Seasonal Dummy Model

- Deterministic seasonality S_t can be written as a function of seasonal dummy variables
- Let s be the seasonal frequency
  - s=4 for quarterly
  - s=12 for monthly
- Let D_{1t} , D_{2t} , D_{3t} ,…, D_{st} be seasonal dummies
  - D_{1t}= 1 if s is the first period, otherwise D_{1t}= 0
  - D_{2t}= 1 if s is the second period, otherwise D_{2t}= 0
- At any time period t, one of the seasonal dummies D_{1t} , D_{2t} , D_{3t} ,…, D_{st} will equal 1, all the others will equal 0.

Seasonal Dummy Model

- Deterministic seasonality

$$S_t = \begin{cases} \gamma_1 & if \quad t = January \\ \gamma_2 & if \quad t = February \\ \vdots & \qquad \vdots \\ \gamma_{12} & if \quad t = December \end{cases}$$

$$= \sum_{i=1}^{s} \gamma_i D_{it}$$

a linear function of the dummy variables

Estimation

- Least squares regression

$$y_{t+h} = \sum_{i=1}^{s} \gamma_i D_{it} + e_t$$

$$= \alpha + \sum_{i=1}^{s-1} \beta_i D_{it} + e_t$$

- You can either
  - Regress y on all the seasonal dummies, omitting the intercept, or
  - Regress y on an intercept and the seasonal dummies, omitting one dummy (one season, e.g. December)
- You cannot regress on both the intercept plus all seasonal dummies, for they would be collinear and redundant.

Interpreting Coefficients

- In the model

$$S_t = \alpha + \sum_{i=1}^{s-1} \beta_i D_{it}$$

  the intercept $\alpha = \gamma_s$ is the seasonality in the omitted season.
- The coefficients $\beta_i = \gamma_i - \gamma_s$ are the difference in the seasonal component from the s'th period.

STATA Programming

- If the time index is *t* and is formatted as a time index, you can determine the period using the commands

```
generate m=month(dofm(t))
generate q=quarter(dofq(t))
```

for monthly and quarterly data, respectively

(See dates and times in STATA Data manual)

Creating Dummies

- If *m* is the month (1 for January, 2 for February, etc.), then
  - **generate m1=(m==1)**
  - This creates a dummy variable "m1" for January
  - Then
  - **regress y m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11**

    or
  - **regress y m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12, noconstant**
- Easier
  - Type "b12.m" in the regressor list
  - **regress y b12.m**
  - This includes dummies for months 1 through 11, omits 12
  - Same as mechanically listing the eleven dummies, but easier.
  - It is important that "m" be the numerical month (1 for January, 2 for February, etc.)

Estimation

```
. use "C:\Users\Bruce Hansen\Documents\docs\classdocs\390\housingstarts.dta"

. regress total b12.m
```

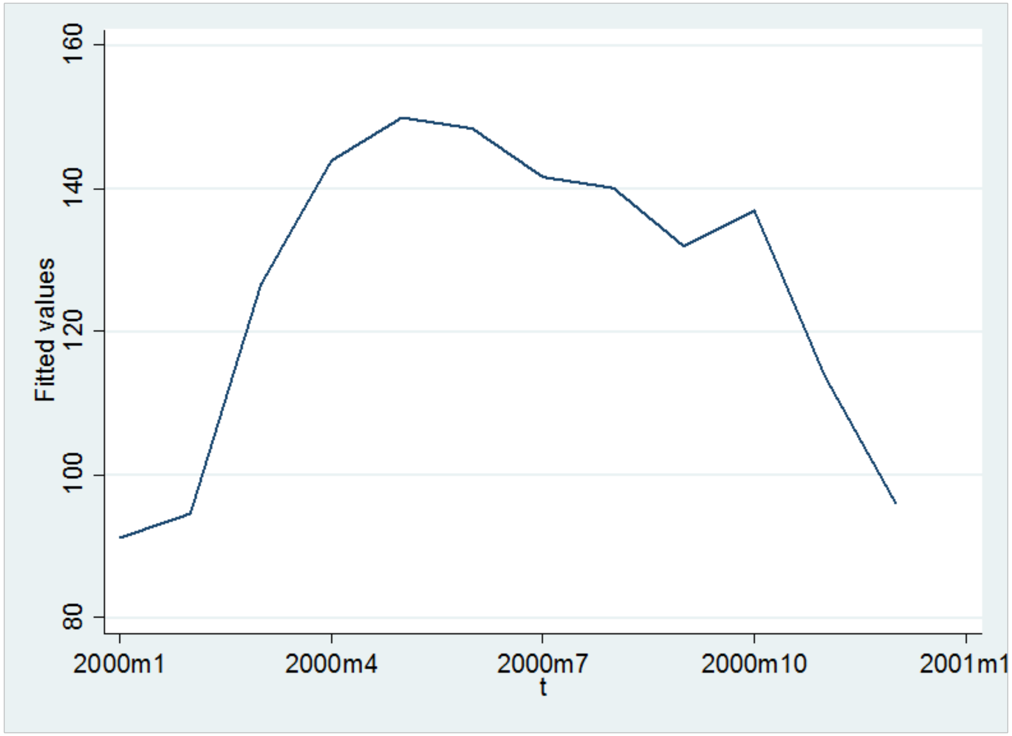
Source	SS	df	MS
Model	267331.386	11	24302.8533
Residual	557738.603	600	929.564339
Total	825069.989	611	1350.36005

Number of obs	=	612
F(11, 600)	=	26.14
Prob > F	=	0.0000
R-squared	=	0.3240
Adj R-squared	=	0.3116
Root MSE	=	30.489

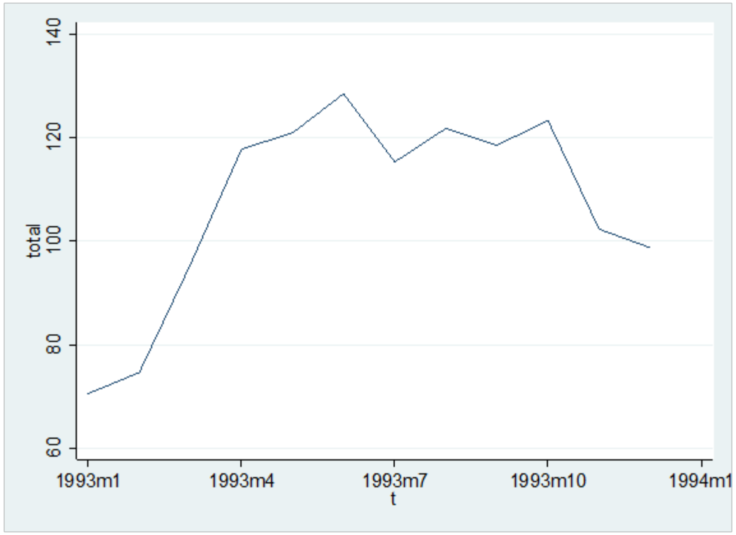
total	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
m						
1	-4.931373	6.037674	-0.82	0.414	-16.78891	6.92617
2	-1.547058	6.037674	-0.26	0.798	-13.4046	10.31048
3	30.51765	6.037674	5.05	0.000	18.66011	42.37519
4	47.82353	6.037674	7.92	0.000	35.96599	59.68107
5	53.87255	6.037674	8.92	0.000	42.01501	65.73009
6	52.31569	6.037674	8.66	0.000	40.45815	64.17323
7	45.55294	6.037674	7.54	0.000	33.6954	57.41048
8	43.95294	6.037674	7.28	0.000	32.0954	55.81048
9	35.82745	6.037674	5.93	0.000	23.96991	47.68499
10	40.84902	6.037674	6.77	0.000	28.99148	52.70656
11	17.64706	6.037674	2.92	0.004	5.789517	29.5046
_cons	96.07843	4.26928	22.50	0.000	87.69388	104.463

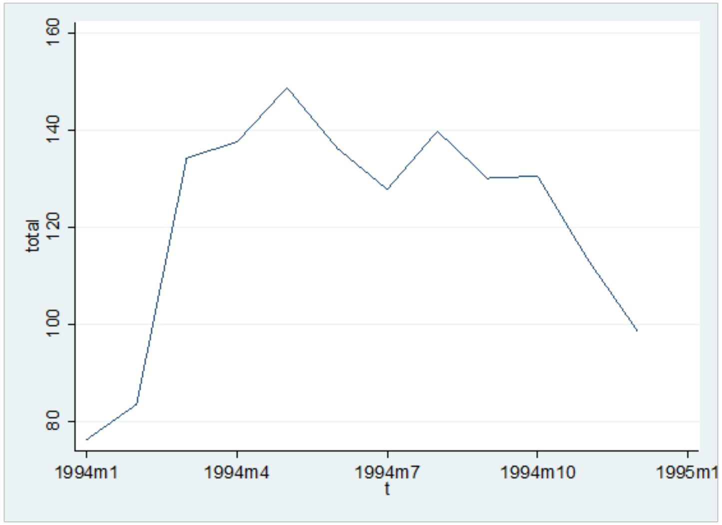
Estimated Seasonality – Housing Starts

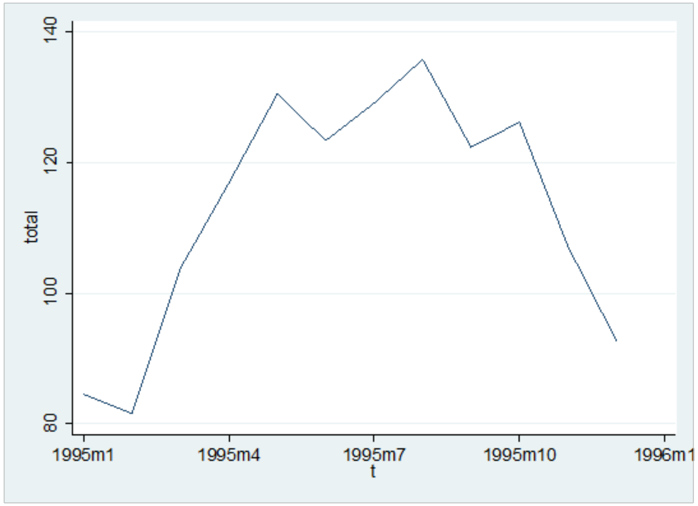
January	91
February	95
March	127
April	144
May	150
June	148
July	142
August	140
September	132
October	137
November	114
December	96



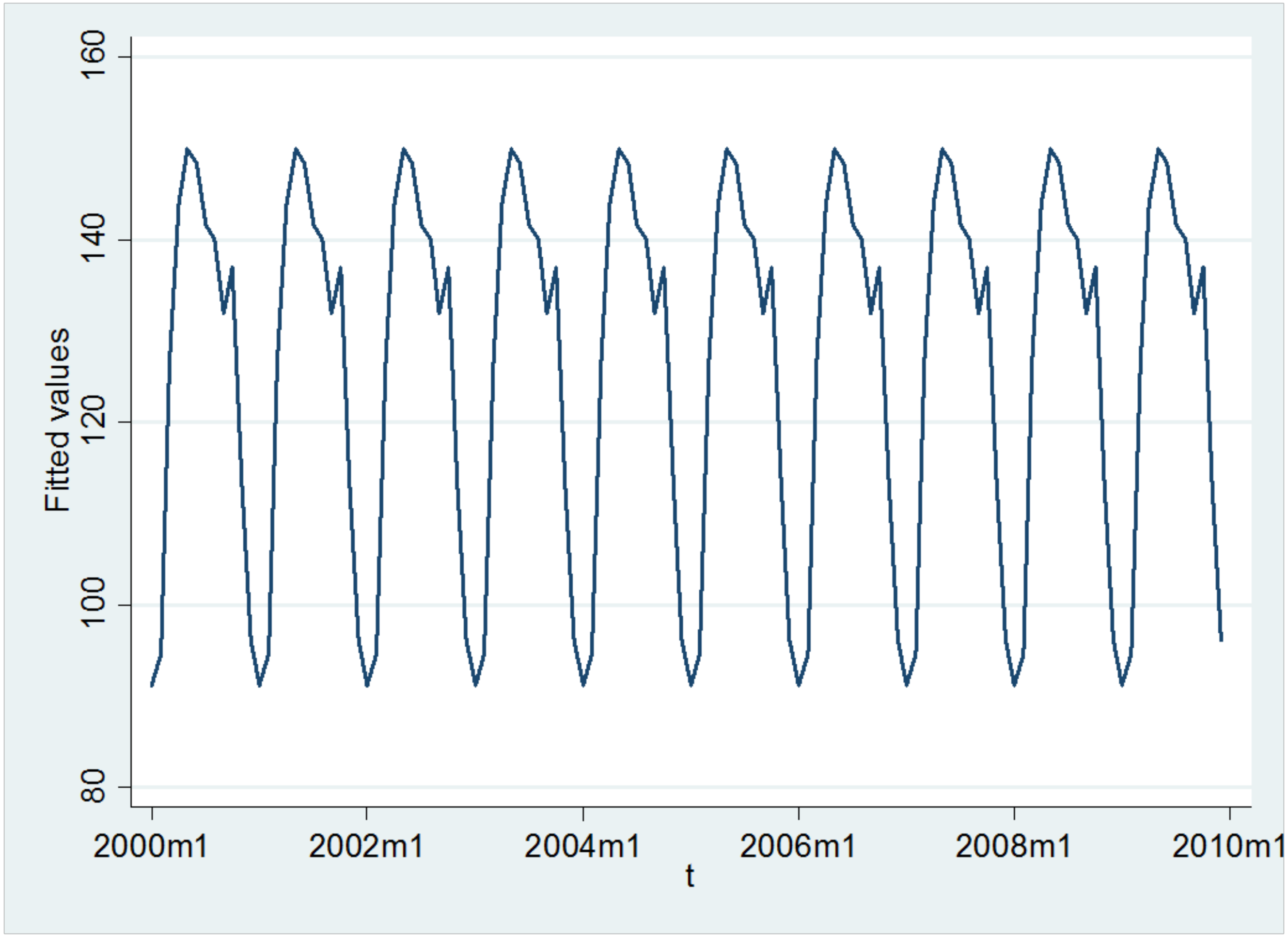
Housing Starts, by year, and estimated seasonality







