Trend Models

- A trend model is

$$T_t = g(Time_t)$$

  where $Time_t$ is the time index.
- In STATA, $Time_t$ is an integer sequence, normalized to be zero at first observation of 1960.
- Most common models
  - Linear Trend
  - Exponential Trend
  - Quadratic Trend
  - Trends with Changing SLope

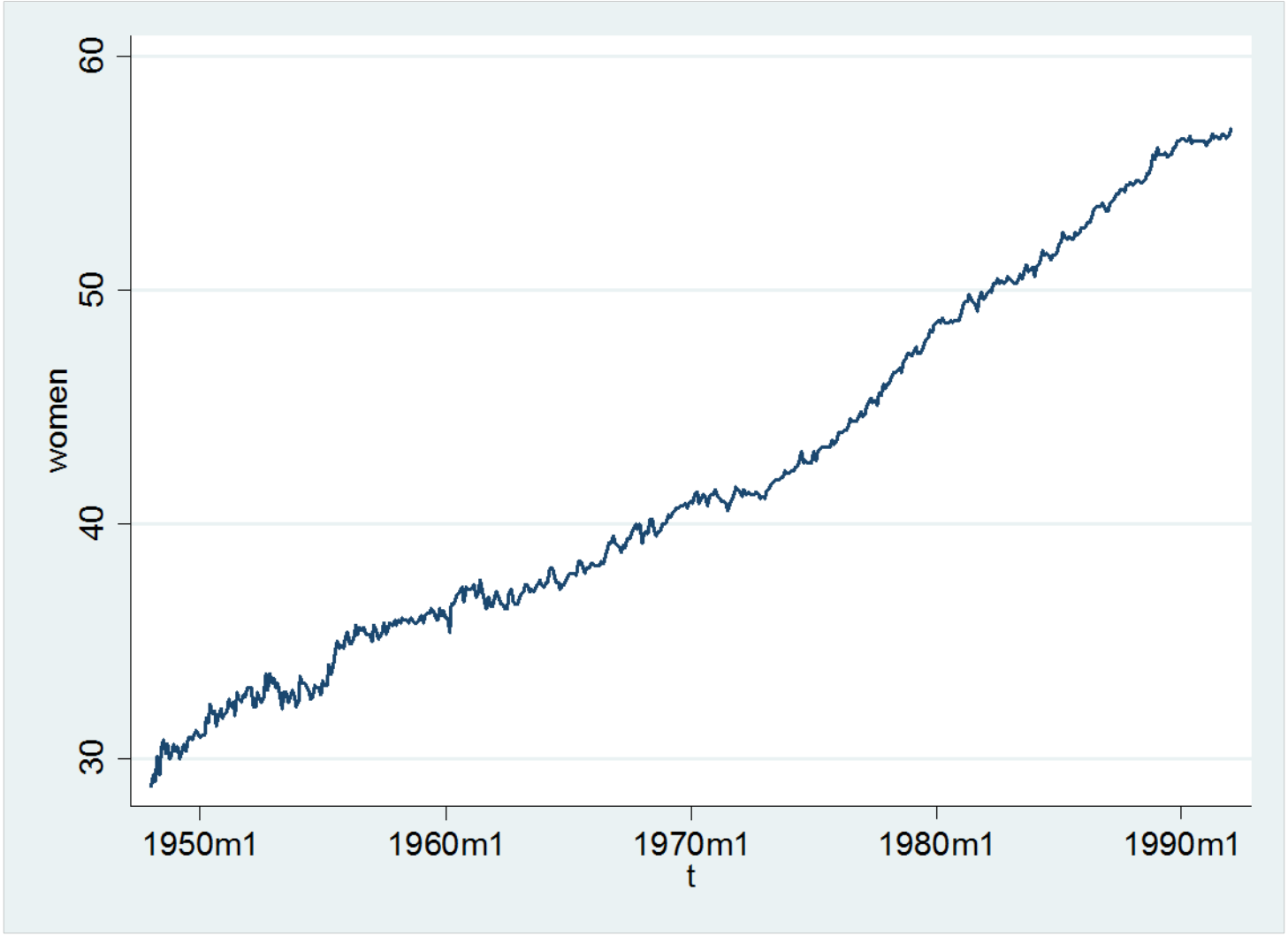
Warning:
Be skeptical of Trend Models

- While in some cases, trend forecasting can be useful.
- In many cases, it can be hazardous.
- We will examine some of the trend examples in Chapter 5 of Diebold's text
- They did not forecast well out of sample.
- A constructive alternative is to forecast growth rates, as we did for consumption expenditure.

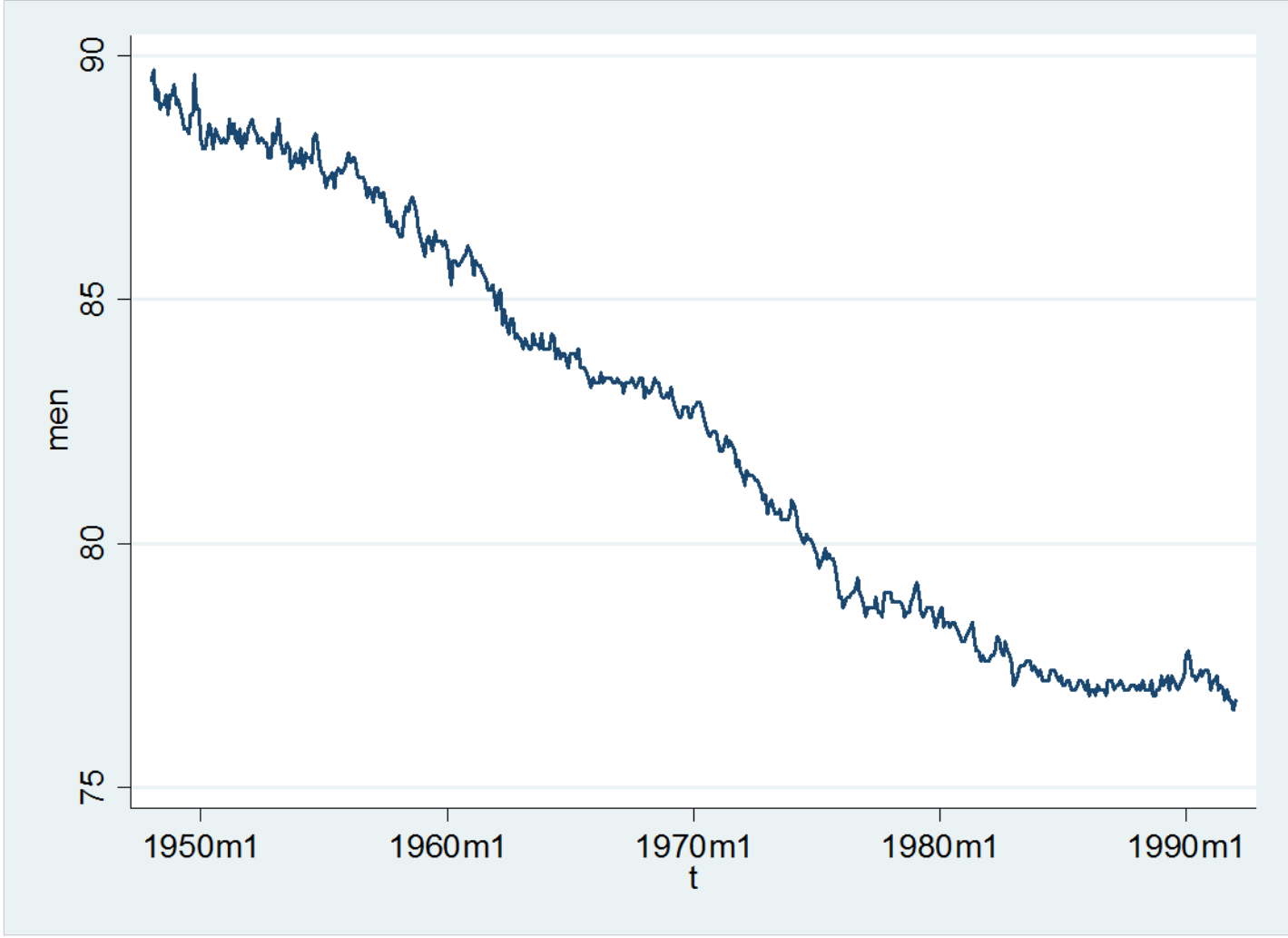
Example 1
Labor Force Participation Rate

- From BLS
- Monthly, 1948-2009, Seasonally adjusted
- Men and Women, ages 25+
- Percentage of population in labor force (employed plus unemployed divided by population)
- Diebold estimates on 1948-1992
- We will estimate on 1948-1992, forecast 1993-2009

Women's Labor Participation Rate 1948-1992



Men's Labor Participation Rate 1948-1992



Linear Trend Model

- The labor force participation rates have been smoothly and linearly increasing (for women) and smoothly and linearly decreasing (for men) over 1948-1992
- This suggests a linear trend

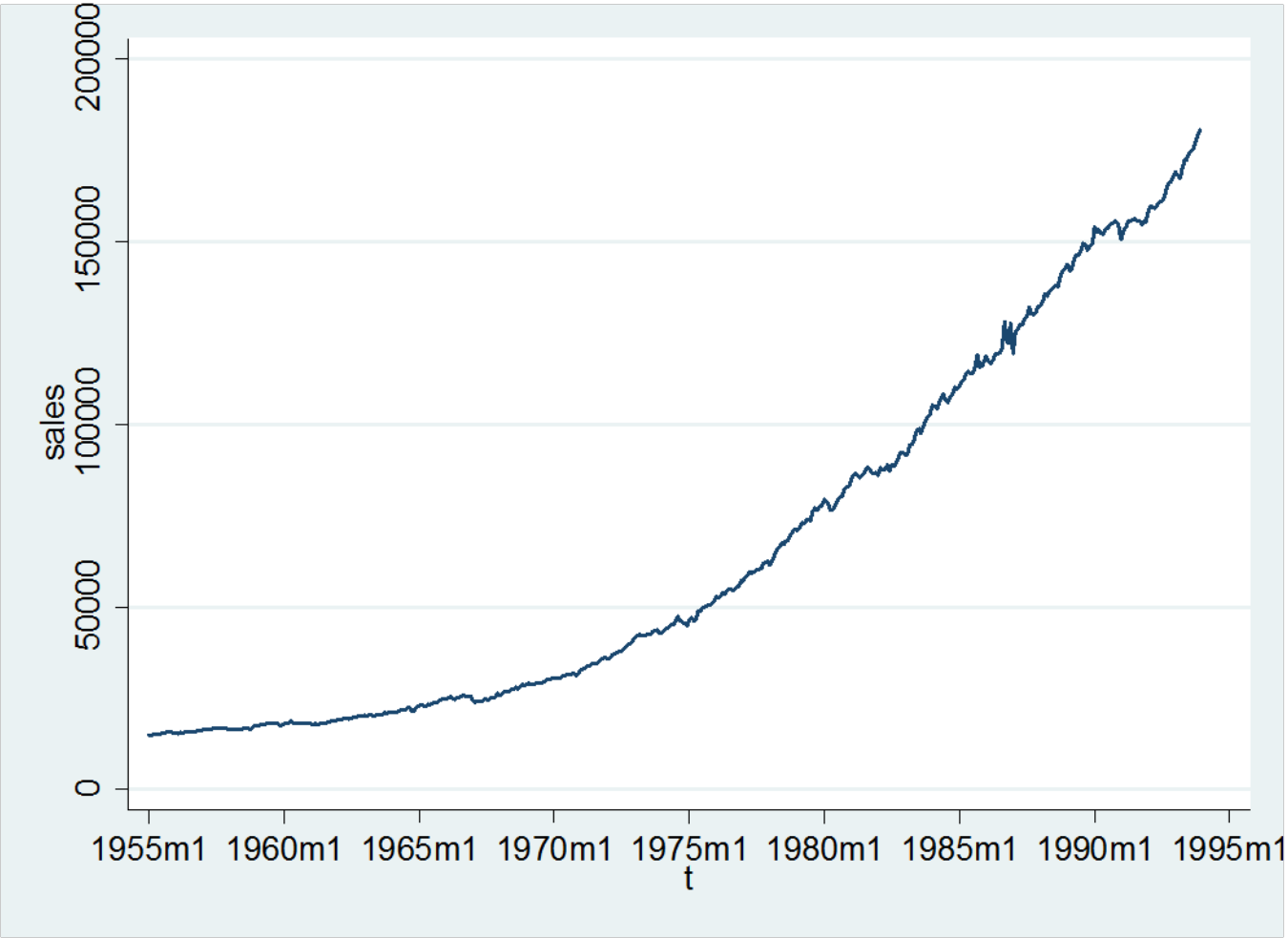
$$T_t = \beta_0 + \beta_1 Time_t$$

- In this model, β_1 is the expected period-to-period change in the trend T_t

Example 2
Retail Sales, Current Dollars

- From Census Bureau
- Monthly, 1955-2001, seasonally adjusted
  - This particular series discontinued after 2001
- Diebold estimates up to 1991
- We will forecast 1992-current

Retail Sales
1955-1993



Quadratic Trends

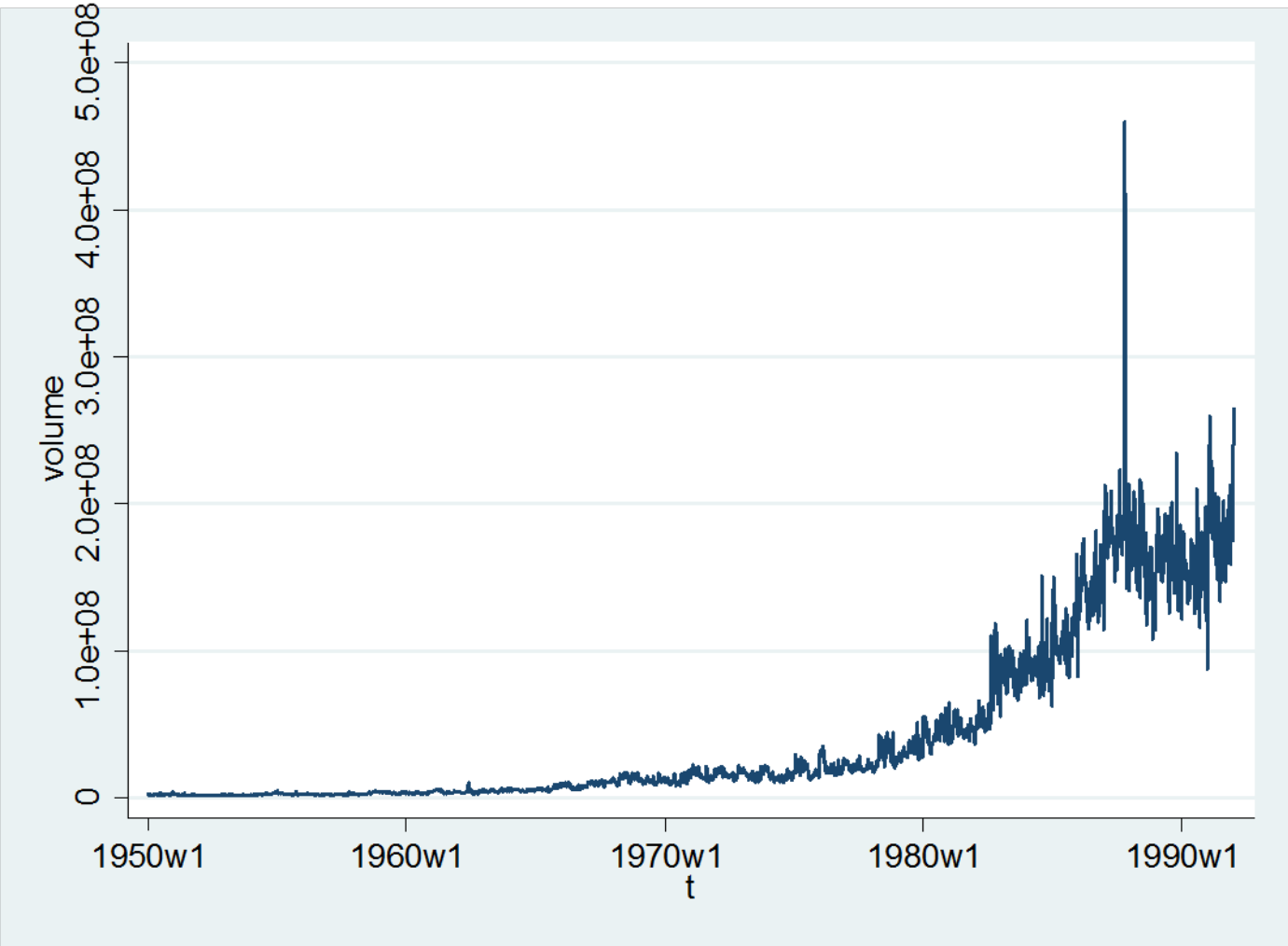
- The retail sales series has been increasing smoothly over 1955-1993, but not linearly.
- To model this we will use a quadratic trend

$$T_t = \beta_0 + \beta_1 Time_t + \beta_2 Time_t^2$$

Example 3
Transaction Volume, S&P Index

- From Yahoo Finance
- (Similar to NYSE series in Diebold)
- Weekly, 1950-current
- Diebold estimates on 1955-1993, forecasts 1994
- We will forecast 1994-2001

Transaction Volume



Exponential Trend

- To model this we will use an exponential trend

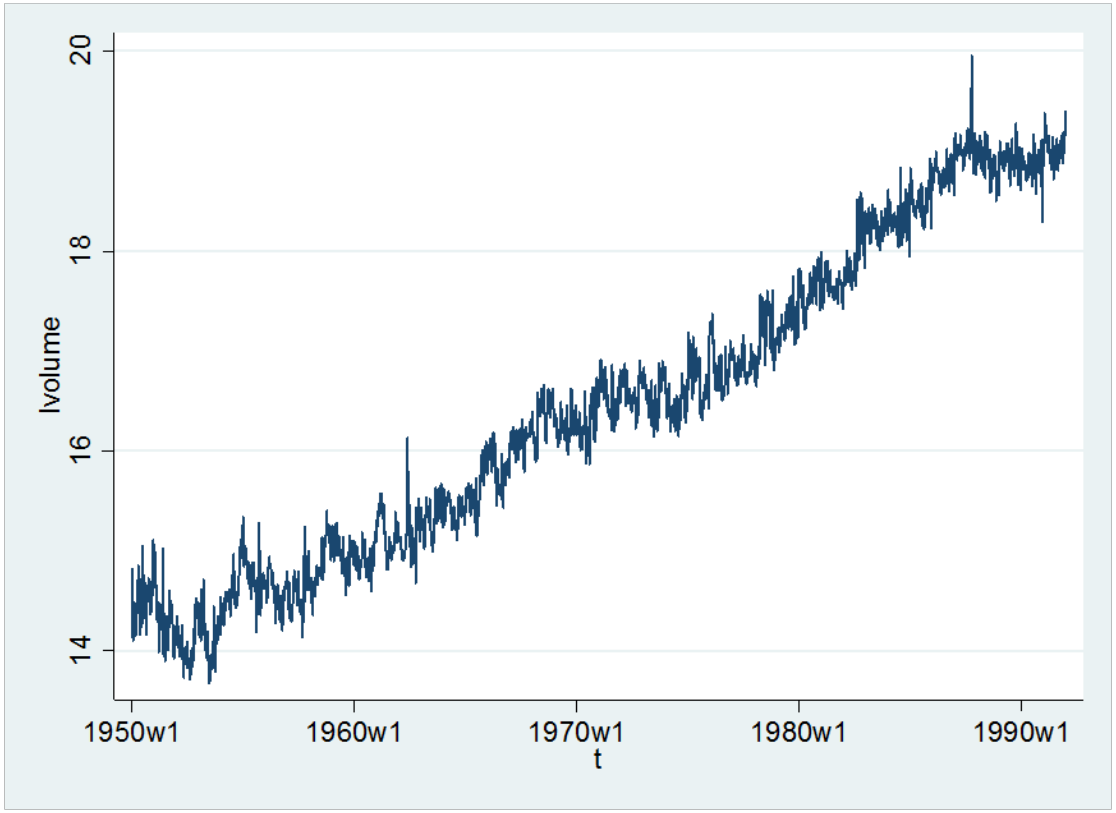
$$T_t = e^{\beta_0 + \beta_1 Time_t}$$

- The exponential trend is linear after taking (natural) logarithms

$$\ln(T_t) = \beta_0 + \beta_1 Time_t$$

- This is typically estimated by a linear model after taking logs of the variable to forecast

Ln(Volume)



- In logarithms, trend is roughly linear.

