# Components Model

- Remember that we said that it was useful to think about the components representation

$$y_t = T_t + S_t + C_t$$

- Suppose that $C_t$ is an AR(p) process
- What model does this imply for $y_t$?

# Trend+Cycle Model

- For simplicity, we start with the trend-cycle model

$$y_t = T_t + C_t$$

- And specify the cycle as an AR(1)

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

# Intercept only

- Suppose the trend is just an intercept

$$T_t = \mu$$

- The model is

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

# Partial Differencing

- Lag the first equation, multiply by β and subtract

$$y_t = \mu + C_t$$

$$y_{t-1} = \mu + C_{t-1}$$

$$y_t - \beta y_{t-1} = (1 - \beta)\mu + C_t - \beta C_{t-1}$$

Then use

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

to find

$$y_t = (1 - \beta)\mu + \beta y_{t-1} + e_t$$

# Equivalence with AR(1)

- Thus

$$y_t = \mu + C_t$$

implies

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

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

with

$$\alpha = (1 - \beta)\mu$$

- The model is just an AR(1) with intercept

# Linear Trend

- Suppose the trend is a linear time trend

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

- Then

$$y_t = \mu_1 + \mu_2 t + C_t$$

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

# Partial Differencing

- Lag the first equation, multiply by β and subtract, and use AR(1) equation

$$y_t = \mu_1 + \mu_2 t + C_t$$
$$y_{t-1} = \mu_1 + \mu_2(t-1) + C_{t-1}$$
$$y_t = \beta y_{t-1} + (1-\beta)\mu_1 + \beta\mu_2 + (1-\beta)\mu_2 t + e_t$$

- We find

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 t + e_t$$

# Summary : Trend+AR(1) Cycle

- When the trend is an intercept or a time trend
- The components model

$$y_t = T_t + C_t$$

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

is equivalent with

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

- The components model is equivalent with a regression on the trend variables and the lag

# AR(p)+Trend

- A linear trend plus AR(p)

$$y_t = \mu_1 + \mu_2 t + C_t$$

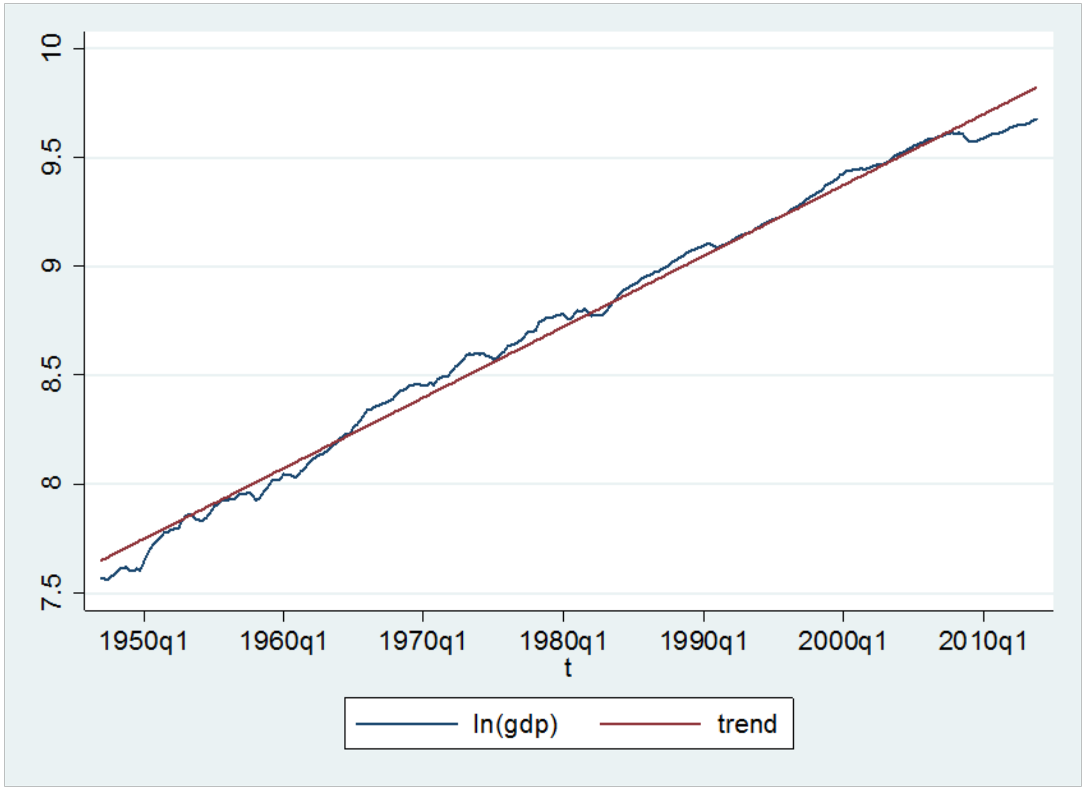
$$C_t = \beta_1 C_{t-1} + \cdots + \beta_p C_{t-p} + e_t$$

is equivalent to

$$y_t = \alpha + \gamma t + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + e_t$$

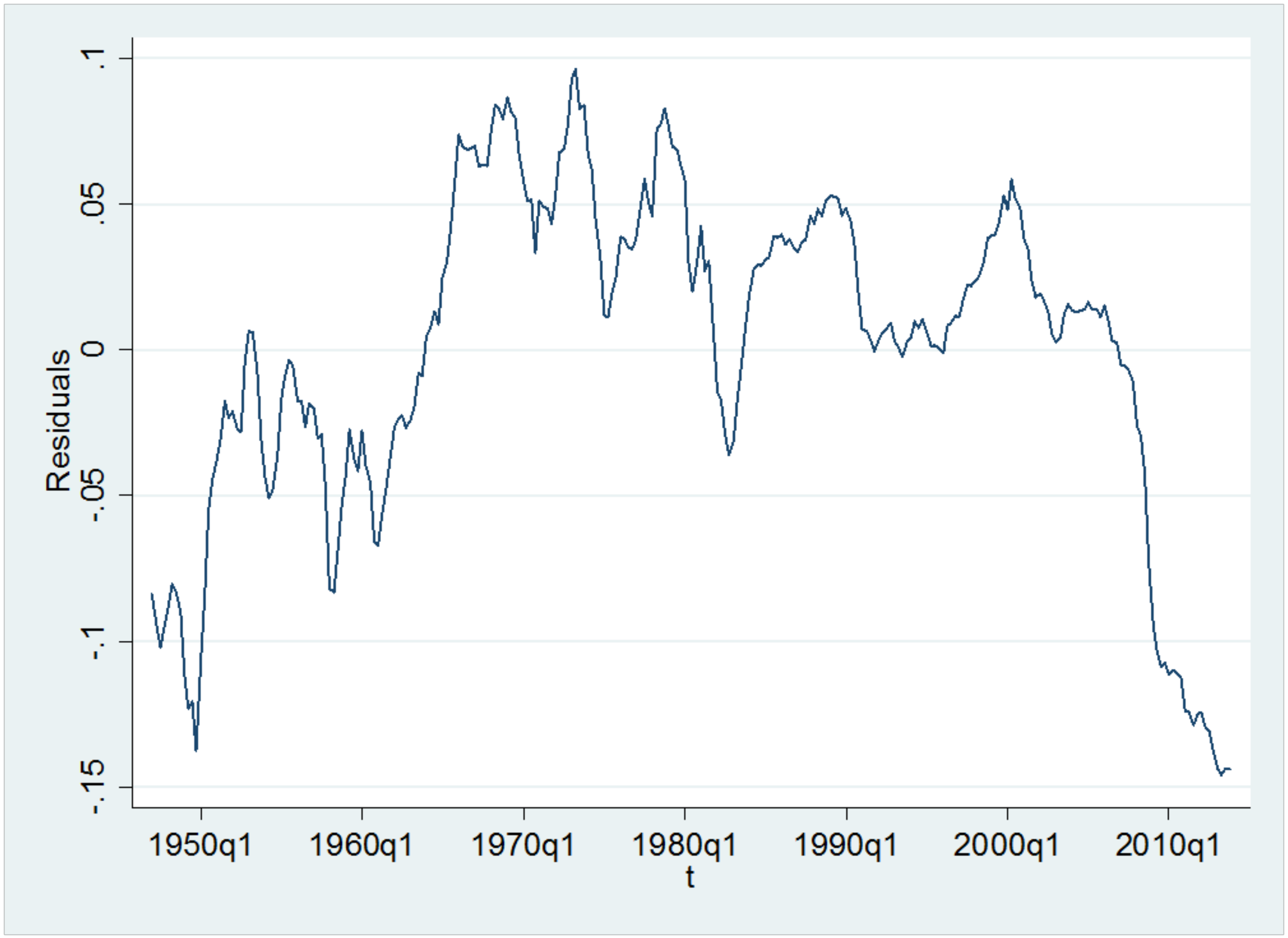
- A regression on a time trend plus *p* lags of *y*