Predicted Values



Example 1
Unemployment Rate

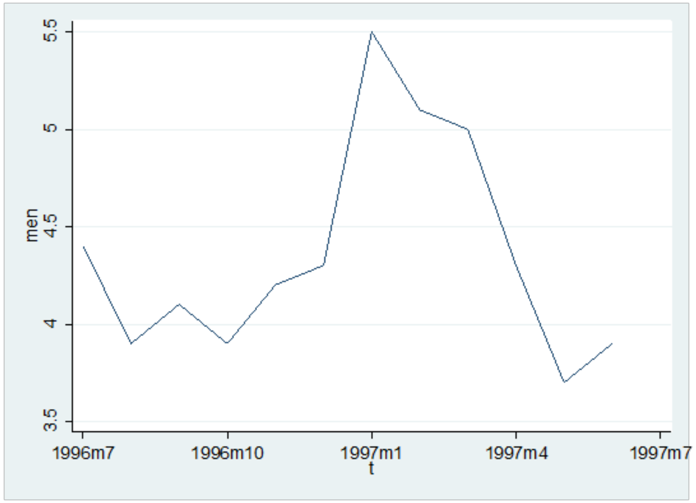
```
. use ur_nsa

. regress ur b12.m
```

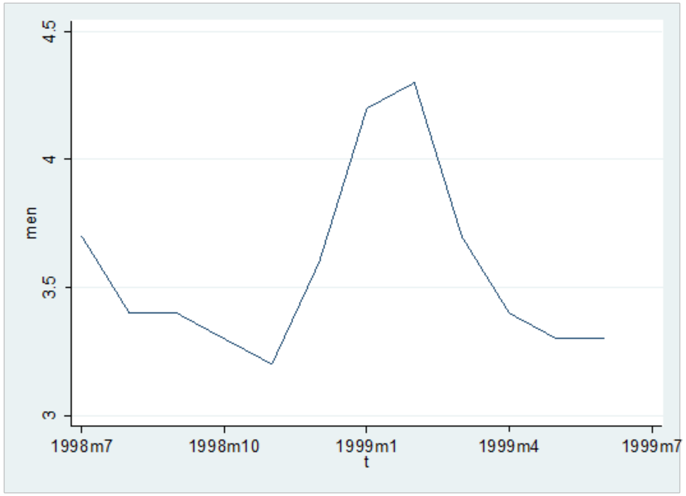
Source	SS	df	MS
Model	105.325822	11	9.57507469
Residual	1571.36666	733	2.14374715
Total	1676.69248	744	2.25361893

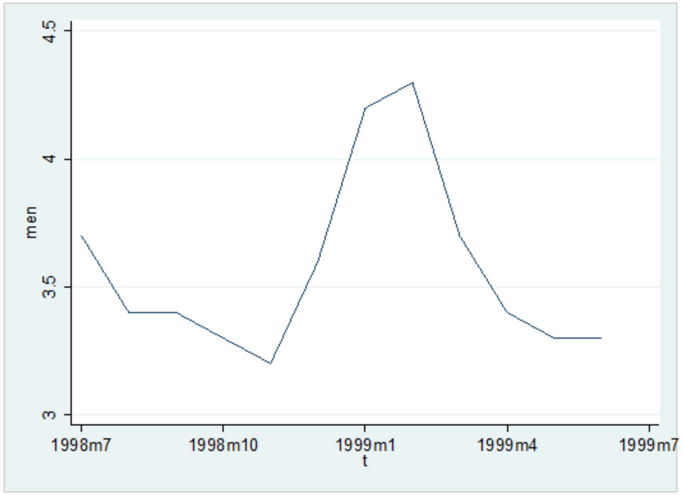
Number of obs	=	745
F(11, 733)	=	4.47
Prob > F	=	0.0000
R-squared	=	0.0628
Adj R-squared	=	0.0488
Root MSE	=	1.4642

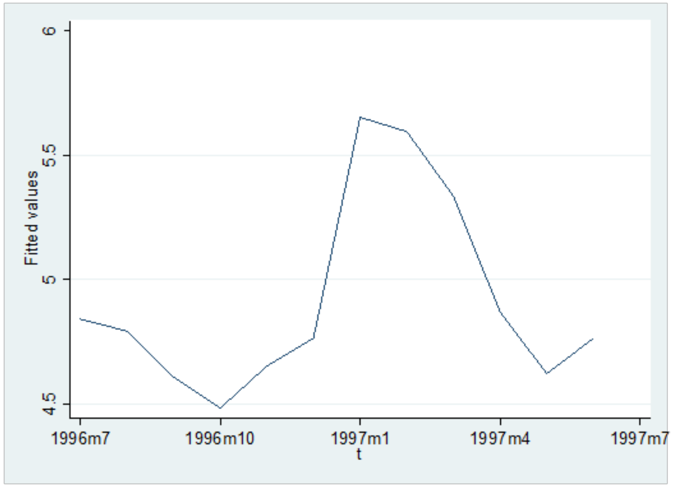
ur	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
m						
1	.8878648	.2619242	3.39	0.001	.3736537	1.402076
2	.8306451	.2629698	3.16	0.002	.3143813	1.346909
3	.5709677	.2629698	2.17	0.030	.0547039	1.087232
4	.1048387	.2629698	0.40	0.690	-.4114251	.6211026
5	-.1435484	.2629698	-0.55	0.585	-.6598123	.3727154
6	-.0016129	.2629698	-0.01	0.995	-.5178768	.514651
7	.0758064	.2629698	0.29	0.773	-.4404574	.5920703
8	.0274194	.2629698	0.10	0.917	-.4888445	.5436832
9	-.1580645	.2629698	-0.60	0.548	-.6743284	.3581993
10	-.2822581	.2629698	-1.07	0.283	-.7985219	.2340058
11	-.1129032	.2629698	-0.43	0.668	-.6291671	.4033606
_cons	4.764516	.1859478	25.62	0.000	4.399462	5.12957

Unemployment Rate, by year, and estimated seasonality

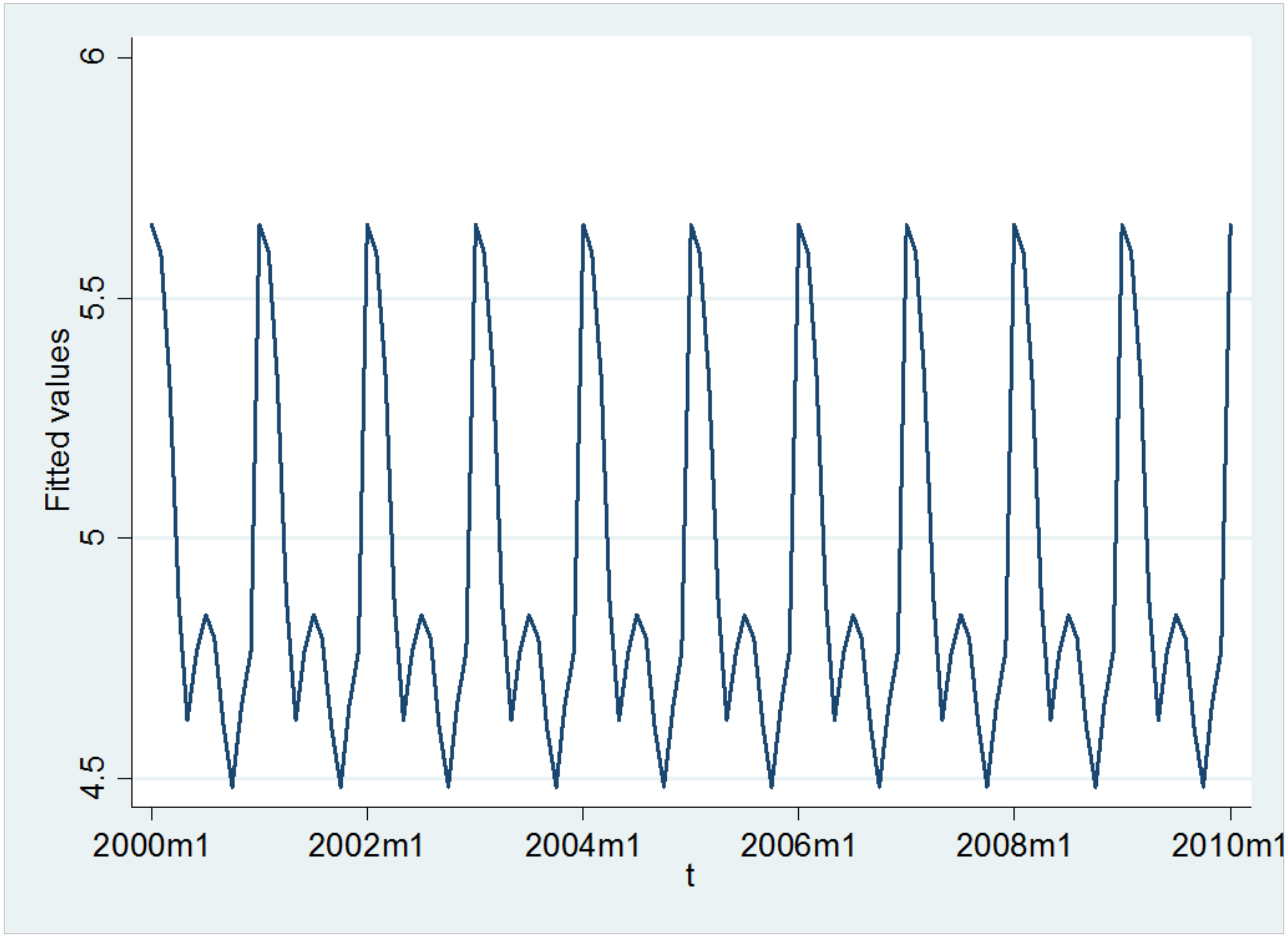








Predicted Values



Example 2
Gasoline Sales

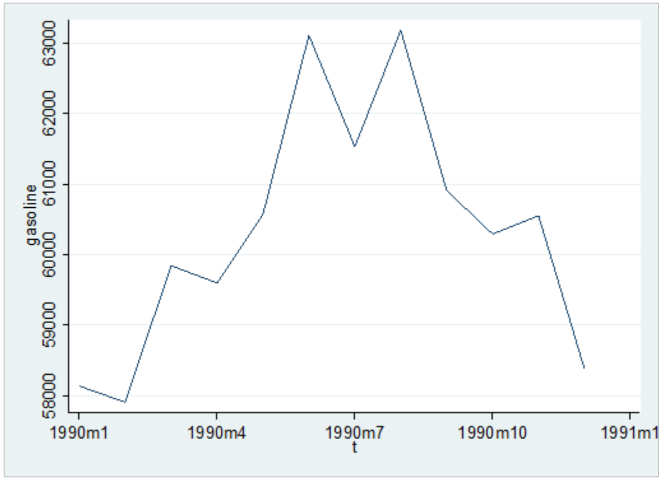
```
. use gasoline

. regress gasoline b12.m
```

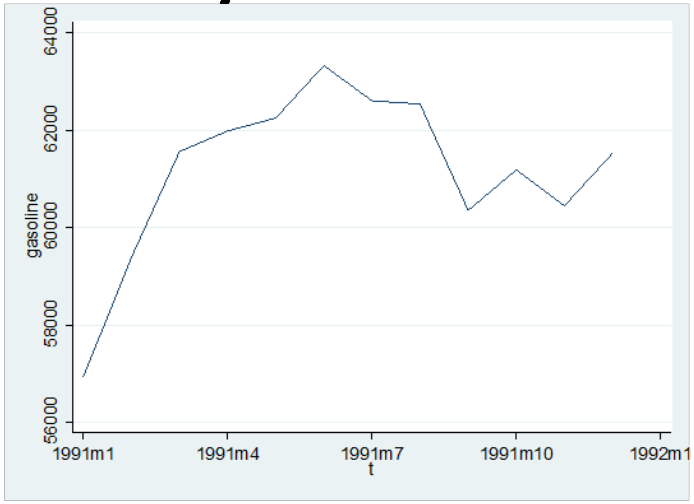
Source	SS	df	MS
Model	636131968	11	57830178.9
Residual	4.1919e+09	311	13478698.6
Total	4.8280e+09	322	14993811.3

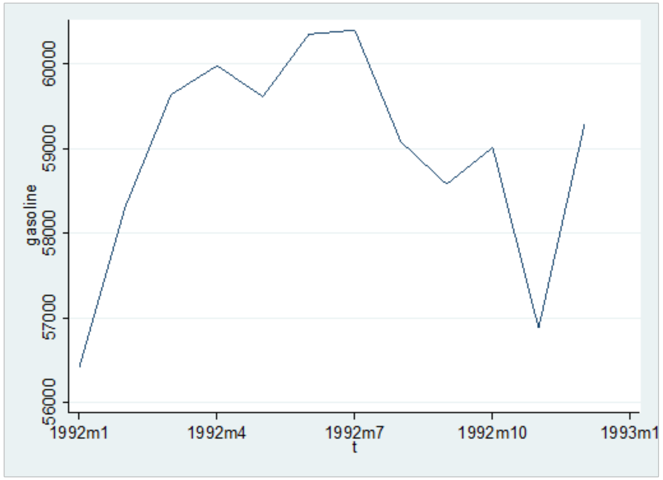
Number of obs	=	323
F(11, 311)	=	4.29
Prob > F	=	0.0000
R-squared	=	0.1318
Adj R-squared	=	0.1010
Root MSE	=	3671.3

