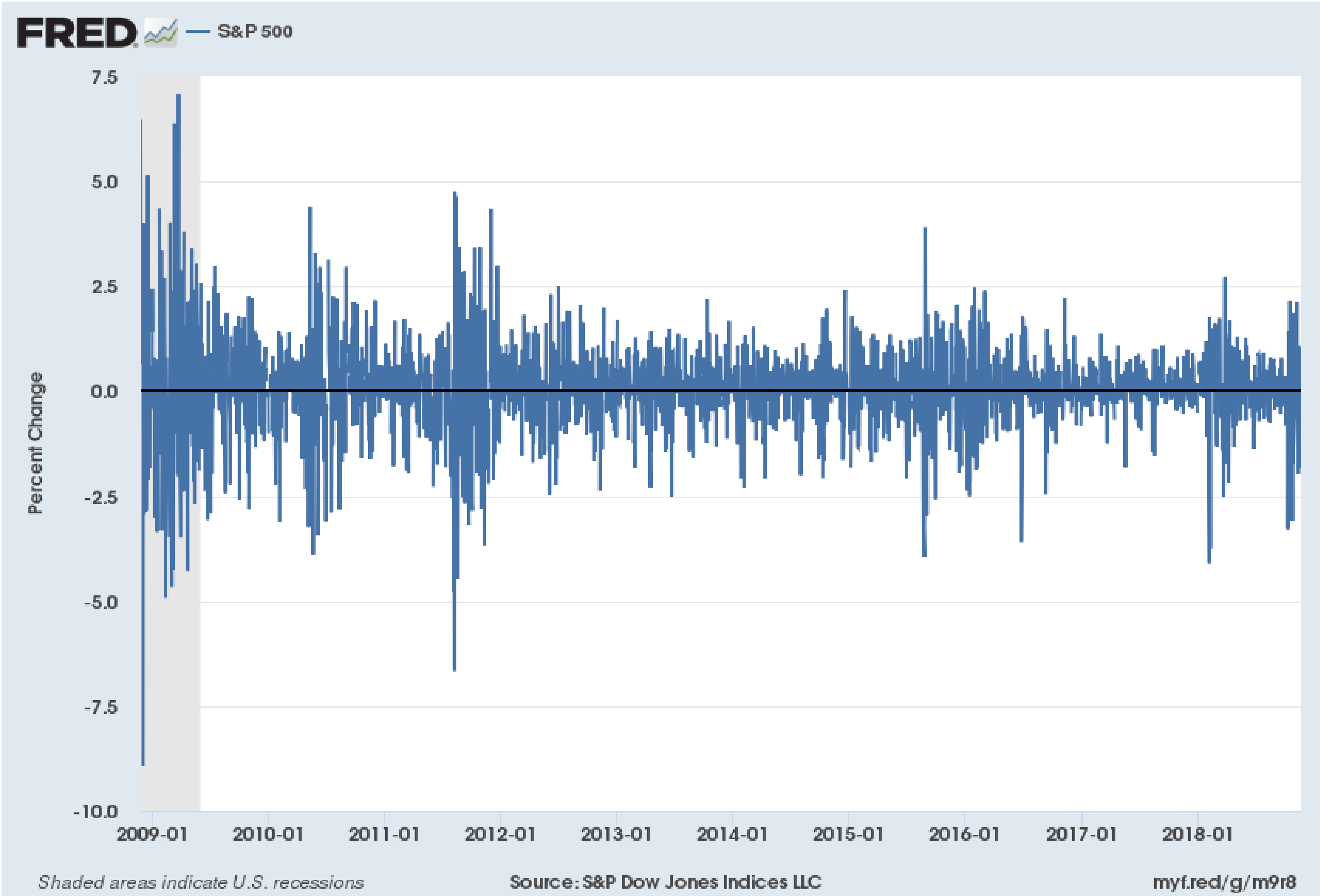
FRED
Civilian Unemployment Rate
11
10
9
8
7
6
5
4
3
2
Percent
1950
1960
1970
1980
1990
2000
2010
Shaded areas indicate U.S. recessions
Source: U.S. Bureau of Labor Statistics
myf.red/g/lQTW


FRED
Real Gross Domestic Product
4
3
2
1
0
-1
-2
-3
Percent Change
1950
1960
1970
1980
1990
2000
2010
Shaded areas indicate U.S. recessions
Source: U.S. Bureau of Economic Analysis
myf.red/g/m9qu