# Example: Real GDP




- Ln(rgdp) and linear trend

# Residuals from Linear Trend



# AR(4) on residuals

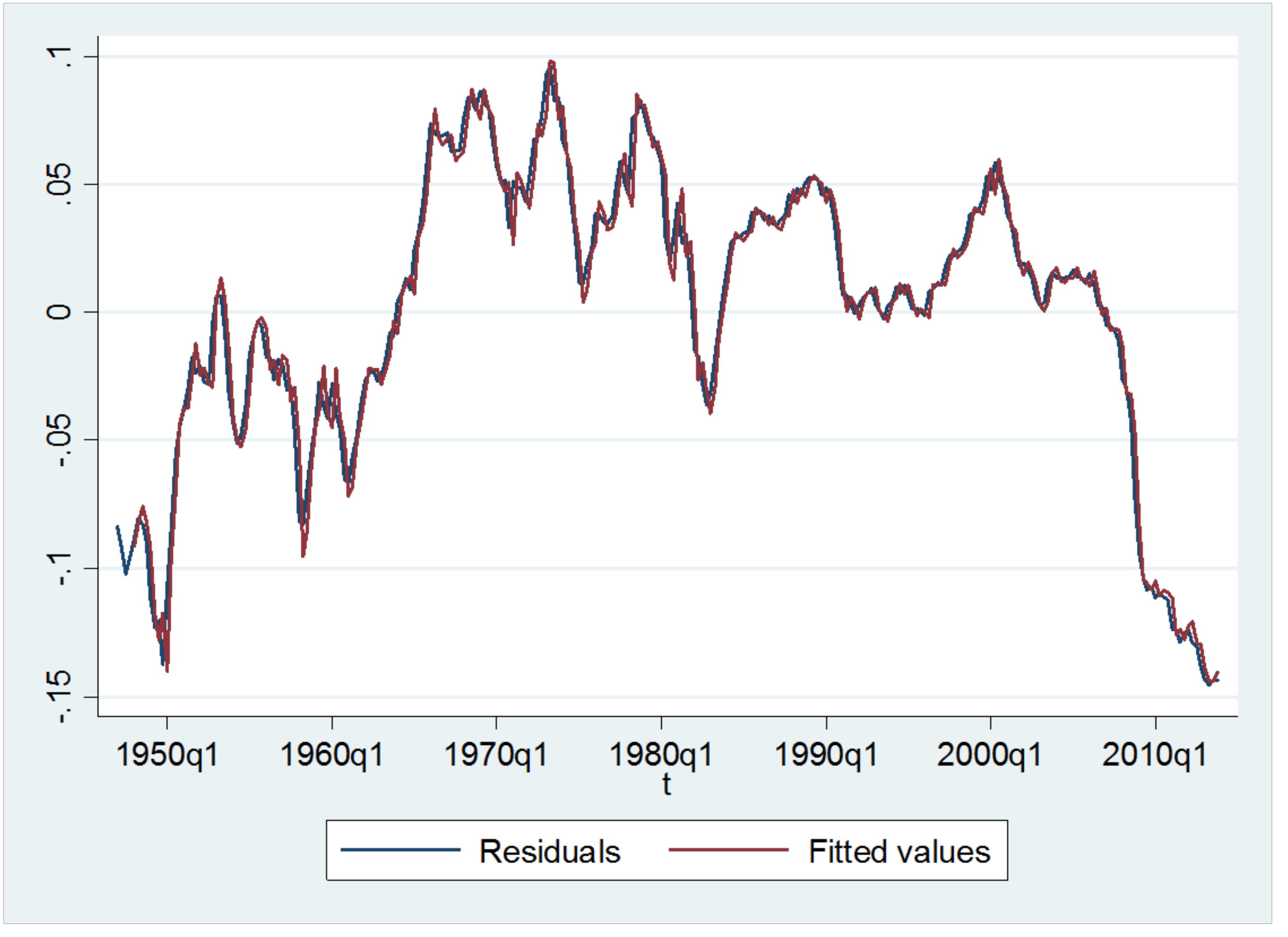
```
. reg u L(1/4).u
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | .800747324 | 4 | .200186831 |
| Residual | .020349748 | 259 | .00007857 |
| Total | .821097073 | 263 | .003122042 |

| | |
|---|---|
| Number of obs | = 264 |
| F( 4, 259) | = 2547.86 |
| Prob > F | = 0.0000 |
| R-squared | = 0.9752 |
| Adj R-squared | = 0.9748 |
| Root MSE | = .00886 |

| u | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| u | | | | | | |
| L1. | 1.337073 | .0618057 | 21.63 | 0.000 | 1.215368 | 1.458779 |
| L2. | -.2144409 | .1023036 | -2.10 | 0.037 | -.4158936 | -.0129882 |
| L3. | -.2211563 | .102324 | -2.16 | 0.032 | -.4226492 | -.0196634 |
| L4. | .0817946 | .0621732 | 1.32 | 0.189 | -.0406348 | .204224 |
| _cons | -.0000945 | .0005461 | -0.17 | 0.863 | -.00117 | .0009809 |

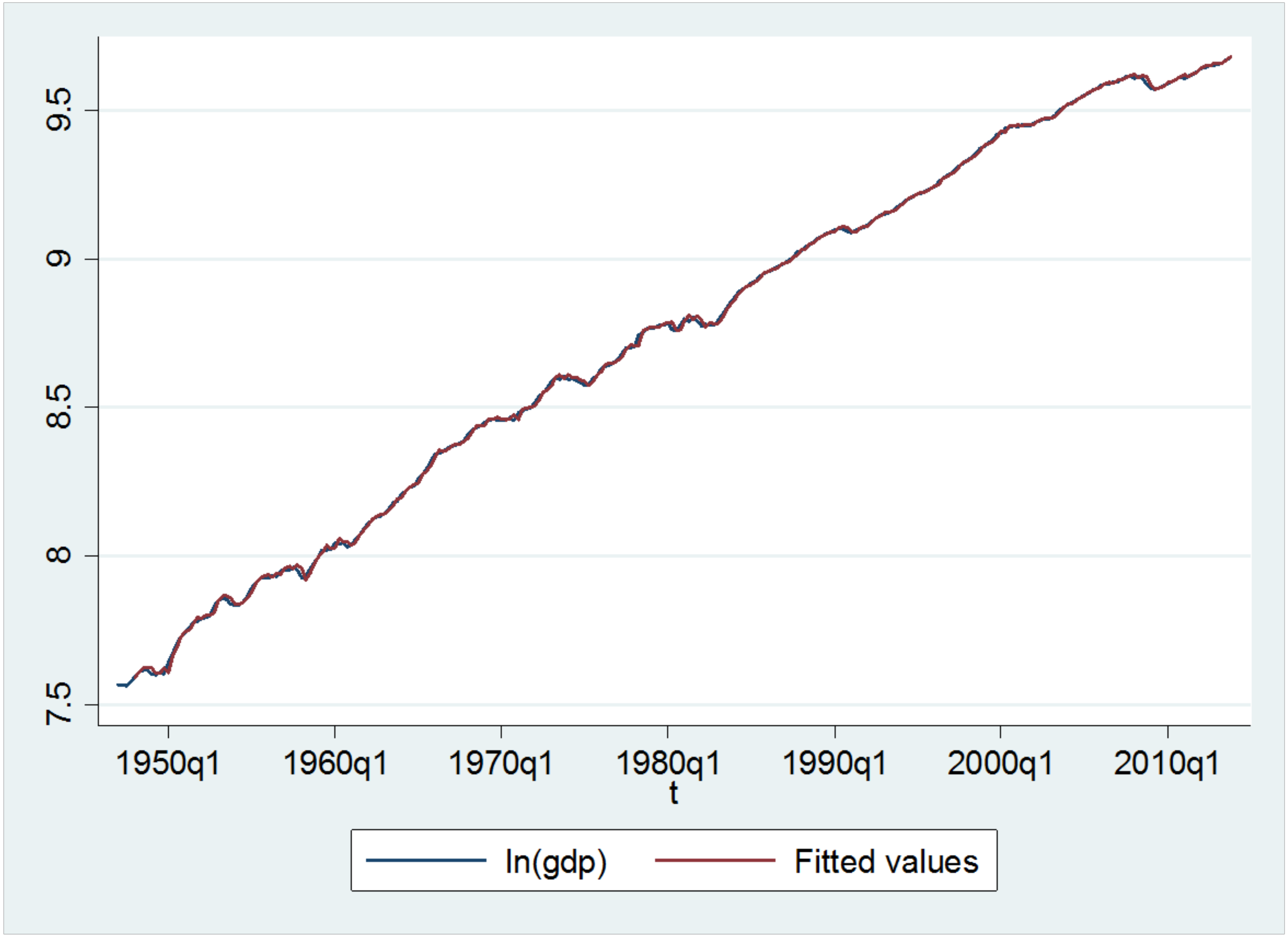
# Fitted Values



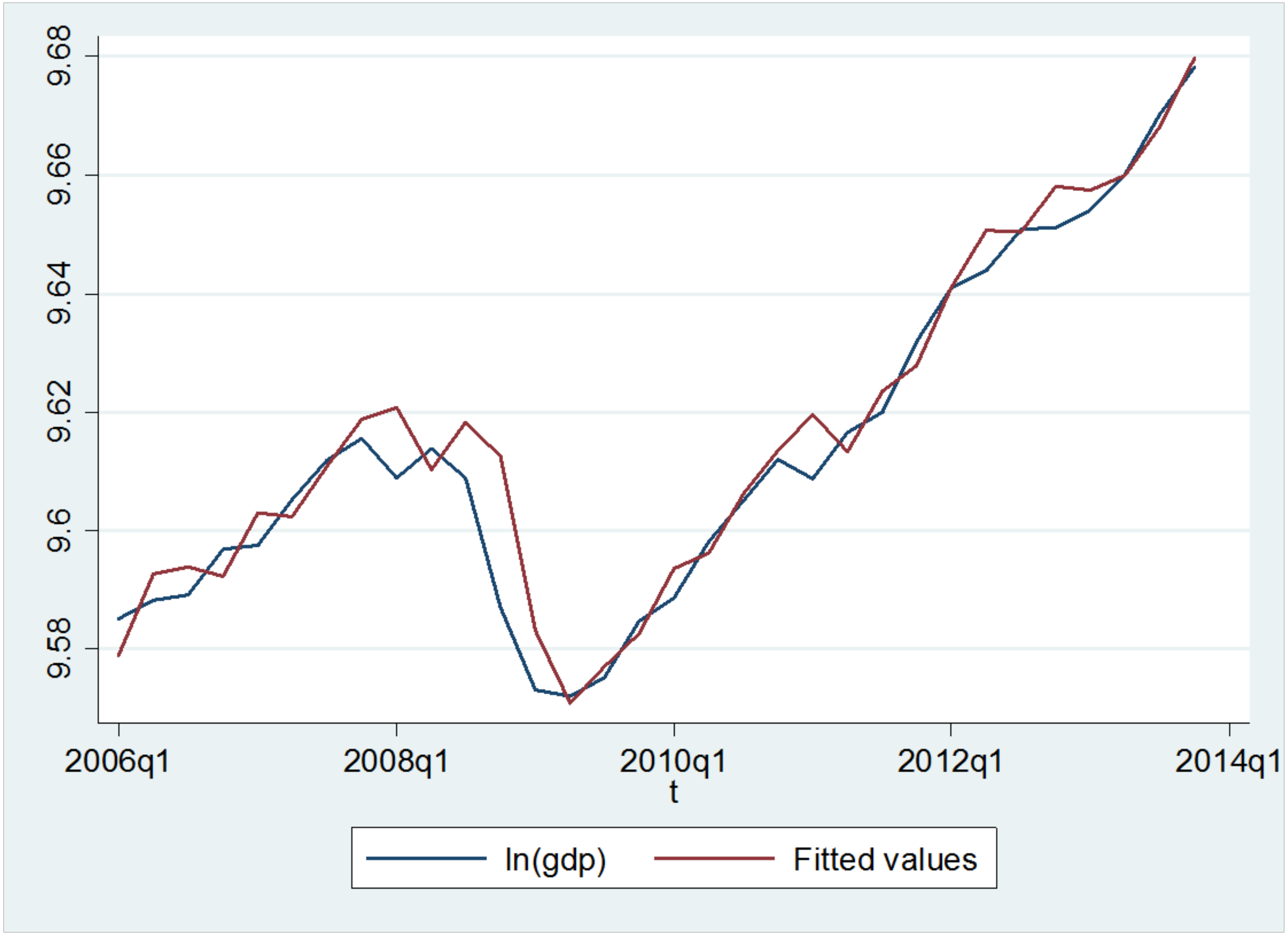
# AR(4) with trend

```
. reg y t L(1/4).y

      Source |       SS       df       MS              Number of obs =     264
-------------+------------------------------           F(  5,   258) =       .
       Model |  101.418159     5  20.2836318           Prob > F      =  0.0000
    Residual |  .020112173   258  .000077954           R-squared     =  0.9998
-------------+------------------------------           Adj R-squared =  0.9998
       Total |  101.438271   263  .385696849           Root MSE      =  .00883

