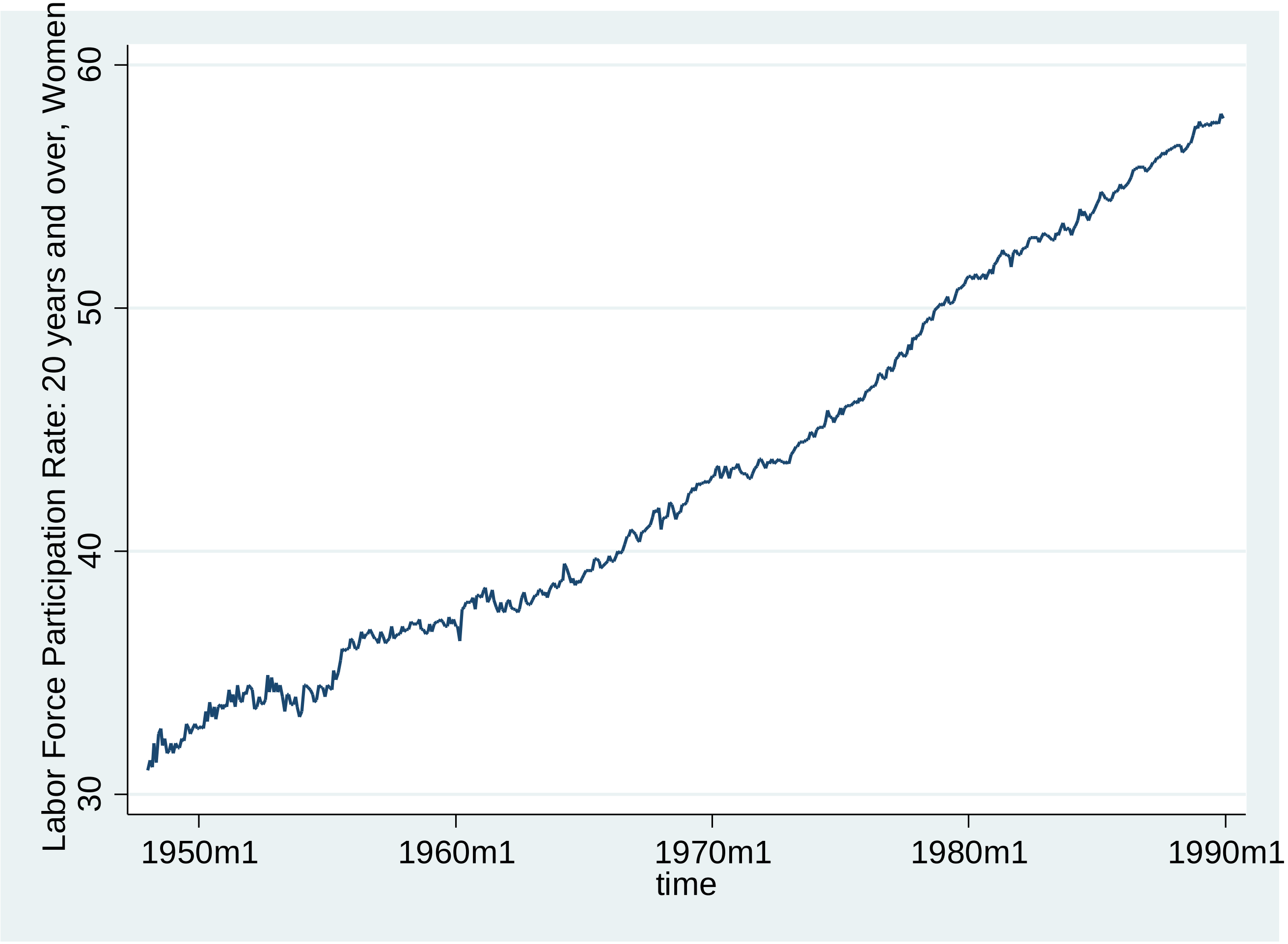
Forecast with Trend Model

- The point forecast is the linear function with estimated coefficients

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

- Estimate coefficients using **regress**
- Compute forecasts with **predict**
- Forecast intervals:
  - Compute standard deviation of forecast with **predict**
  - Add and subtract, multiplied by normal quantile