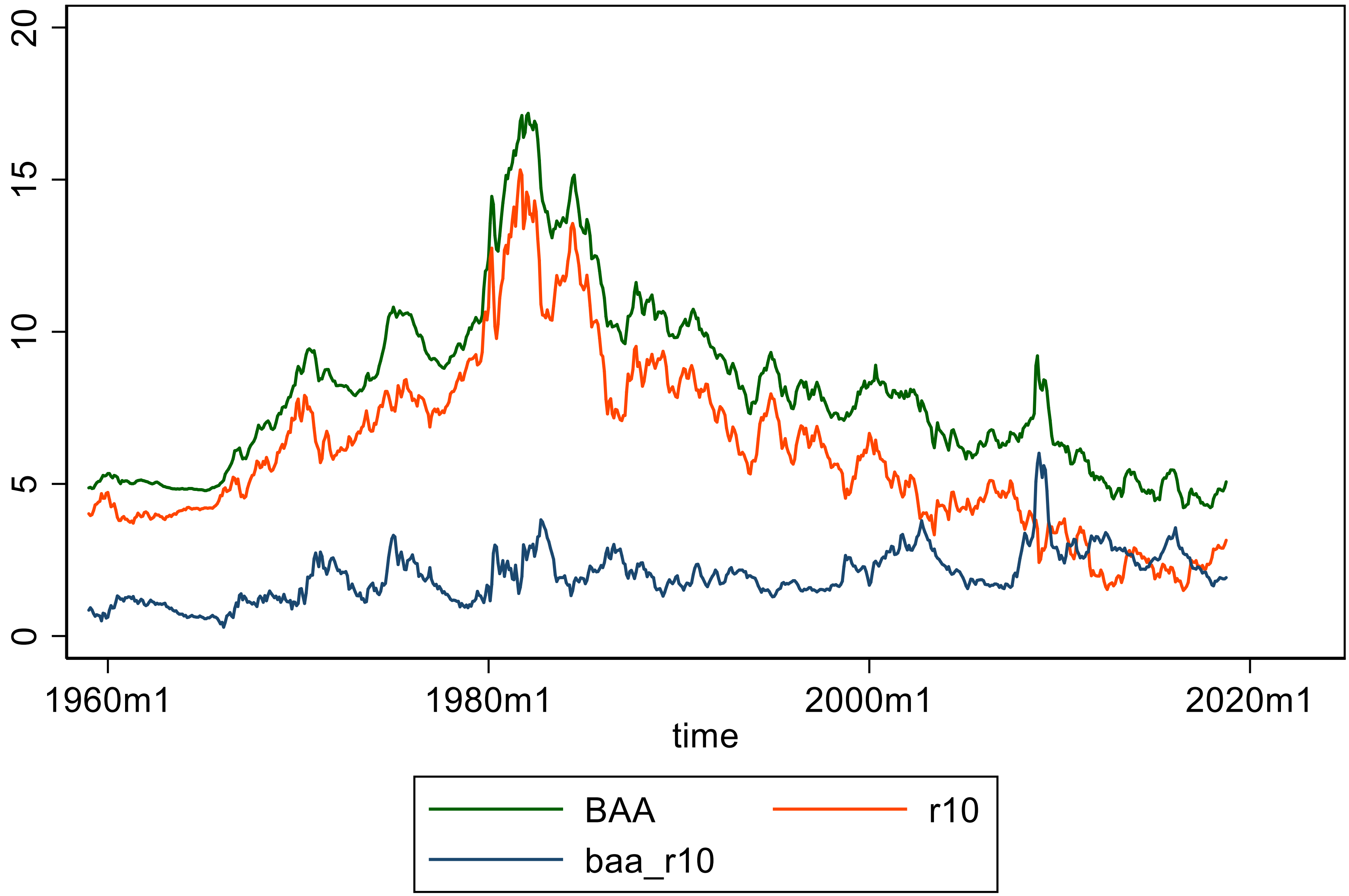

FRED
Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index)
Compounded Annual Rate of Change
12.5
10.0
7.5
5.0
2.5
0.0
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
Shaded areas indicate U.S. recessions
Source: U.S. Bureau of Economic Analysis
myf.red/g/m9r0