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0001173   .0000836     1.40   0.162    -.0000472    .0002819
             |
           y |
         L1. |   1.325836   .0618984    21.42   0.000     1.203946    1.447727
         L2. |  -.2114016   .1019164    -2.07   0.039    -.4120956   -.0107077
         L3. |  -.2231381   .1019282    -2.19   0.029    -.4238553   -.0224208
         L4. |   .0927073   .0622436     1.49   0.138    -.0298629    .2152775
             |
       _cons |   .1350218   .0820877     1.64   0.101    -.0266254     .296669
------------------------------------------------------------------------------
```

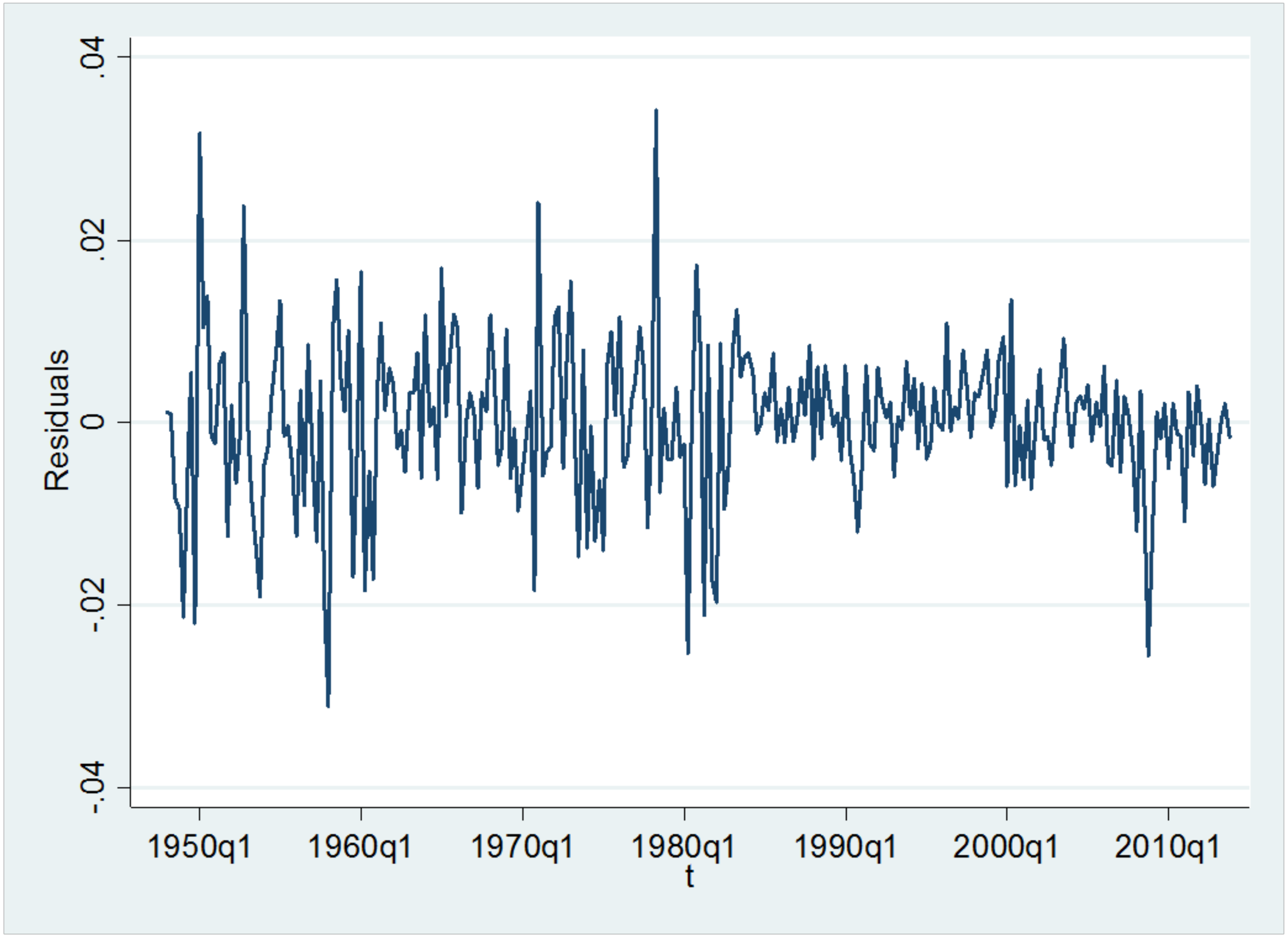
# Fitted Values from AR(4) with Trend



# Last 7 years


9.68
9.66
9.64
9.62
9.6
9.58
2006q1
2008q1
2010q1
2012q1
2014q1
t
ln(gdp)
Fitted values

# Residuals



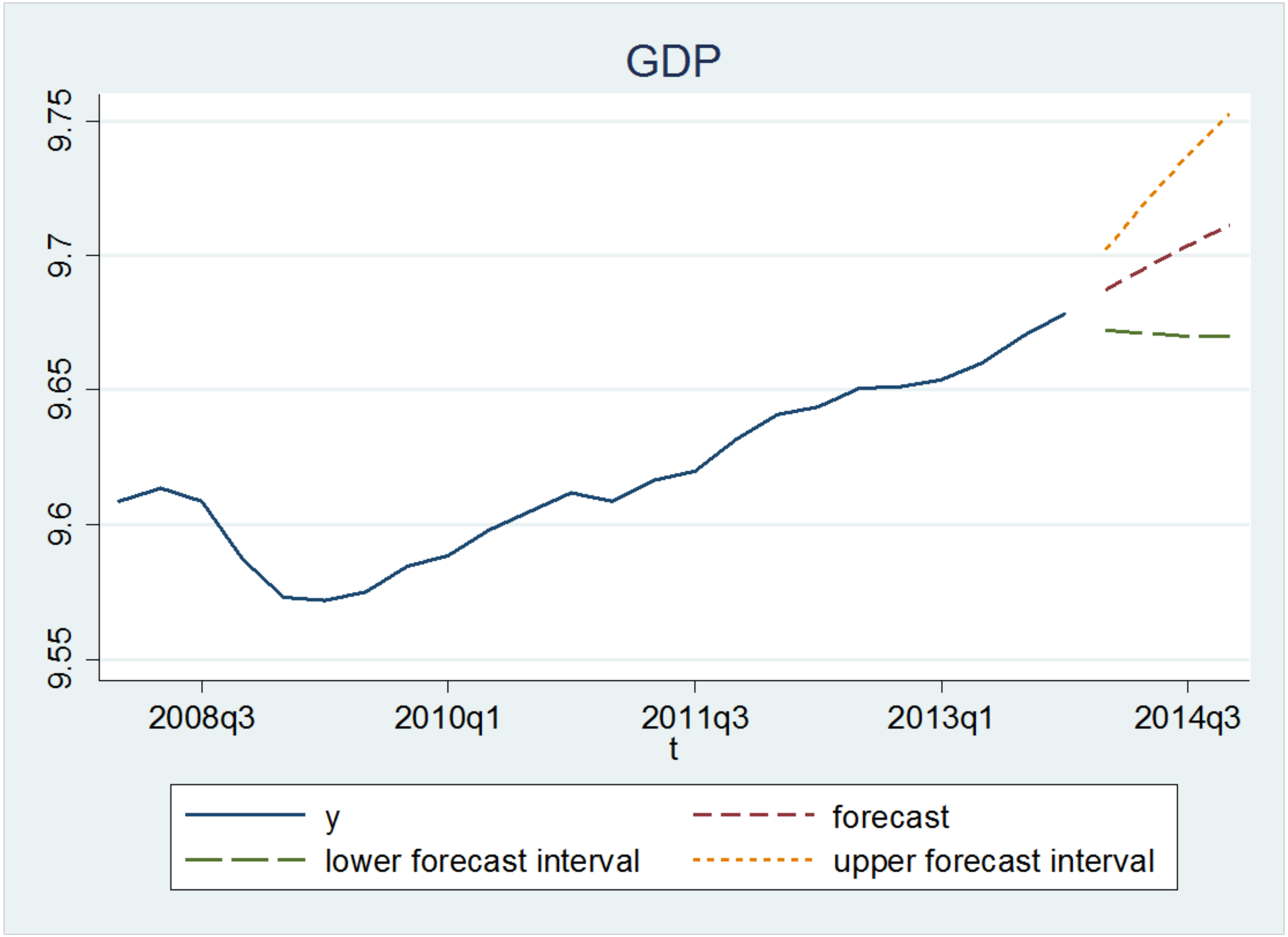
Residuals
.04
.02
0
-.02
-.04
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
t

# Forecasts

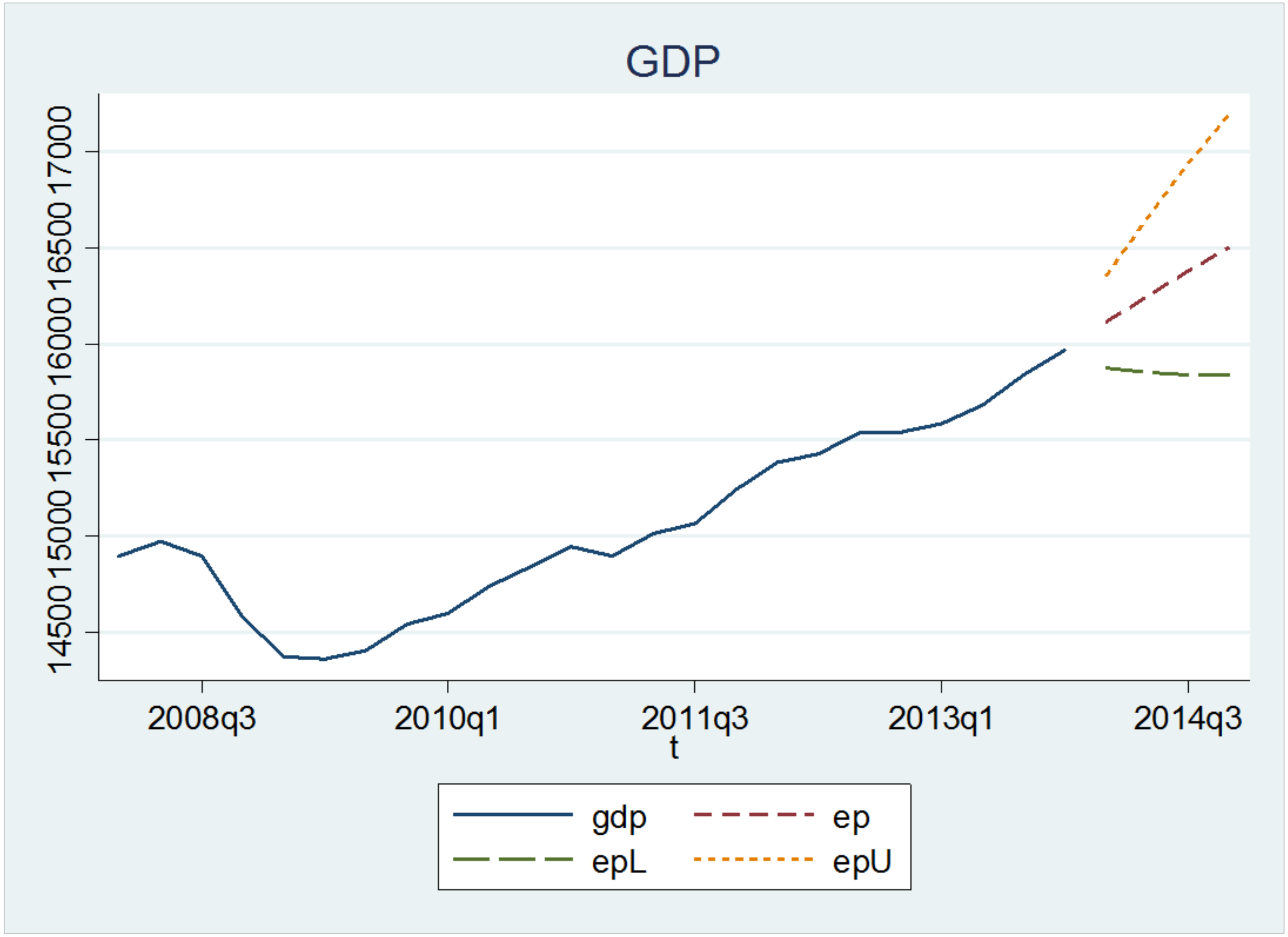
- Same as for AR(p) models, but include time trend as a regressor
- h-step forecast based on

$$y_t = \alpha + \gamma t + \beta_1 y_{t-h} + \cdots + \beta_p y_{t-h-p+1} + e_t$$

# Forecast for ln(GDP) using AR(4)+trend



# Forecast for GDP (using exponential)

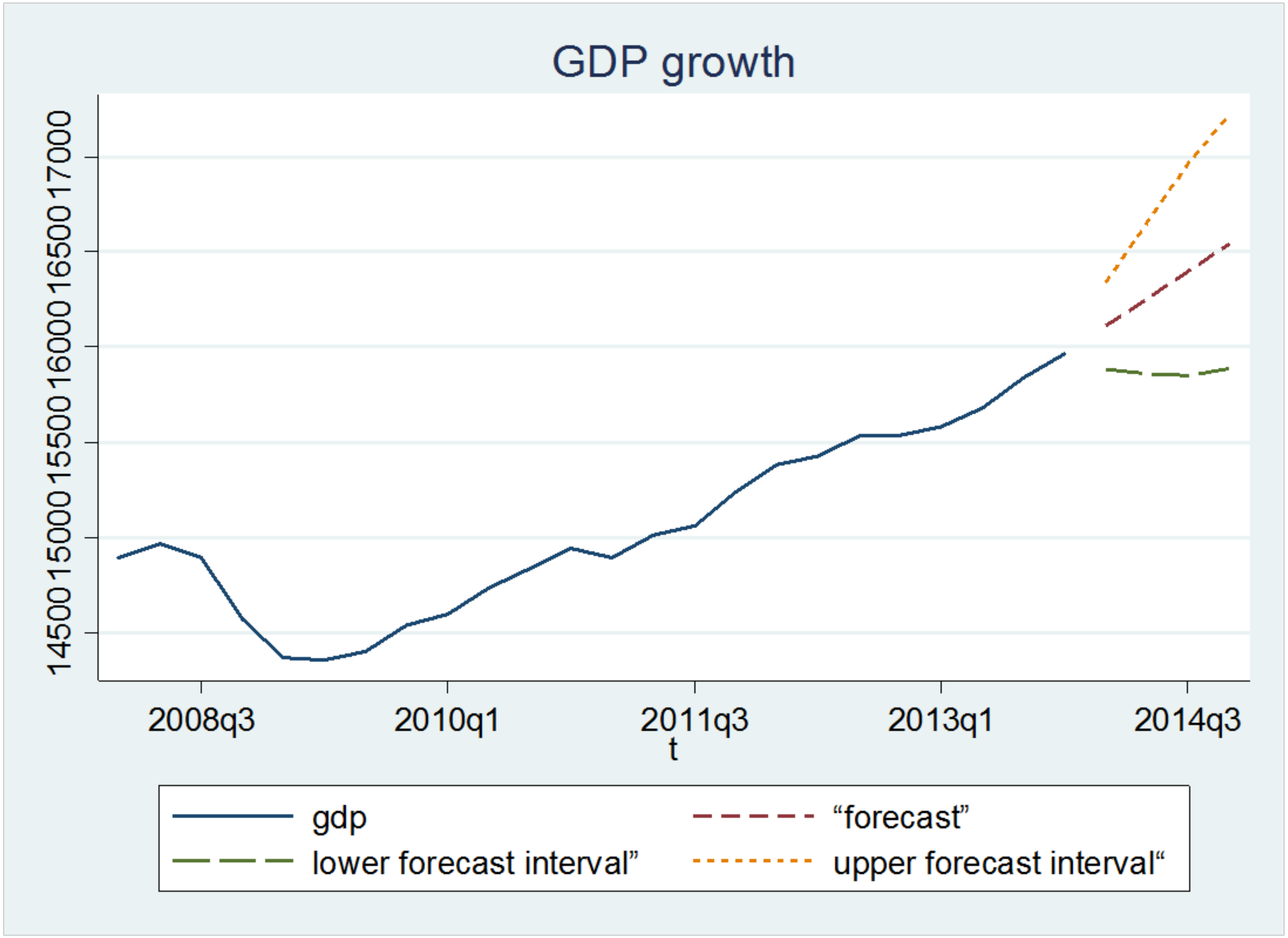


# Direct Interval Forecasts