Exponential Trends

- Most economic series which are growing (aggregate output, such as GDP, investment, consumption) are exponentially increasing
  - Percentage changes are stable in the long run
- These series cannot be fit by a linear trend
- We can fit a linear trend to their (natural) logarithm

Linear Models

- The linear and quadratic trends are both linear regression models of the form

$$\mu_t = \beta_0 + \beta_1 x_{1t}$$

or

$$\mu_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t}$$

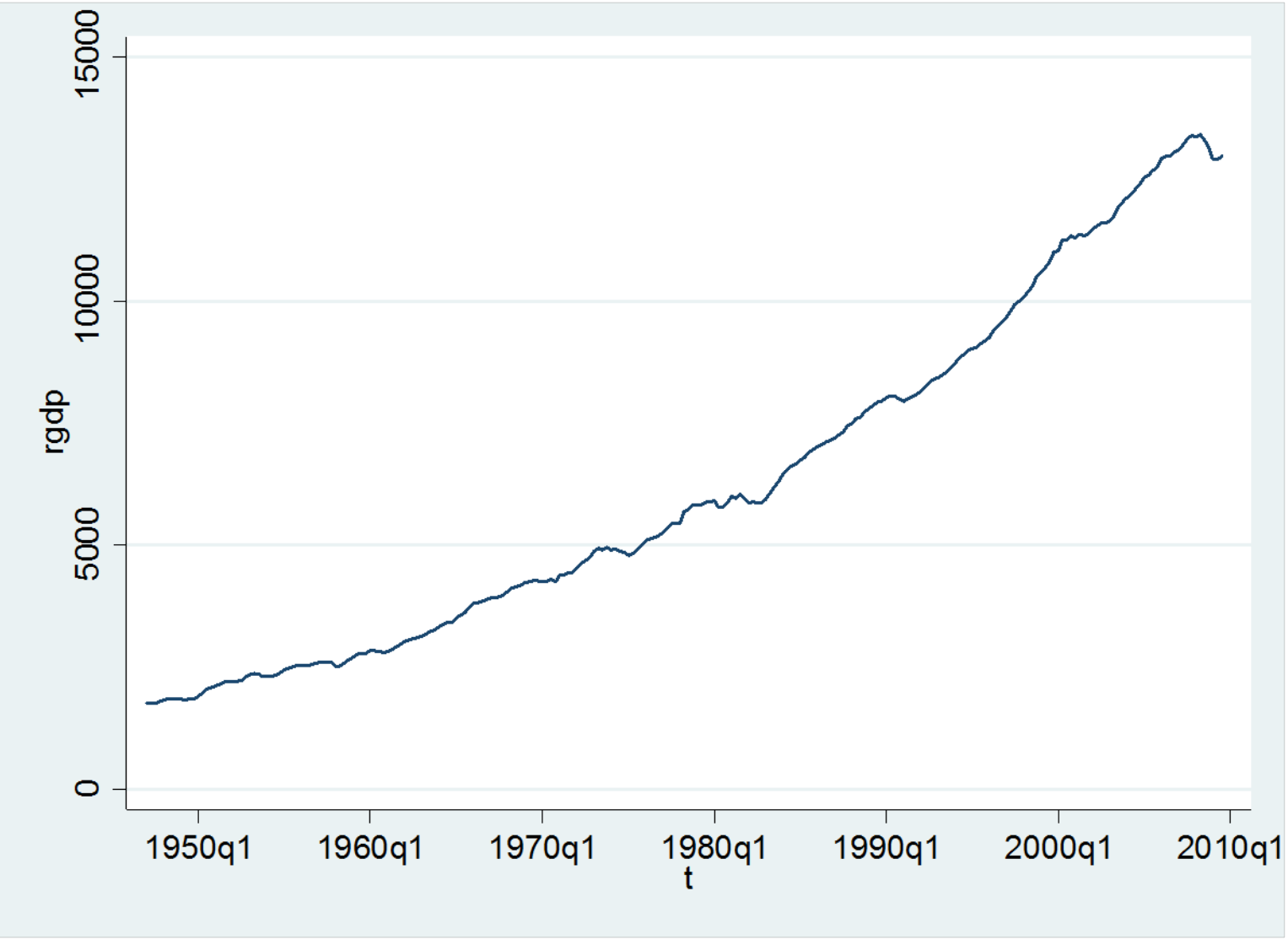
where

  - $x_{1t} = Time_t$
  - $x_{2t} = Time_t^2$

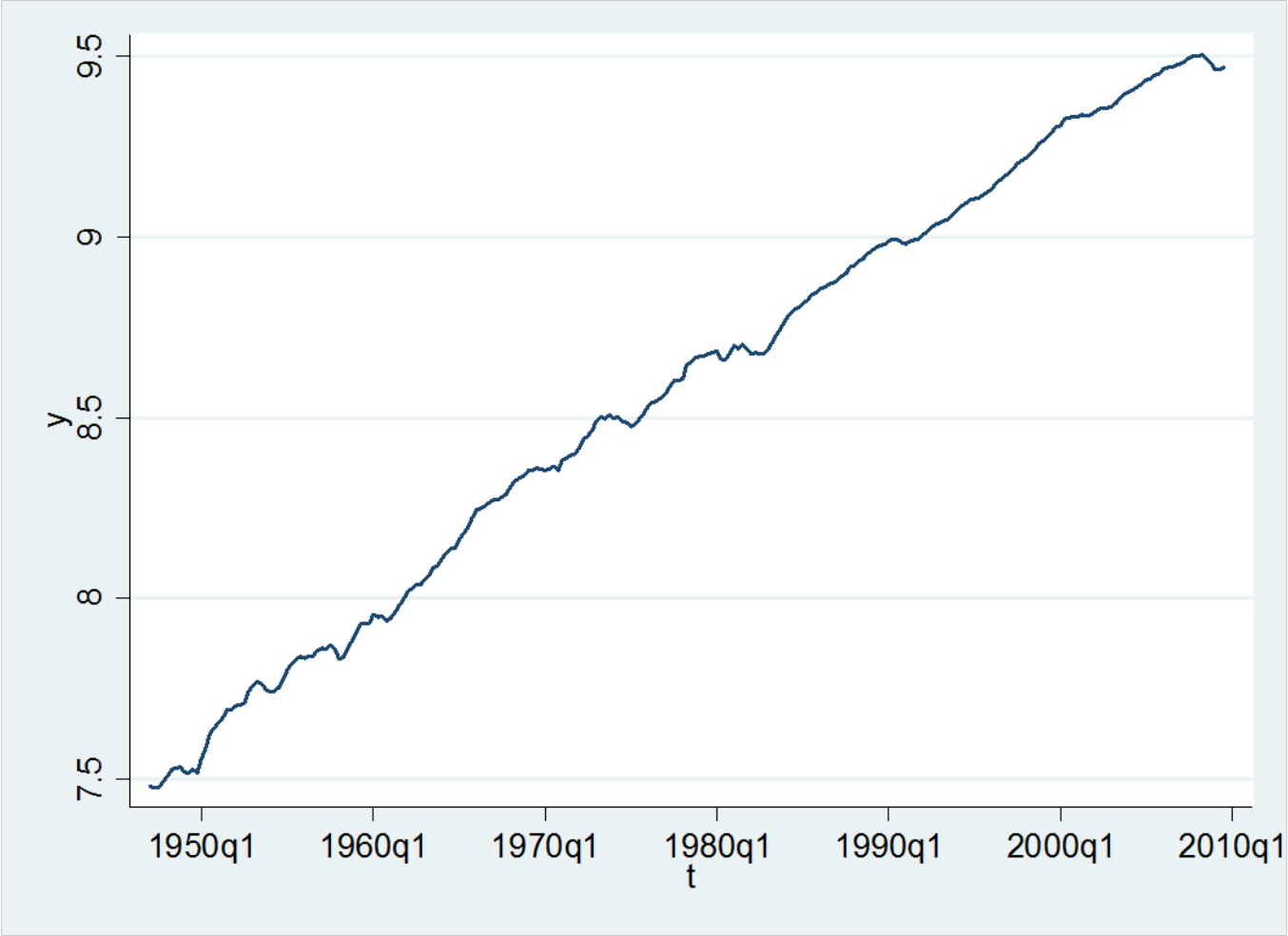
Example 4
Real GDP

- From BEA
- Quarterly, 1947-2009
- We will estimate on 1947-1990, forecast 1991-2009
- Also use an exponential trend

Real GDP


15000
10000
5000
0
rgdp
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
t

Ln(Real GDP)


y
7.5
8
8.5
9
9.5
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
t

Linear Forecasting

- The goal is to forecast future observations given a linear function of observables
- In the case of trend estimation, these observables are functions of the time index
- In other cases, they will be other functions of the data
- In the model $\mu_t = \beta_0 + \beta_1 x_t$

  the forecast for y_{t+h} is $\hat{y}_{t+h}=b_0+b_1x_t$ where b_0 and b_1 are estimates

Estimation

- How should we select b_0 and b_1 ?
- The goal is to produce a forecast with low mean square error (MSE)
- The best linear forecast is the linear function $\beta_0+\beta_1 x_t$ that minimizes the MSE

$$E\left(y_{t+h}-\hat{y}_{t+h}\right)^2 = E\left(y_{t+h}-\beta_0-\beta_1 x_t\right)^2$$

- We do not know the MSE, but we can estimate it by a sample average

Sum of Squared Errors

- Sample estimate of mean square error is the sum of squared errors

$$S_n(\beta_0, \beta_1) = \frac{1}{n}\sum_{t=1}^{n}\left(y_{t+h} - \beta_0 - \beta_1 x_t\right)^2$$

- The best linear forecast is the linear function $\beta_0 + \beta_1 x_t$ that minimizes the MSE, or expected sum of squared errors.
- Our sample estimate of the best linear forecast is the linear function which minimizes the (sample) sum of squared errors.
- This is called the least-squares estimator

Least Squares

- The least-squares estimates (b_0,b_1) are the values which minimize the sum of squared errors

$$S_n(\beta_0, \beta_1) = \frac{1}{n}\sum_{t=1}^{n}\left(y_{t+h} - \beta_0 - \beta_1 x_t\right)^2$$

- This produces estimates of the best linear predictor – the linear function $\beta_0+\beta_1 x_t$ that minimizes the MSE

Multiple Regressors

- There are multiple regressors

$$\mu_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t}$$

- For example, the quadratic trend

$$T_t = \beta_0 + \beta_1 Time_t + \beta_2 Time_t^2$$

- The best linear predictor is the linear function $\beta_0+\beta_1 x_{1t}+\beta_2 x_{2t}$ that minimizes the MSE

$$E\left(y_{t+h} - \hat{y}_{t+h}\right)^2 = E\left(y_{t+h} - \beta_0 - \beta_1 x_{1t} - \beta_2 x_{2t}\right)^2$$

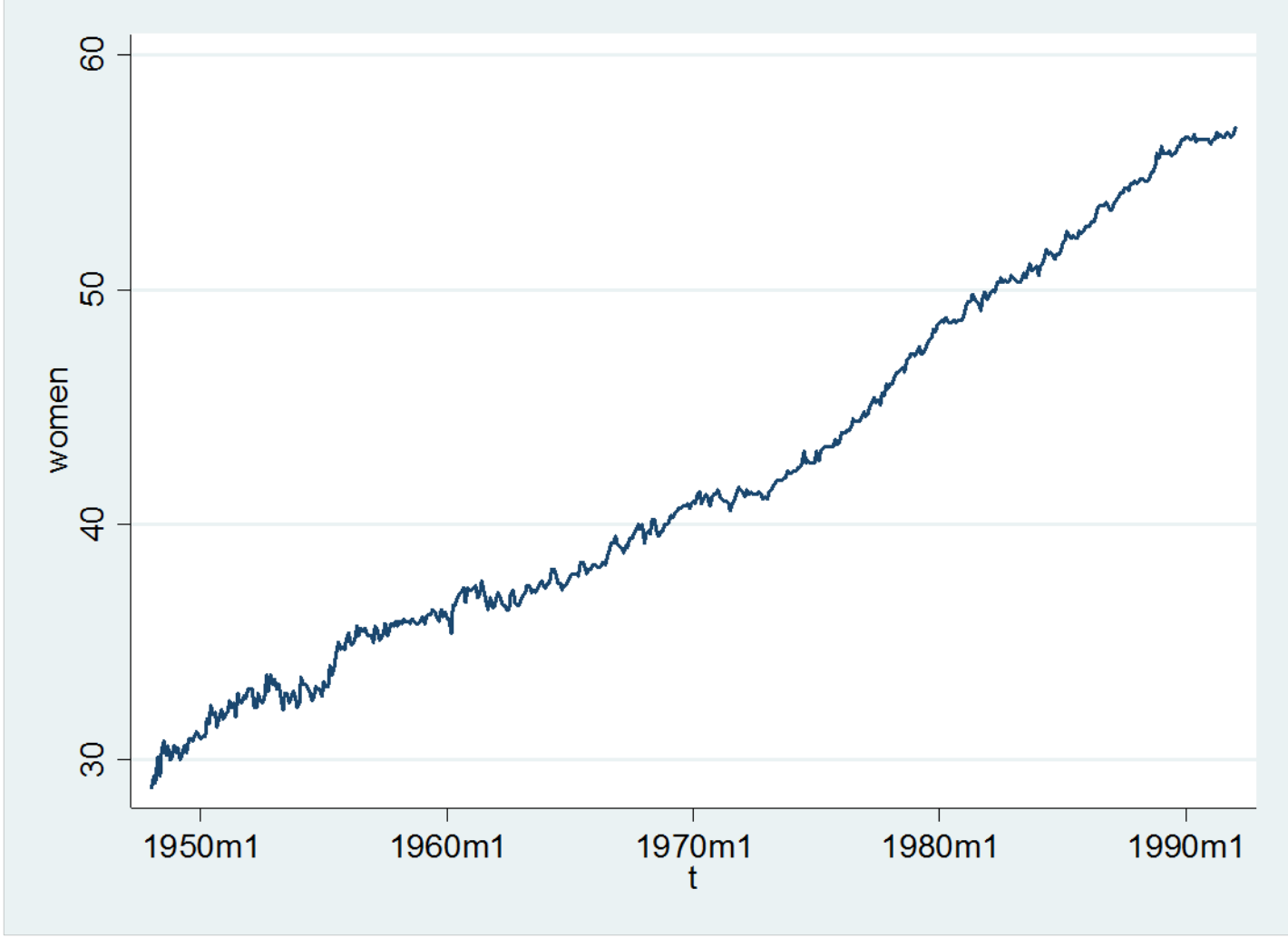
Multiple Regression

- The sample estimate of the best linear predictor are the values (b_0,b_1,b_2) which minimize the sum of squared errors

$$S_n(\beta_0, \beta_1, \beta_2) = \frac{1}{n}\sum_{t=1}^{n}\left(y_{t+h} - \beta_0 - \beta_1 x_{1t} - \beta_2 x_{2t}\right)^2$$