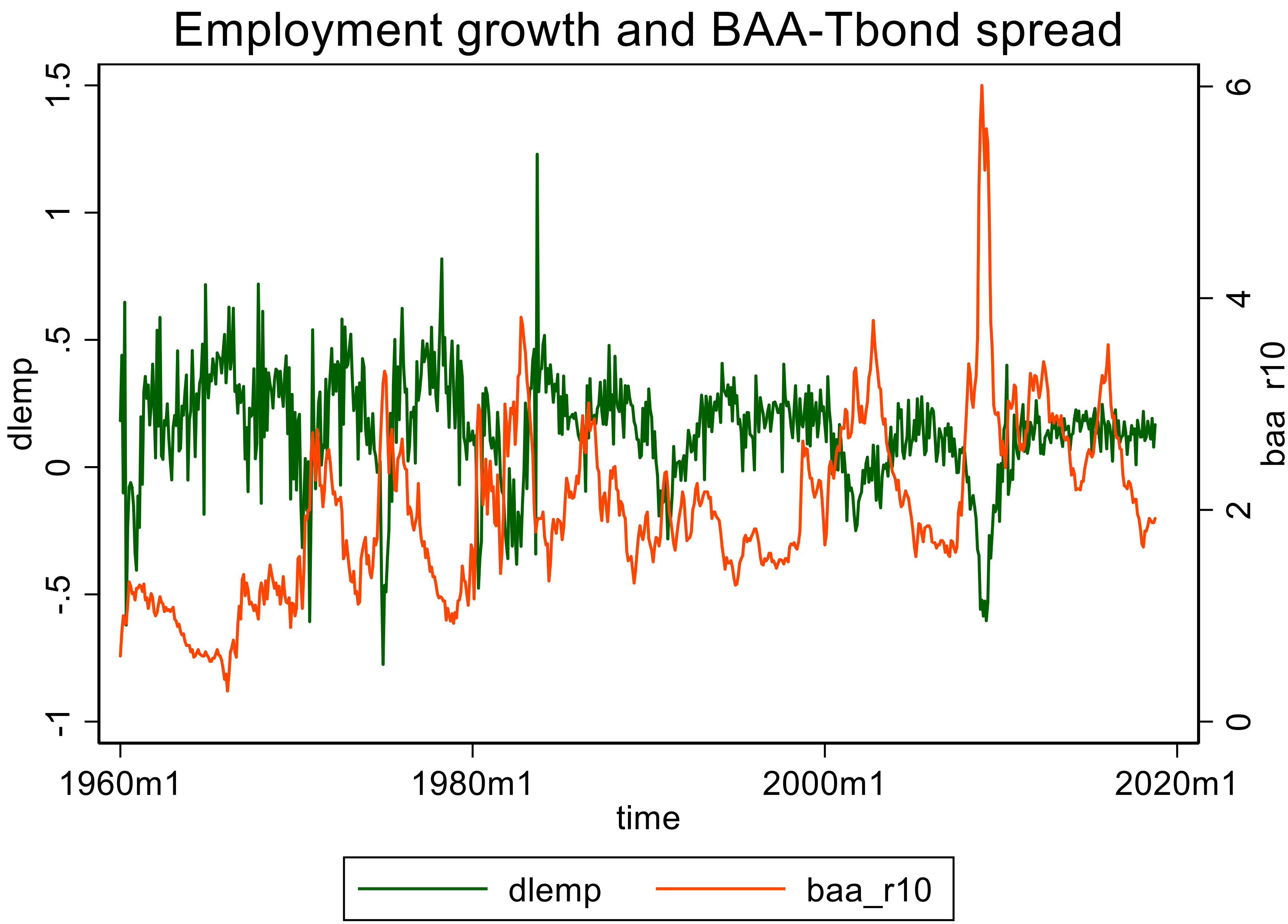

FRED
10-Year Treasury Constant Maturity Rate
3-Month Treasury Bill: Secondary Market Rate
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Percent
1960
1970
1980
1990
2000
2010
Shaded areas indicate U.S. recessions
Source: Board of Governors of the Federal Reserve System (US)
myf.red/g/m9r4


S&P 500©
15
10
5
0
-5
-10
Percent Change
2008
2010
2012
2014
2016
Source: S&P Dow Jones Indices LLC
fred.stlouisfed.org
myf.red/g/bdCm


FRED
S&P 500
Percent Change
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0
2009-01
2010-01
2011-01
2012-01
2013-01
2014-01
2015-01
2016-01
2017-01
2018-01
Shaded areas indicate U.S. recessions
Source: S&P Dow Jones Indices LLC
myf.red/g/m9r8

Baa bonds, 10-yr Tbond rates, & spread




Employment growth and BAA-Tbond spread
dlemp
1.5
1
.5
0
-.5
-1
baa_r10
6
4
2
0
1960m1
1980m1
2000m1
2020m1
time
dlemp