```
tsappend, add(4)
gen y=ln(gdp)
reg y t L(1/4).y
predict y1
predict sf1,stdf
gen y1L=y1-1.645*sf1
gen y1U=y1+1.645*sf1
reg y t L(2/5).y
predict y2
predict sf2,stdf
gen y2L=y2-1.645*sf2
gen y2U=y2+1.645*sf2
reg y t L(3/6).y
predict y3
predict sf3,stdf
gen y3L=y3-1.645*sf3
gen y3U=y3+1.645*sf3
```

```
reg y t L(4/7).y
predict y4
predict sf4,stdf
gen y4L=y4-1.645*sf4
gen y4U=y4+1.645*sf4
egen p=rowfirst(y1 y2 y3 y4) if t>=tq(2014q1)
egen pL=rowfirst(y1L y2L y3L y4L) if t>=tq(2014q1)
egen pU=rowfirst(y1U y2U y3U y4U) if t>=tq(2014q1)
label variable p "forecast"
label variable pL "lower forecast interval"
label variable pU "upper forecast interval"
tsline y p pL pU if t>=tq(2008q1), title(GDP) lpattern
(solid dash longdash shortdash)
gen ep=exp(p)
gen epL=exp(pL)
gen epU=exp(pU)
tsline gdp ep epL epU if t>=tq(2008q1), title(GDP)
lpattern (solid dash longdash shortdash)
```

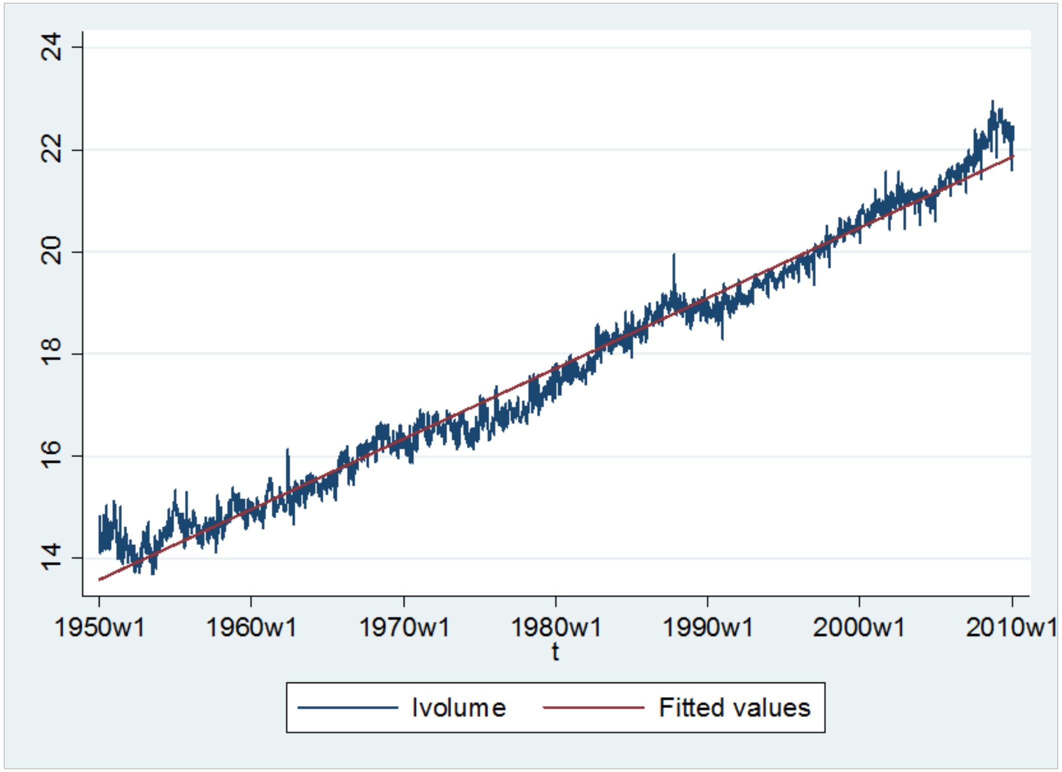
# Iterated Forecast for GDP



# Direct Interval Forecasts

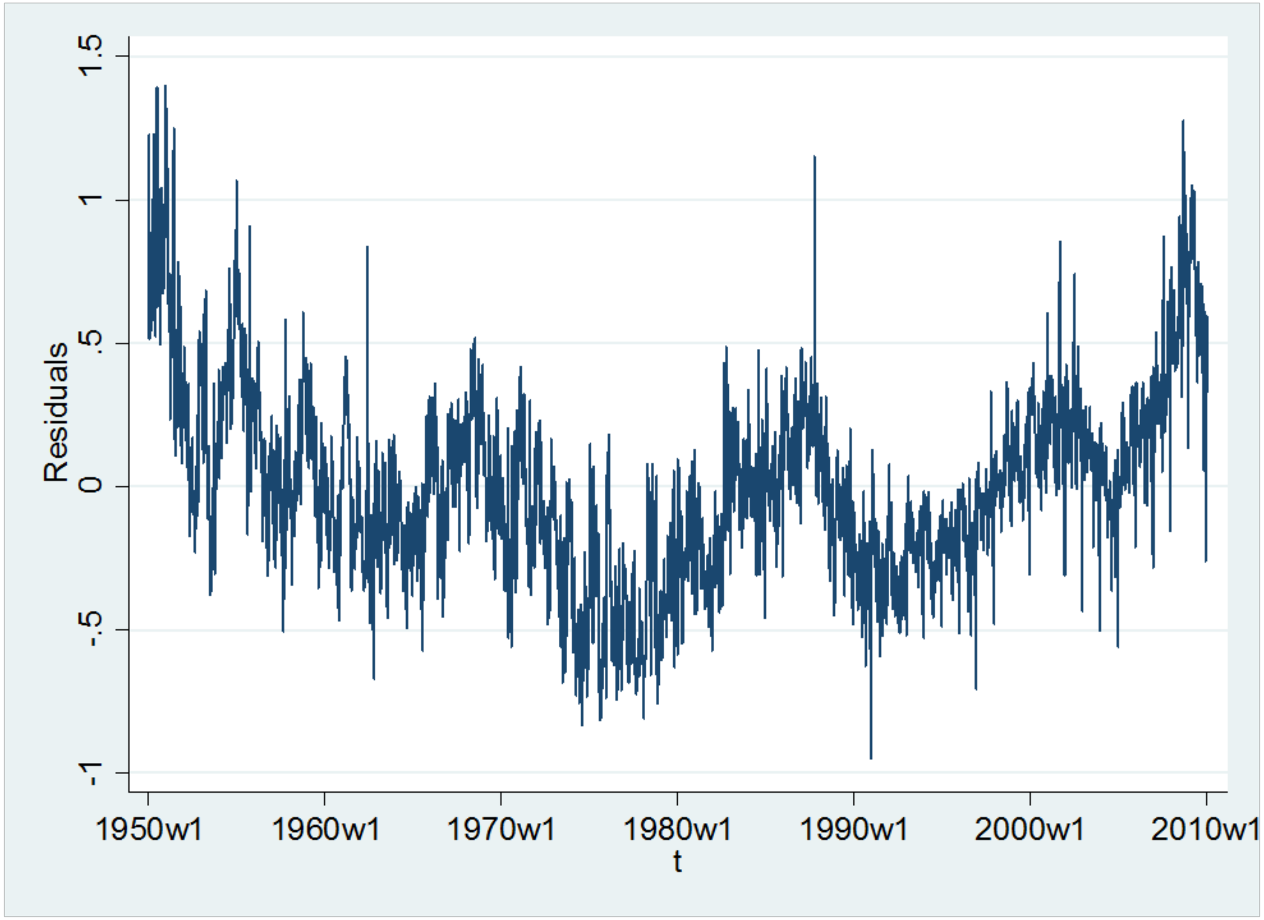
```
tsappend, add(4)
gen y=ln(gdp)
reg y t L(1/4).y
forecast create ar4
estimate store model1
forecast estimates model1
forecast solve, simulate(errors,statistic(stddev,prefix(sd_)) reps(1000) )
gen p=exp(f_y) if t>=tq(2014q1)
gen pL=exp(f_y-1.645*sd_y)
gen pU=exp(f_y+1.645*sd_y)
label variable p “forecast”
label variable pL "lower forecast interval”
label variable pU "upper forecast interval“