Example 1
Women's Labor Force Participation Rate



Regression on Time Trend

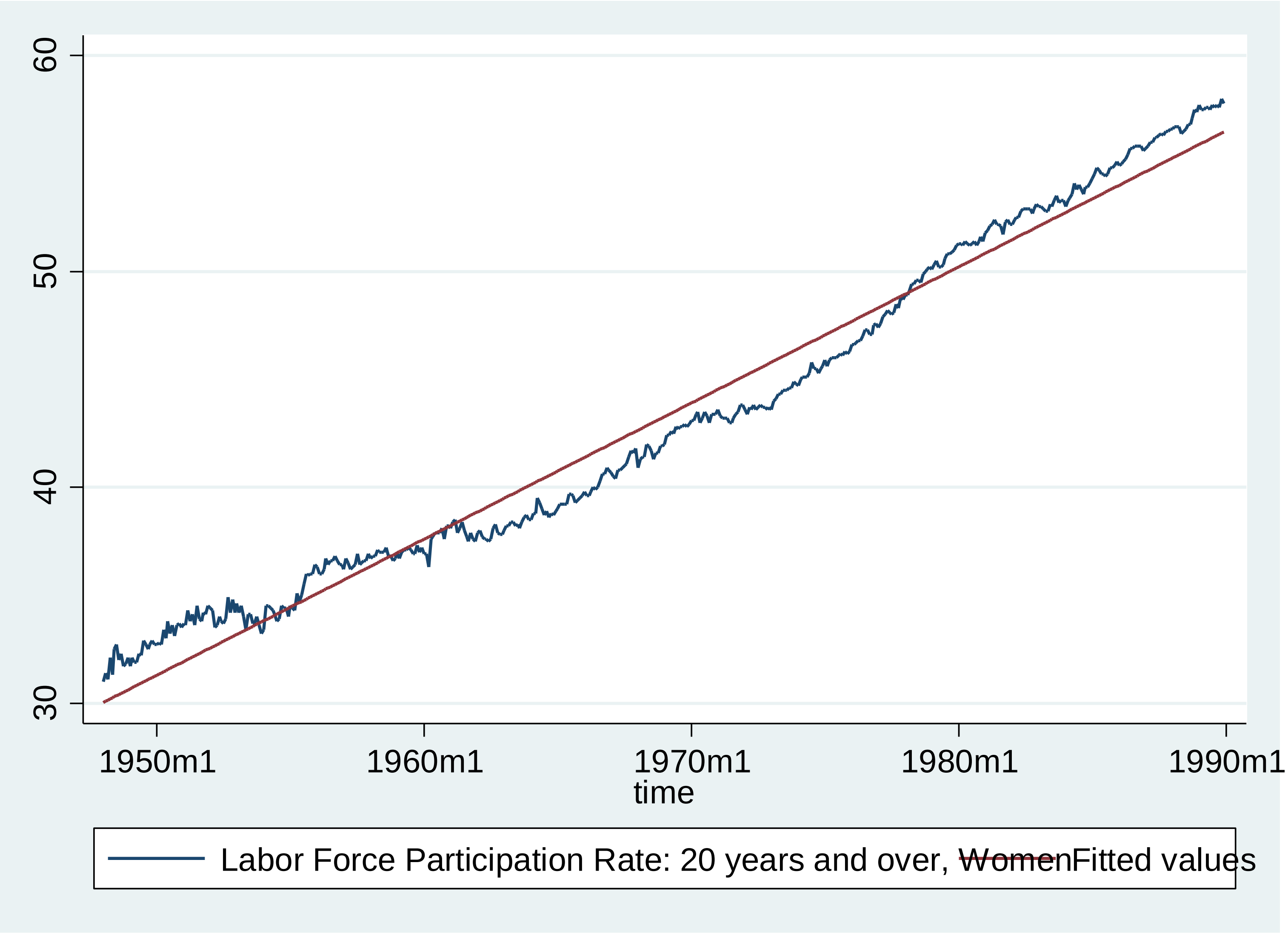
- `regress women time if time<tm(1990m1)`
- `predict wp if time<tm(1990m1)`
- `tsline women wp if time<tm(1990m1)`

Source	SS	df	MS
Model	29468.1115	1	29468.1115
Residual	708.123471	502	1.41060452
Total	30176.2349	503	59.9925148

Number of obs	=	504
F(1, 502)	=	20890.41
Prob > F	=	0.0000
R-squared	=	0.9765
Adj R-squared	=	0.9765
Root MSE	=	1.1877

women	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
time	.0525559	.0003636	144.54	0.000	.0518415	.0532703
_cons	37.59706	.0657783	571.57	0.000	37.46783	37.7263