baa_r10

Lags, percent changes, and logarithms

Example: Quarterly rate of inflation at an annual rate (U.S.)

CPI = Consumer Price Index (Bureau of Labor Statistics)

- CPI in the first quarter of 2004 (2004:I) = 186.57
- CPI in the second quarter of 2004 (2004:II) = 188.60
- Percentage change in CPI, 2004:I to 2004:II

$$= 100 \times \left(\frac{188.60 - 186.57}{186.57} \right) = 100 \times \left(\frac{2.03}{186.57} \right) = 1.088\%$$

- Percentage change in CPI, 2004:I to 2004:II, *at an annual rate* =

  4×1.088 = **4.359**% ≈ **4.4**% (percent per year)

- Like interest rates, inflation rates are (as a matter of convention) reported at an annual rate.
- Using the logarithmic approximation to percent changes yields

  $4 \times 100 \times [\log(188.60) - \log(186.57)]$ = **4.329**%

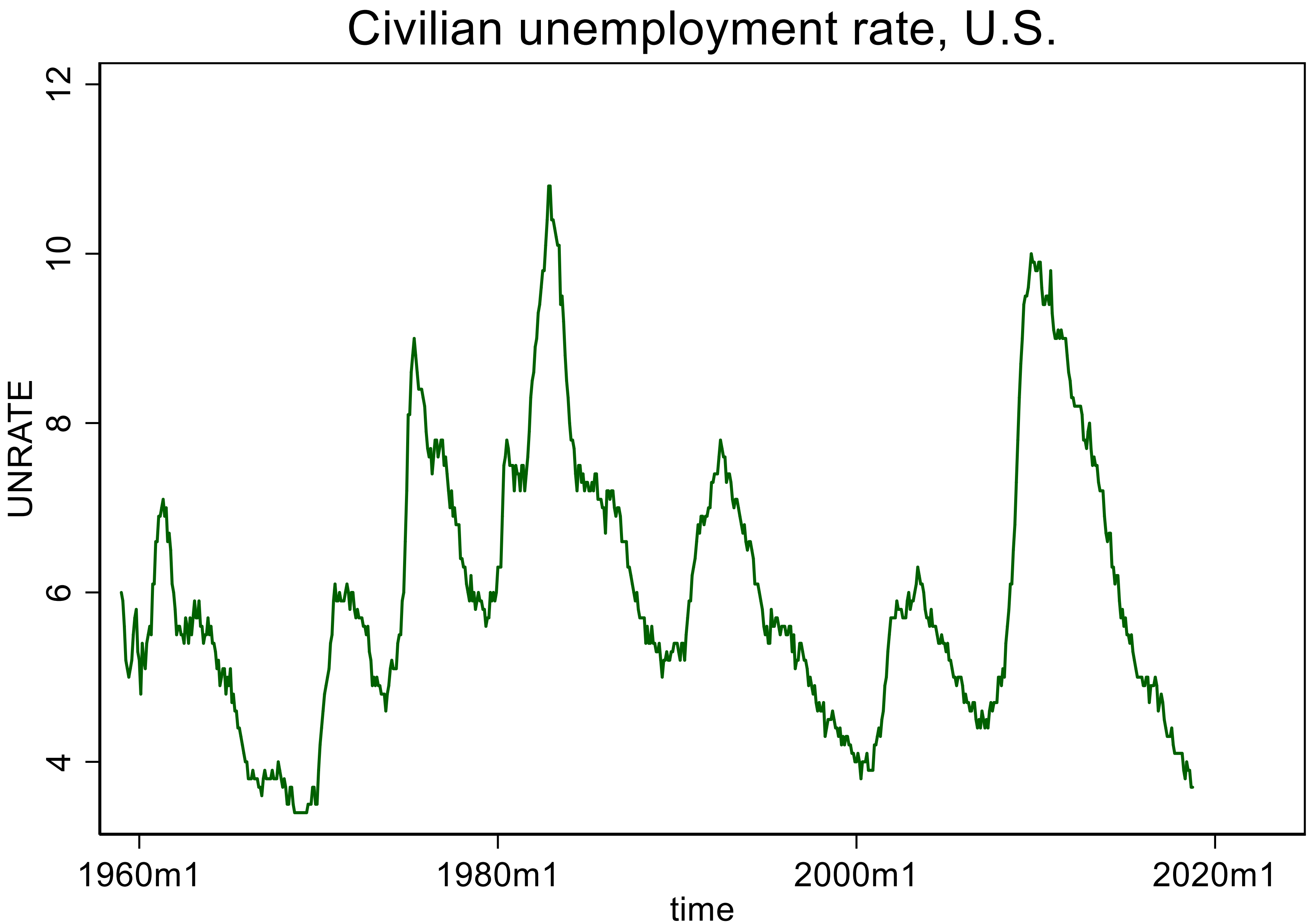
Example: US CPI inflation – its first lag and its change