tsline gdp p pL pU if t>=tq(2008q1), title(GDP growth) lpattern (solid dash longdash shortdash)
```

# Example: Log Stock Volume




- Log Volume and Linear Trend

# Residuals

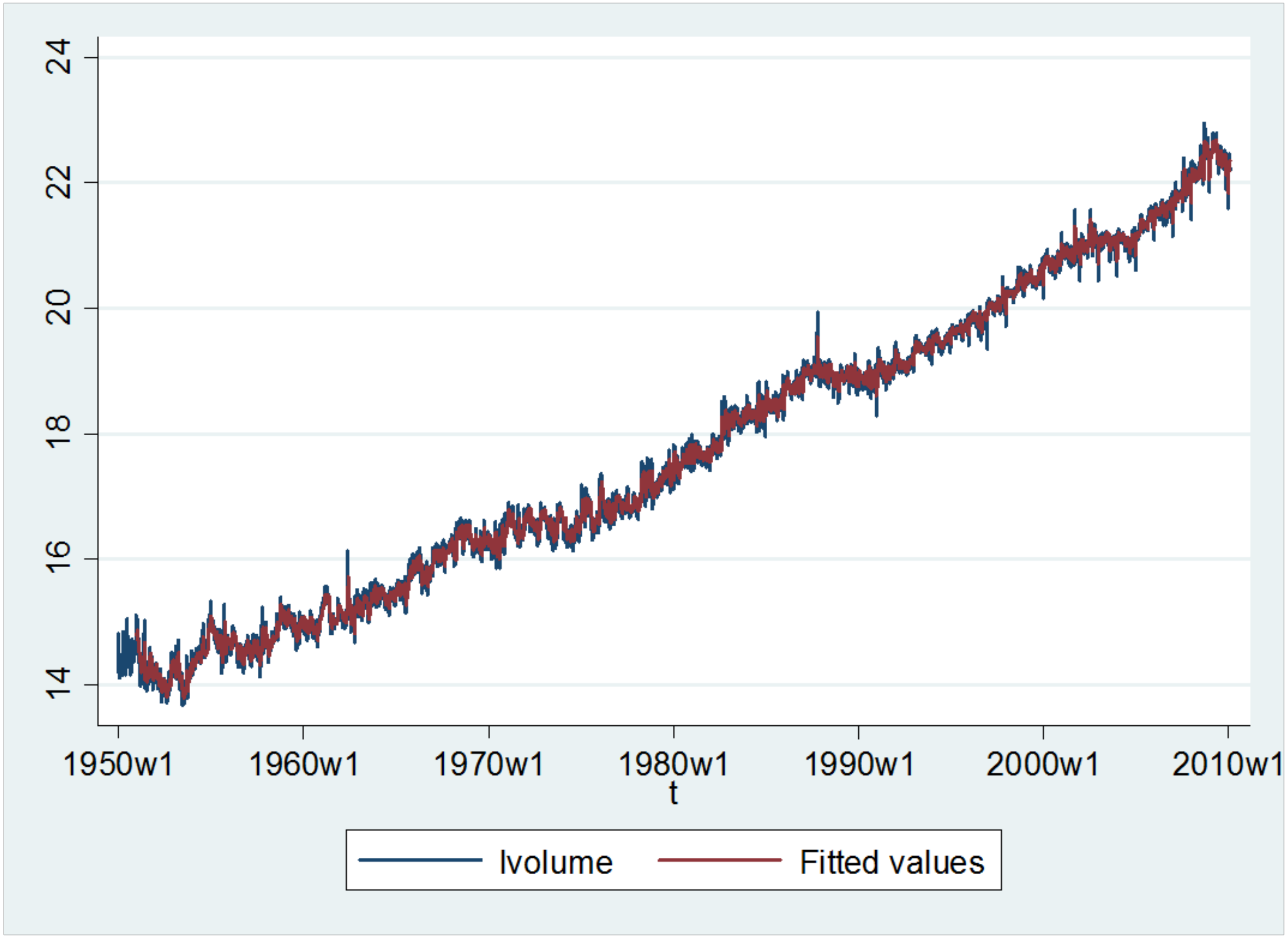


Residuals
1.5
1
.5
0
-.5
-1
1950w1
1960w1
1970w1
1980w1
1990w1
2000w1
2010w1
t

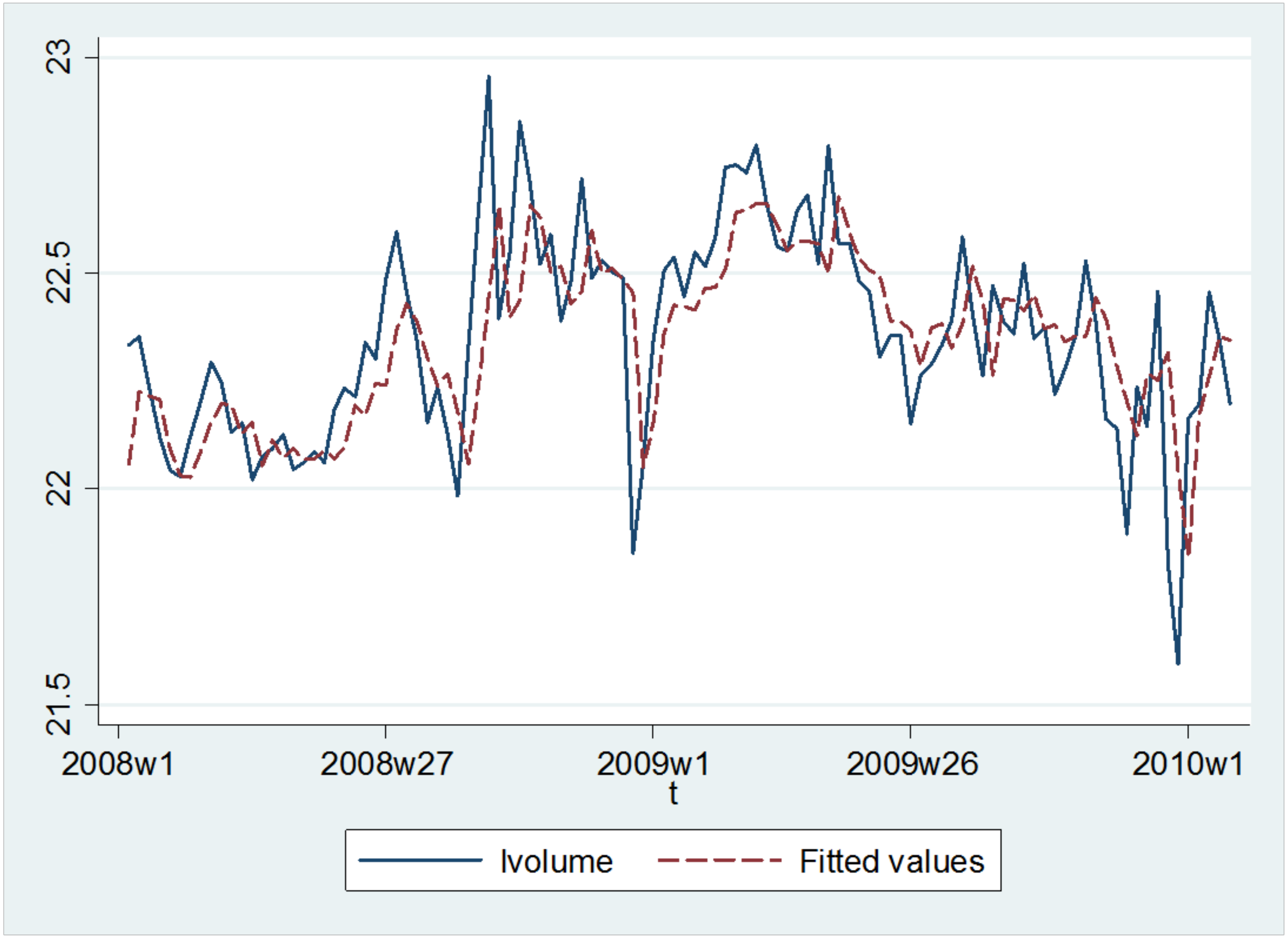
# Model

- Weekly Data through 2009
- 3082 observations
- Fit AR(52)+trend

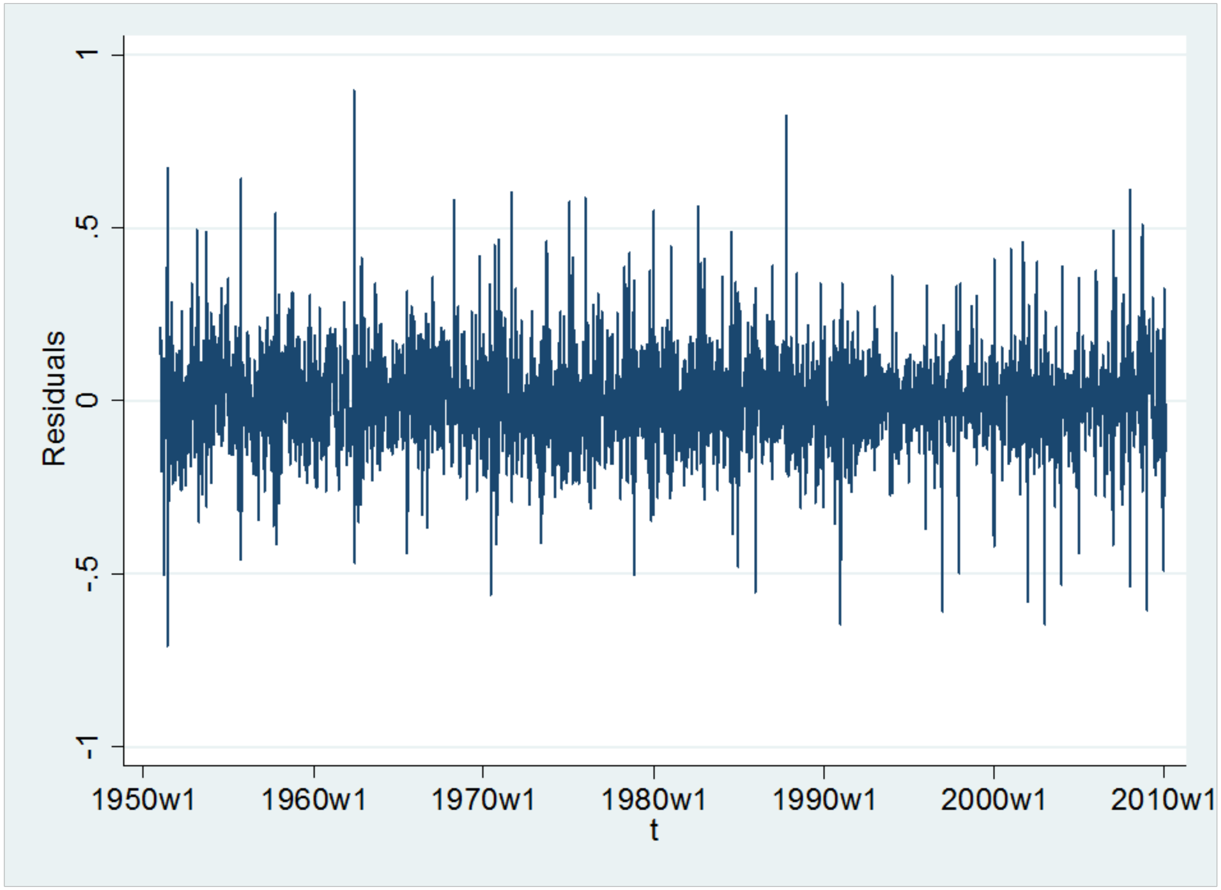
# Data and Fitted



# Last Two Years of Sample



# Residuals



1
.5
0
-.5
-1
Residuals
1950w1
1960w1
1970w1
1980w1
1990w1
2000w1
2010w1
t

# Trend Omission

- Suppose the truth is that the data have a trend, but you fit an AR model without a trend.
- What happens?
- Suppose

$$y_t = \mu_1 + \mu_2 t$$

- Then

$$y_t = y_{t-1} + \mu_2$$

# Example

- Since

$$y_t = y_{t-1} + \mu_2$$

- If you estimate an AR(1), you obtain

$$y_t = \hat{\alpha} + \hat{\beta} y_{t-1} + \hat{e}_t$$

$$= \mu_2 + y_{t-1}$$

$$\hat{\alpha} = \mu_2$$

$$\hat{\beta} = 1$$

- You estimate a unit coefficient on the AR lag

# General Effect of Trend Omission

- If the truth is

$$y_t = \mu_1 + \mu_2 t + \beta y_{t-1} + e_t$$

- But you estimate an AR(1) **without** a trend

$$y_t = \hat{\alpha} + \hat{\beta} y_{t-1} + \hat{e}_t$$

- Then you tend to find

$$\hat{\beta} \approx 1$$

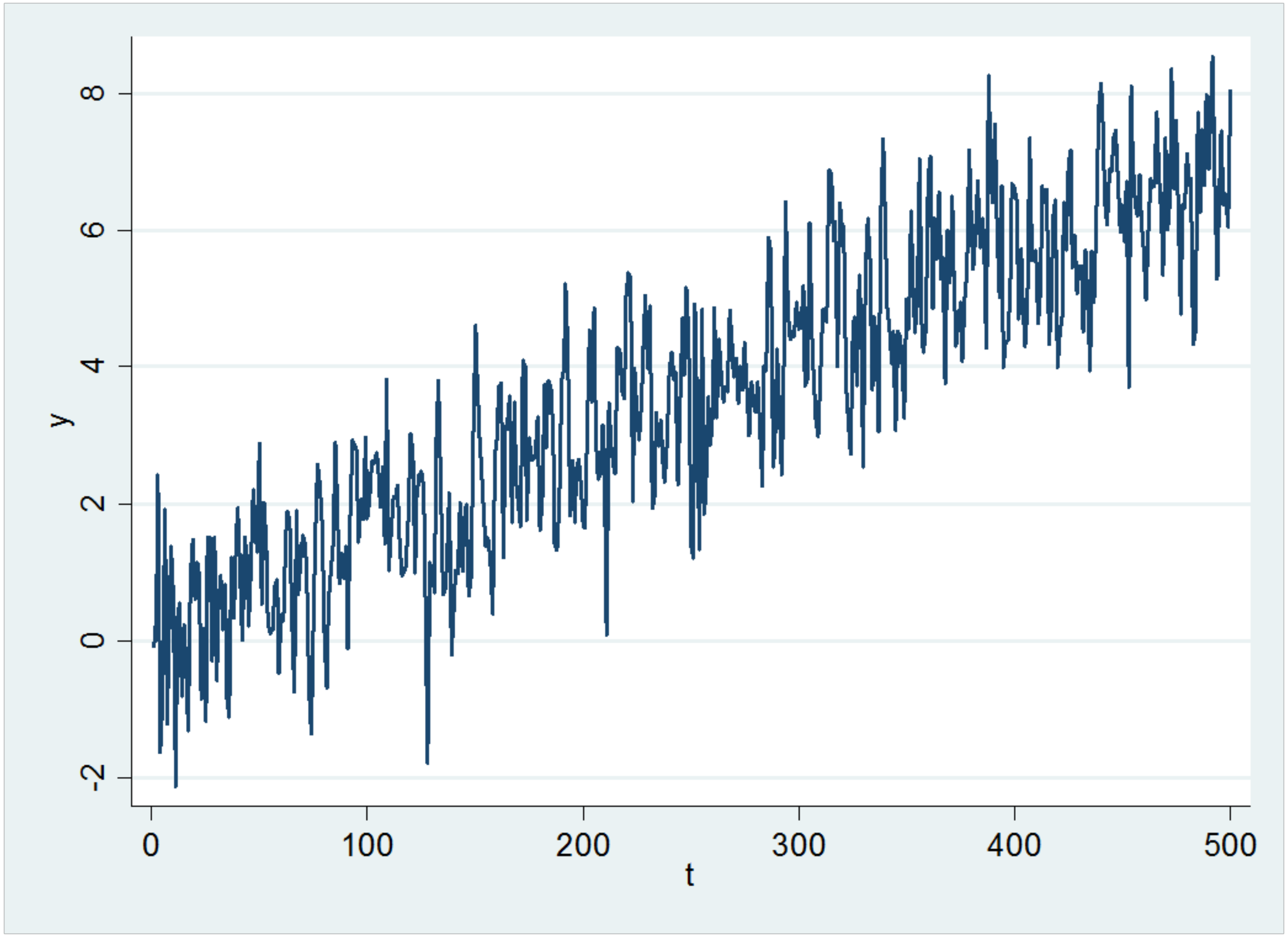
- This is due to model misspecification

# Simulated Example

$$y_t = .01t + .3y_{t-1} + e_t$$