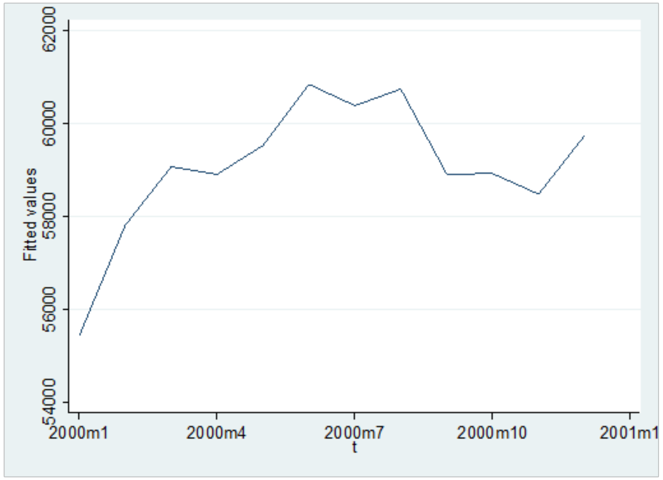
gasoline	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
m						
1	-4309.025	1008.773	-4.27	0.000	-6293.908	-2324.142
2	-1928.984	1008.773	-1.91	0.057	-3913.867	55.8983
3	-671.3583	1008.773	-0.67	0.506	-2656.241	1313.524
4	-829.025	1008.773	-0.82	0.412	-2813.908	1155.858
5	-210.0881	1008.773	-0.21	0.835	-2194.971	1774.795
6	1102.401	1008.773	1.09	0.275	-882.4817	3087.284
7	644.1563	1008.773	0.64	0.524	-1340.726	2629.039
8	1003.964	1008.773	1.00	0.320	-980.9188	2988.847
9	-822.2439	1008.773	-0.82	0.416	-2807.127	1162.639
10	-817.0473	1008.773	-0.81	0.419	-2801.93	1167.835
11	-1260.781	1008.773	-1.25	0.212	-3245.663	724.102
_cons	59735.28	720.008	82.96	0.000	58318.57	61151.98

Gasoline Sales, by year, and estimated seasonality

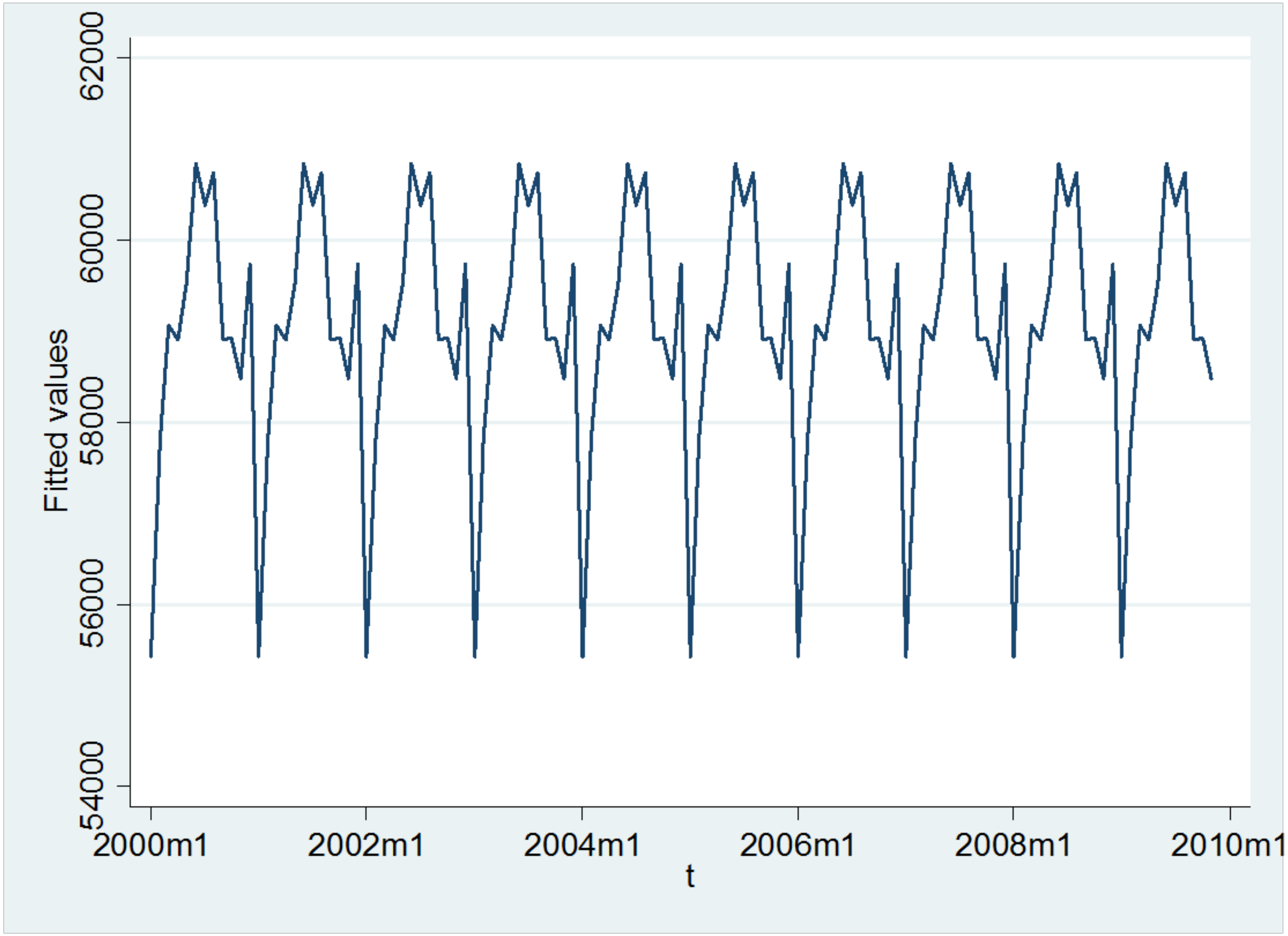








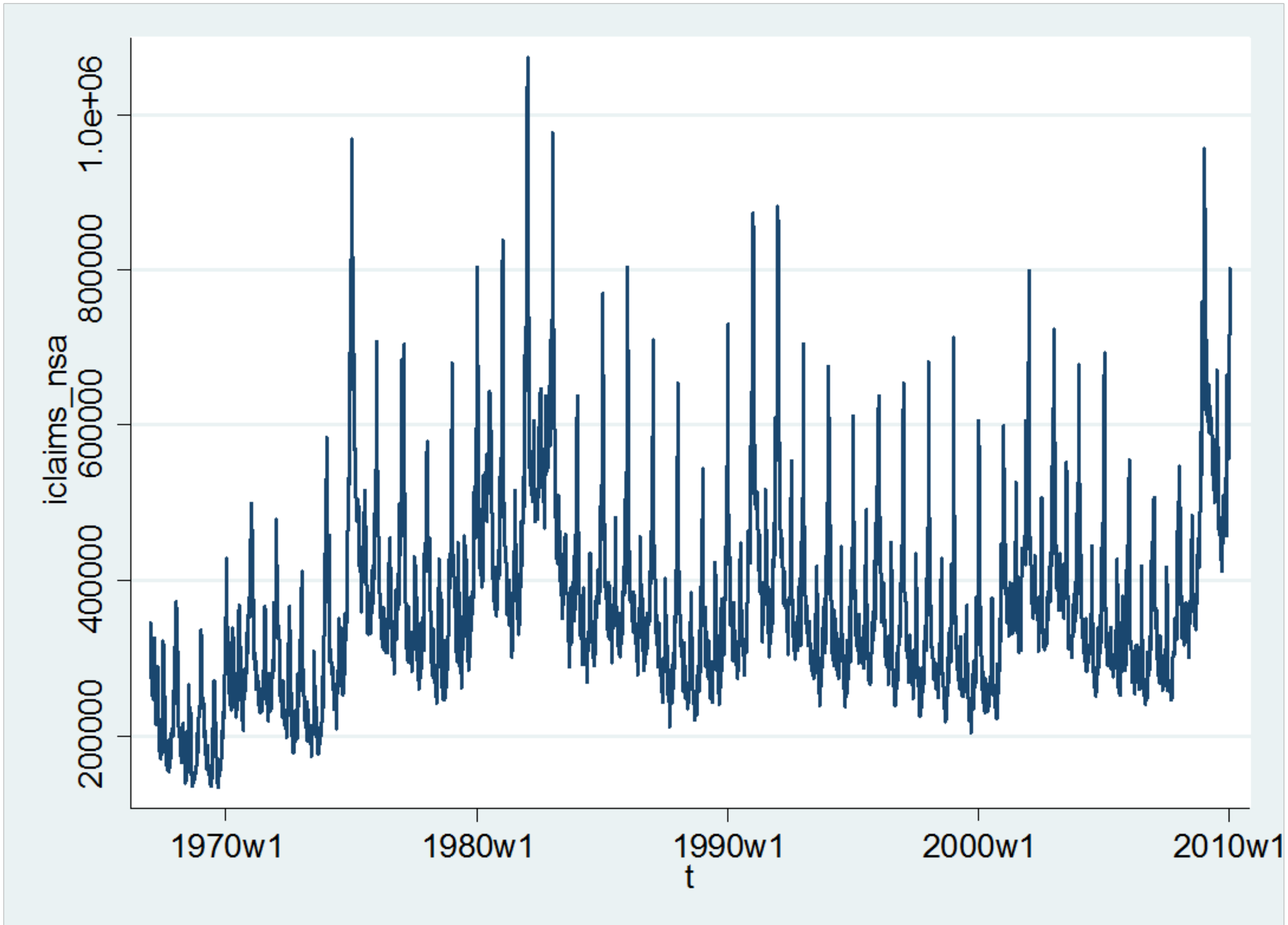
Predicted Values



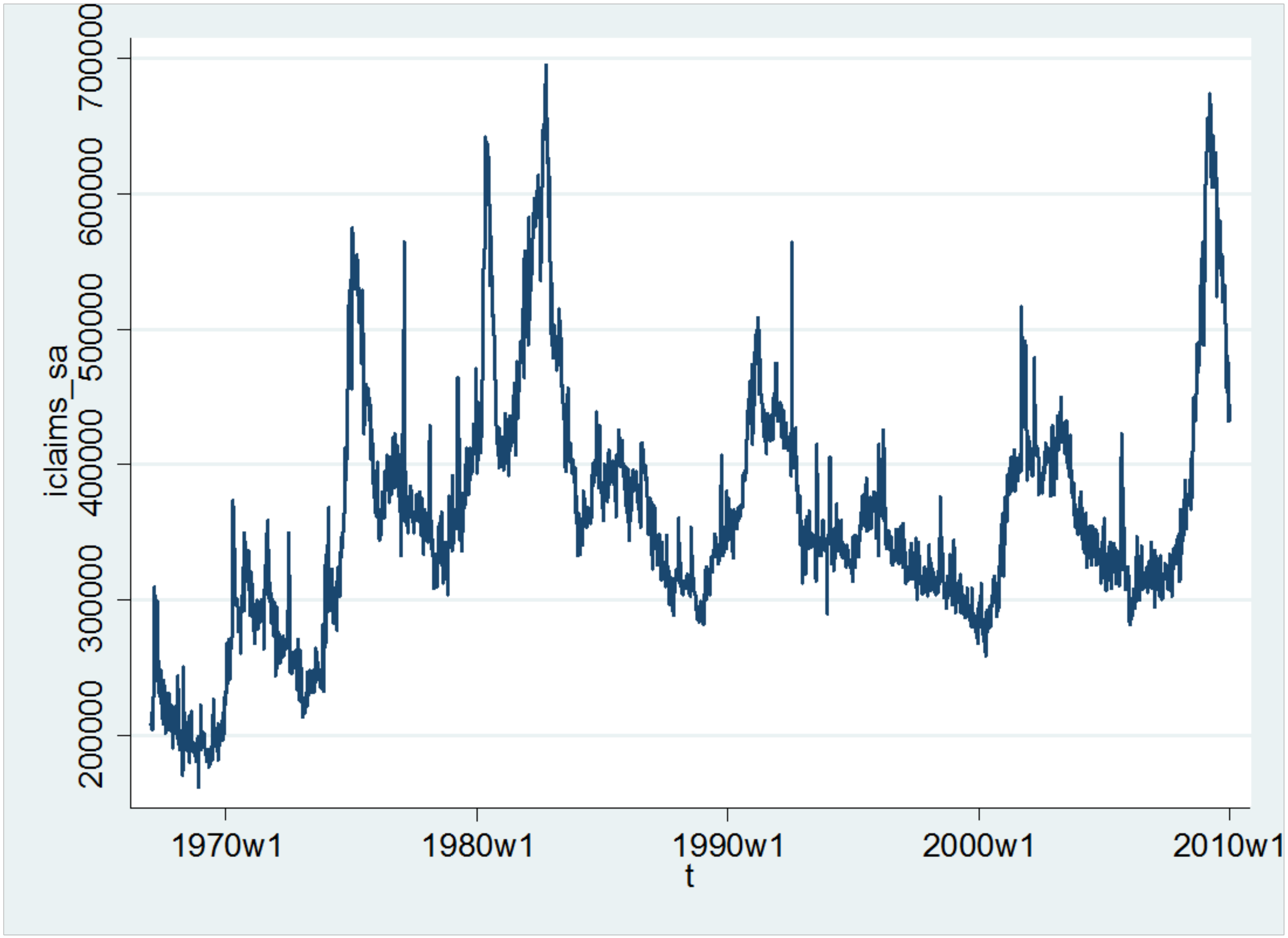
Application – Weekly Data

- Unemployment Insurance Claims
- Department of Labor
- Issued Weekly
- Important indicator for unemployment