- In STATA, use the **regress** command

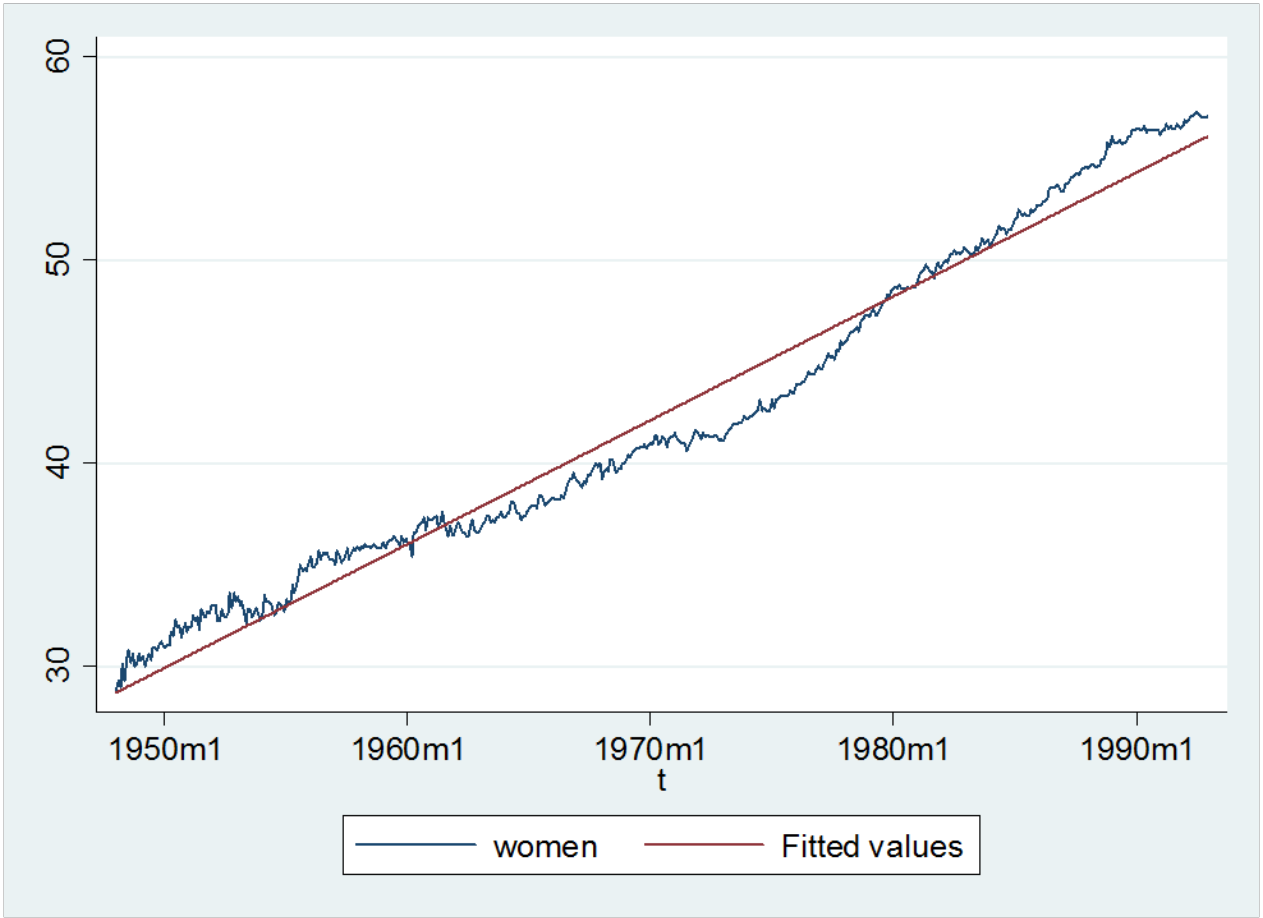
Example 1
Women's Labor Force Participation Rate



Regression Estimation

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

. regress women t if t<=tm(1992m12)

      Source |       SS       df       MS              Number of obs =     540
-------------+------------------------------           F(  1,   538) =19575.55
       Model |  33879.9082     1  33879.9082           Prob > F      =  0.0000
    Residual |  931.130559   538  1.73072595           R-squared     =  0.9733
-------------+------------------------------           Adj R-squared =  0.9732
       Total |  34811.0387   539  64.5844874           Root MSE      =  1.3156

------------------------------------------------------------------------------
       women |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0508126   .0003632   139.91   0.000     .0500992    .0515261
       _cons |   36.02153   .0726804   495.62   0.000     35.87876     36.1643
------------------------------------------------------------------------------
```

In-Sample Fit



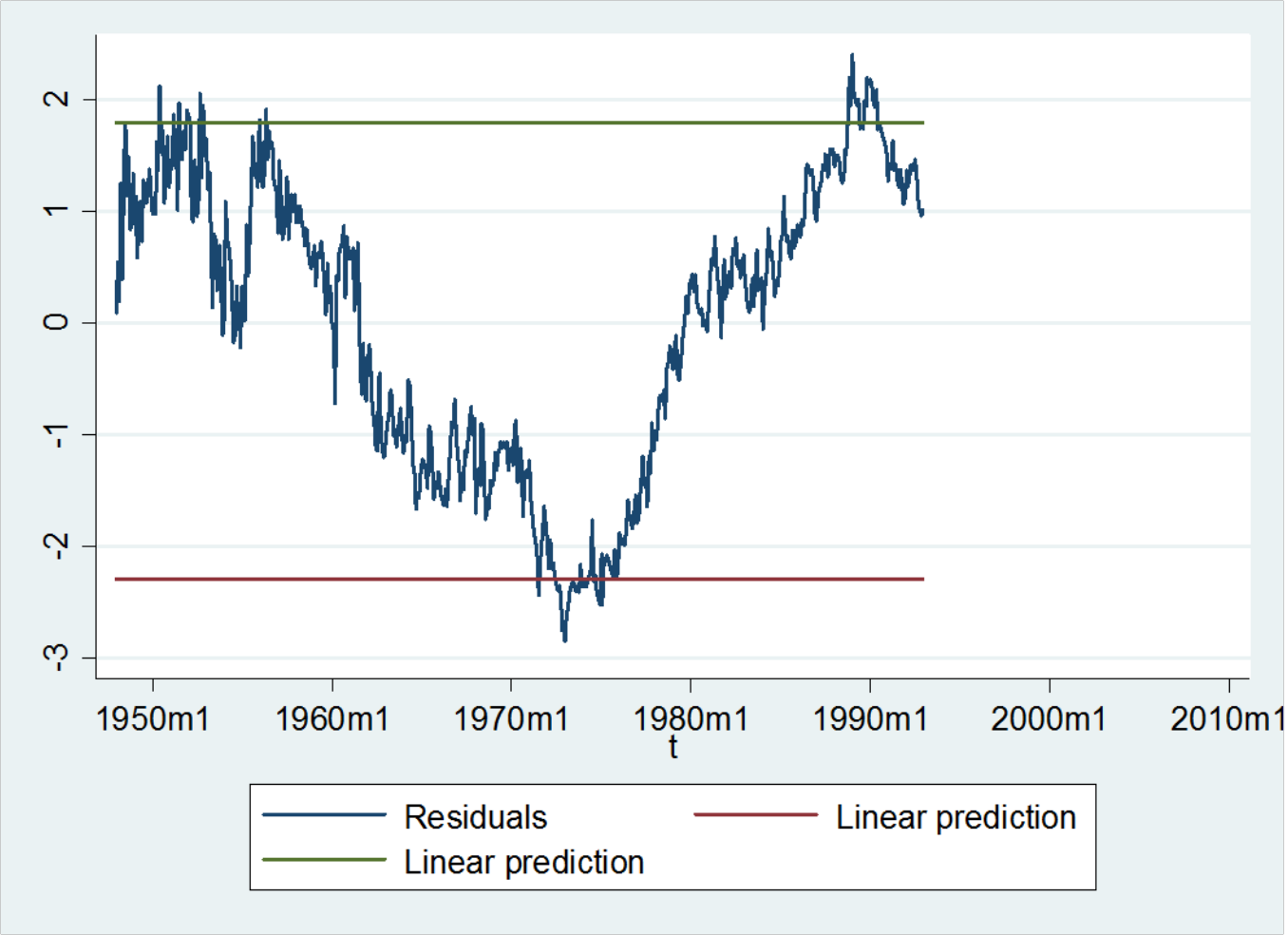
Residuals

- Residuals are difference between data and fitted regression line

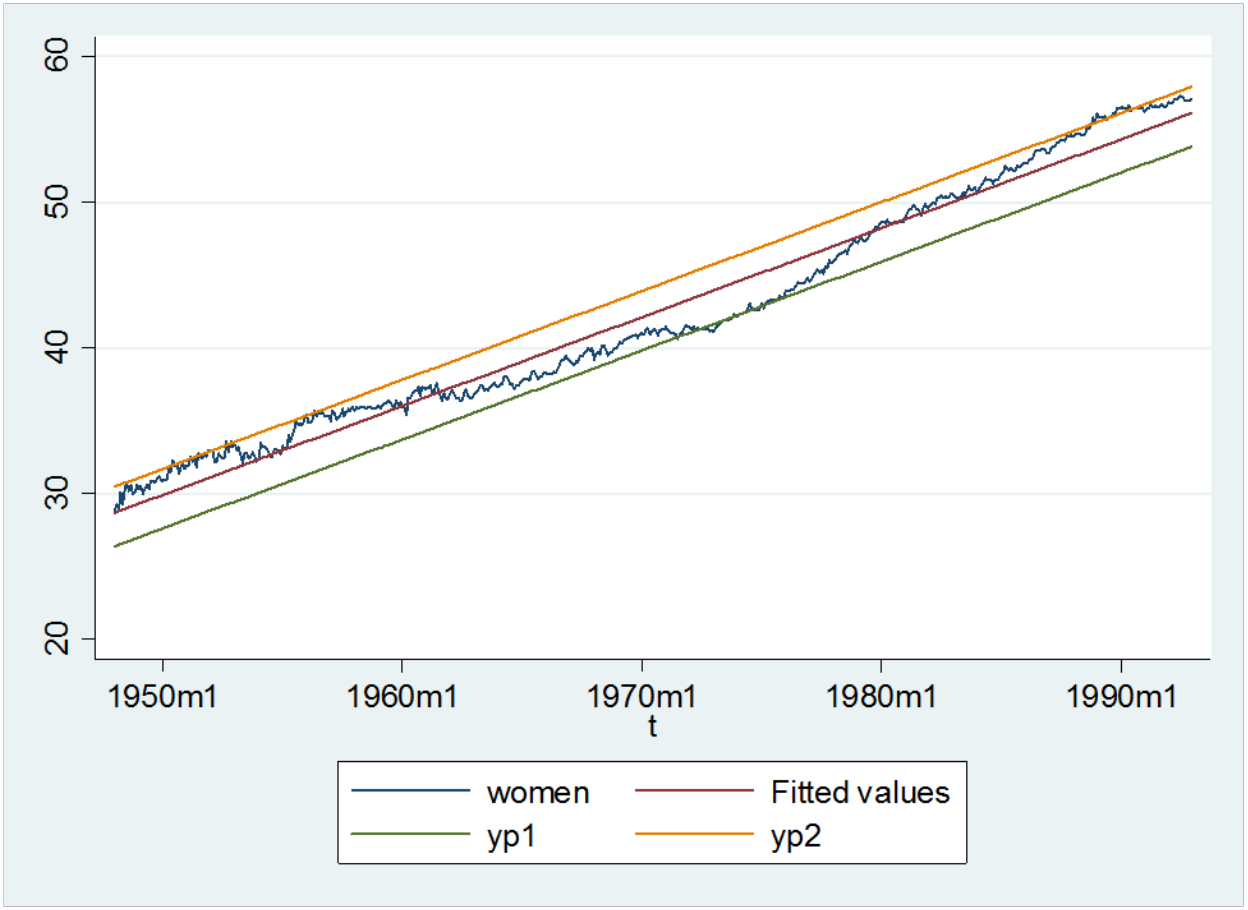
$$\hat{e}_t = y_{t+h} - T_t$$
$$= y_{t+h} - b_0 - b_1 Time_t$$