```
. gen e=rnormal(0)

. gen y=e

. replace y=.01*t+.3*L.y+e if t>1
(499 real changes made)
```

# Simulated Process



# Estimate AR(1) without Trend

```
. reg y L.y

      Source |       SS       df       MS              Number of obs =     499
-------------+------------------------------           F(  1,   497) = 1212.27
       Model |  2031.45086     1  2031.45086           Prob > F      =  0.0000
    Residual |  832.842361   497  1.67573916           R-squared     =  0.7092
-------------+------------------------------           Adj R-squared =  0.7086
       Total |  2864.29322   498  5.75159281           Root MSE      =  1.2945

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           y |
         L1. |   .8423331   .0241927    34.82   0.000     .7948006    .8898657
             |
       _cons |   .5786825   .1039191     5.57   0.000     .3745076    .7828573
------------------------------------------------------------------------------
```

- The estimated AR(1) coefficient is 0.84, much too large (true value was 0.3)

# Estimate AR(1) with Trend

```
. reg y t L.y

      Source |       SS       df       MS              Number of obs =     499
-------------+------------------------------           F(  2,   496) = 1032.23
       Model |  2309.43596     2  1154.71798           Prob > F      =  0.0000
    Residual |  554.857262   496  1.11866383           R-squared     =  0.8063
-------------+------------------------------           Adj R-squared =  0.8055
       Total |  2864.29322   498  5.75159281           Root MSE      =  1.0577