TABLE 14.1 Inflation in the United States in 2004 and the First Quarter of 2005

Quarter	U.S. CPI	Rate of Inflation at an Annual Rate (Inf_t)	First Lag (Inf_{t-1})	Change in Inflation (ΔInf_t)
2004:I	186.57	3.8	0.9	2.9
2004:II	188.60	4.4	3.8	0.6
2004:III	189.37	1.6	4.4	−2.8
2004:IV	191.03	3.5	1.6	1.9
2005:I	192.17	2.4	3.5	−1.1

The annualized rate of inflation is the percentage change in the CPI from the previous quarter to the current quarter, times four. The first lag of inflation is its value in the previous quarter, and the change in inflation is the current inflation rate minus its first lag. All entries are rounded to the nearest decimal.

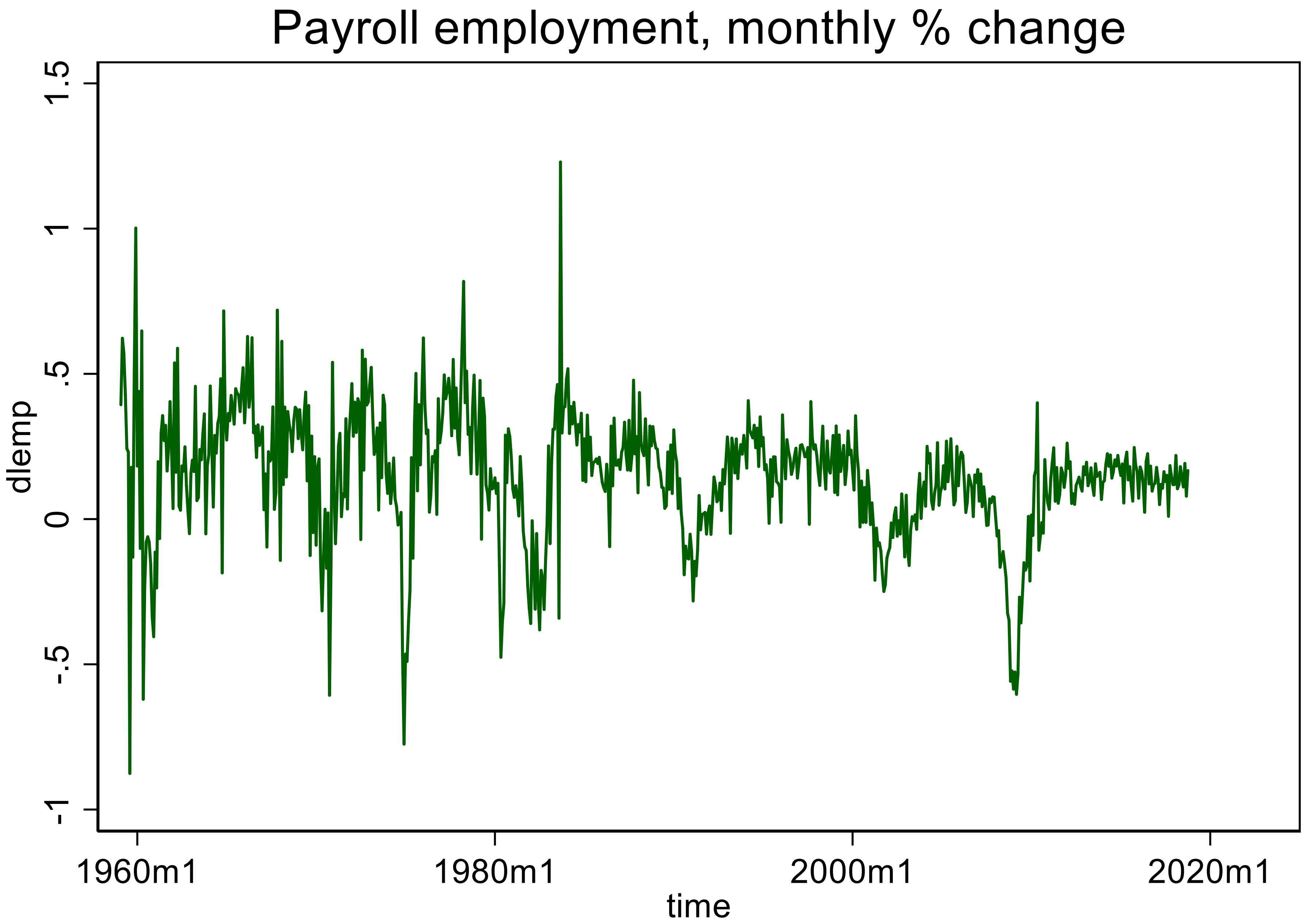
Autocorrelations: examples



```
. corrgram unrate

                                                      -1       0       1 -1       0       1
LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.9921   0.9951   709.68  0.0000          |-------           |-------
2        0.9827  -0.1244   1406.9  0.0000          |-------           |
3        0.9701  -0.2613   2087.3  0.0000          |-------         --|
4        0.9550  -0.1656   2747.6  0.0000          |-------          -|
5        0.9371  -0.1603   3384.2  0.0000          |-------          -|
6        0.9168  -0.0839   3994.5  0.0000          |-------           |
7        0.8947  -0.0546   4576.5  0.0000          |-------           |
8        0.8715  -0.0037   5129.5  0.0000          |------            |
9        0.8470  -0.0100   5652.5  0.0000          |------            |
10       0.8210  -0.0380   6144.7  0.0000          |------            |
11       0.7948   0.0421   6606.6  0.0000          |------            |
12       0.7670  -0.0751   7037.4  0.0000          |------            |
```