Unemployment Claims
Not Seasonally Adjusted



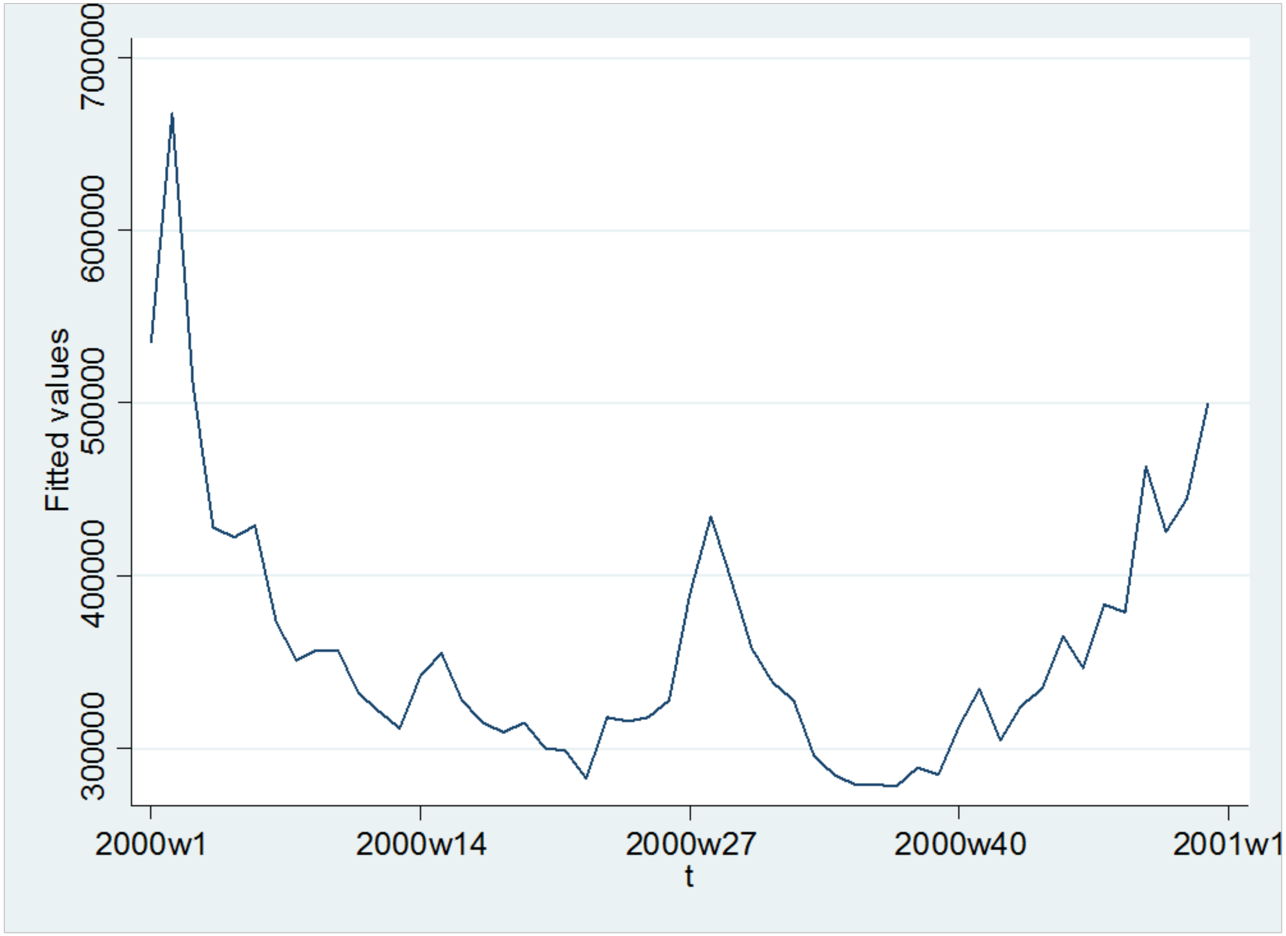
Unemployment Claims
Official Seasonally Adjusted Series



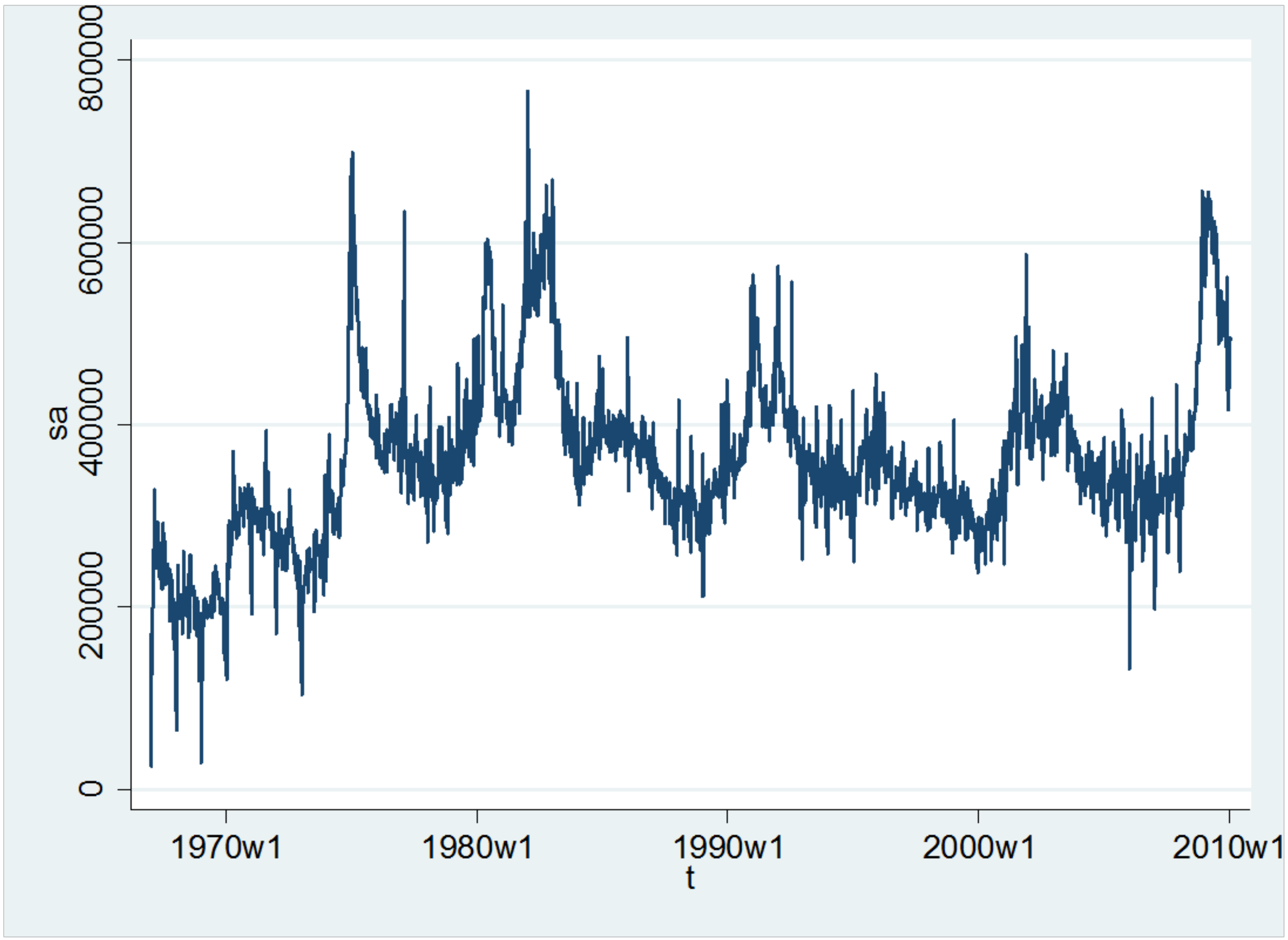
Estimation

```
. regress iclaims_nsa b52.w

      Source |       SS       df       MS              Number of obs =    2238
-------------+------------------------------           F( 51,  2186) =   29.09
       Model |  1.2804e+13    51  2.5105e+11           Prob > F      =  0.0000
    Residual |  1.8865e+13  2186  8.6297e+09           R-squared     =  0.4043
-------------+------------------------------           Adj R-squared =  0.3904
       Total |  3.1668e+13  2237  1.4157e+10           Root MSE      =   92896

------------------------------------------------------------------------------
 iclaims_nsa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |
          1  |   35954.91   19920.38     1.80   0.071    -3109.952    75019.78
          2  |   168185.3   19920.38     8.44   0.000     129120.4    207250.2
          3  |   11166.58   20034.54     0.56   0.577    -28122.15    50455.32
          4  |  -71213.05   20034.54    -3.55   0.000    -110501.8   -31924.31
          5  |  -77421.58   20034.54    -3.86   0.000    -116710.3   -38132.85
          6  |  -70397.88   20034.54    -3.51   0.000    -109686.6   -31109.15
          7  |  -125853.5   20034.54    -6.28   0.000    -165142.3    -86564.8
          8  |  -148580.9   20034.54    -7.42   0.000    -187869.6   -109292.2
          9  |  -142809.6   20034.54    -7.13   0.000    -182098.4   -103520.9
         10  |  -142668.5   20034.54    -7.12   0.000    -181957.2   -103379.8
         11  |  -167656.8   20034.54    -8.37   0.000    -206945.6   -128368.1
         12  |  -178125.4   20034.54    -8.89   0.000    -217414.1   -138836.7
         13  |  -187898.8   20034.54    -9.38   0.000    -227187.6   -148610.1
         14  |  -157631.2   20034.54    -7.87   0.000    -196919.9   -118342.5
         15  |  -144329.8   20034.54    -7.20   0.000    -183618.5   -105041.1
         16  |  -171520.5   20034.54    -8.56   0.000    -210809.2   -132231.8
```

Estimated Seasonal Process



Seasonally Adjusted (by Dummy Variable Method)



Other types of seasonality

- Daily data
  - Day of the week
  - Handle by including dummy variables for each day
- High-frequency data
  - Include hourly or time-of-day indicators
- Holiday effects
  - Flower sales big on Valentines Day, Mothers Day, Easter, yet these days can move around
  - Trading-day/business-day variation
    - Number of trading days/business days varies across months
    - Can divide by number of trading days, or include as a regressor

Seasonal + Cycle

- Consider a components model with seasonal and AR(1) cycle

$$y_t = S_t + C_t$$

$$C_t = \beta C_{t-1} + e_t$$

- The seasonal S_t is a set of seasonal dummies

$$S_t = \sum_{i=1}^{s} \alpha_i D_{it}$$

Transformation

$$y_t = S_t + C_t$$
$$C_t = \beta C_{t-1} + e_t$$

- Take the first equation and lag it once

$$y_{t-1} = S_{t-1} + C_{t-1}$$

- Multiply it by β

$$\beta y_{t-1} = \beta S_{t-1} + \beta C_{t-1}$$

- Then subtract it from the first equation to find

$$y_t = \beta y_{t-1} + S_t - \beta S_{t-1} + e_t$$

Seasonal Representation

- We find

$$y_t = \beta y_{t-1} + S_t - \beta S_{t-1} + e_t$$

- When the seasonal S_t is a set of seasonal dummies, one for each season, this equation suggests a regression on
  - y_{t-1}
  - Seasonal dummies
  - Lagged Seasonal dummies

Redundant

- But lagged seasonal dummies are redundant with the original seasonal dummies
- The set of lagged dummy variables are collinear with the current dummy variables
- Given that you know this month is February, there is no information in knowing that last month was January.
- The lagged dummies can be (should be) omitted

Seasonal + Cycle

- We have found that the regression model is

$$y_t = \sum_{i=1}^{s} \alpha_i D_{it} + \beta y_{t-1} + e_t$$

or

$$y_t = \alpha_0 + \sum_{i=1}^{s-1} \alpha_i D_{it} + \beta y_{t-1} + e_t$$

AR(p) Case

- If the cycle is an AR(p) we have

$$y_t = \sum_{i=1}^{s} \alpha_i D_{it} + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + e_t$$