```
. predict e if t<=tm(1992m12), residuals
(204 missing values generated)
```

Residual Plot



In-Sample Fit



- Compute with **predict** command
- Fit looks good

Forecast

- Forecast is the linear function with estimated coefficients

$$T_{T+h} = b_0 + b_1 Time_{T+h}$$

- Compute with **predict** command

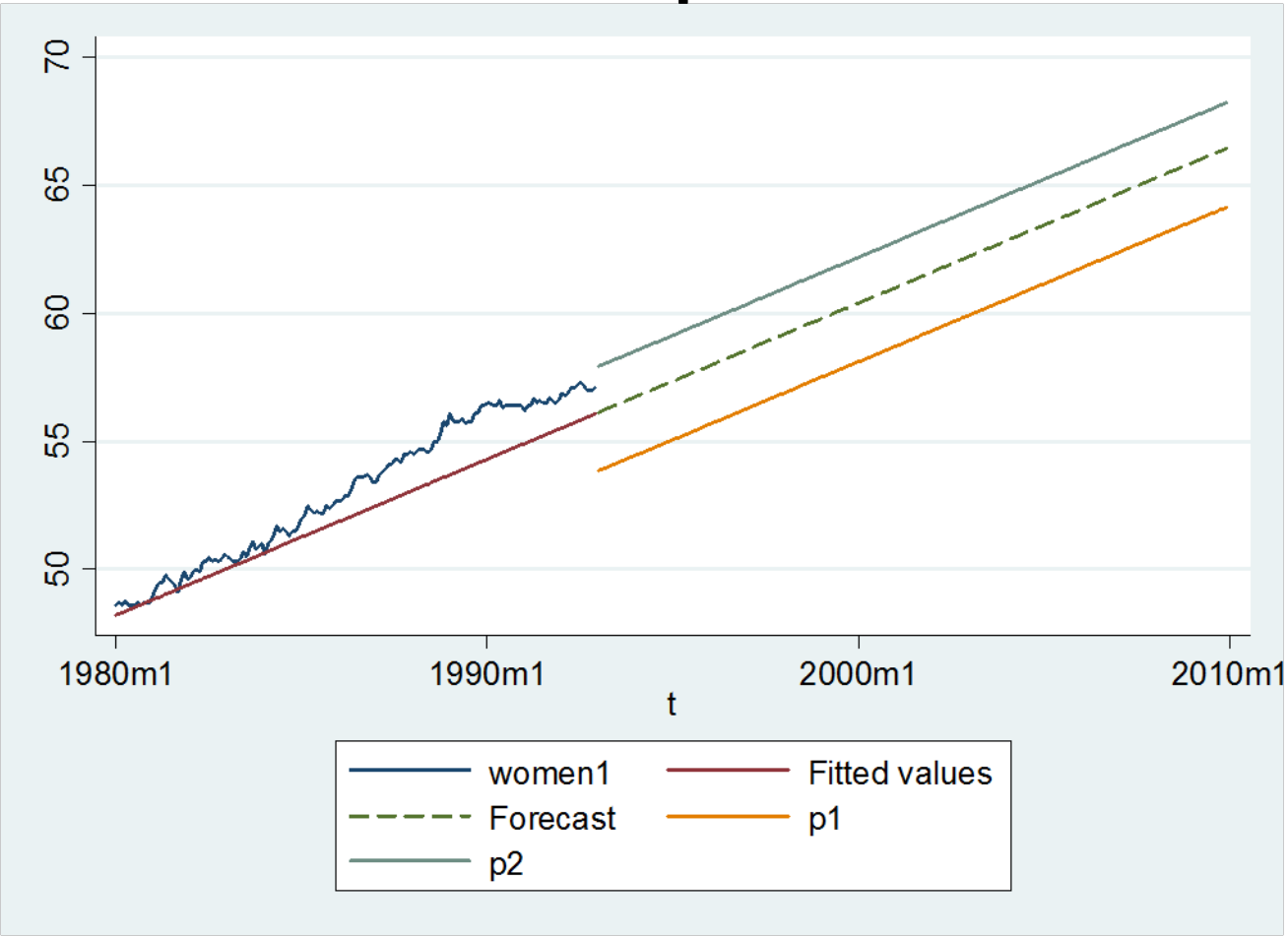
Forecast Intervals

- Compute residuals

$$\begin{aligned}\hat{e}_t &= y_{t+h} - \hat{\mu}_t \\ &= y_{t+h} - b_0 - b_1 Time_t\end{aligned}$$

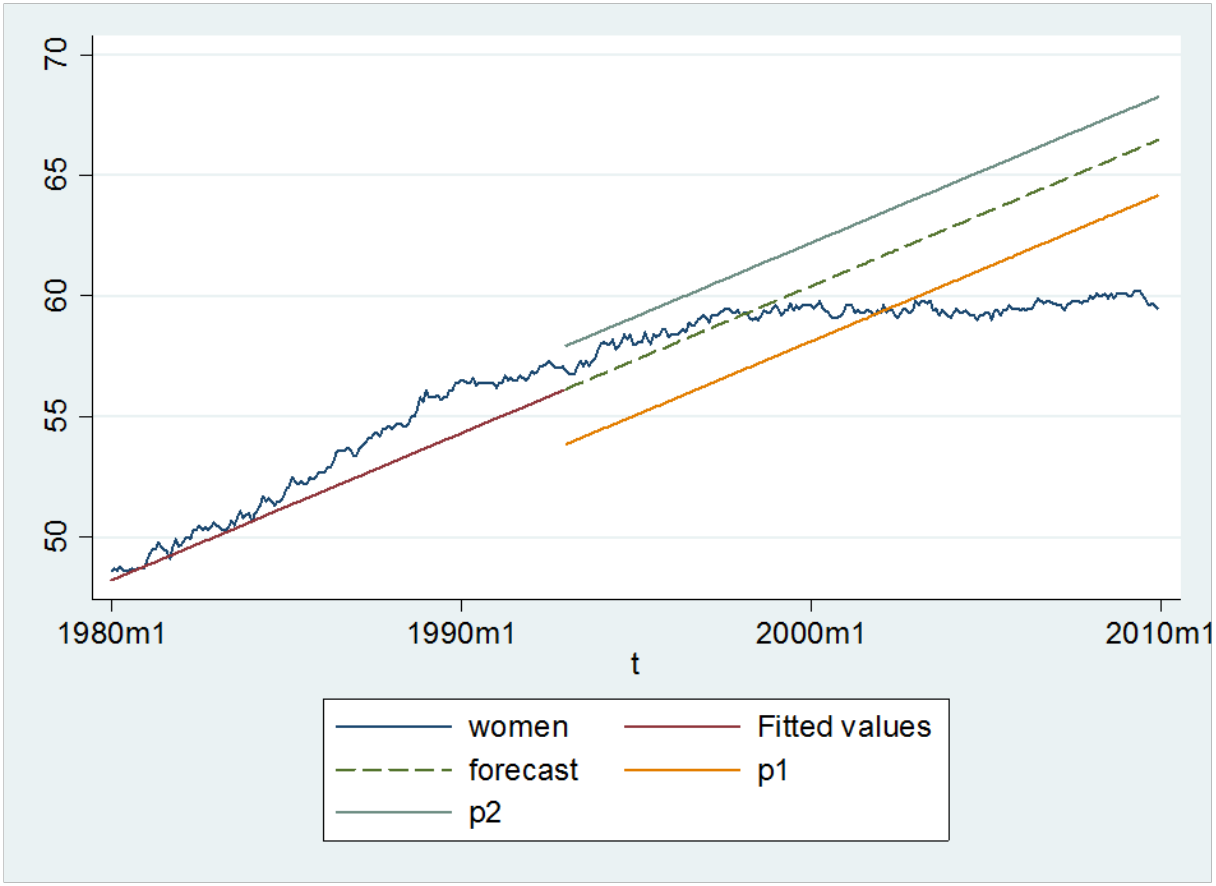
- Compute quantiles of residuals
  - These are constant over time
- Add to predicted values
  - Identical to constant mean case

Out-of-Sample Forecast



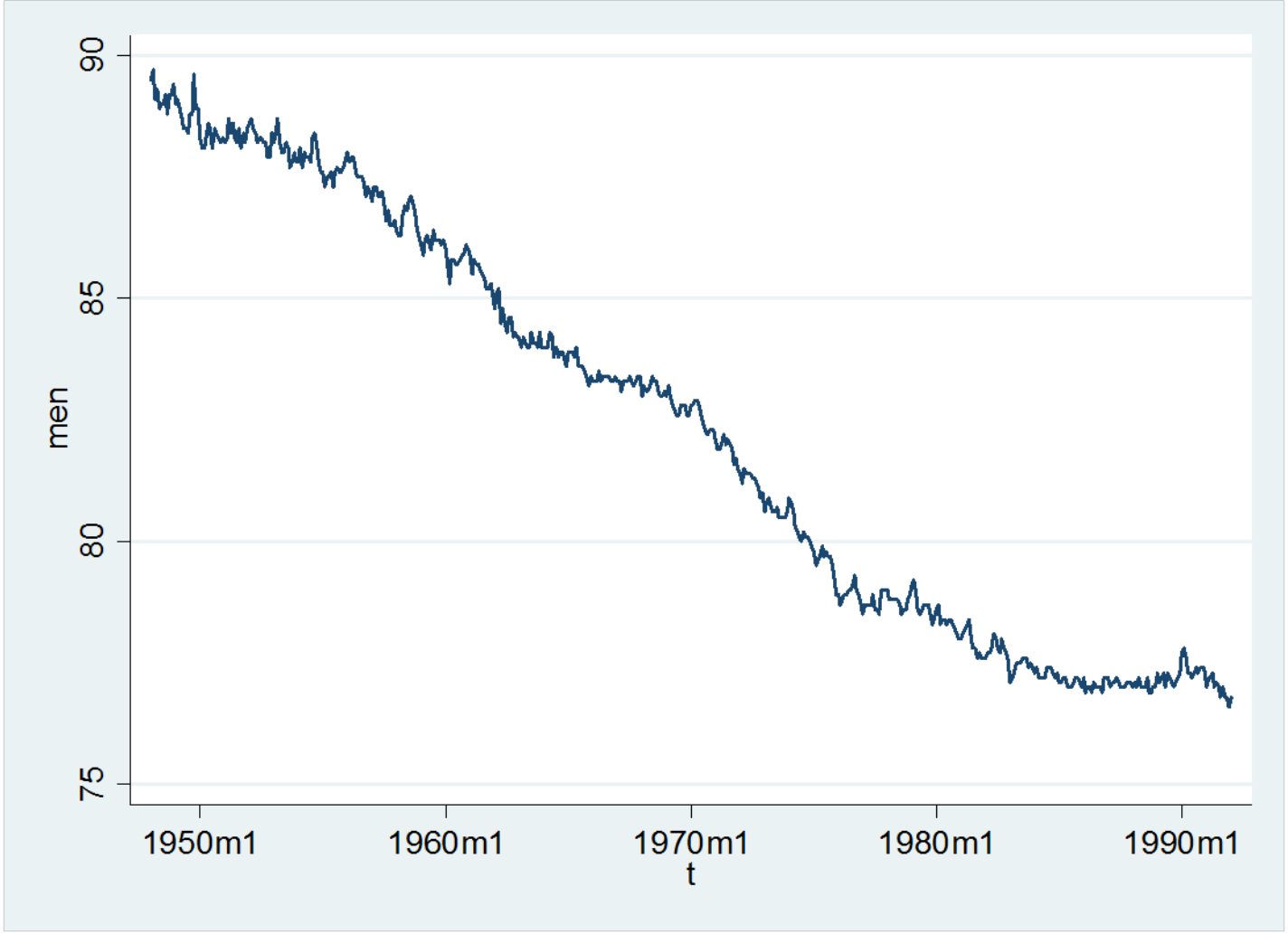
- Out of sample prediction might be too low.

Out-of-Sample
Women's Labor Force Participation



- No: Prediction was way too high!

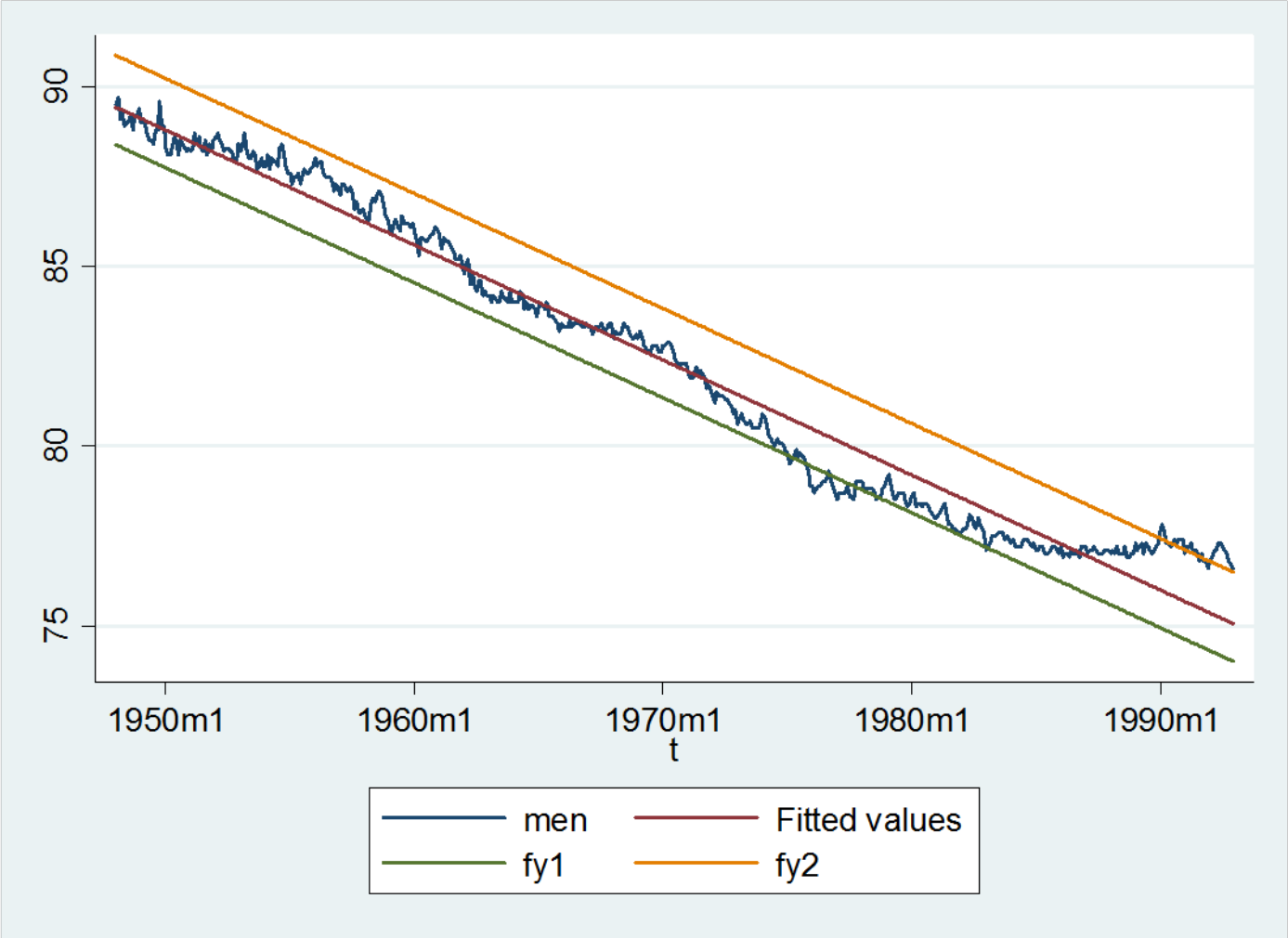
Men's Labor Force Participation Rate



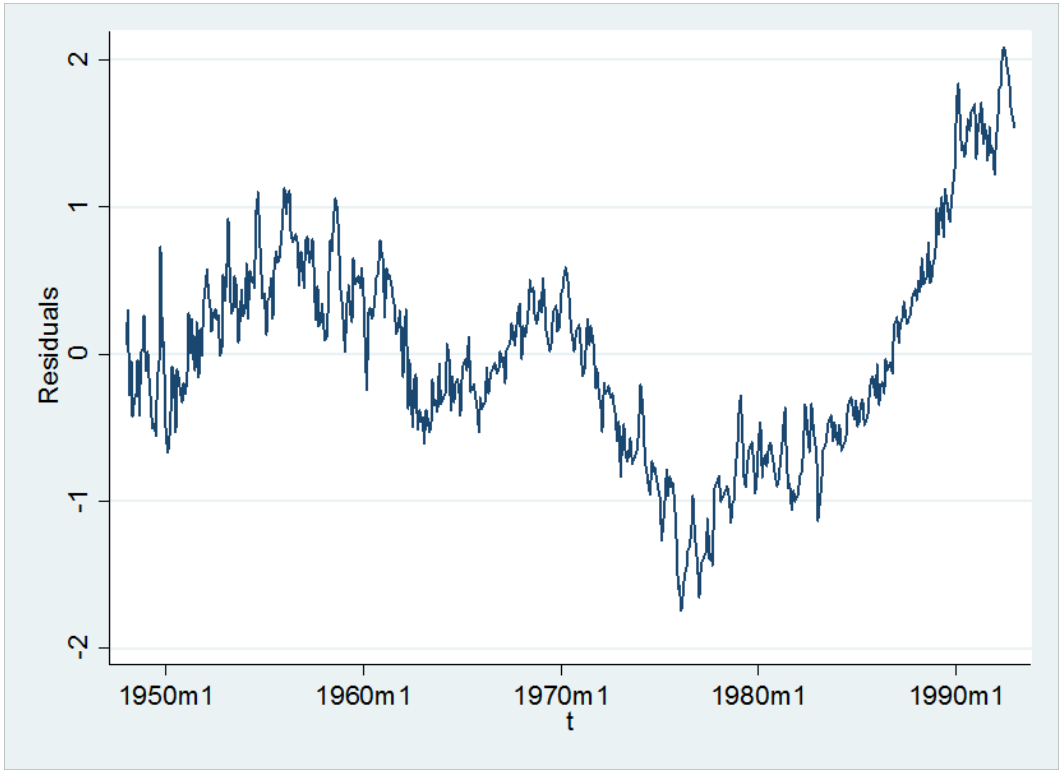
Estimation

```
. regress men t if t<=tm(1992m12)

      Source |       SS       df       MS              Number of obs =     540
-------------+------------------------------           F(  1,   538) =17920.70
       Model |  9332.01915     1  9332.01915           Prob > F      =  0.0000
    Residual |  280.157915   538  .520739619           R-squared     =  0.9709
-------------+------------------------------           Adj R-squared =  0.9708
       Total |  9612.17706   539  17.8333526           Root MSE      =  .72162