Note: the only results from this command we will use are the autocorrelations (AC).


Payroll employment, monthly % change
dlemp
1.5
1
.5
0
-.5
-1
1960m1
1980m1
2000m1
2020m1
time

```
. corrgram dlemp

                                          -1       0       1 -1       0       1
LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1       0.5449   0.5449   213.75  0.0000           |----              |----
2       0.5837   0.4102   459.41  0.0000           |----              |---
3       0.4952   0.1493   636.46  0.0000           |---               |-
4       0.4471   0.0492   780.99  0.0000           |---               |
5       0.4169   0.0484   906.83  0.0000           |---               |
6       0.3633   0.0010   1002.5  0.0000           |--                |
7       0.3532   0.0201   1093.1  0.0000           |--                |
8       0.3122   0.0021     1164  0.0000           |--                |
9       0.3281   0.0691   1242.4  0.0000           |--                |
10      0.2372  -0.0733   1283.4  0.0000           |-                 |
11      0.2184  -0.0417   1318.2  0.0000           |-                 |
12      0.1395  -0.0983   1332.5  0.0000           |-                 |
```

AR(1) model of employment growth: STATA

Create monthly employment growth.

The name of the employment series is "payems"

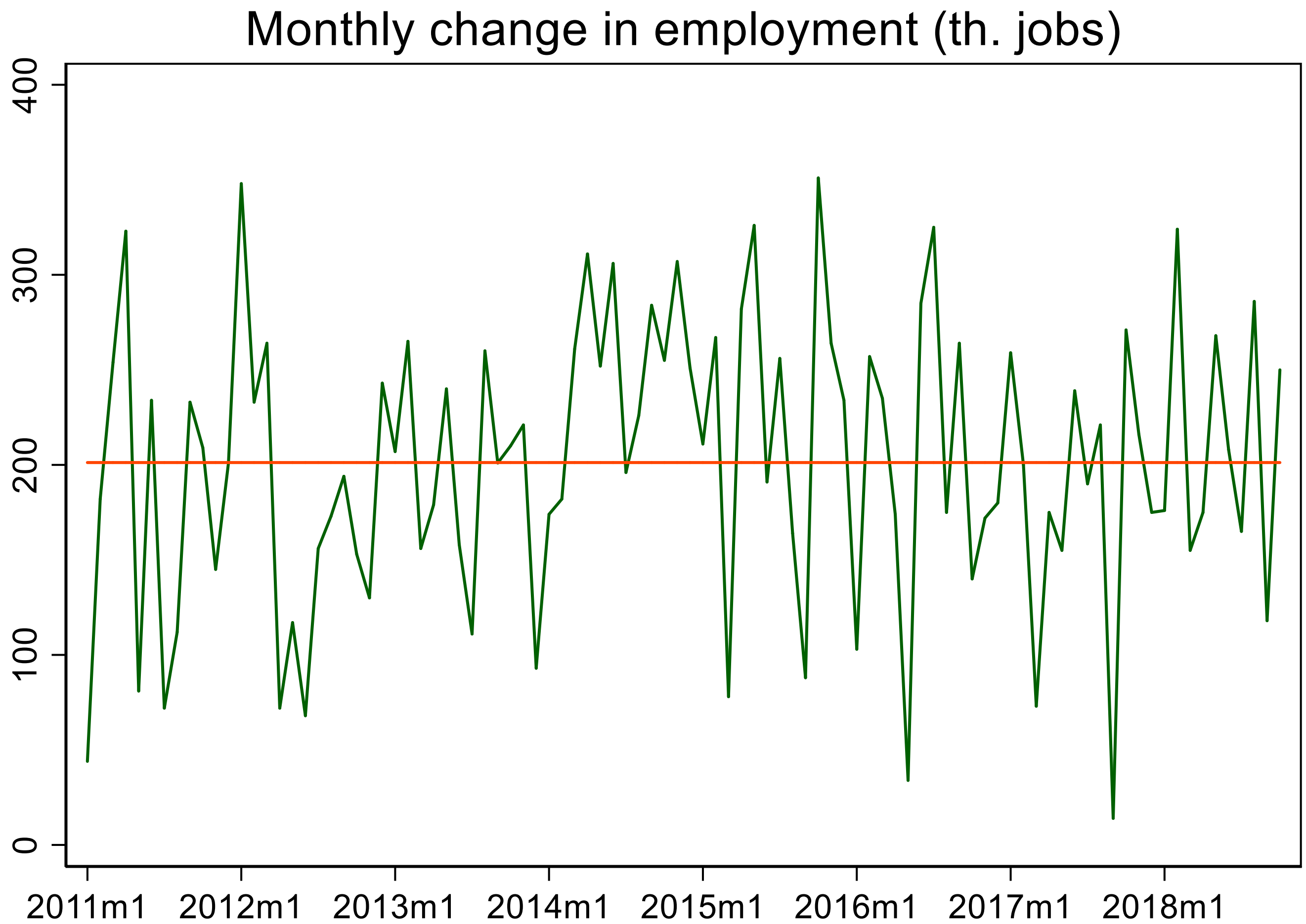
```
. gen lemp = log(payems);                    create log employment
. gen dlemp = 100*(lemp[_n]-lemp[_n-1]);  dlemp is 100*first difference of log
                                          employment, which is approximately the
                                         monthly percentage growth of employment

. global tn "2018m10"
. list time payems lemp dlemp if tin(2016m1,$tn)

      +--------------------------------------------------+
      |    time   payems   demp       lemp       dlemp  |
      |--------------------------------------------------|
 709. |  2018m1   147801    176   11.90362    .1191139  |
 710. |  2018m2   148125    324   11.90581     .219059  |
 711. |  2018m3   148280    155   11.90686    .1045227  |
 712. |  2018m4   148455    175   11.90804    .1179695  |
 713. |  2018m5   148723    268   11.90984    .1803398  |
      |--------------------------------------------------|
 714. |  2018m6   148931    208   11.91124    .1398087  |
 715. |  2018m7   149096    165   11.91235    .1107216  |
 716. |  2018m8   149382    286   11.91426    .1915932  |
 717. |  2018m9   149500    118   11.91505    .0789642  |
 718. | 2018m10   149750    250   11.91672    .1670837  |
```

Plot the series

```
su demp if tin(2011q1,$tn)
gen demp_mn = r(mean)
tsline demp demp_mn if tin(2011q1,$tn), ///
 title("Monthly change in employment (th. jobs)") legend(off) ttitle("")
```



.

2011q1 - 2018q10: mean = 201k, SD = 75k

Estimate an AR(1) model: regress $100\Delta\ln(Emp_t)$ against $100\Delta\ln(Emp_t)$ using data from 1962m1 – 2018m9

```
. global tnm1 "2018m9"
. reg dlemp L.dlemp if tin(1962m1,$tnm1), r

Linear regression                               Number of obs     =        657
                                                F(1, 655)         =     152.58
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3521
                                                Root MSE          =     .17117

------------------------------------------------------------------------------
             |               Robust
       dlemp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dlemp |
         L1. |    .593337   .0480337    12.35   0.000     .4990185    .6876556
             |
       _cons |   .0600784   .0108659     5.53   0.000     .0387423    .0814146
------------------------------------------------------------------------------
```

$$y_t = \underset{(.011)}{0.060} + \underset{(.048)}{0.593}y_{t-1}, \qquad y_t = 100\Delta\ln(Emp_t)$$

Compute the forecast for 2016m10

```
.  predict yhat_ar1_6216
. list time payems demp dlemp yhat_ar1_6216 if tin(2016m1,$tn), noobs

  +-------------------------------------------------+
  |    time   payems   demp       dlemp   y~1_6215 |
  |-------------------------------------------------|
  |    2018m5   148723    268    .1803398   .1300741 |
  |  2018m6   148931    208    .1398087   .1670807 |
  |  2018m7   149096    165    .1107216   .1430321 |
  |  2018m8   149382    286    .1915932   .1257737 |
  |  2018m9   149500    118    .0789642   .1737577 |
  |-------------------------------------------------|
  | 2018m10   149750    250    .1670837   .1069308 |
  | 2018m11        .      .           .   .1592154 |
```

$$100\Delta \ln(Emp_{2016m10}) = 0.060 + 0.593\times 0.0789642 = 0.107 \text{ (\%)}$$

= increase in employment of 160,000 jobs

Actual was +250k, forecast error was -90k (ouch)

Example: AR(4) model of employment growth – STATA

```
. reg dlemp L(1/4).dlemp if tin(1962m1,$tnm1), r

Linear regression                                      Number of obs   =       657
                                                       F(4, 652)       =    108.45
                                                       Prob > F        =    0.0000
                                                       R-squared       =    0.4935
                                                       Root MSE        =     .1517

------------------------------------------------------------------------------
             |               Robust
       dlemp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dlemp |
         L1. |   .2367451    .069784     3.39   0.001     .0997167    .3737735
         L2. |    .307935   .0561869     5.48   0.000     .1976058    .4182642
         L3. |   .1906997   .0498195     3.83   0.000     .0928737    .2885257
         L4. |    .095125   .0396045     2.40   0.017     .0173573    .1728927
             |
       _cons |   .0247342    .008958     2.76   0.006     .0071442    .0423243
------------------------------------------------------------------------------
```