- Estimate by least squares
- Linear Forecasting

Trend+Seasonal+Cycle

- A full model is

$$y_t = T_t + S_t + C_t$$

$$T_t = \mu_1 + \mu_2 t$$

$$S_t = \sum_{i=1}^{s} \alpha_i D_{it}$$

$$C_t = \beta_1 C_{t-1} + \cdots + \beta_p C_{t-p} + e_t$$

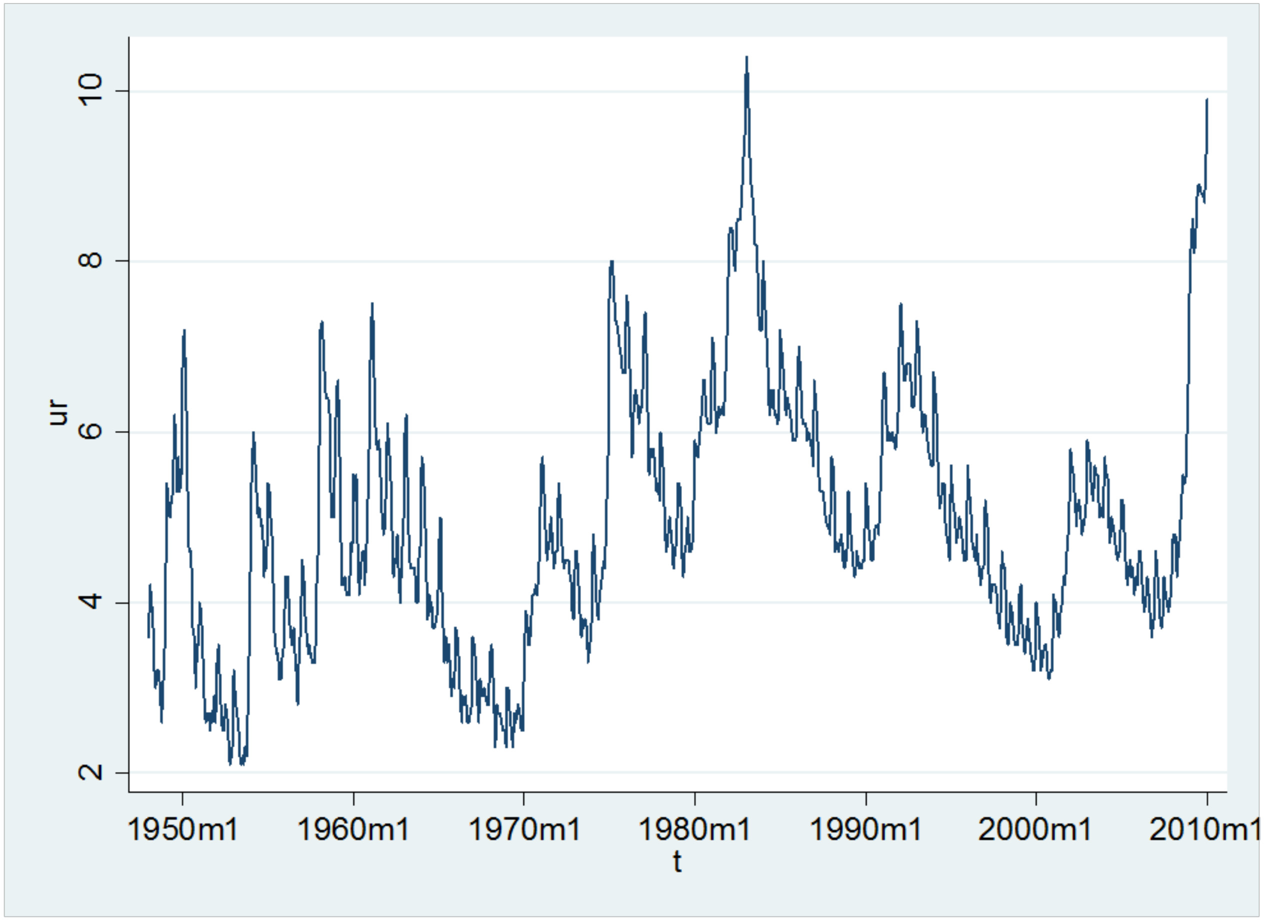
Regression Model

- The implied regression model is

$$y_t = \sum_{i=1}^{s} \alpha_i D_{it} + \gamma t + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + e_t$$

- This can be estimated by least-squares
- It is a complete forecasting model

Example: NSA Unemployment Rate



Regress on Dummies plus AR(12)

```
. reg ur b12.m L(1/12).ur
```

Source	SS	df	MS
Model	1609.97357	23	69.9988508
Residual	33.8239198	709	.047706516
Total	1643.79749	732	2.24562498

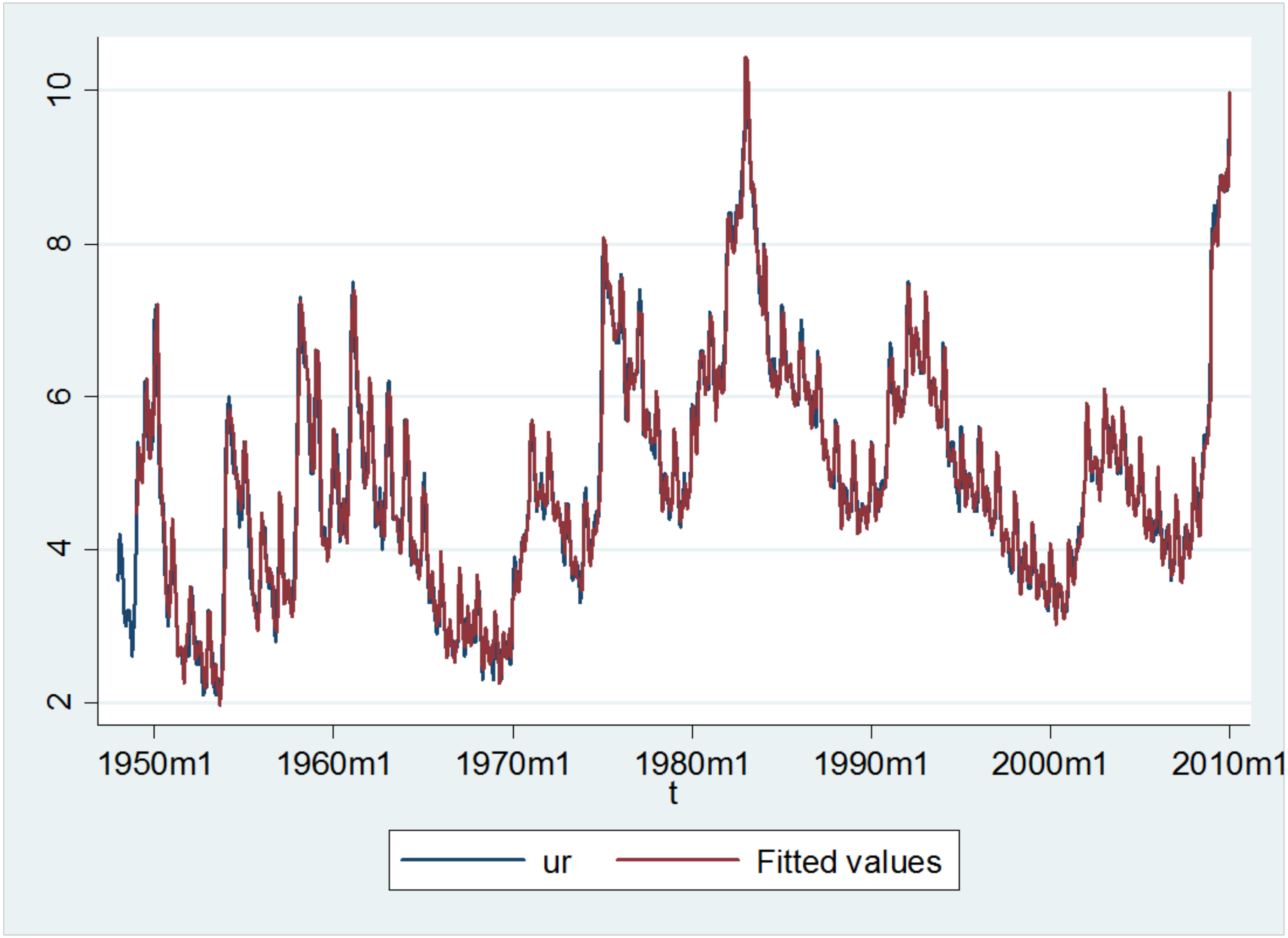
Number of obs =	733
F(23, 709) =	1467.28
Prob > F =	0.0000
R-squared =	0.9794
Adj R-squared =	0.9788
Root MSE =	.21842

ur	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
m						
1	.8614153	.0577449	14.92	0.000	.7480438	.9747868
2	-.1534497	.0777527	-1.97	0.049	-.3061028	-.0007967
3	-.468174	.087469	-5.35	0.000	-.6399033	-.2964448
4	-.5474595	.0821548	-6.66	0.000	-.7087554	-.3861636
5	-.1594984	.0684777	-2.33	0.020	-.2939418	-.025055
6	.2175848	.0633983	3.43	0.001	.0931139	.3420557
7	.1527194	.0691765	2.21	0.028	.0169041	.2885346
8	-.0525233	.0835277	-0.63	0.530	-.2165145	.1114679
9	-.2668832	.089909	-2.97	0.003	-.4434029	-.0903635
10	-.1902467	.0812989	-2.34	0.020	-.3498622	-.0306312
11	.1247499	.0685836	1.82	0.069	-.0099012	.2594011

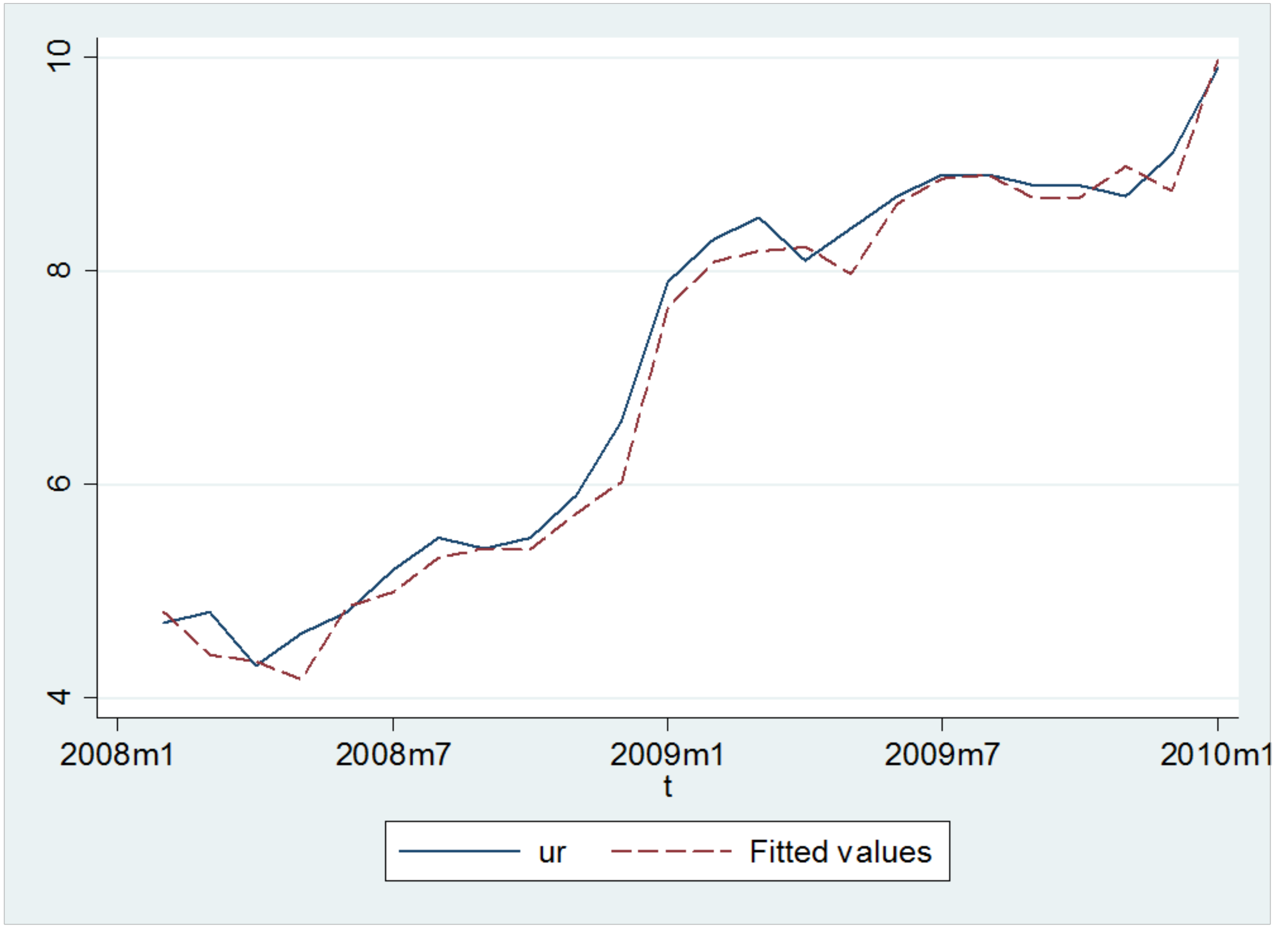
AR Coefficients

ur	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
L1.	1.119934	.0373309	30.00	0.000	1.046642	1.193227
L2.	.0911447	.0562169	1.62	0.105	-.0192268	.2015162
L3.	-.1434849	.0563492	-2.55	0.011	-.2541162	-.0328536
L4.	-.1024096	.0566111	-1.81	0.071	-.213555	.0087359
L5.	.0776225	.0566159	1.37	0.171	-.0335323	.1887774
L6.	-.1096277	.0566984	-1.93	0.054	-.2209446	.0016892
L7.	.0013858	.0565946	0.02	0.980	-.1097274	.1124989
L8.	.0828428	.0565074	1.47	0.143	-.0280991	.1937846
L9.	-.0202508	.0564409	-0.36	0.720	-.1310621	.0905605
L10.	-.0057293	.0563117	-0.10	0.919	-.116287	.1048284
L11.	.0887878	.0562091	1.58	0.115	-.0215683	.199144
L12.	-.1025115	.0373434	-2.75	0.006	-.1758284	-.0291946
_cons	.1530973	.059812	2.56	0.011	.0356676	.2705271