------------------------------------------------------------------------------
         men |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0266679   .0001992  -133.87   0.000    -.0270592   -.0262766
       _cons |   85.59182    .039867  2146.94   0.000     85.51351    85.67013
------------------------------------------------------------------------------
```

In-Sample Fit



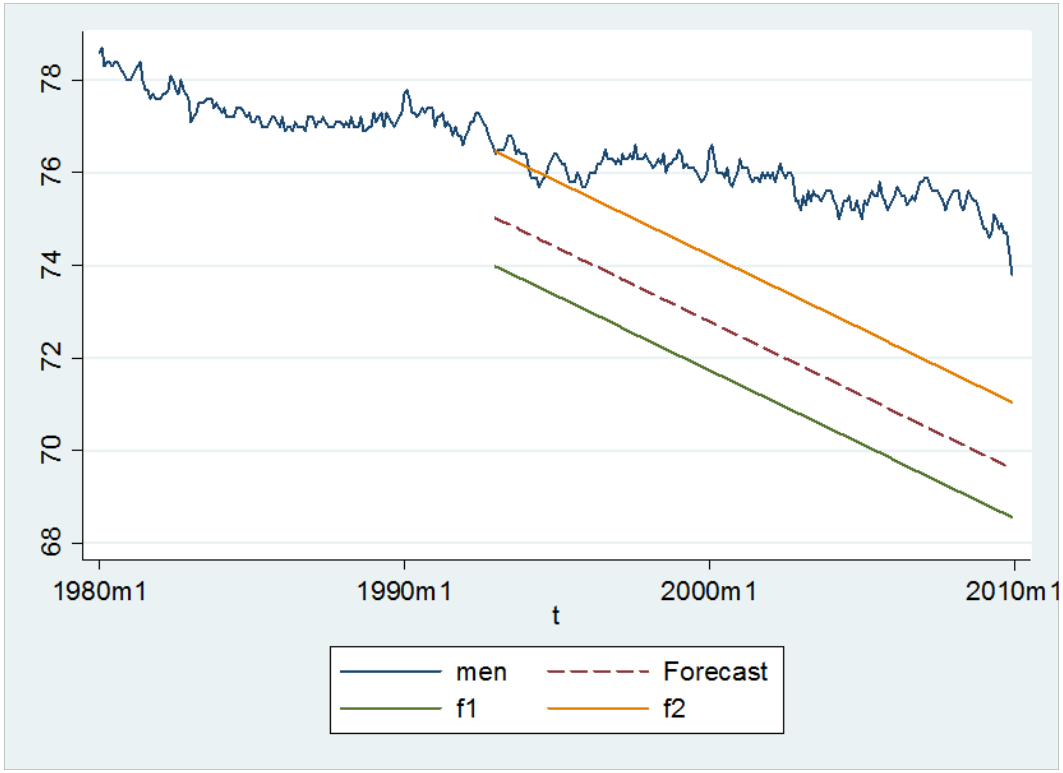
Residuals



Forecast

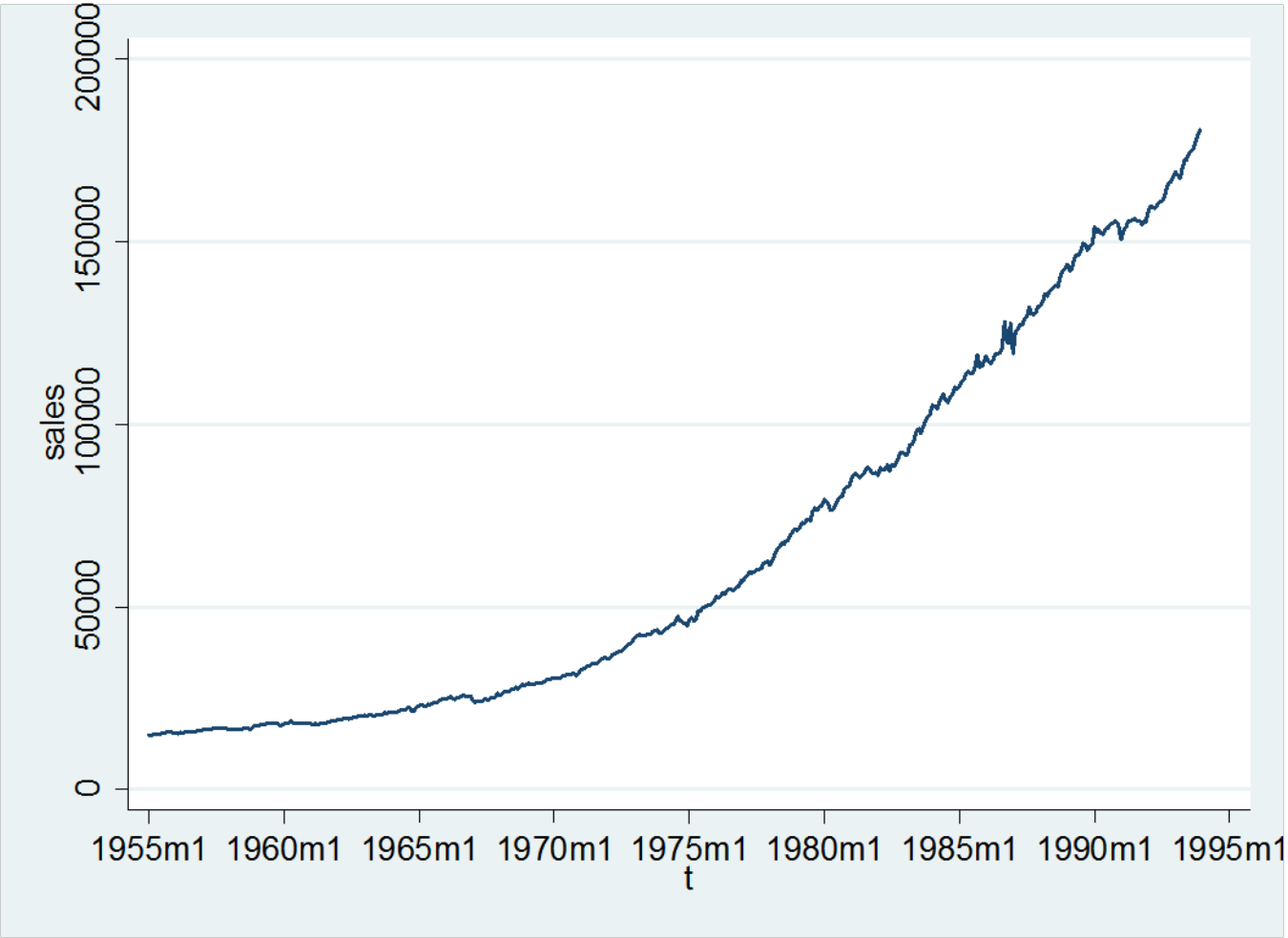
- End of Sample looks worrying

Out-of-Sample
Men's Labor Force Participation



- Linear Trend Terrible

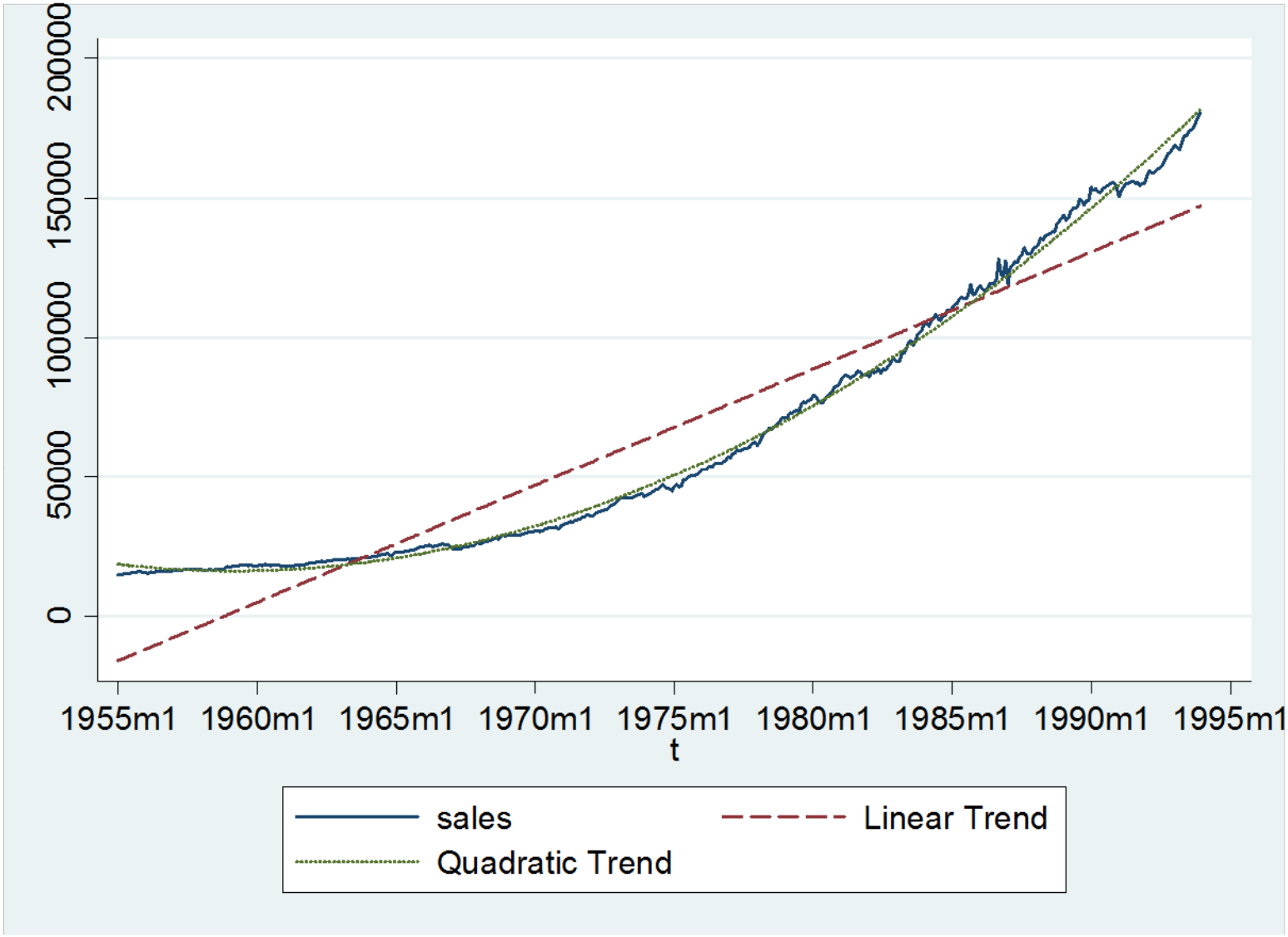
Example 2
Retail Sales



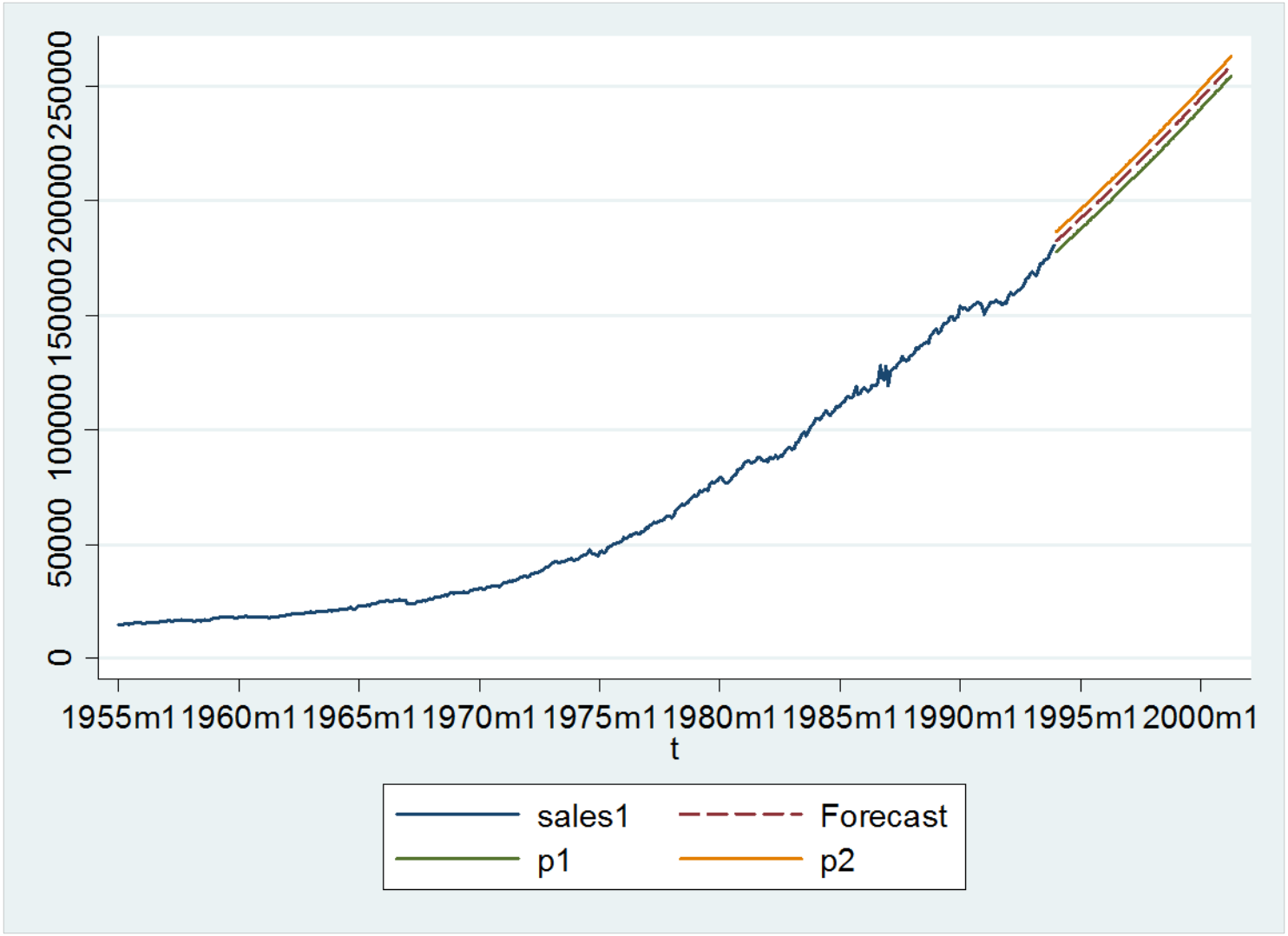
Linear and Quadratic Trend

```
. regress sales t if t<=tm(1993m12)

      Source |       SS       df       MS              Number of obs =     468
-------------+------------------------------           F(  1,   466) = 4167.23
       Model |  1.0423e+12     1  1.0423e+12           Prob > F      =  0.0000
    Residual |  1.1655e+11   466   250107680           R-squared     =  0.8994
-------------+------------------------------           Adj R-squared =  0.8992
       Total |  1.1588e+12   467  2.4814e+09           Root MSE      =   15815

------------------------------------------------------------------------------
       sales |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   349.3091    5.41111    64.55   0.000     338.6759    359.9423
       _cons |   4983.413    1189.88     4.19   0.000     2645.218    7321.609
------------------------------------------------------------------------------

. generate t2=t^2

. regress sales t t2 if t<=tm(1993m12)

      Source |       SS       df       MS              Number of obs =     468
-------------+------------------------------           F(  2,   465) =75689.11
       Model |  1.1553e+12     2  5.7763e+11           Prob > F      =  0.0000
    Residual |  3.5487e+09   465  7631593.27           R-squared     =  0.9969
-------------+------------------------------           Adj R-squared =  0.9969
       Total |  1.1588e+12   467  2.4814e+09           Root MSE      =  2762.5