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0114629   .0007272    15.76   0.000     .0100342    .0128916
             |
           y |
         L1. |   .2274365   .0437293     5.20   0.000      .141519     .313354
             |
       _cons |  -.1060309    .095372    -1.11   0.267    -.2934137     .081352
------------------------------------------------------------------------------
```

- The estimated AR coef is 0.23, close to the true 0.3
- The estimated trend coef is 0.11, close to the true 0.10
- The root MSE decreases from 1.29 to 1.06

# Seasonality

- Recall that we said that it can be useful to describe the mean of a time series as the sum of components

$$\mu_t = T_t + S_t + C_t$$

  where $S_t$ is the seasonal component.
- The seasonal component $S_t$ is a repetitive cycle over the calendar year
- Seasonality $S_t$ can be deterministic (predictable) or stochastic

# Seasonality – Examples

- Gasoline consumption rises in summer due to increased auto travel
- International airline prices rise in summer due to increased tourism
- Natural gas consumption and prices rise in winter due to heating
- Electricity consumption increases in summer due to air conditioning
- Construction activity and jobs decrease in winter in the Midwest
- Consumer spending increases in November and December due to holiday shopping

# Deterministic vs Stochastic Seasonality

- If the seasonal pattern repeats year after year, it is deterministic and predictable.
  - Christmas is always in December
- If the seasonal pattern roughly repeats itself, but evolves over the years, it is stochastic and only partially predictable
  - Holiday shopping as a percentage of income is not a fixed constant
- Seasonal patterns can change dramatically as the economy evolves
  - The spread of air conditioning shifted the seasonal pattern of residential electricity consumption from winter to summer

# Seasonal Adjustment

- Most economic indicators reported by the government are **seasonally adjusted**.
- Roughly, the component $S_t$ is estimated, and then what is reported is