Fitted Values

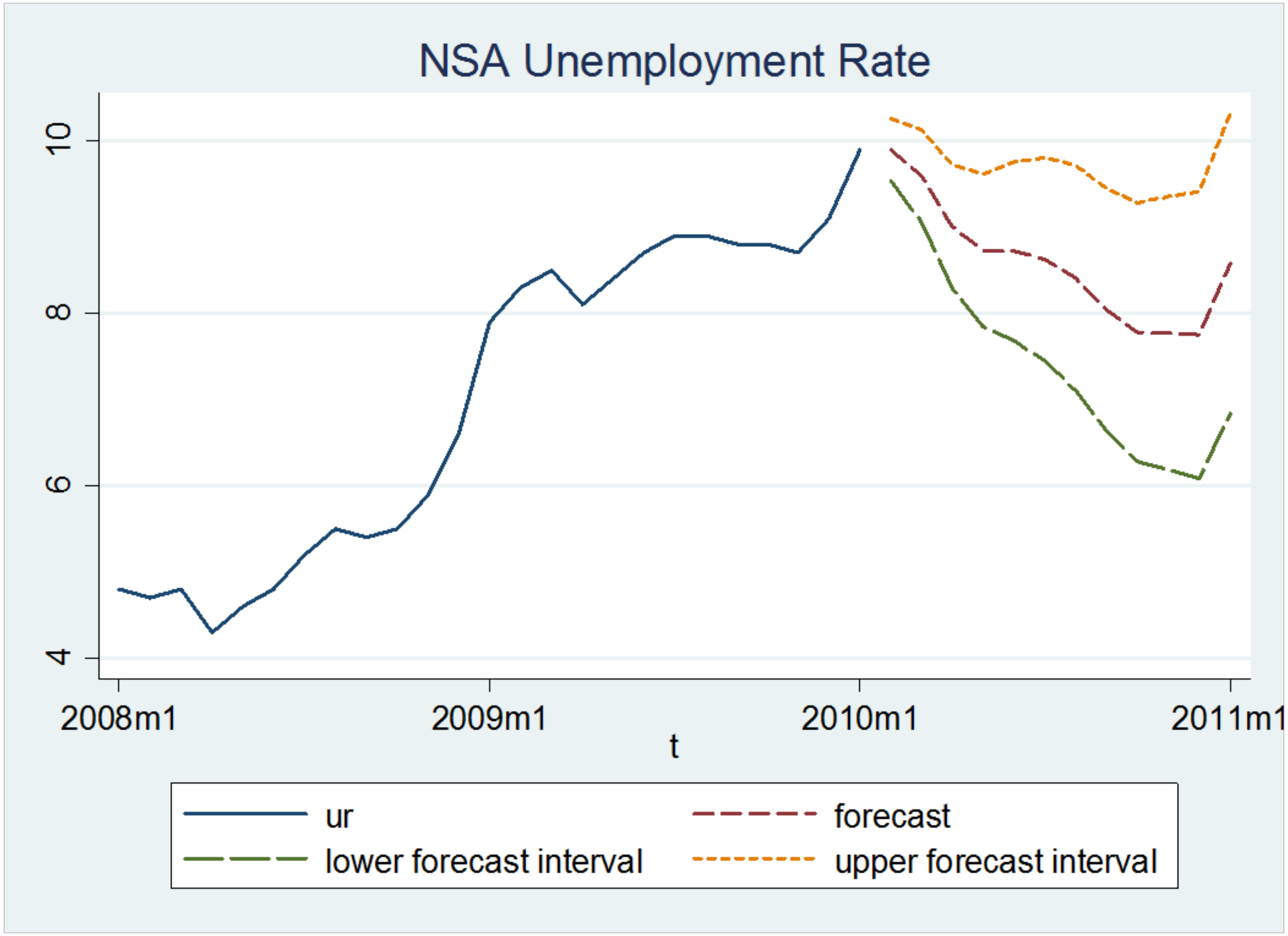


Last 2 years



10
8
6
4
2008m1
2008m7
2009m1
2009m7
2010m1
t
ur
Fitted values

12-month forecast



Forecasting with Seasonal Dummy

- To forecast in STATA with seasonal dummies, the dummy variables must be defined for the forecast period
- **After** you use the **tsappend** command, you create the month variable

  .gen m=month(dofm(t))

  or

  .replace m=month(dofm(t))
- Otherwise m will have missing values for the forecast period

Example: Retail Sales

- U.S. Census Bureau
  - Monthly Retail Sales
  - Not Seasonally Adjusted and Seasonally Adjusted
  - Sales listed by variety of categories
  - 1992-2009

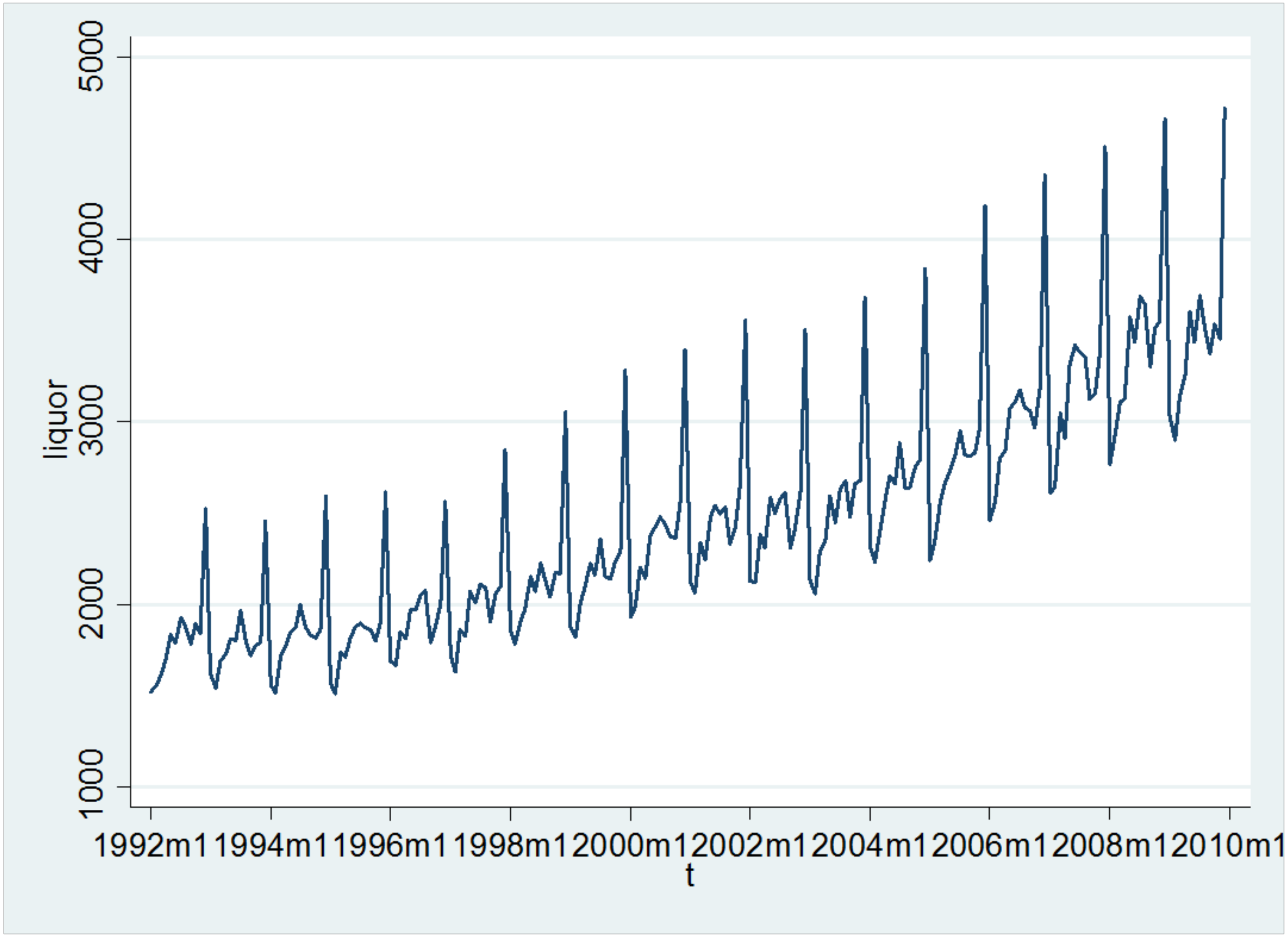
From Census Bureau Spreadsheet

	Jan. 2009	Feb. 2009	Mar. 2009	Apr. 2009	May 2009	Jun. 2009
NOT ADJUSTED						
Retail and food services sales, total	313,593	304,056	334,149	336,155	354,668	351,418
Retail sales and food services excl motor vehicle and parts	262,237	252,560	275,000	277,938	294,559	289,242
Retail sales, total	277,402	269,015	295,520	298,119	313,979	312,547
Retail sales, total (excl. motor vehicle and parts dealers)	226,046	217,519	236,371	239,902	253,870	250,371
GAFO(1)	83,323	82,918	87,477	87,248	93,069	88,330
Motor vehicle and parts dealers	51,356	51,496	59,149	58,217	60,109	62,176
Automobile and other motor vehicle dealers	45,385	45,544	52,498	51,558	53,528	55,239
Automobile dealers	41,790	41,632	46,834	45,373	46,601	48,405
New car dealers	36,032	35,304	40,575	39,492	40,941	42,530
Used car dealers	5,758	6,328	6,259	5,881	5,660	5,875
Automotive parts, acc., and tire stores	5,971	5,952	6,651	6,659	6,581	6,937
Furniture, home furn, electronics, and appliance stores	16,069	15,700	15,530	14,545	15,362	15,519
Furniture and home furnishings stores	7,428	7,219	7,601	7,293	7,689	7,728
Furniture stores	4,230	4,296	4,262	3,969	4,261	4,148
Home furnishings stores	3,198	2,923	3,339	3,324	3,428	3,580
Floor covering stores	1,381	1,335	1,448	1,475	1,462	1,673
All other home furnishings stores	1,698	1,480	1,762	1,725	1,851	1,790
Electronics and appliance stores	8,641	8,481	7,929	7,252	7,673	7,791
Appl.,TV, and other elect. stores	6,706	6,615	5,936	5,374	5,827	5,857
Household appliance stores	1,306	1,201	1,249	1,252	1,317	1,349
Radio, T.V., and other elect. stores	5,400	5,414	4,687	4,122	4,510	4,508
Computer and software stores	1,734	1,722	1,790	1,679	1,612	1,706

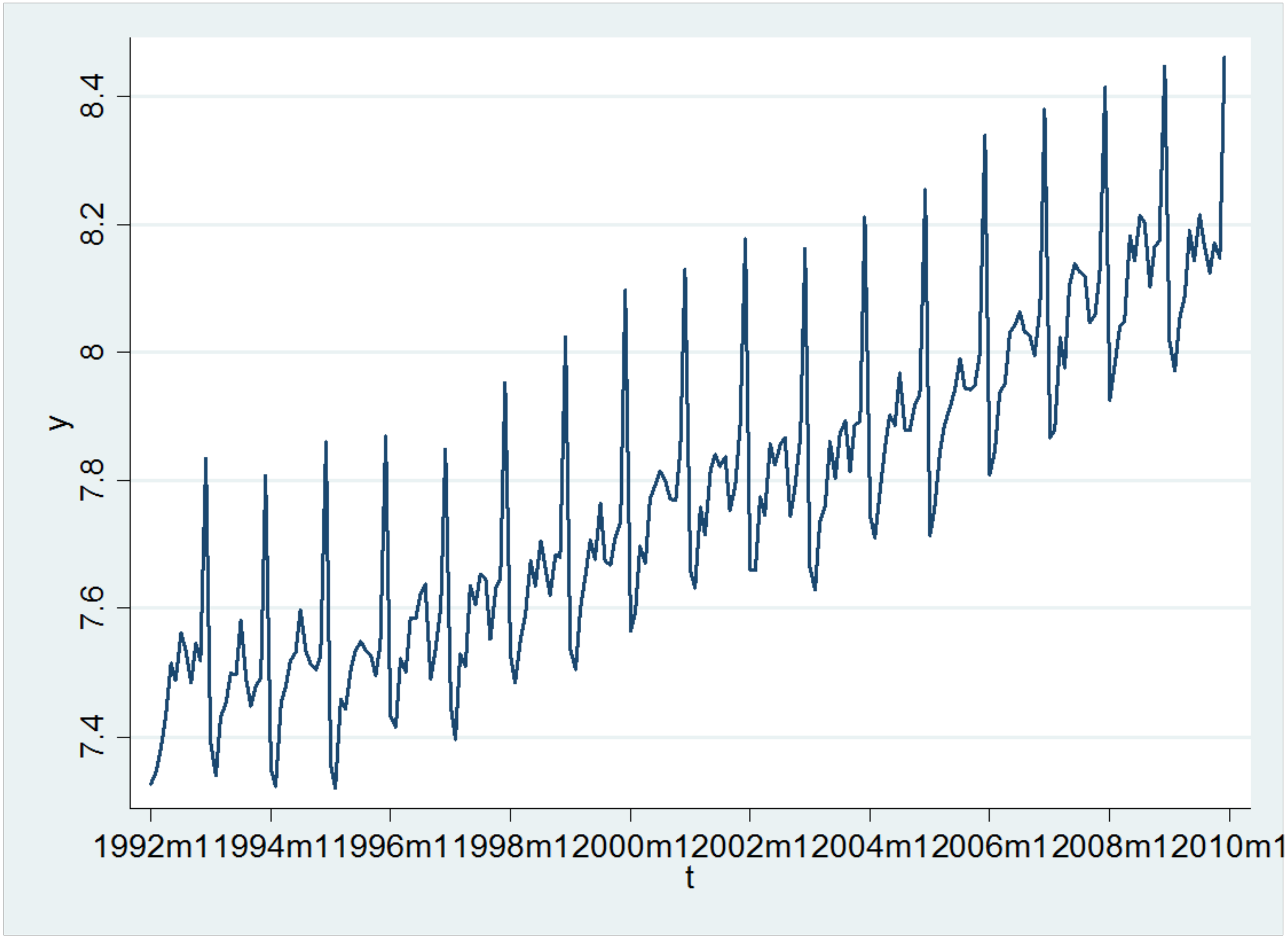
Liquor Sales

- Beer, wine and liquor
- Sample: 1992-2009
- Not Seasonally Adjusted
- Big spike in December