------------------------------------------------------------------------------
       sales |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   19.01707   2.874205     6.62   0.000     13.36903    24.66511
          t2 |   .9518503   .0078223   121.68   0.000     .9364789    .9672217
       _cons |   16263.16   227.5825    71.46   0.000     15815.94    16710.37
------------------------------------------------------------------------------
```

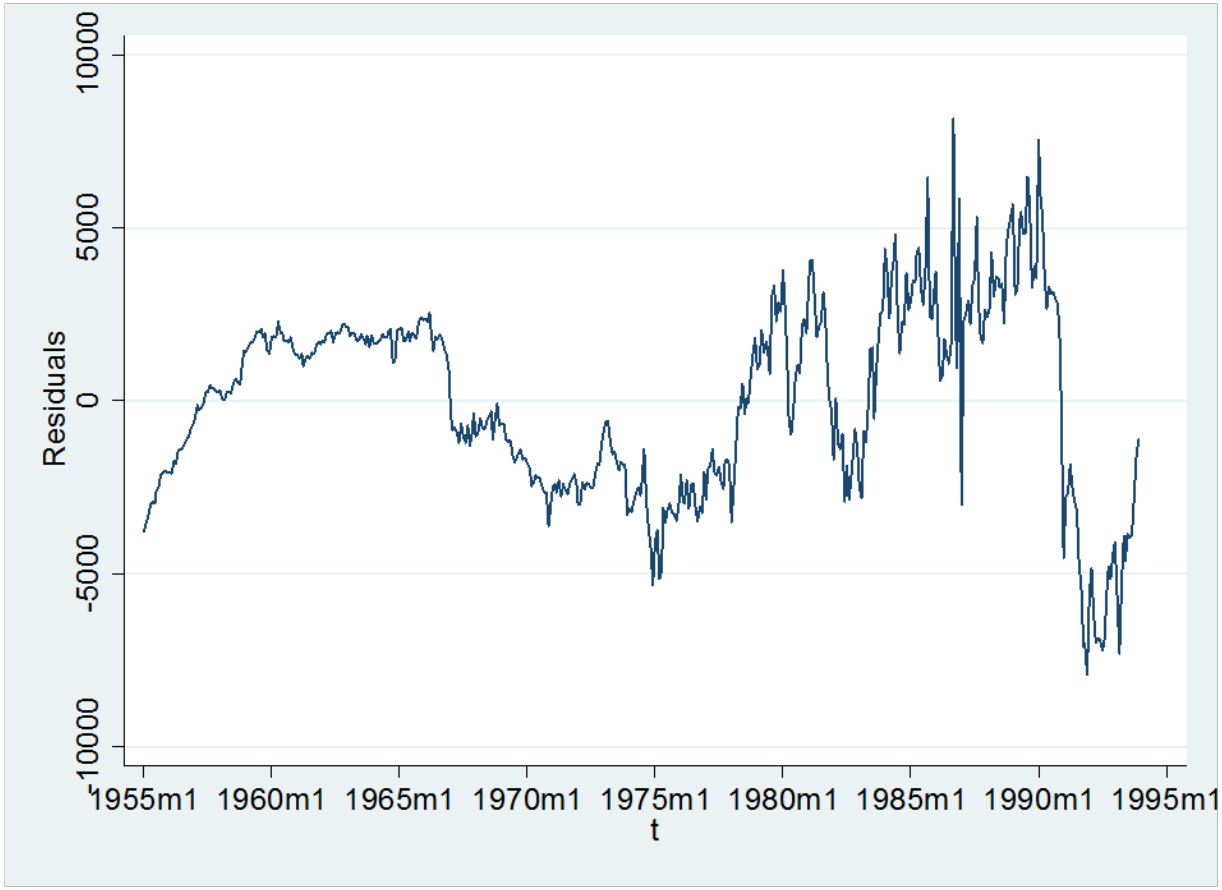
Linear and Quadratic Trend



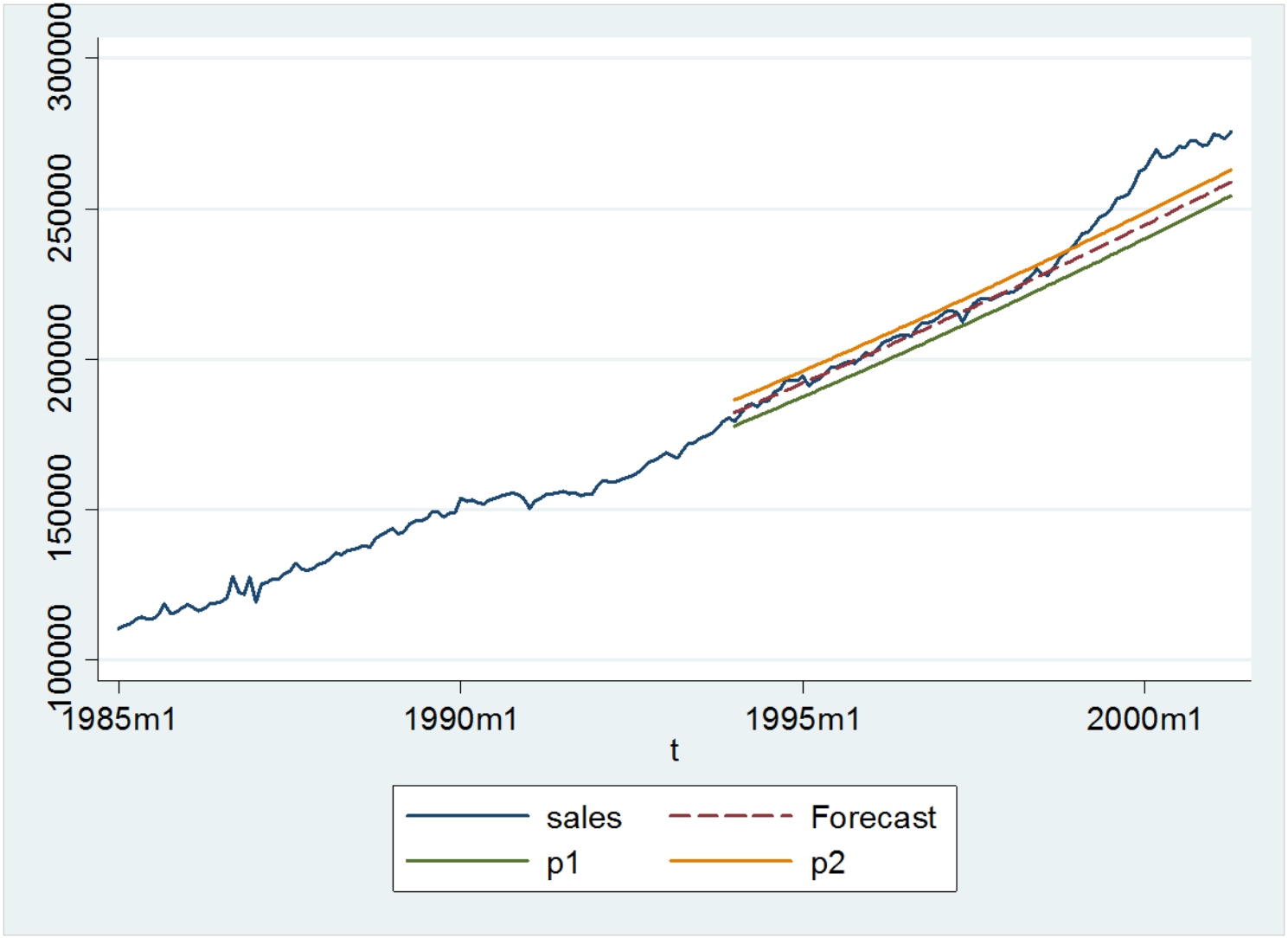
Forecast


0
50000
100000
150000
200000
250000
1955m1
1960m1
1965m1
1970m1
1975m1
1980m1
1985m1
1990m1
1995m1
2000m1
t
sales1
Forecast
p1
p2

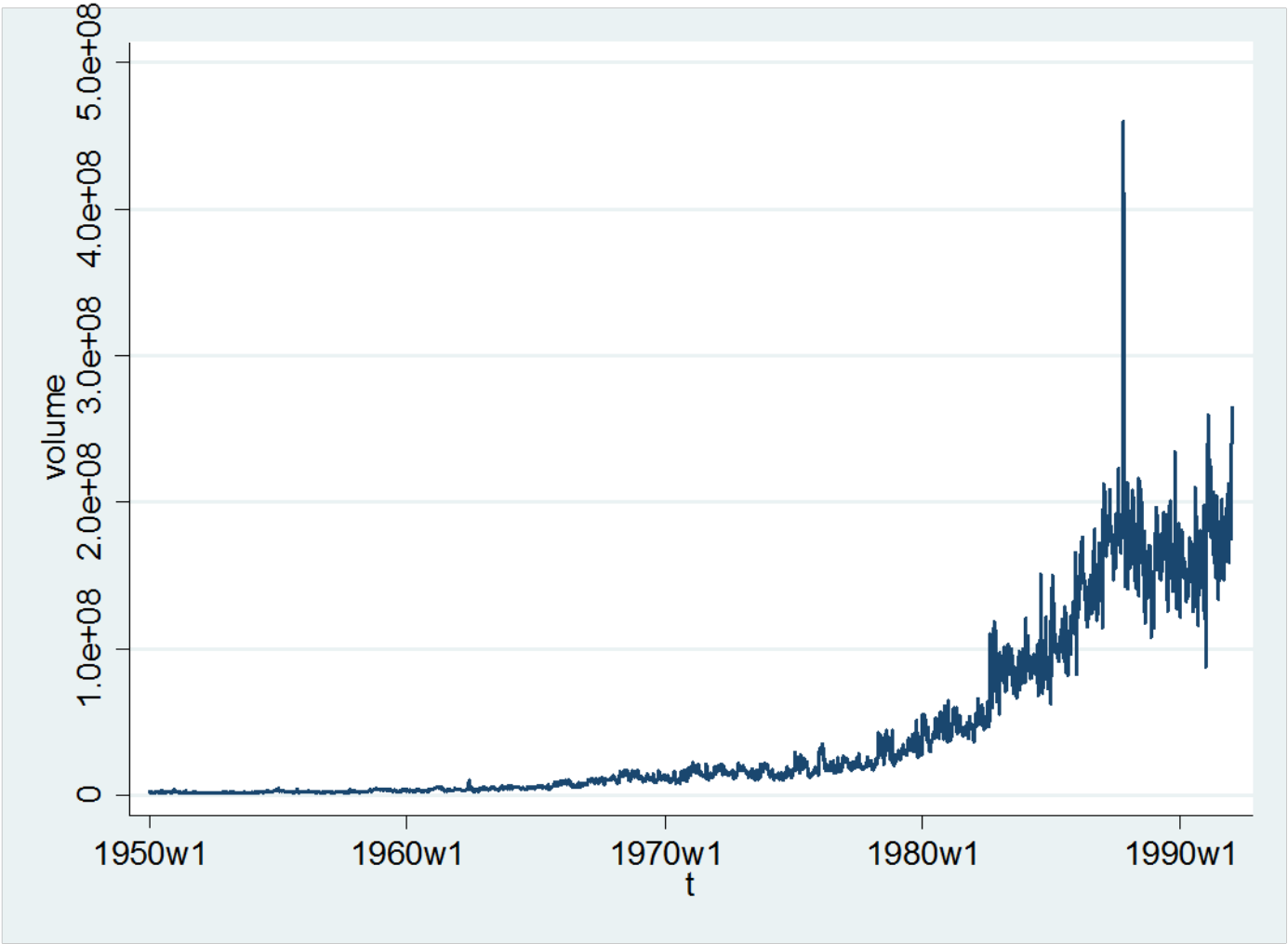
Residuals


Residuals
10000
5000
0
-5000
-10000
1955m1
1960m1
1965m1
1970m1
1975m1
1980m1
1985m1
1990m1
1995m1
t

Actual Values



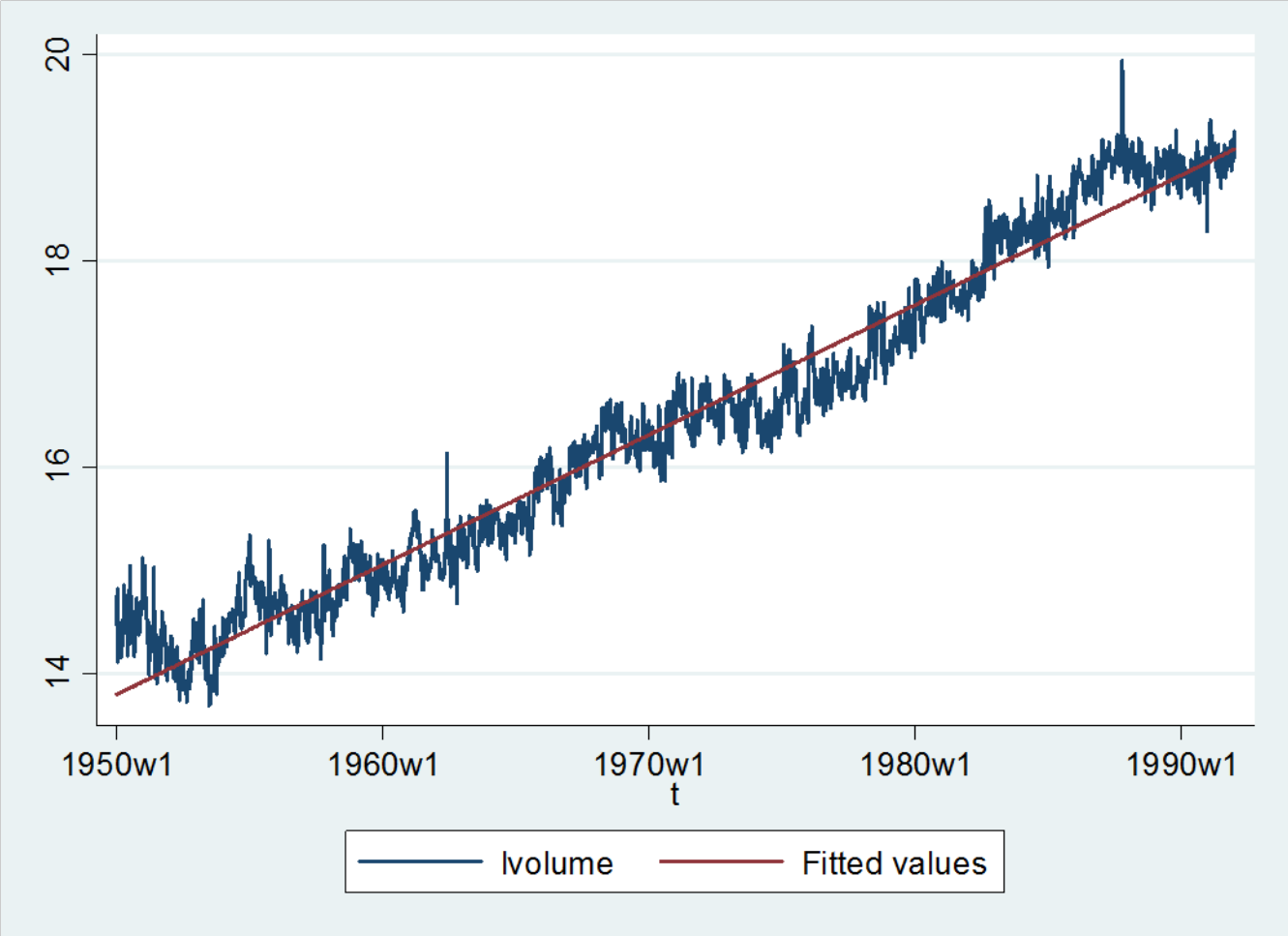
Example 3: Volume



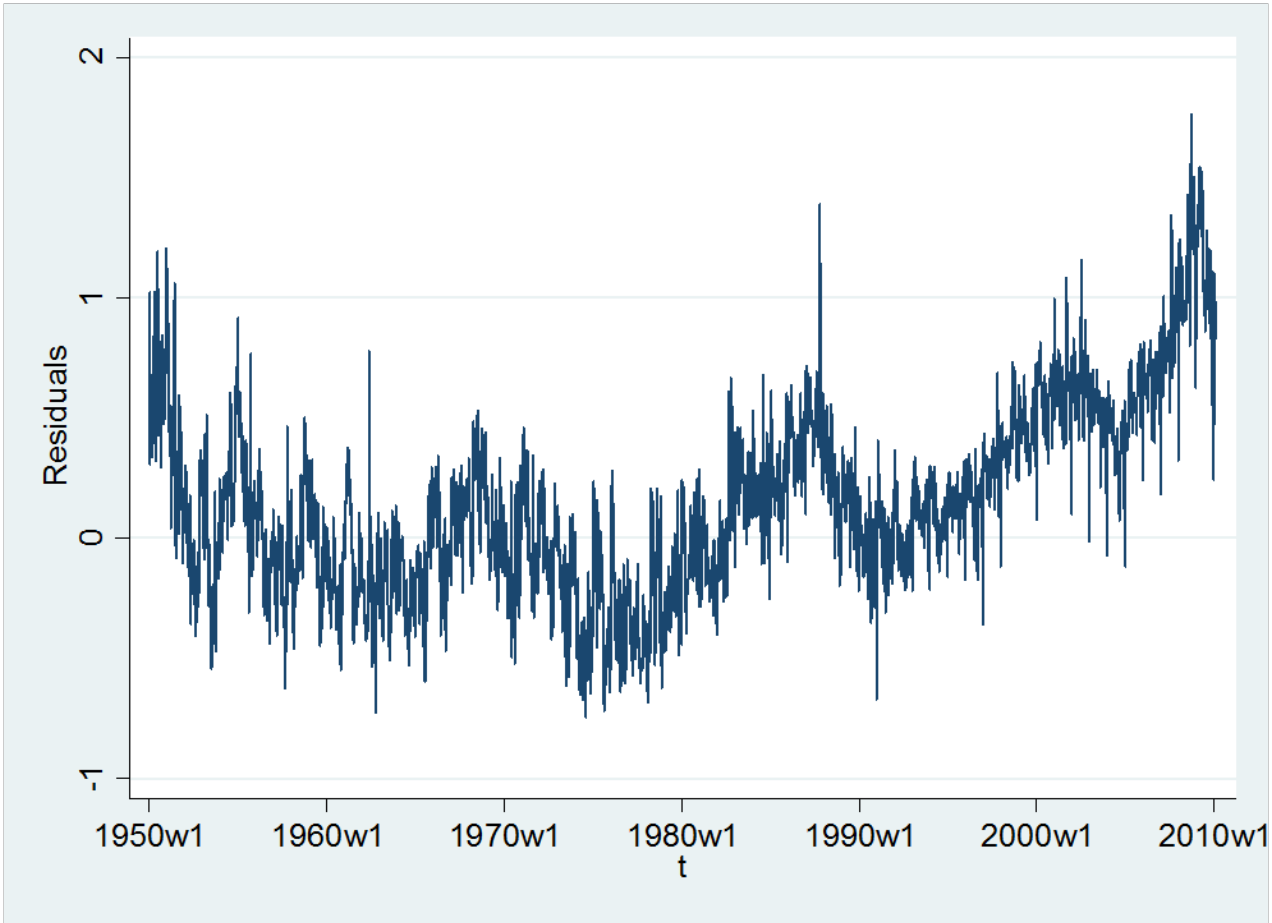
Estimating Logarithmic Trend

```
. use "C:\Users\Bruce Hansen\Documents\docs\classdocs\390\s&p.dta"

. generate lvolume=ln(volume)

. regress lvolume t if t<tw(1992w1)

      Source |       SS       df       MS              Number of obs =    2192
-------------+------------------------------           F(  1,  2190) =56150.82
       Model |  5122.49788     1  5122.49788           Prob > F      =  0.0000
    Residual |   199.78817  2190  .091227475           R-squared     =  0.9625
-------------+------------------------------           Adj R-squared =  0.9624
       Total |  5322.28605  2191   2.4291584           Root MSE      =  .30204