NOTES

- L(1/4).dlemp *generates the list: L1.dlemp L2.dlemp L3.dlemp L4.dlemp*
- *L1,…,L4 refer to the first, second,… 4th lags of dlemp*

AR(4) model of employment growth ctd.

```
.  dis "Adjusted Rsquared = " e(r2_a)       e(r2_a) is the rbar-squared
   of the most recently run regression
Adjusted Rsquared = .49067357

.  test L2.dlemp L3.dlemp L4.dlemp


 ( 1)  L2.dlemp = 0
 ( 2)  L3.dlemp = 0
 ( 3)  L4.dlemp = 0

       F(  3,   652) =   21.19
            Prob > F =    0.0000
```

ADL: forecast employment growth using BAA-10yr Treasury spread

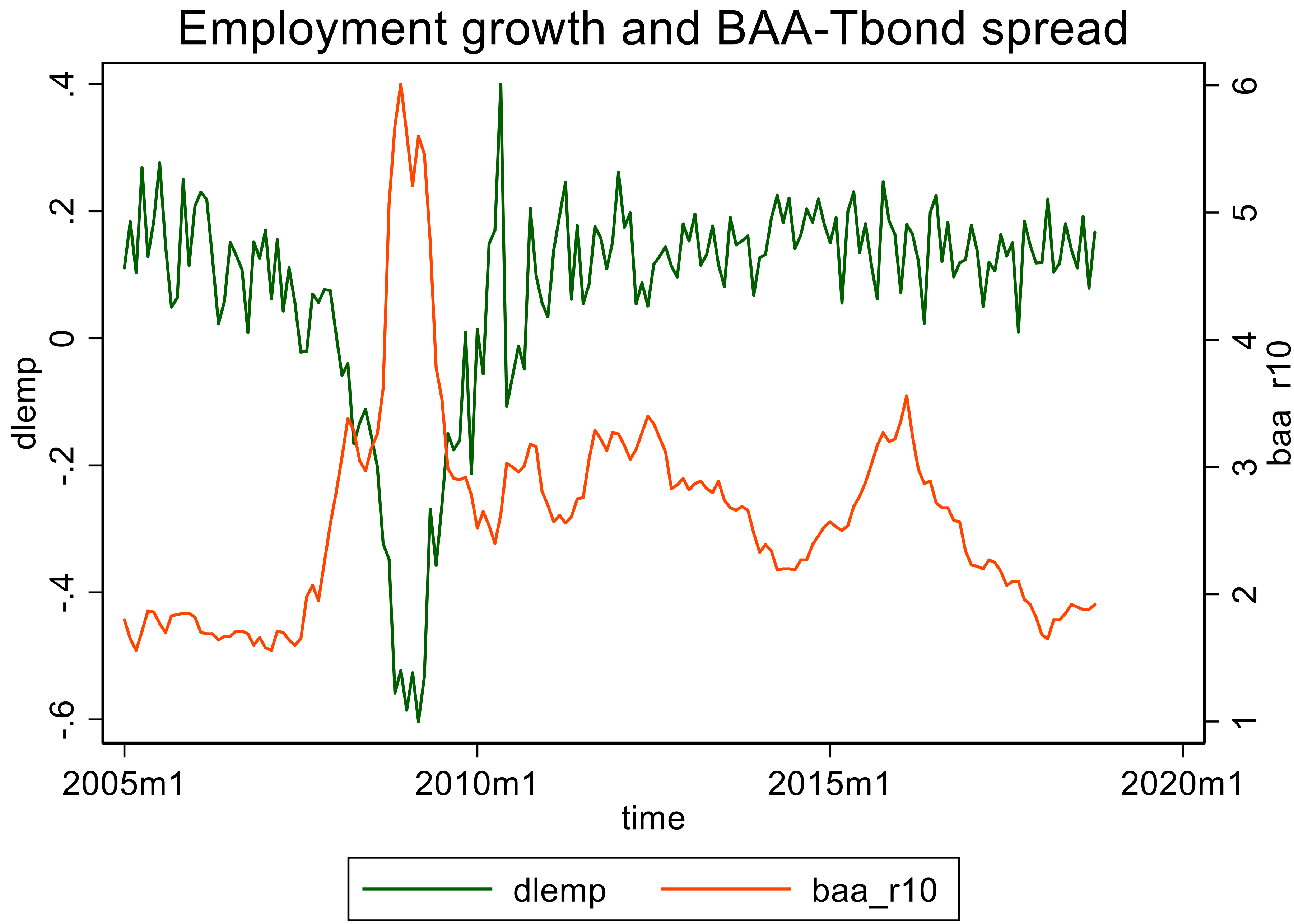
ADL(1,1), then ADL(4,4)

```
. reg dlemp L1.dlemp L1.baa_r10 if tin(1962m1,$tn), r

Linear regression                               Number of obs     =        658
                                                F(2, 655)         =     188.85
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4101
                                                Root MSE          =     .16331

------------------------------------------------------------------------------
             |               Robust
       dlemp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dlemp |
         L1. |   .4423618   .0557219     7.94   0.000     .3329466     .551777
             |
     baa_r10 |
         L1. |  -.0735542   .0111138    -6.62   0.000    -.0953771   -.0517313
             |
       _cons |   .2320476   .0317115     7.32   0.000     .1697792     .294316
------------------------------------------------------------------------------

.  dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = .40832468
```