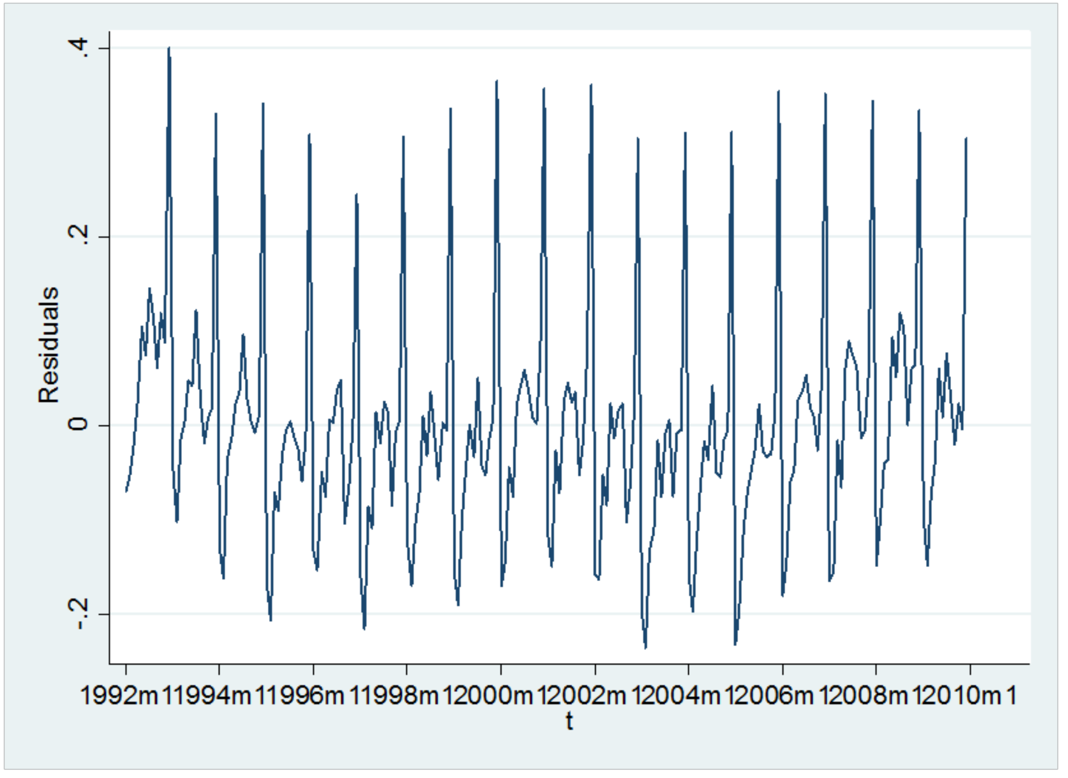
Liquor Sales (Millions of $)



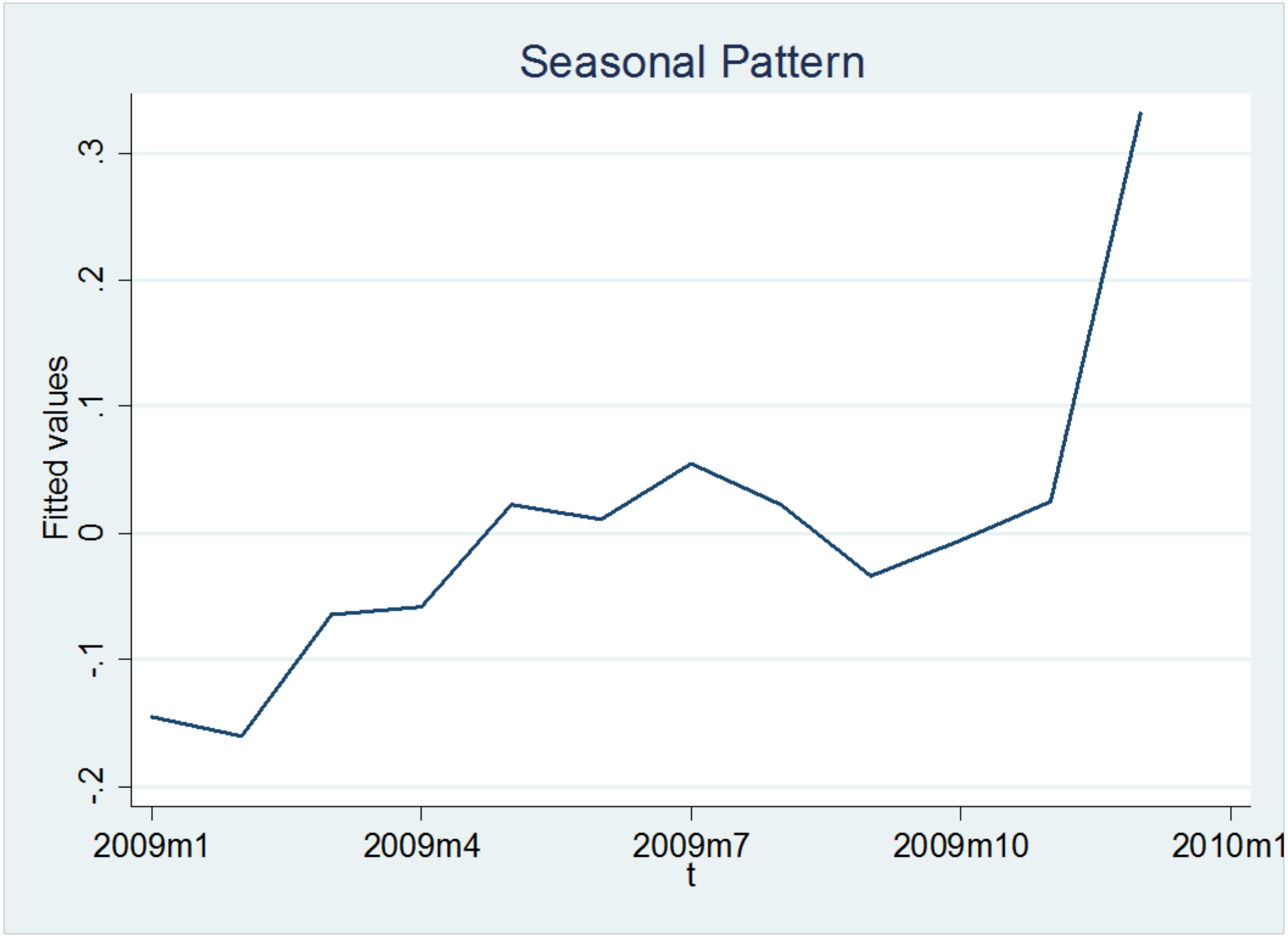
ln(Liquor Sales)



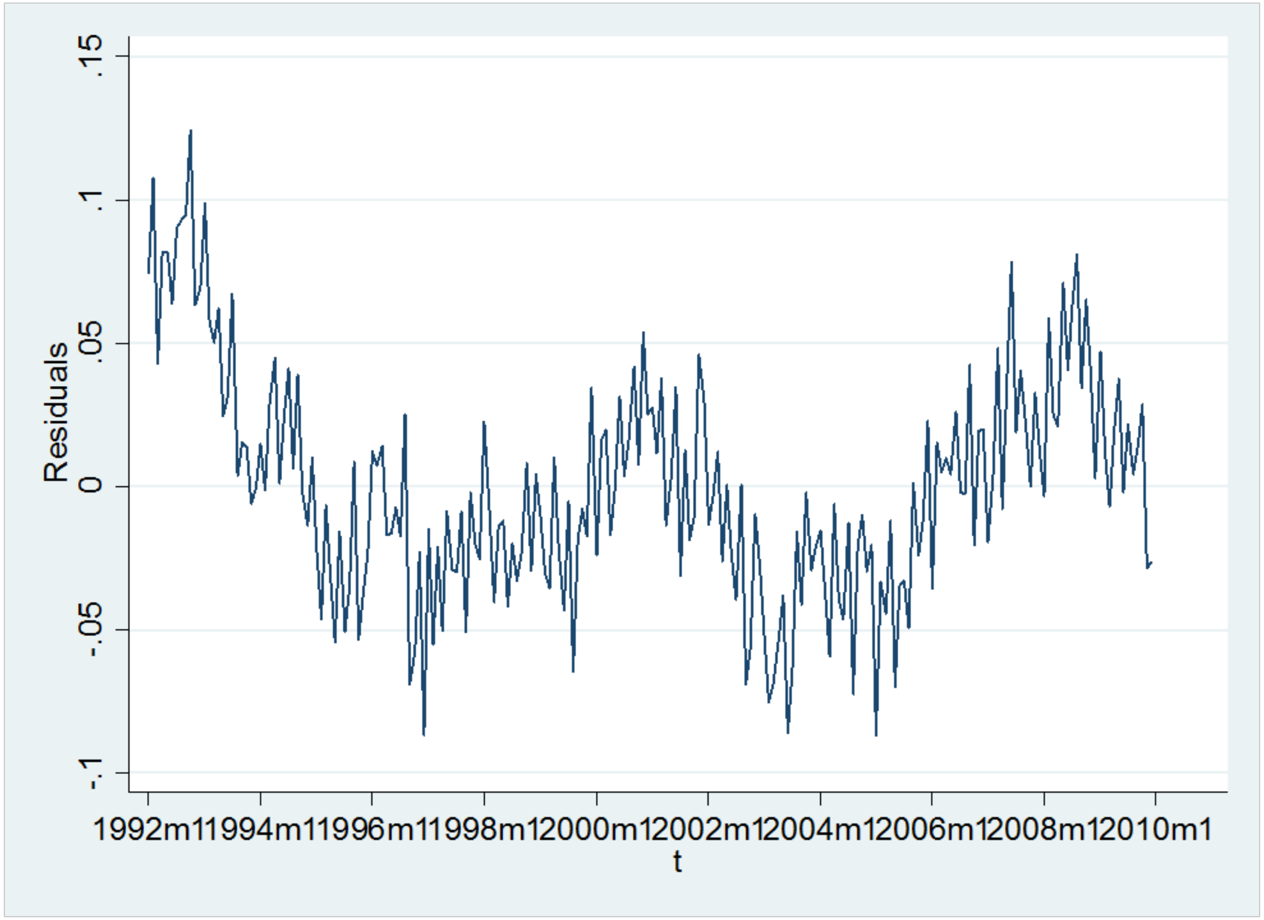
Residual from Linear Trend



Seasonal Dummy



Residuals after Seasonal Dummies



Full Estimation

```
. reg y t b12.m L(1/12).y
```

Source	SS	df	MS
Model	12.6032786	24	.52513661
Residual	.083727442	179	.000467751
Total	12.6870061	203	.062497567

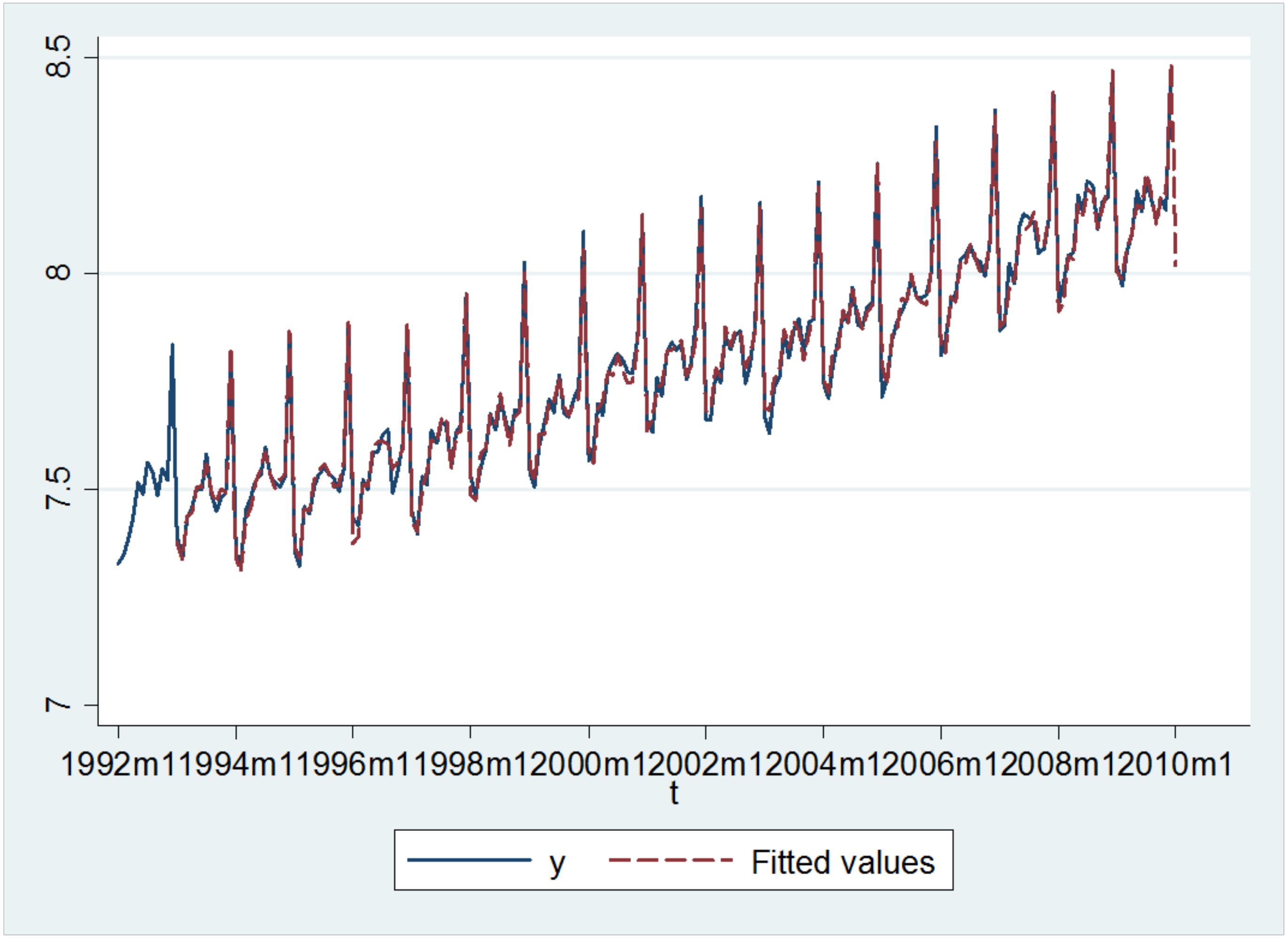
Number of obs =	204
F(24, 179) =	1122.68
Prob > F =	0.0000
R-squared =	0.9934
Adj R-squared =	0.9925
Root MSE =	.02163