------------------------------------------------------------------------------
     lvolume |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0024242   .0000102   236.96   0.000     .0024041    .0024442
       _cons |   15.05281   .0087069  1728.83   0.000     15.03574    15.06989
------------------------------------------------------------------------------
```

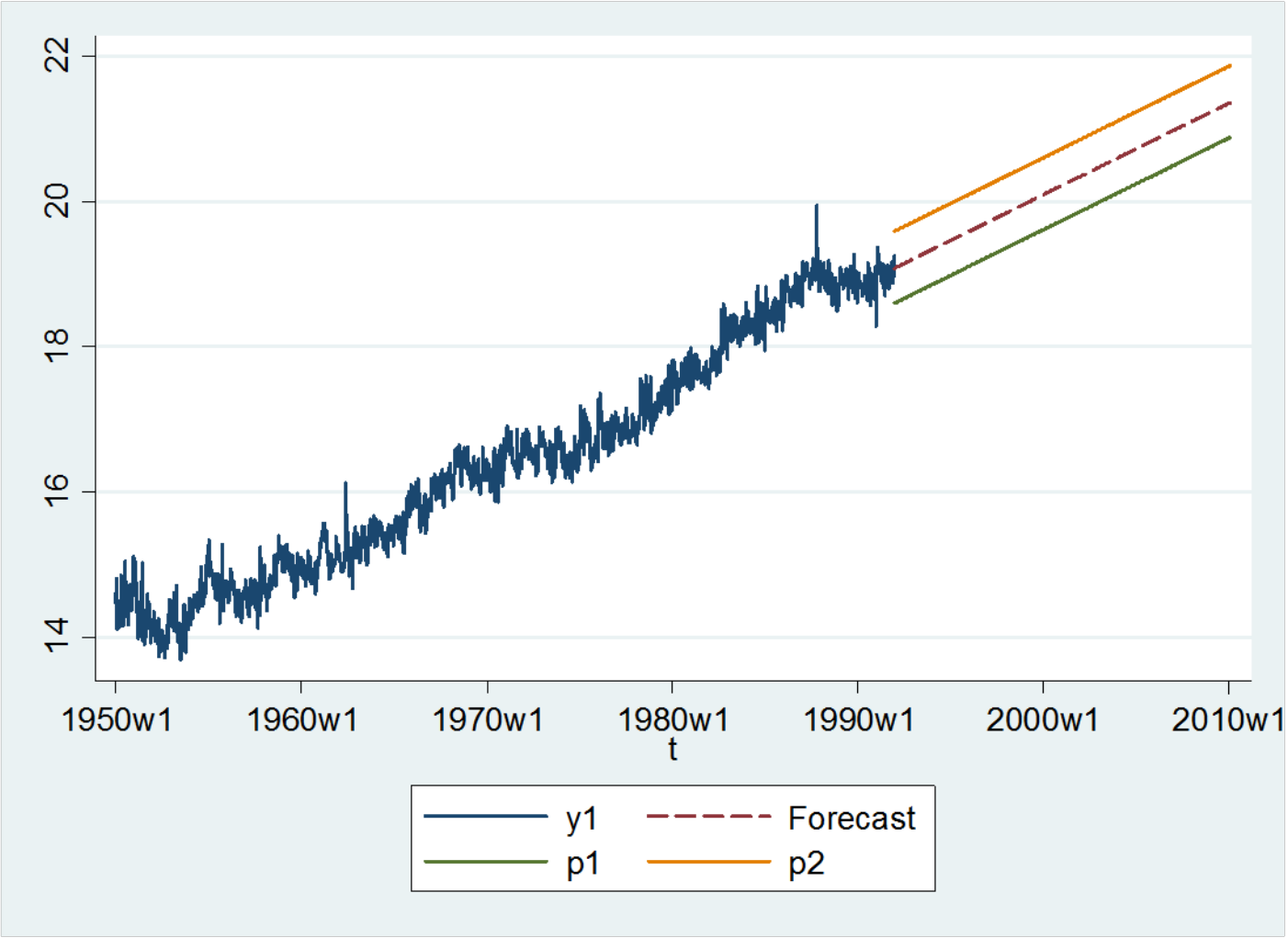
Fitted Trend



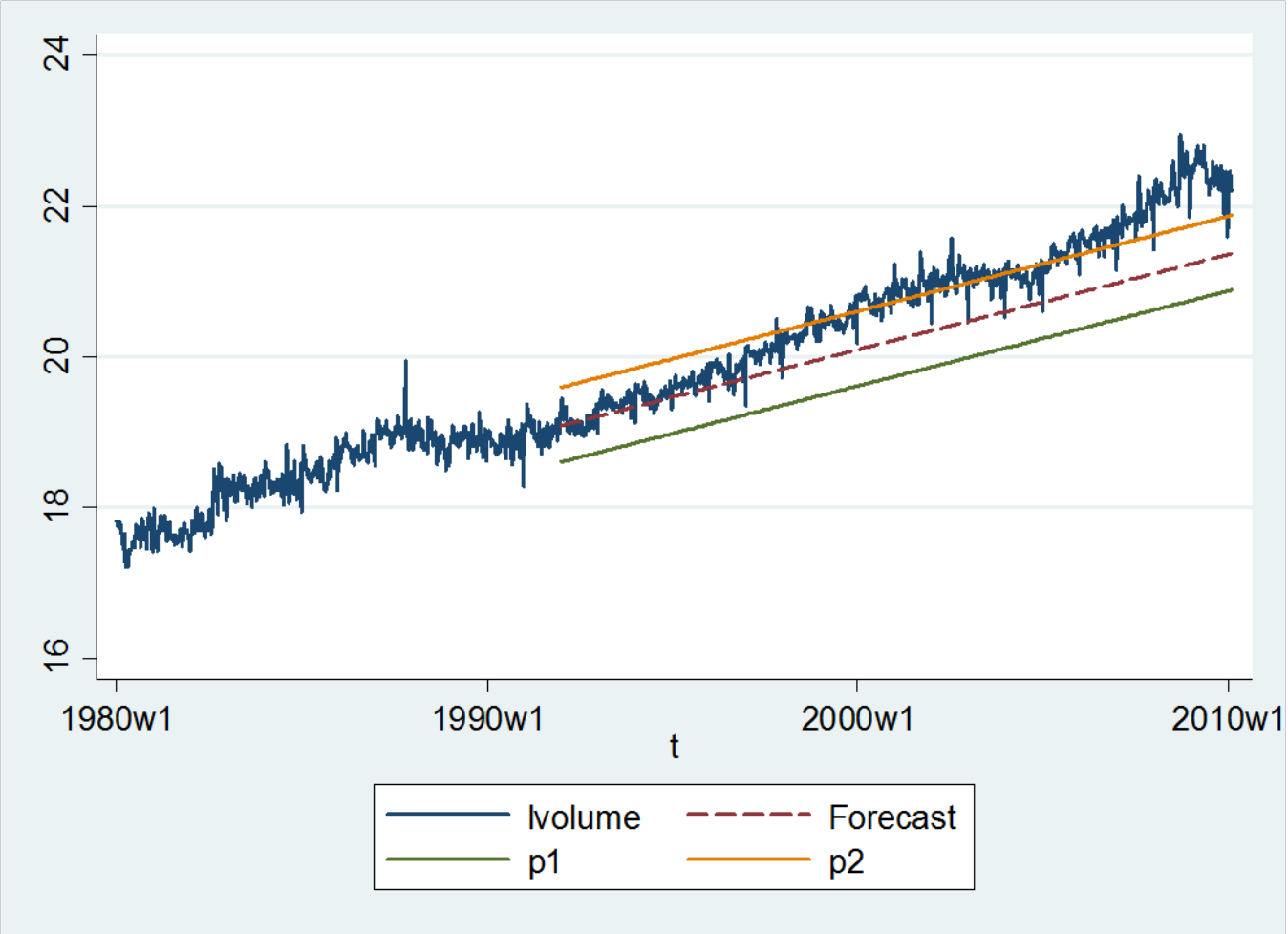
Residuals


Residuals
2
1
0
-1
1950w1
1960w1
1970w1
1980w1
1990w1
2000w1
2010w1
t

Forecast


22
20
18
16
14
1950w1
1960w1
1970w1
1980w1
1990w1
2000w1
2010w1
t
y1
Forecast
p1
p2

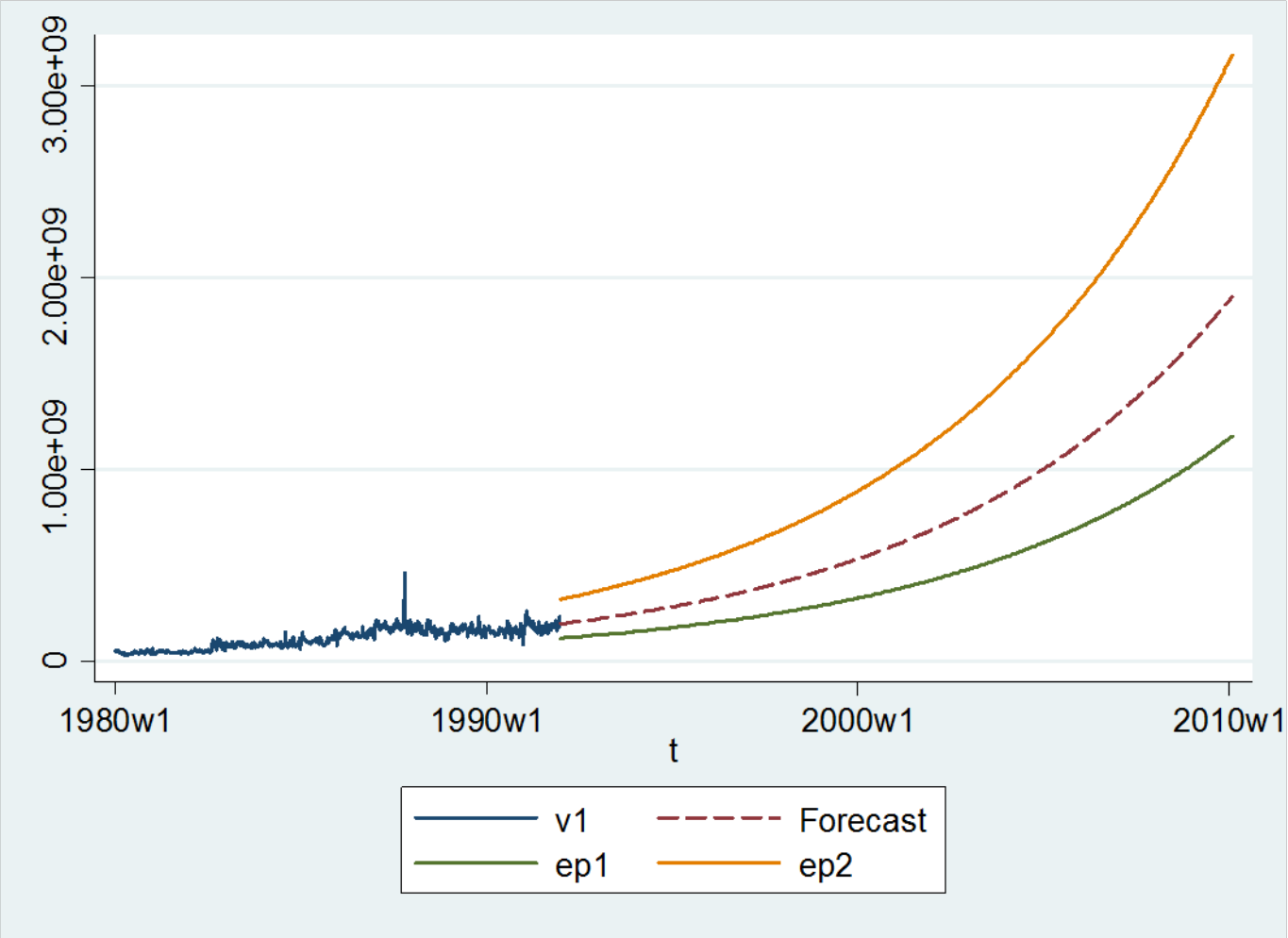
Out-of-Sample



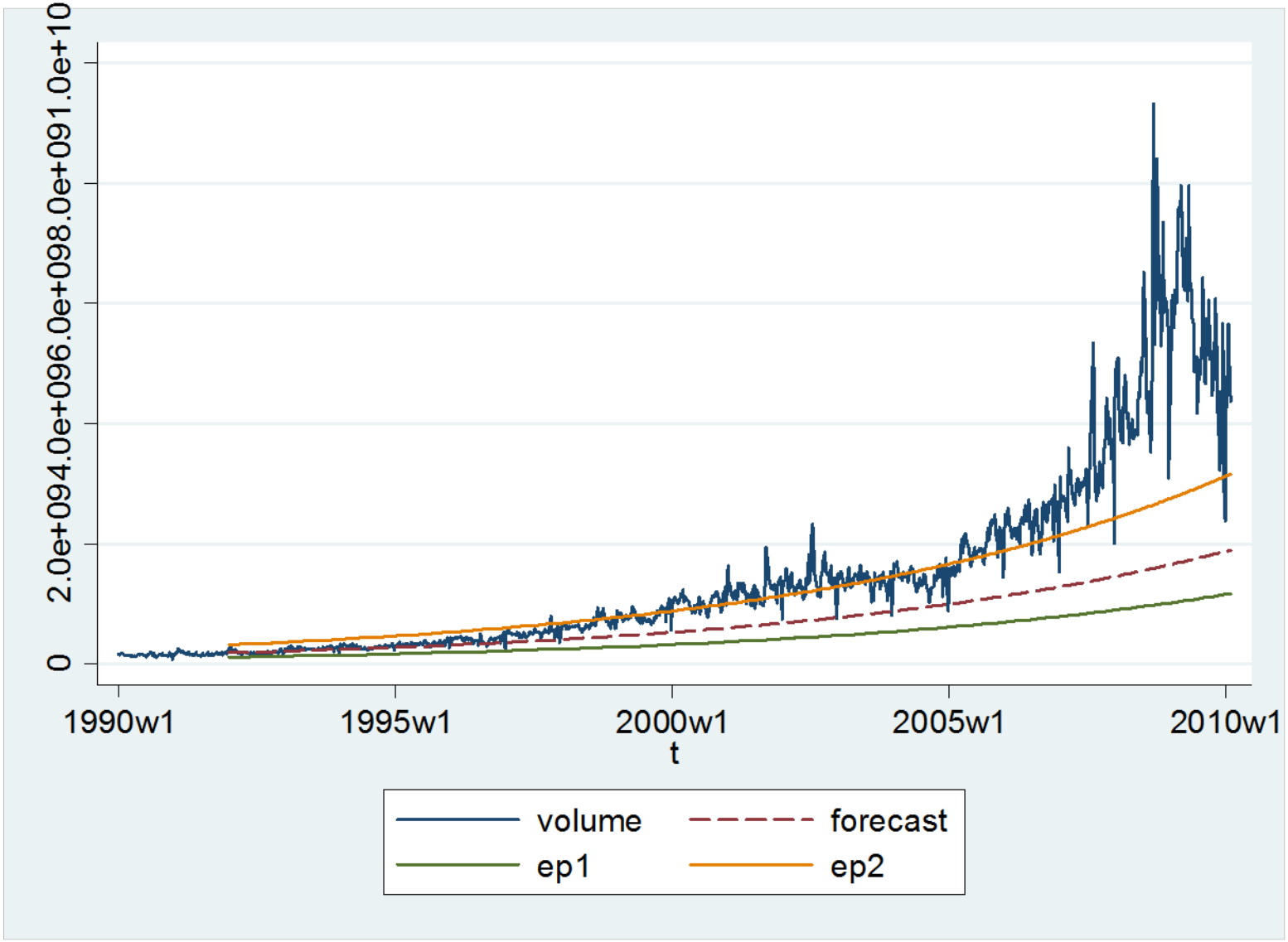
Forecasting Levels from a Forecast of Logs

- Let Y_t be a series and $y_t = \ln(Y_t)$ its logarithm
- Suppose the forecast for the log is a linear trend: $E(y_{t+h} \mid \Omega_t) = T_t = \beta_0 + \beta_1 \, Time_t$
- Then a forecast for Y_t is $\exp(T_t)$
- If $[L_T, U_T]$ is a forecast interval for y_{T+h}
- Then $[\exp(L_T), \exp(U_T)]$ is a forecast interval for Y_{T+h}
- In other words, just take your point and interval forecasts, and apply the exponential function.
  - In STATA, use **generate** command

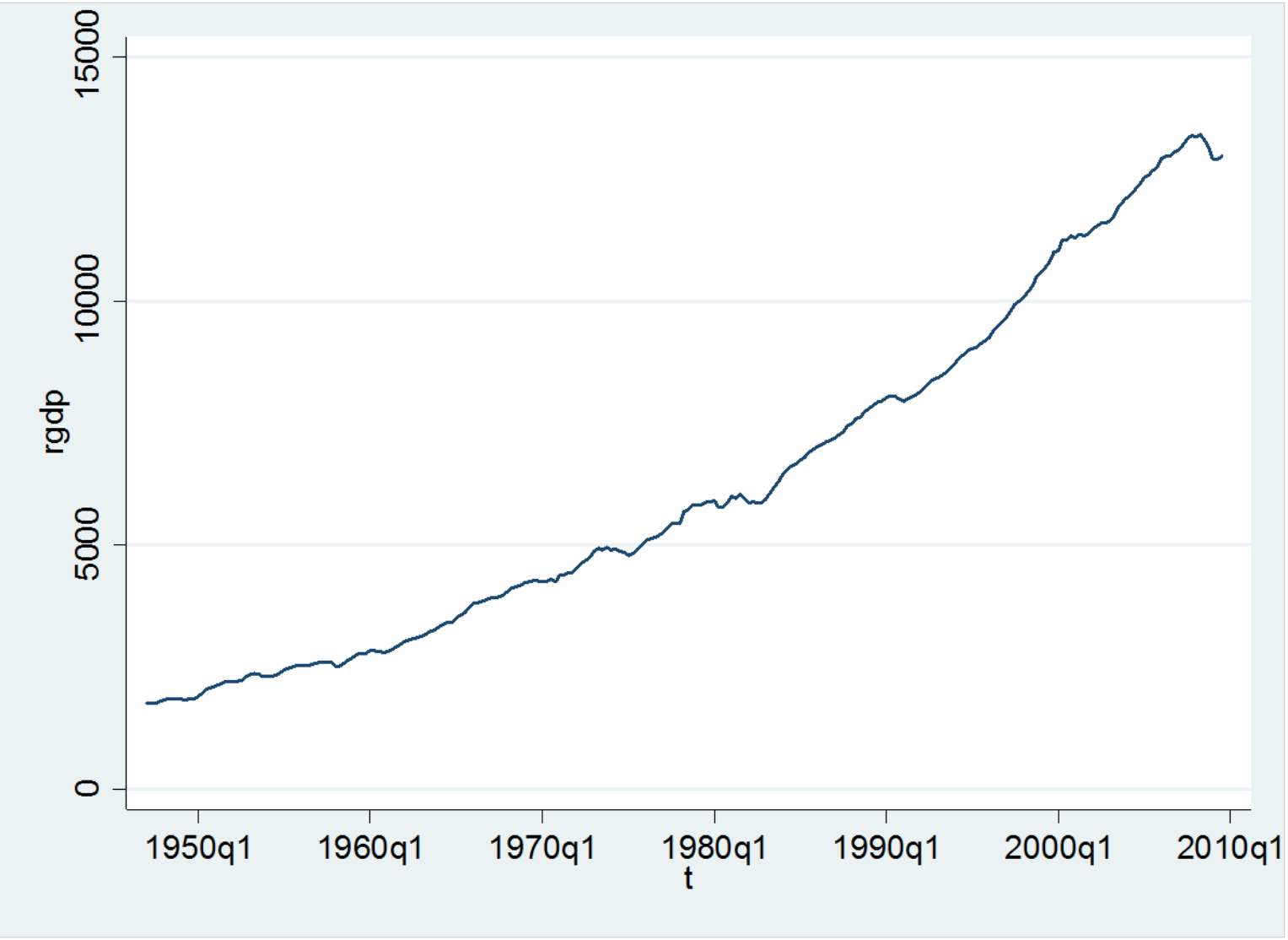
Forecast in Levels



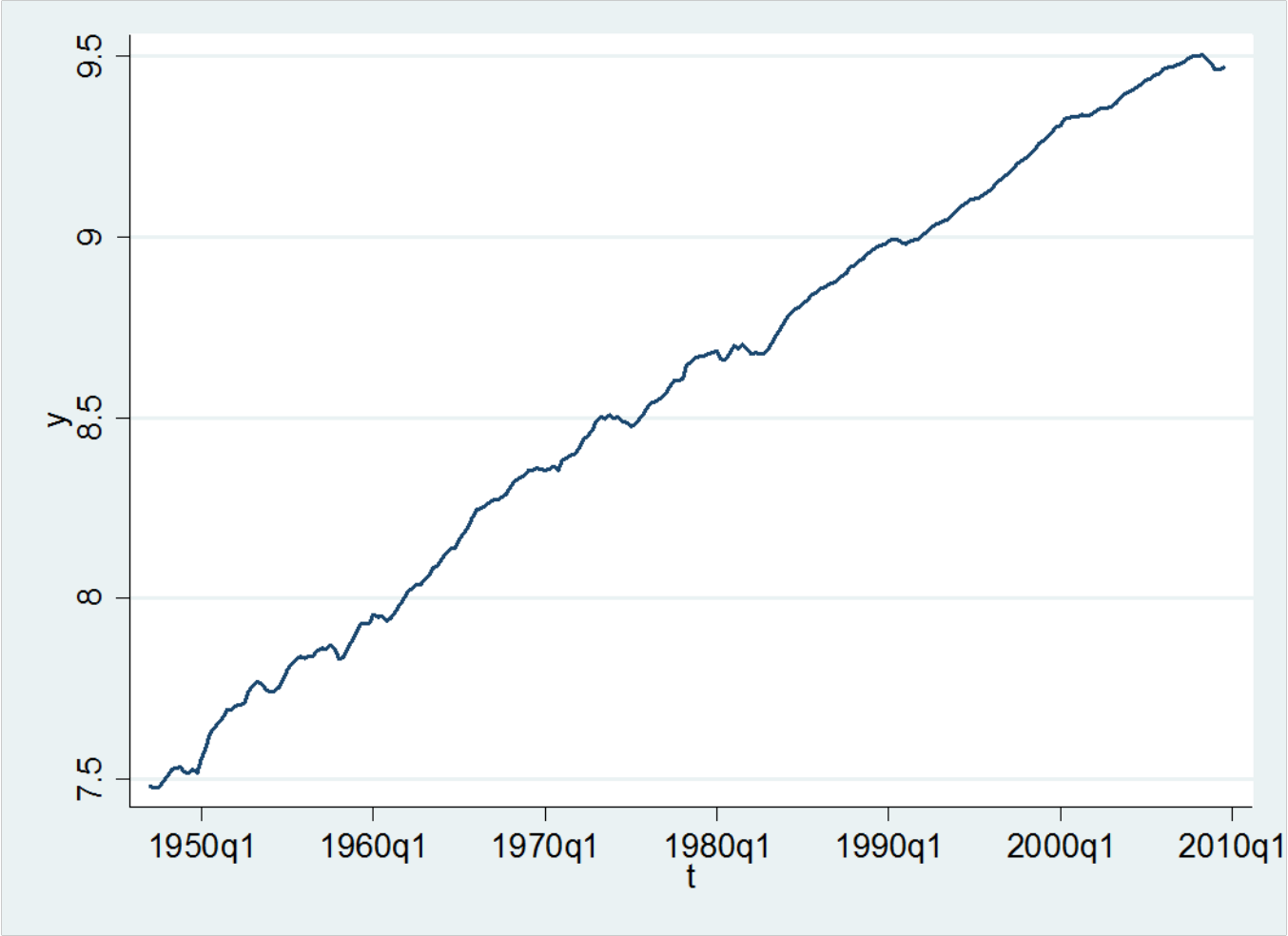
Out-of-Sample



Example 4: Real GDP



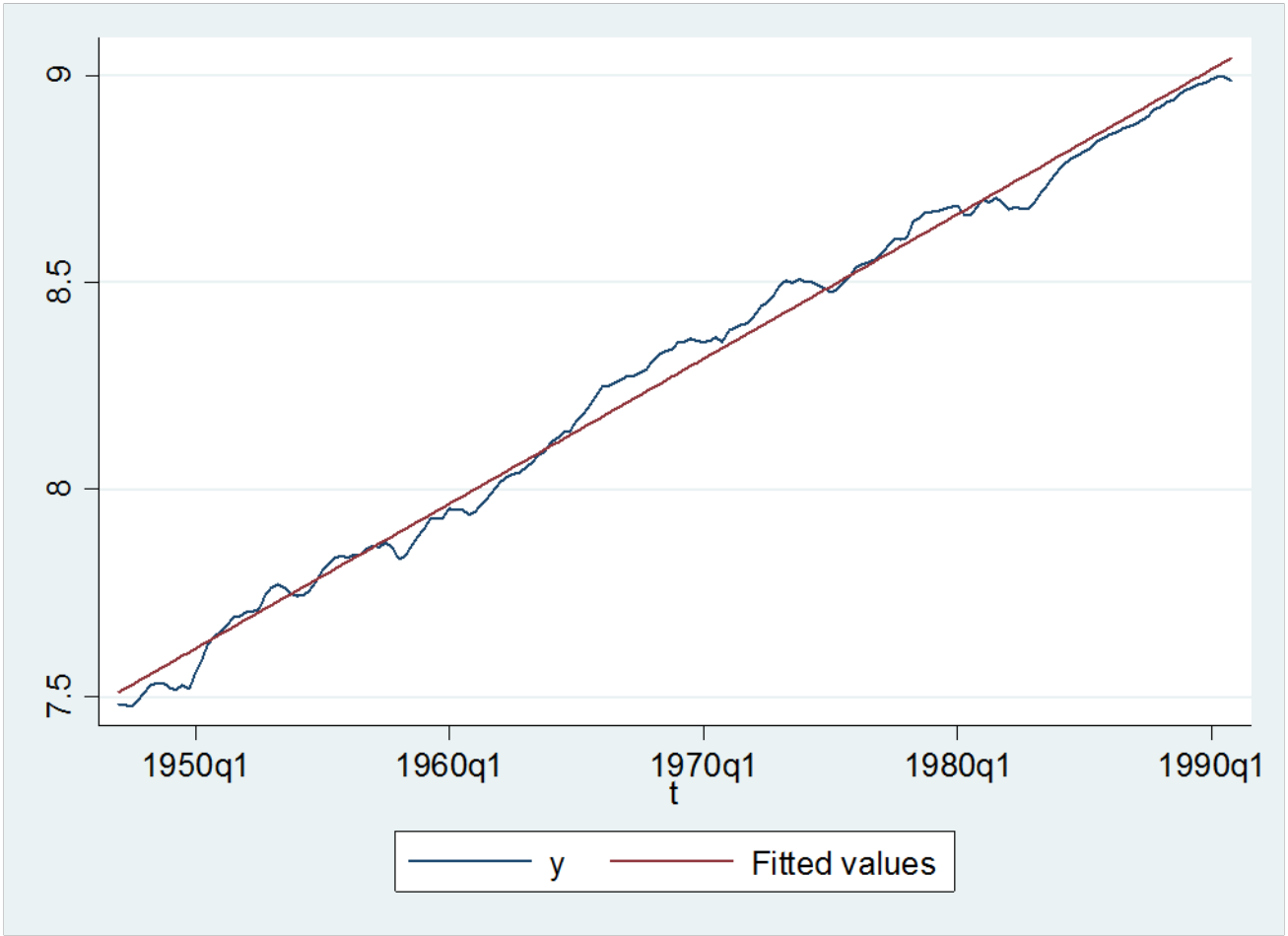
Ln(Real GDP)


y
7.5
8
8.5
9
9.5
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
t

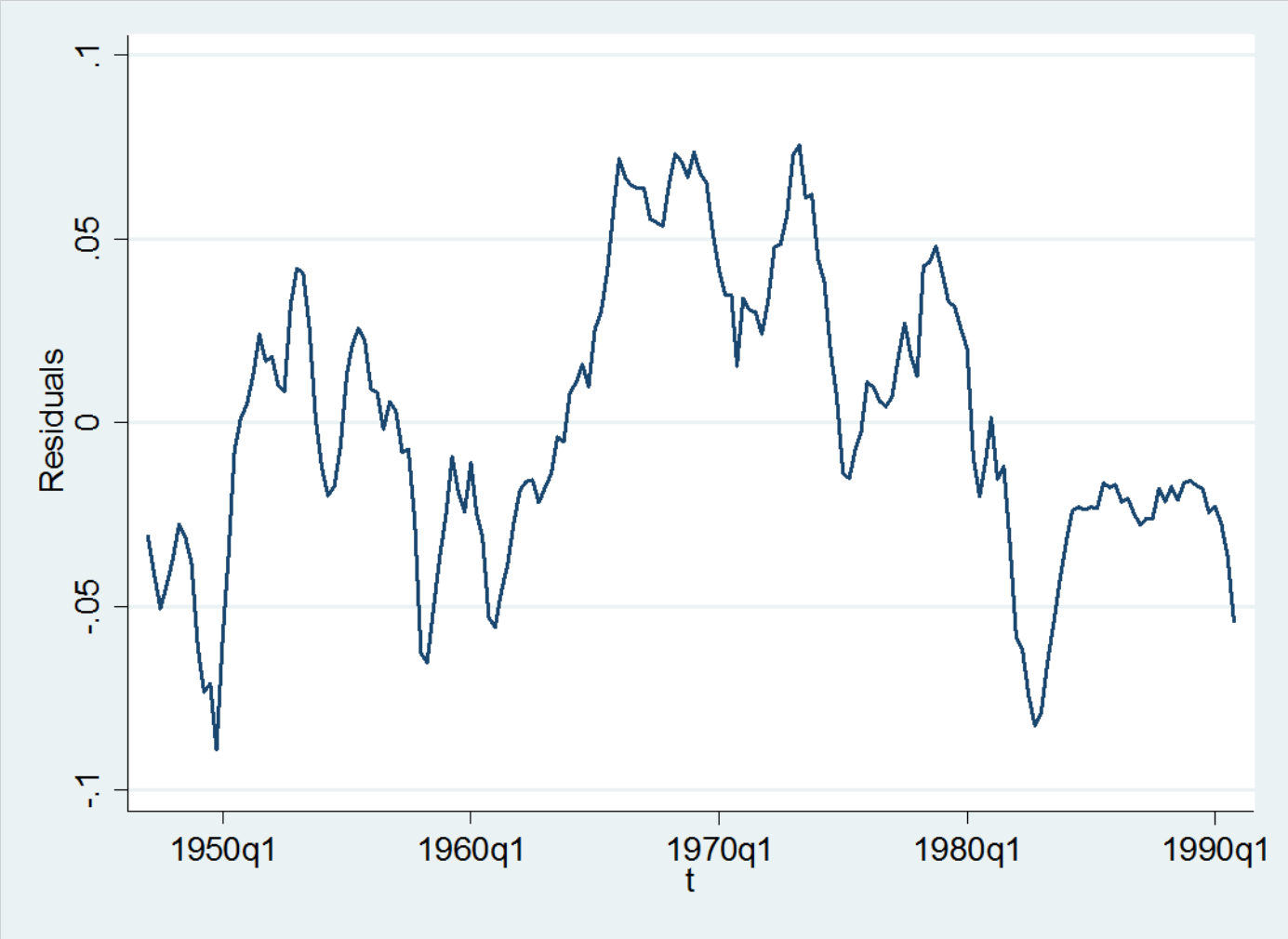
Estimation