Employment growth and BAA-Tbond spread
dlemp
.4
.2
0
-.2
-.4
-.6
baa_r10
6
5
4
3
2
1
2005m1
2010m1
2015m1
2020m1
time
dlemp
baa_r10

```
. reg dlemp L(1/4).dlemp L(1/4).baa_r10 if tin(1962m1,$tnm1), r

Linear regression                               Number of obs     =        682
                                                F(8, 673)         =      70.62
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5247
                                                Root MSE          =     .14477

------------------------------------------------------------------------------
             |               Robust
       dlemp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dlemp |
         L1. |   .1811643   .0686887     2.64   0.009     .0462943    .3160343
         L2. |   .2680536   .0540608     4.96   0.000     .1619054    .3742018
         L3. |   .1892603   .0487062     3.89   0.000     .0936258    .2848947
         L4. |    .111213   .0407134     2.73   0.006     .0312725    .1911535
             |
     baa_r10 |
         L1. |  -.1321763   .0424257    -3.12   0.002     -.215479   -.0488737
         L2. |   .0260404   .0624243     0.42   0.677    -.0965294    .1486102
         L3. |   .0083611   .0621931     0.13   0.893    -.1137548     .130477
         L4. |   .0793762   .0385339     2.06   0.040     .0037151    .1550372
             |
       _cons |   .0742436   .0278167     2.67   0.008     .0196257    .1288615
------------------------------------------------------------------------------
.  dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = .51904216
```

Tests of marginal predictive content

(a) **Granger Causality tests** (do additional predictors matter, given lagged dependent variable)

```
.  testparm L(1/4).baa_r10

 ( 1)  L.baa_r10 = 0
 ( 2)  L2.baa_r10 = 0
 ( 3)  L3.baa_r10 = 0
 ( 4)  L4.baa_r10 = 0

       F(  4,   673) =    7.34
            Prob > F =    0.0000
```

Note: Granger causality tests can be done in bivariate equations (y on lagged y and lagged x, testing lags of x) or multivariate equations (y on lagged x1, x2, testing lags of x1 and x2) – here it is the former

(b) Tests of lag length

```
.  testparm L(2/4).baa_r10

 ( 1)  L2.baa_r10 = 0
 ( 2)  L3.baa_r10 = 0
 ( 3)  L4.baa_r10 = 0

       F(  3,   673) =    5.52
            Prob > F =    0.0009

.  testparm L(2/4).dlemp

 ( 1)  L2.dlemp = 0
 ( 2)  L3.dlemp = 0
 ( 3)  L4.dlemp = 0

       F(  3,   673) =   17.54
            Prob > F =    0.0000
```

How did the forecasts do? Data released 8:30am 11/2/18

```
. gen demp_ar1_62tn = payems[_n-1]*(yhat_ar1_62tn/100)
. gen demp_ar2_62tn = payems[_n-1]*(yhat_ar2_62tn/100)
. gen demp_adl44_62tn = payems[_n-1]*(yhat_adl44_62tn/100)
. list time demp demp_ar1_62tn demp_ar2_62tn demp_ar4_62tn demp_adl44_62tn if
tin(2016m1,$tn), noobs
```

time	demp	d~1_62tn	d~2_62tn	d~dl44~n
2018m1	176	192.6066	202.1776	241.3281
2018m2	324	193.2544	185.524	255.4362
2018m3	155	281.5179	237.5632	263.4986
2018m4	175	181.0434	240.4171	224.1948
2018m5	268	193.1015	177.0602	199.5865
2018m6	208	248.4875	217.8466	194.8989
2018m7	165	213.0191	235.7663	199.8463
2018m8	286	187.5235	195.91	203.1242
2018m9	118	259.5628	220.1483	222.6293
2018m10	250	159.8616	212.1044	217.0184

- Units of are thousands of jobs, change from 2018m9 to 2018m10

BIC: AR(6) model of $\Delta\ln(Emp_t)$

```
. reg dlemp L(1/6).dlemp if tin(1962m1,$tn), r

Linear regression                               Number of obs     =        658
                                                F(6, 651)         =      72.46
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4944
                                                Root MSE          =     .15166

------------------------------------------------------------------------------
             |               Robust
       dlemp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dlemp |
         L1. |   .2339684    .070495     3.32   0.001      .095543    .3723938
         L2. |   .3025803    .057986     5.22   0.000     .1887179    .4164428
         L3. |   .1825232   .0534025     3.42   0.001     .0776611    .2873854
         L4. |   .0905325   .0405063     2.24   0.026     .0109936    .1700714
         L5. |   .0351701   .0442325     0.80   0.427    -.0516858    .1220259
         L6. |  -.0104761    .041289    -0.25   0.800     -.091552    .0705999
             |
       _cons |    .024162   .0091706     2.63   0.009     .0061545    .0421695--
------------------------------------------------------------------------------
.  dis "BIC = " ln(e(rss)/e(N)) + e(df_m)*ln(e(N))/e(N)
BIC = -3.4278561
```

Summary table: $\Delta\ln(Emp_t)$, AR(0) – AR(6):

p	BIC	*Adjusted* R^2
0	-3.0994599	0
1	-3.5257636	.35154748
2	-3.6864851	.46310271
3	-3.6838869	.48682197
4	-3.6239168	.49067357
5	-3.5362389	.49044313
6	-3.4278561	.48971109

BIC-selected lag length = 2