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
t	.0005763	.0001903	3.03	0.003	.0002007	.0009519
m						
1	-.4414455	.0475001	-9.29	0.000	-.5351778	-.3477132
2	-.4495358	.0524949	-8.56	0.000	-.5531243	-.3459472
3	-.32453	.0590456	-5.50	0.000	-.441045	-.2080149
4	-.1865844	.0477627	-3.91	0.000	-.2808349	-.0923339
5	-.2373487	.0430689	-5.51	0.000	-.3223367	-.1523606
6	-.1554224	.0503231	-3.09	0.002	-.2547253	-.0561194
7	-.1289456	.050234	-2.57	0.011	-.2280725	-.0298187
8	-.2270499	.0500036	-4.54	0.000	-.3257222	-.1283775
9	-.4130694	.0607582	-6.80	0.000	-.5329639	-.293175
10	-.2073198	.0597305	-3.47	0.001	-.3251863	-.0894532
11	-.2796951	.0484164	-5.78	0.000	-.3752355	-.1841547

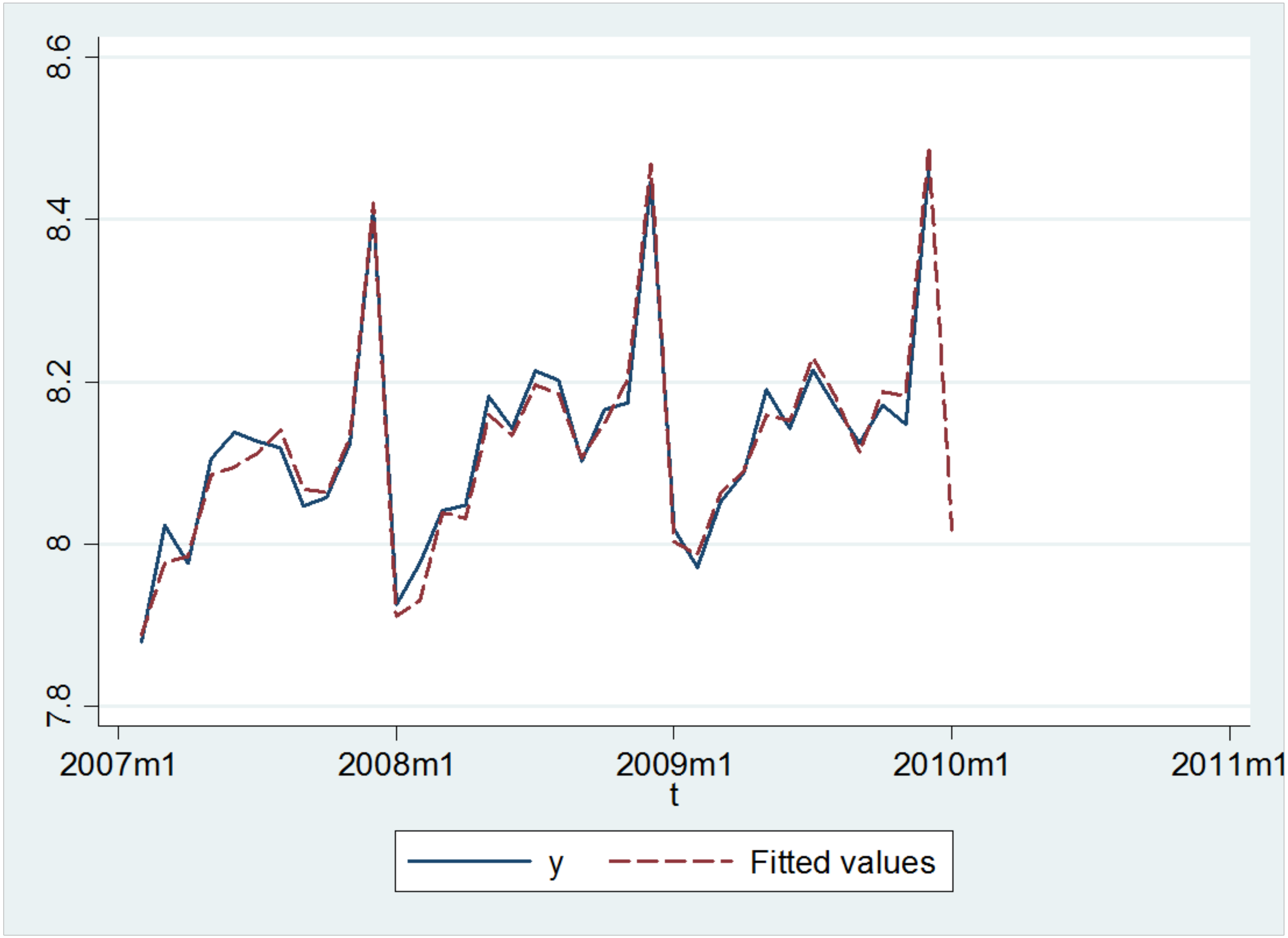
AR Coefficients

y						
L1.	.225336	.0741289	3.04	0.003	.079057	.3716149
L2.	.2636816	.075335	3.50	0.001	.1150227	.4123405
L3.	.3490587	.0749059	4.66	0.000	.2012465	.496871
L4.	-.0818078	.0778903	-1.05	0.295	-.2355091	.0718935
L5.	.2386739	.0769151	3.10	0.002	.0868969	.3904509
L6.	-.0707958	.0780874	-0.91	0.366	-.2248862	.0832946
L7.	-.1447524	.0781988	-1.85	0.066	-.2990625	.0095577
L8.	-.1762223	.0774924	-2.27	0.024	-.3291385	-.0233061
L9.	.2302193	.078365	2.94	0.004	.0755813	.3848574
L10.	-.2566324	.0753442	-3.41	0.001	-.4053095	-.1079553
L11.	.101947	.0733289	1.39	0.166	-.0427534	.2466473
L12.	.1686073	.0720459	2.34	0.020	.0264388	.3107759
_cons	1.164893	.3237073	3.60	0.000	.5261194	1.803666

Fitted Values

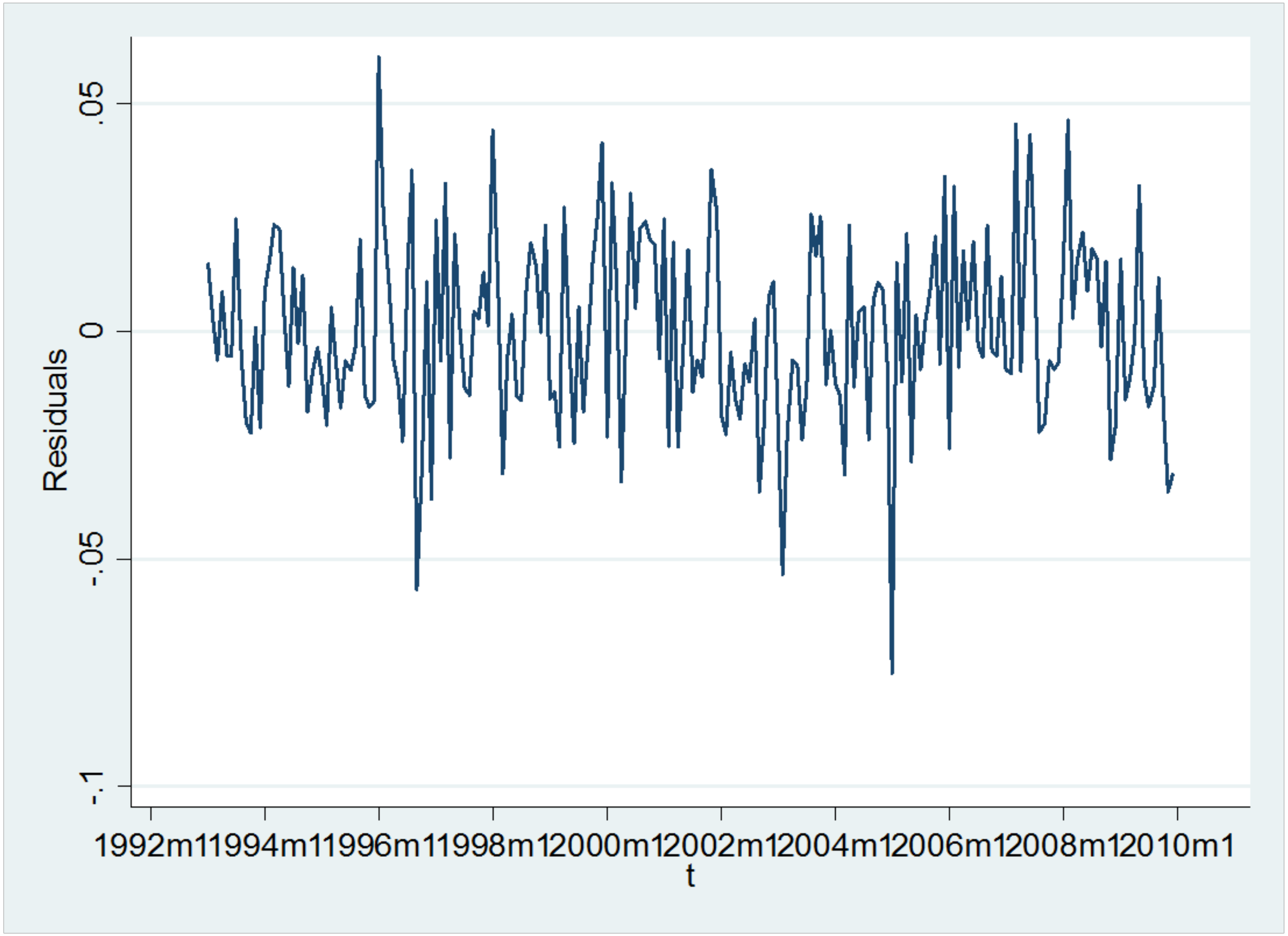


Last 3 years



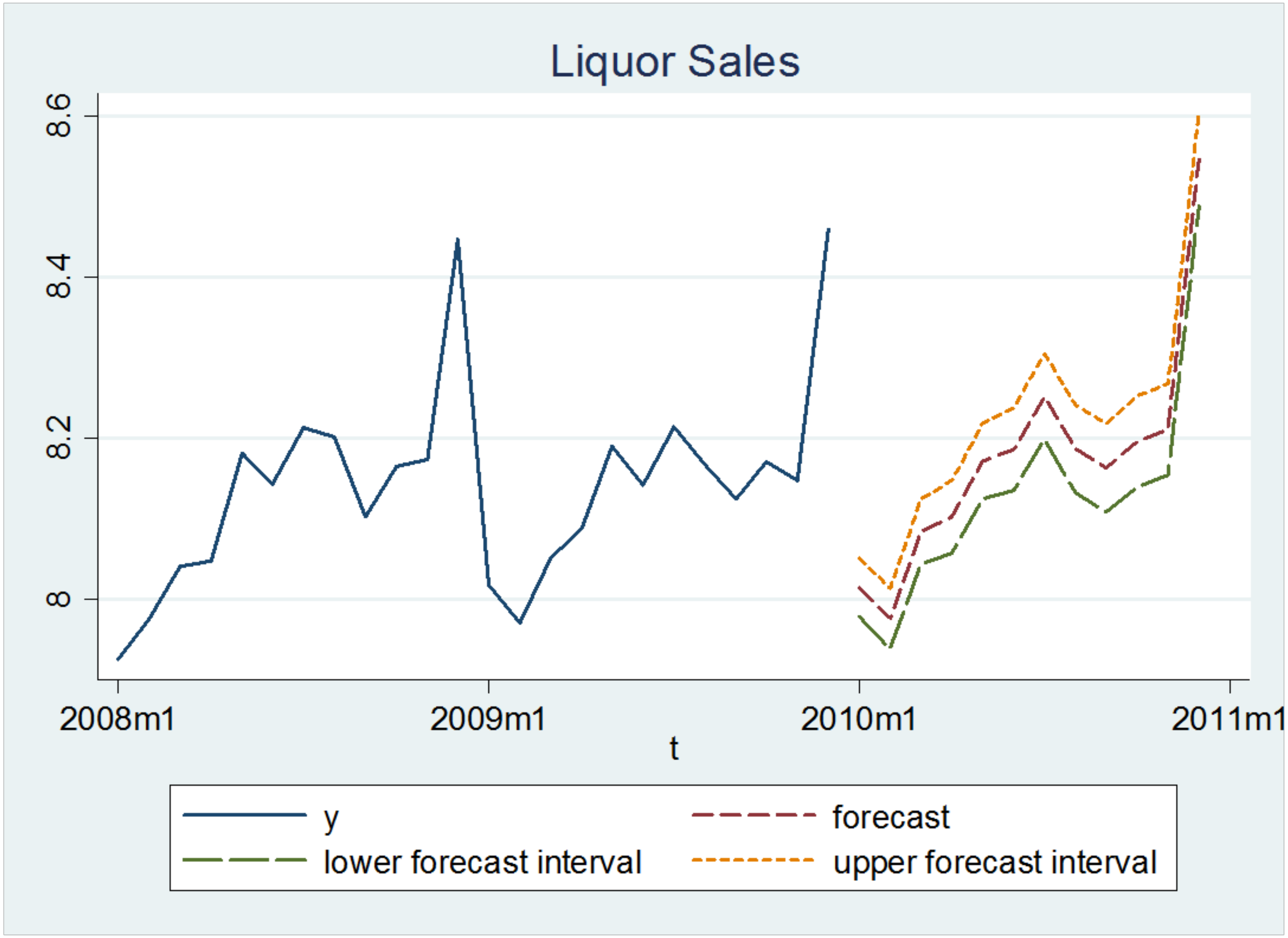
8.6
8.4
8.2
8
7.8
2007m1
2008m1
2009m1
2010m1
2011m1
t
y
Fitted values

Residuals



Residuals
.05
0
-.05
-.1
1992m1
1994m1
1996m1
1998m1
2000m1
2002m1
2004m1
2006m1
2008m1
2010m1
t

12-Month Forecast



12-month forecast

- The big jump down in the forecast for January 2010 is because of the seasonal dummy effect