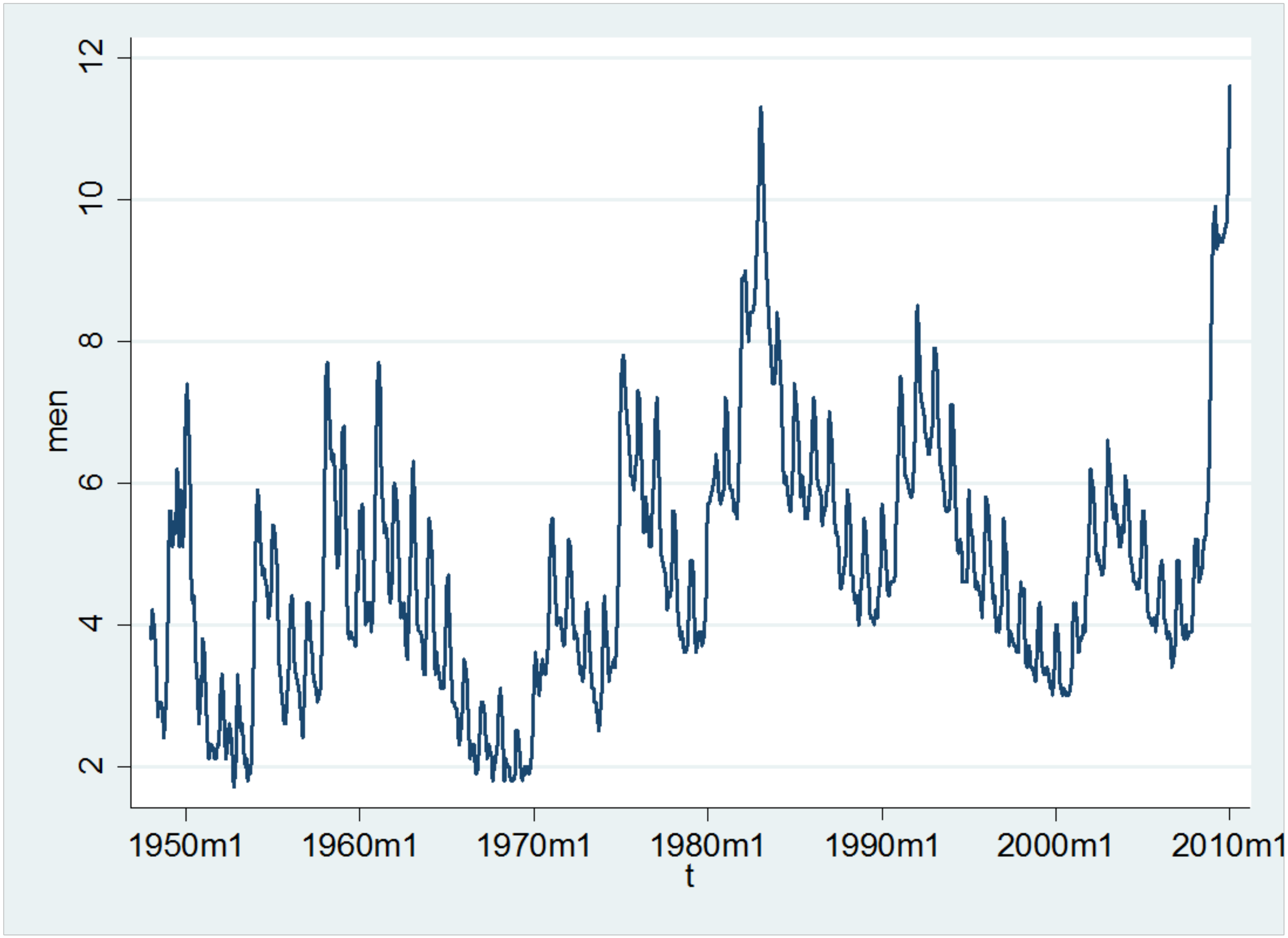
$$\begin{aligned} y_t^* &= y_t - S_t \\ &= T_t + C_t \end{aligned}$$

- The idea is that seasonality distracts from the main reporting purpose
  - Seasonally adjusted data allows users to focus on trend and business cycle movements
- Seasonal adjustment by central statistical agencies is sophisticated, allowing for evolving seasonal patterns.

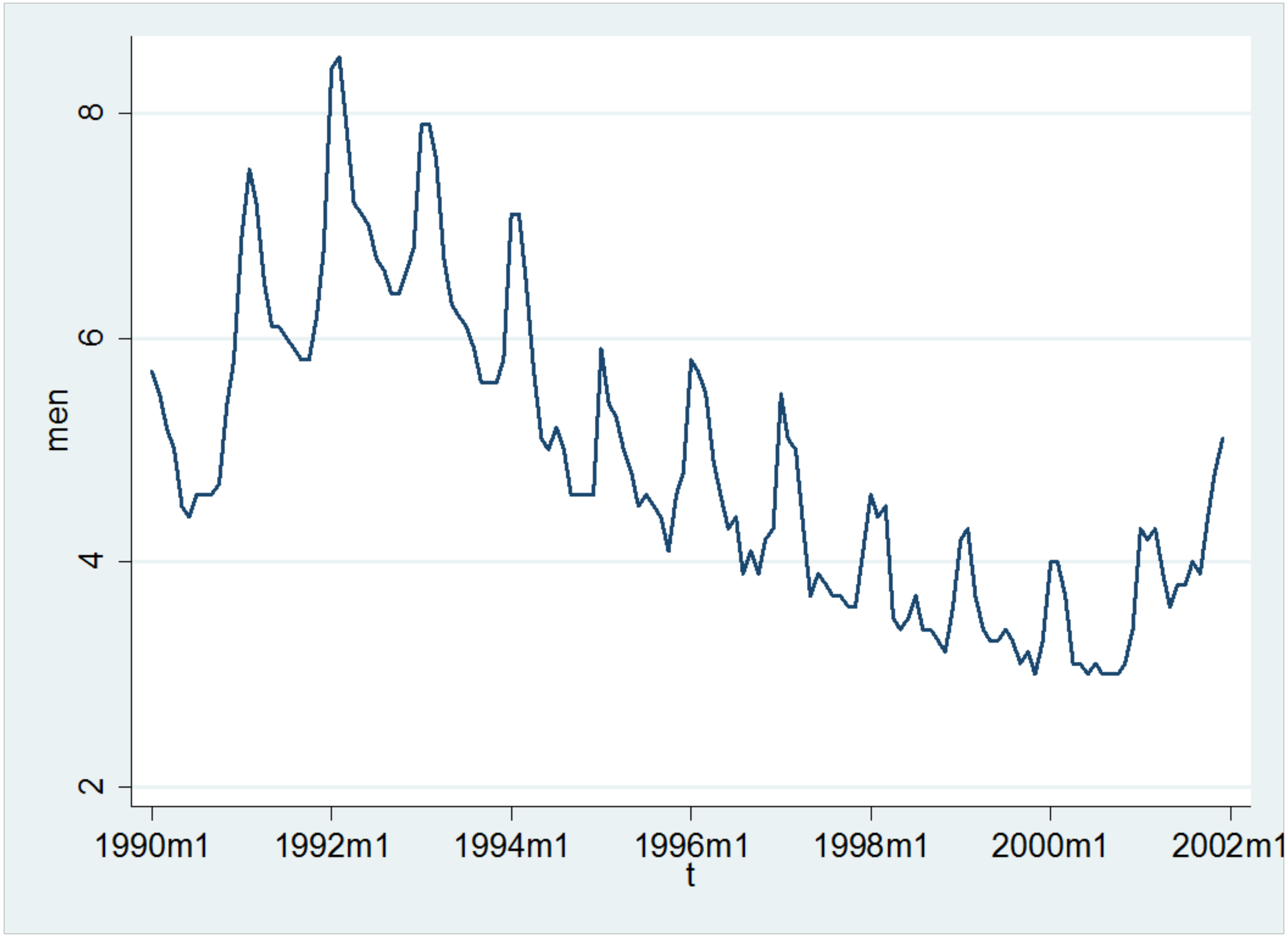
# Examples of Seasonal Time Series

- First Example:
- U.S. Unemployment Rate
  - Men, 20+ years
  - 1948-2009
  - Not seasonally adjusted

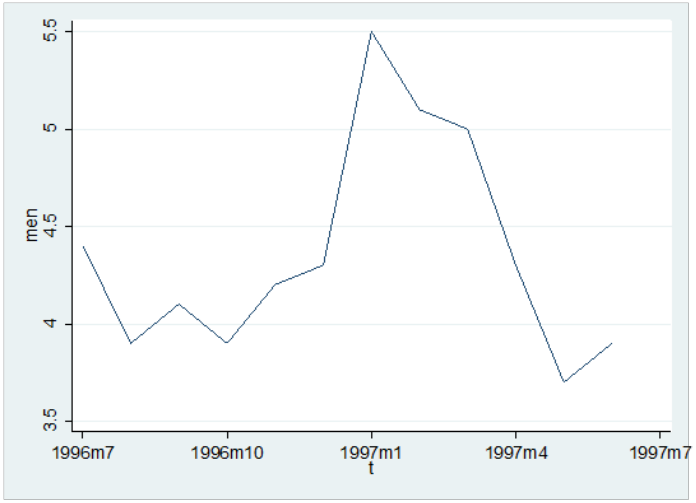
# U.S. Unemployment Rate
# Men, 20+ years, 1948-2009

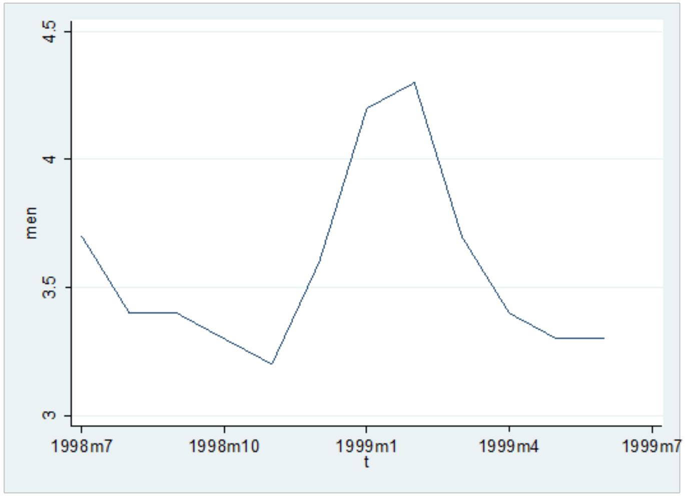


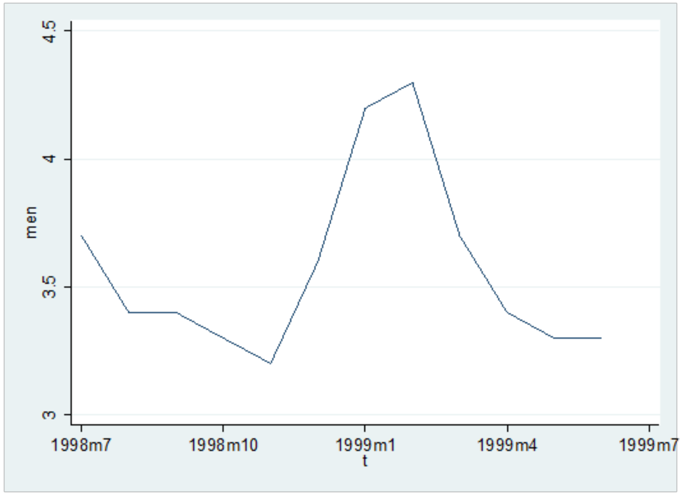
# Unemployment Rate, 1990-2001

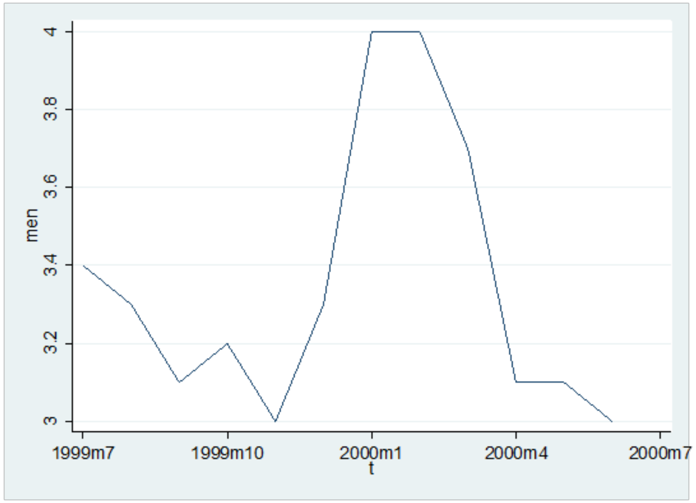


# Unemployment Rate, by year

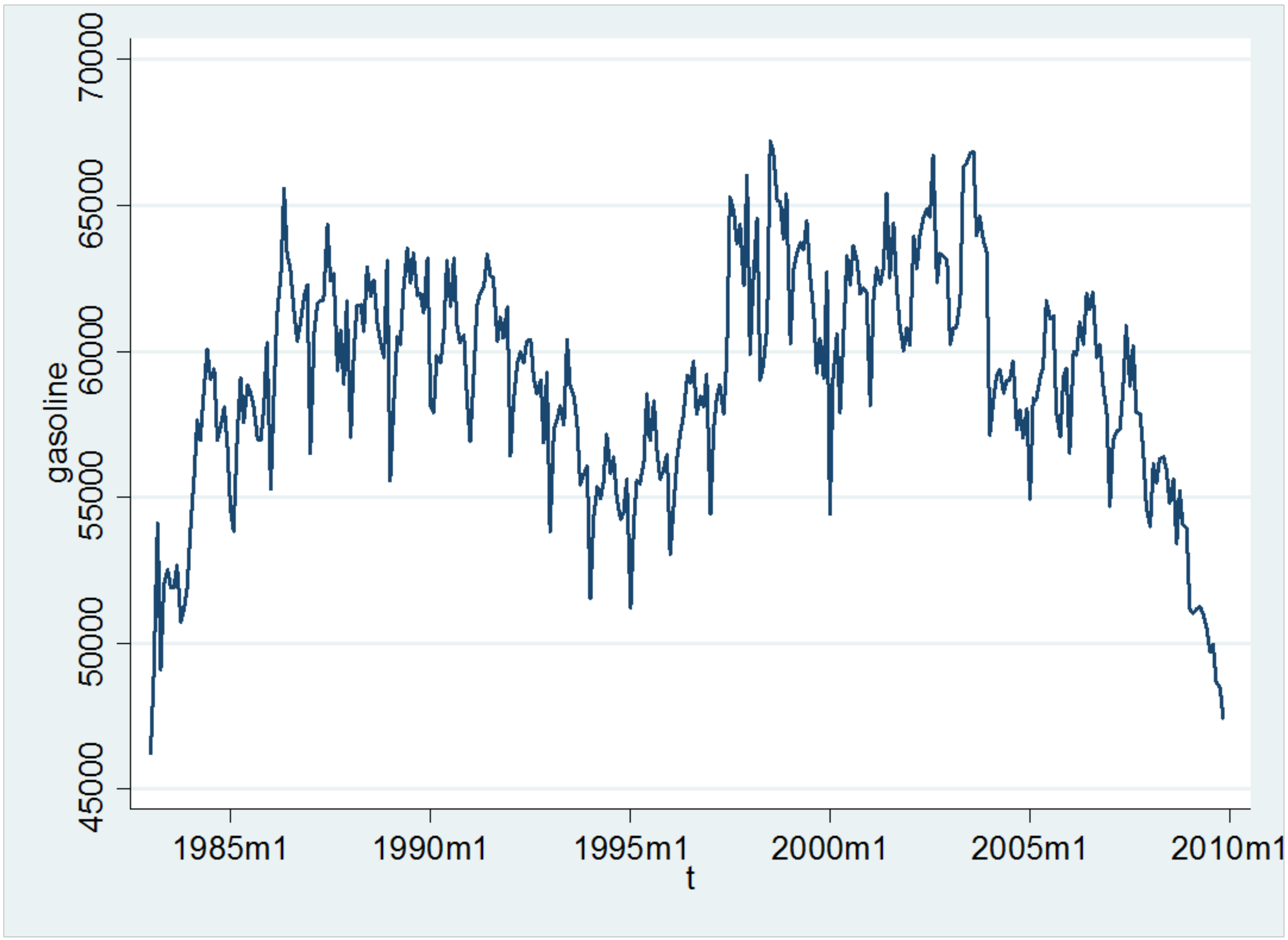


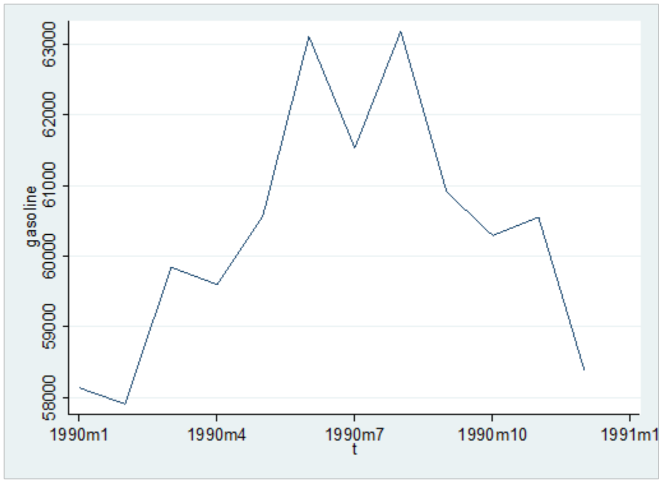



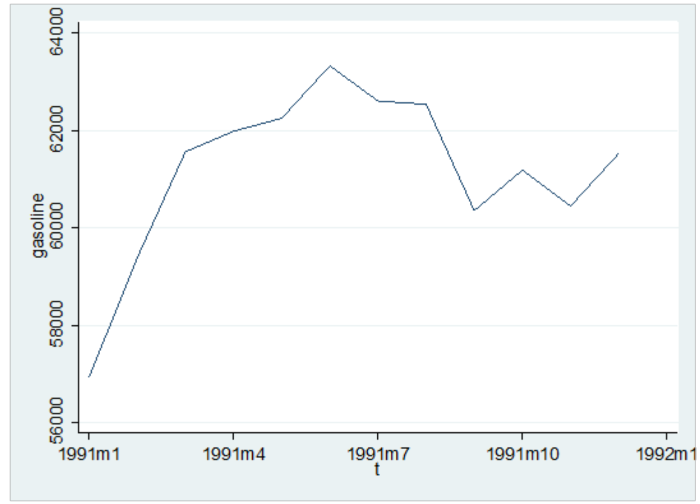



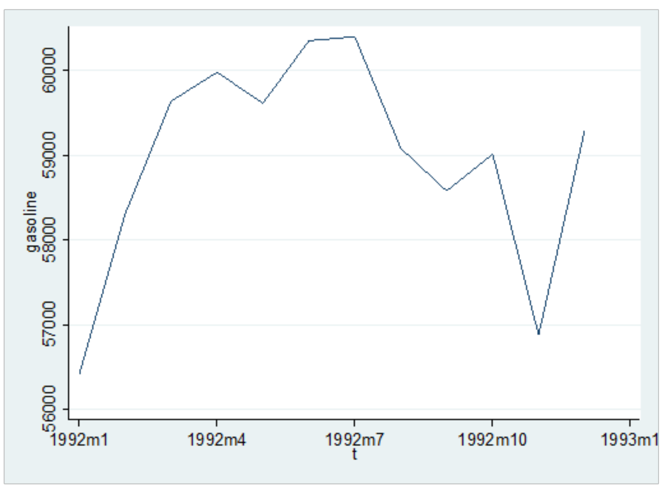
# Example 2
# U.S. Gasoline Sales Volume

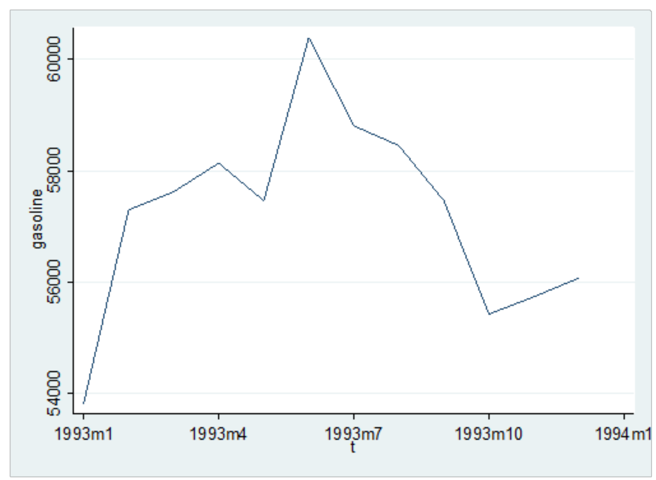


# Gasoline Sales, by year

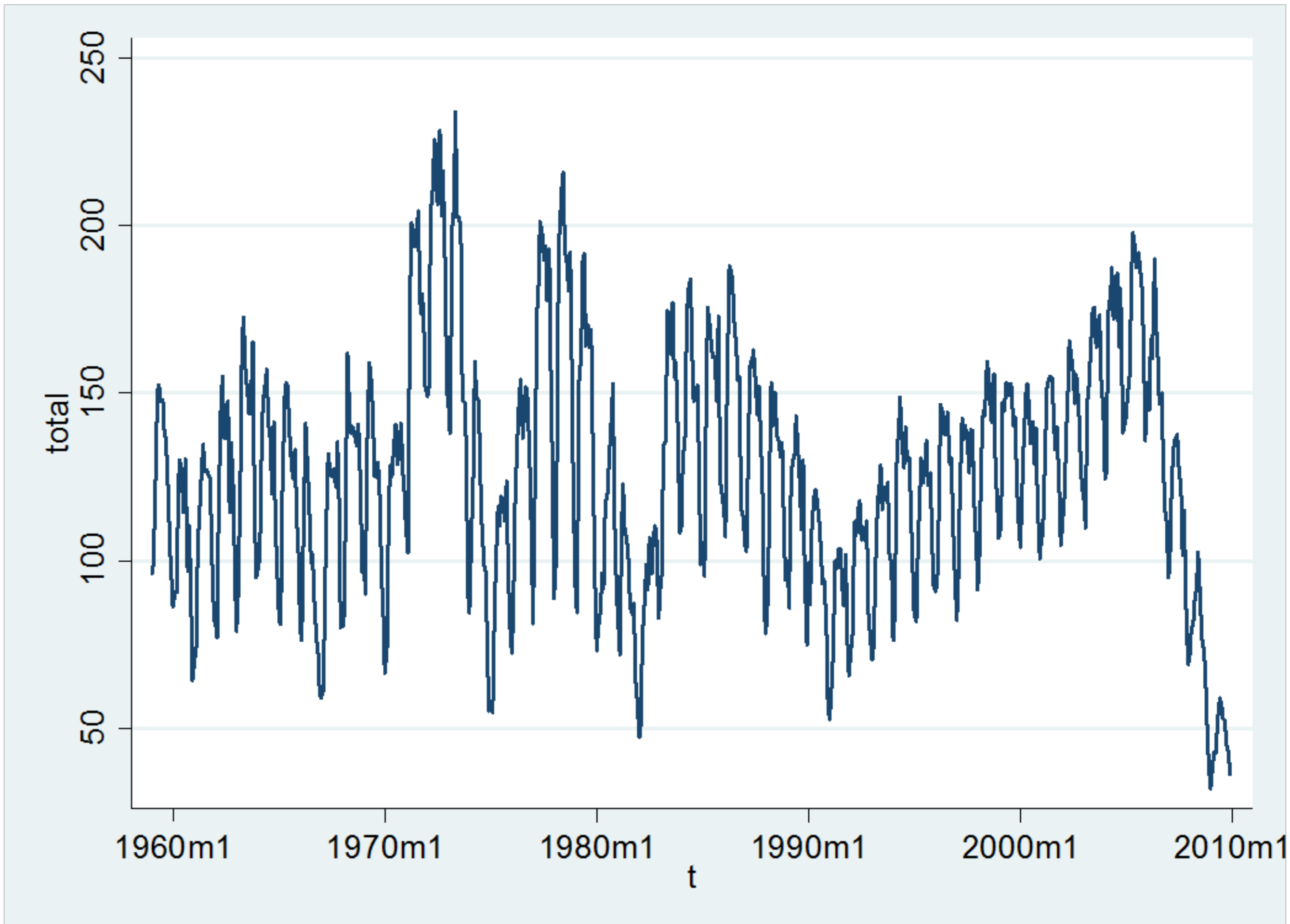


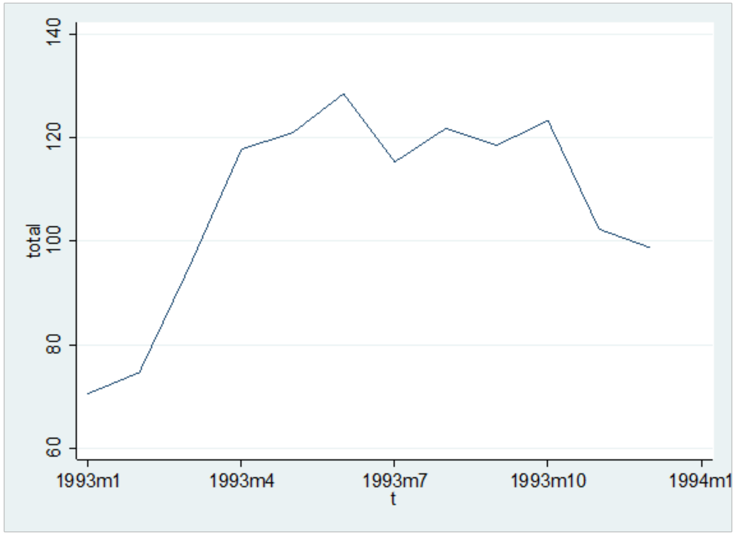



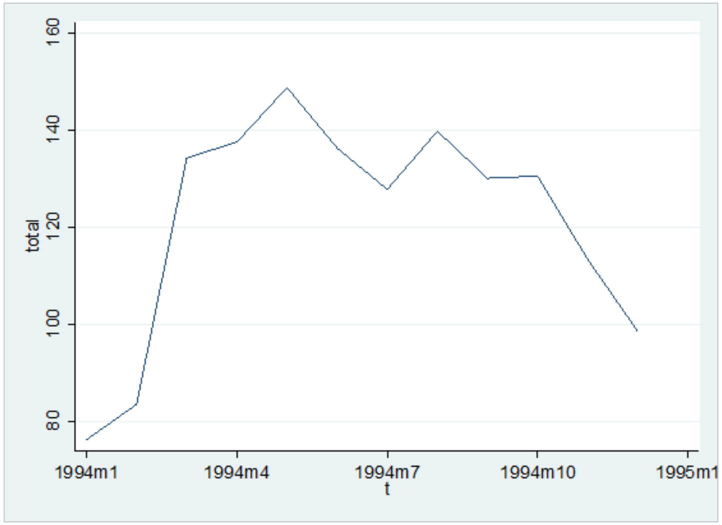



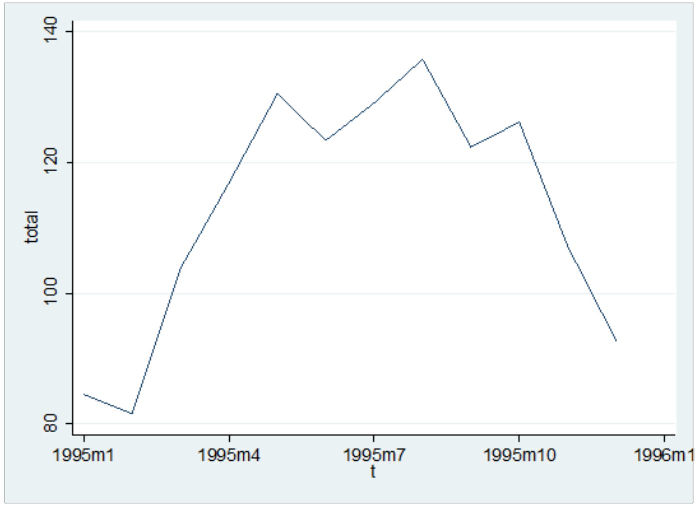
# Example 3
# U.S. Housing Starts
# (New Privately Owned Housing Units)

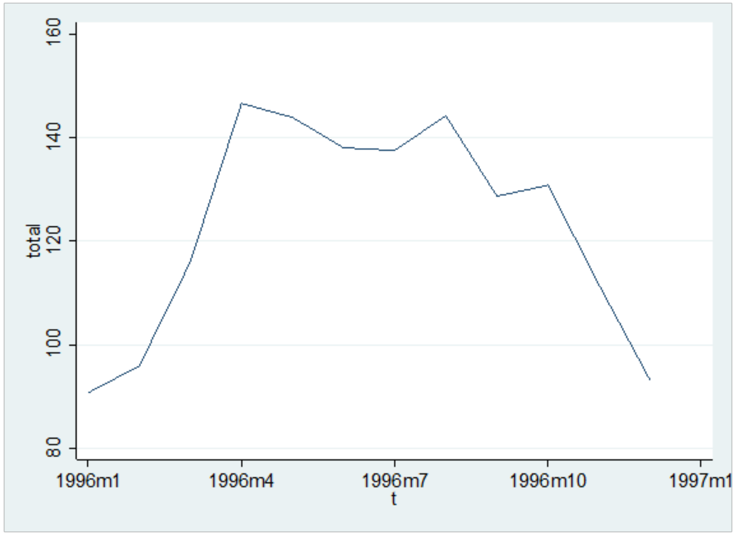


# Housing Starts, by year



# Deterministic Seasonality

- If seasonality is constant and deterministic then $S_t$ is simply a different constant for each period
- For example, for monthly data

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

- Seasonality is a constant which varies by the calendar period (quarter, month, week, day, or time of day)

# Fitted Values and Forecasts
# Pure Deterministic Seasonality

- In the simple pure deterministic seasonality model, fitted values and forecasts are the simple seasonal pattern

# Example – 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 |