```
. regress y t if t<=tq(1990q4)

      Source |       SS       df       MS              Number of obs =     176
-------------+------------------------------           F(  1,   174) =23471.15
       Model |  34.6808119     1  34.6808119           Prob > F      =  0.0000
    Residual |  .257101262   174  .001477593           R-squared     =  0.9926
-------------+------------------------------           Adj R-squared =  0.9926
       Total |  34.9379132   175  .199645218           Root MSE      =  .03844

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0087372    .000057   153.20   0.000     .0086247    .0088498
       _cons |   7.964978   .0035347  2253.35   0.000     7.958001    7.971954
------------------------------------------------------------------------------
```

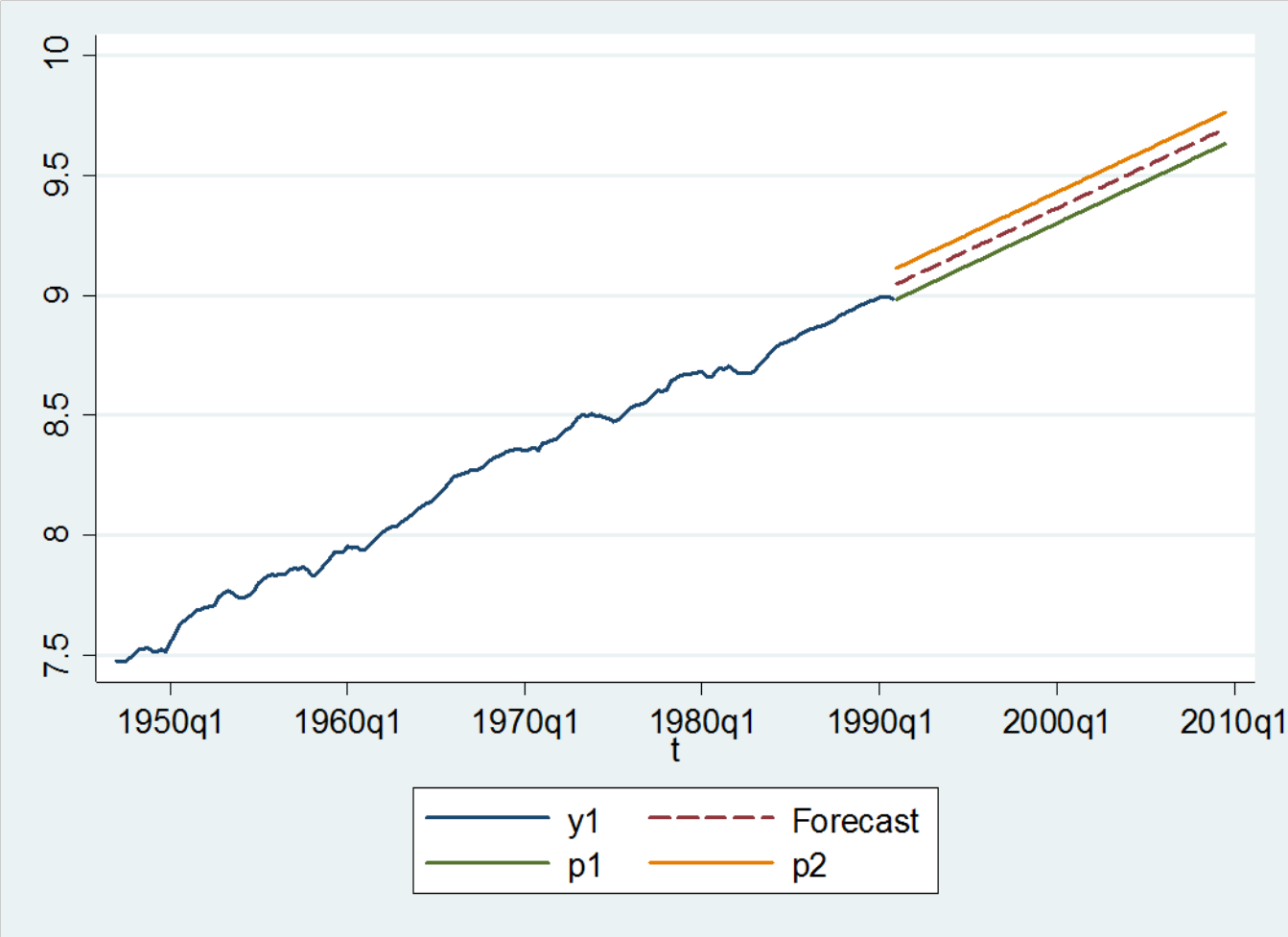
Fitted Trend



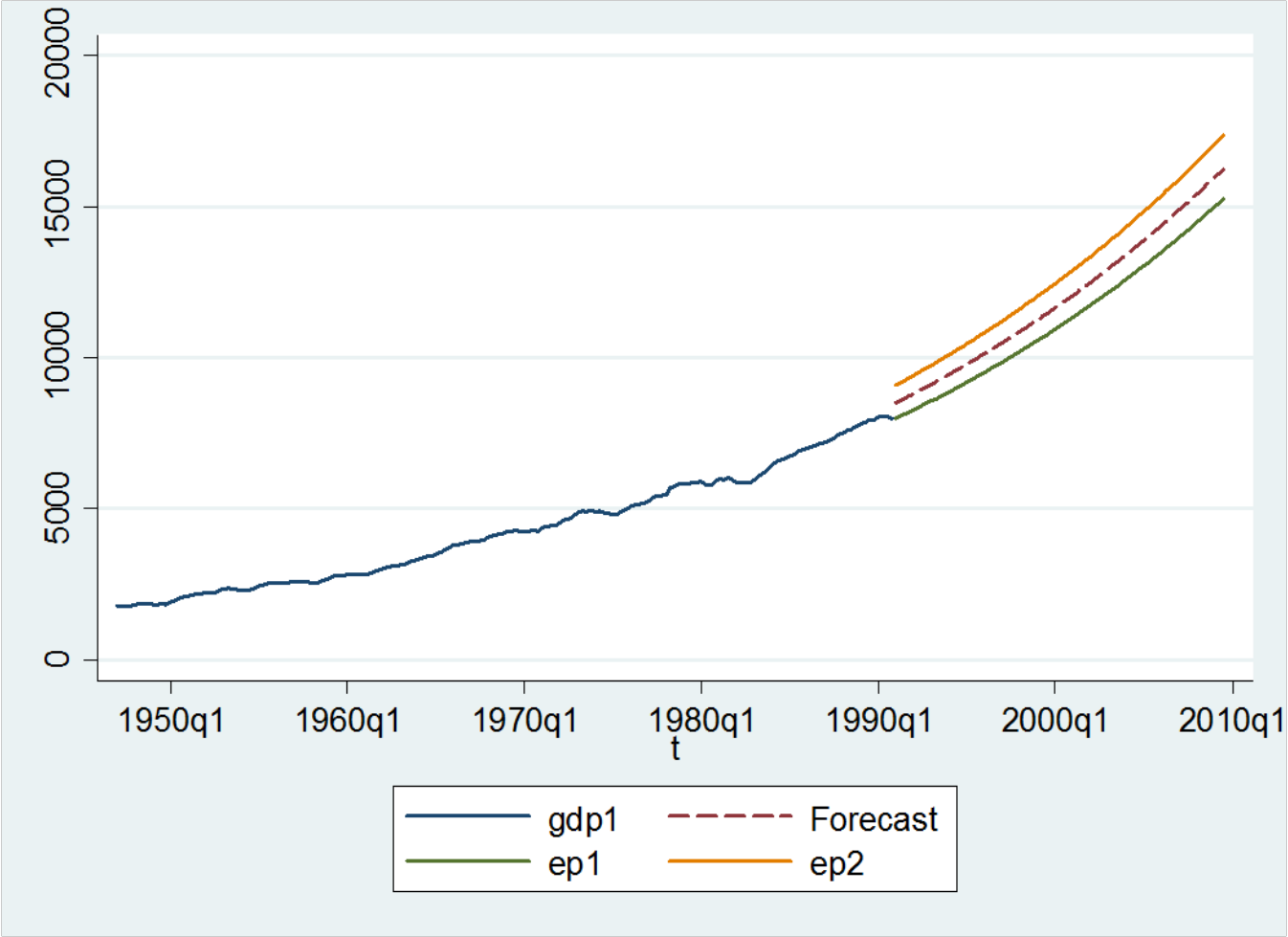
Residuals



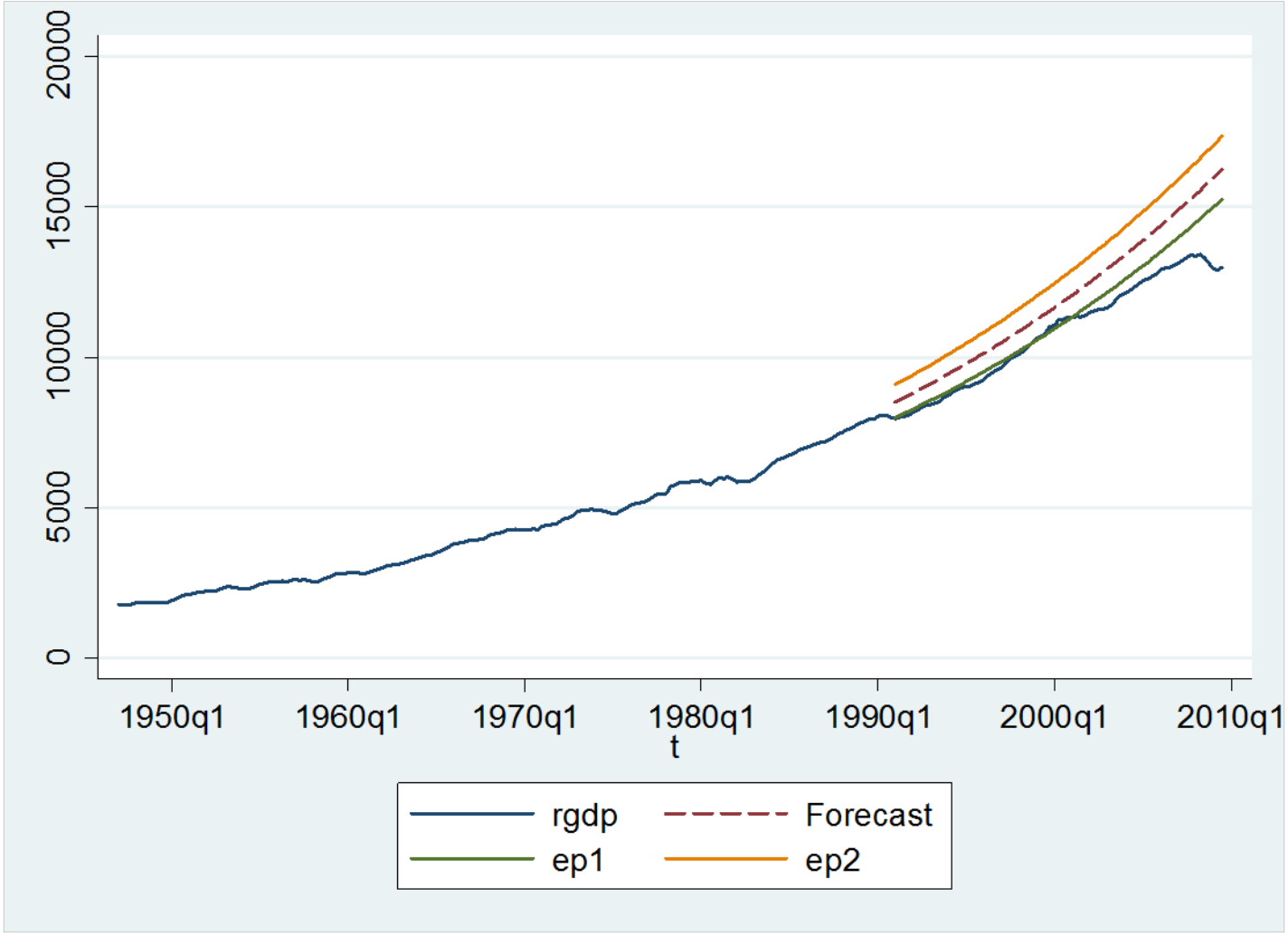
Forecast of ln(RGDP)



Forecast of RGDP (in levels)



Out-of-Sample



Problems with Pure Trend Forecasts

- **Trend forecasts understate uncertainty**
- Actual uncertainty increases at long forecast horizons.
- Short-term trend forecasts can be quite poor unless trend lined up correctly
- Long-term trend forecasts are typically quite poor, as trends change over long time periods
- It is preferred to work with growth rates, and reconstruct levels from forecasted growth rates (more on this later).

Trend Models

- I hope I've convinced you to be skeptical of trend-based forecasting.
- The problem is that there is no economic theory for constant trends, and "changes" in the trend function are not apparent before they occur.
- It is better to forecast growth rates, and build levels from growth.

Final Trend Forecast
World Record – 100 meter sprint