In-Sample Fit

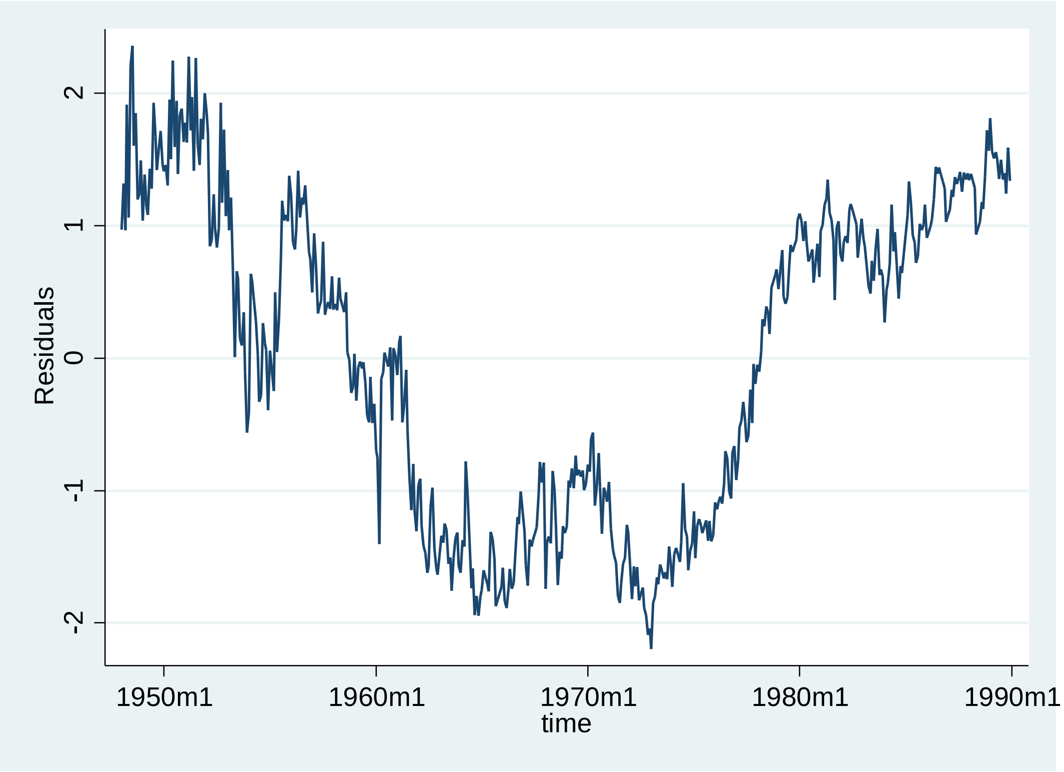


Residuals

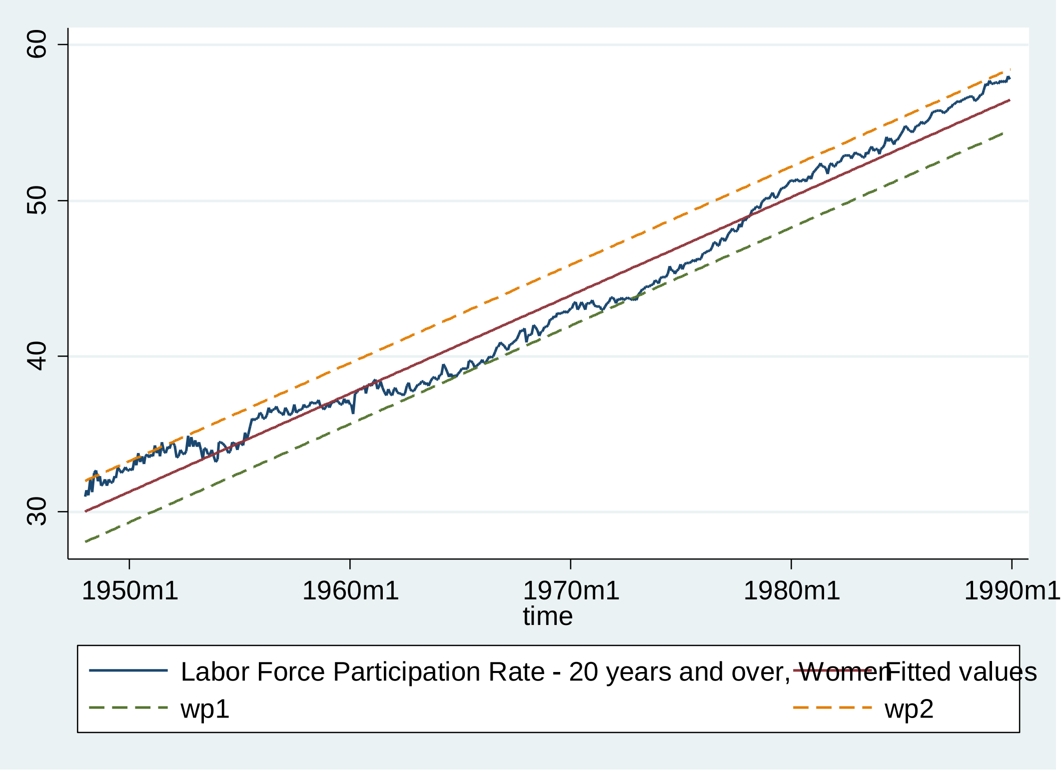
- Residuals are difference between data and fitted regression line
  - `predict e if time<tm(1990m1), residuals`

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

Residual Plot



In-Sample Fit

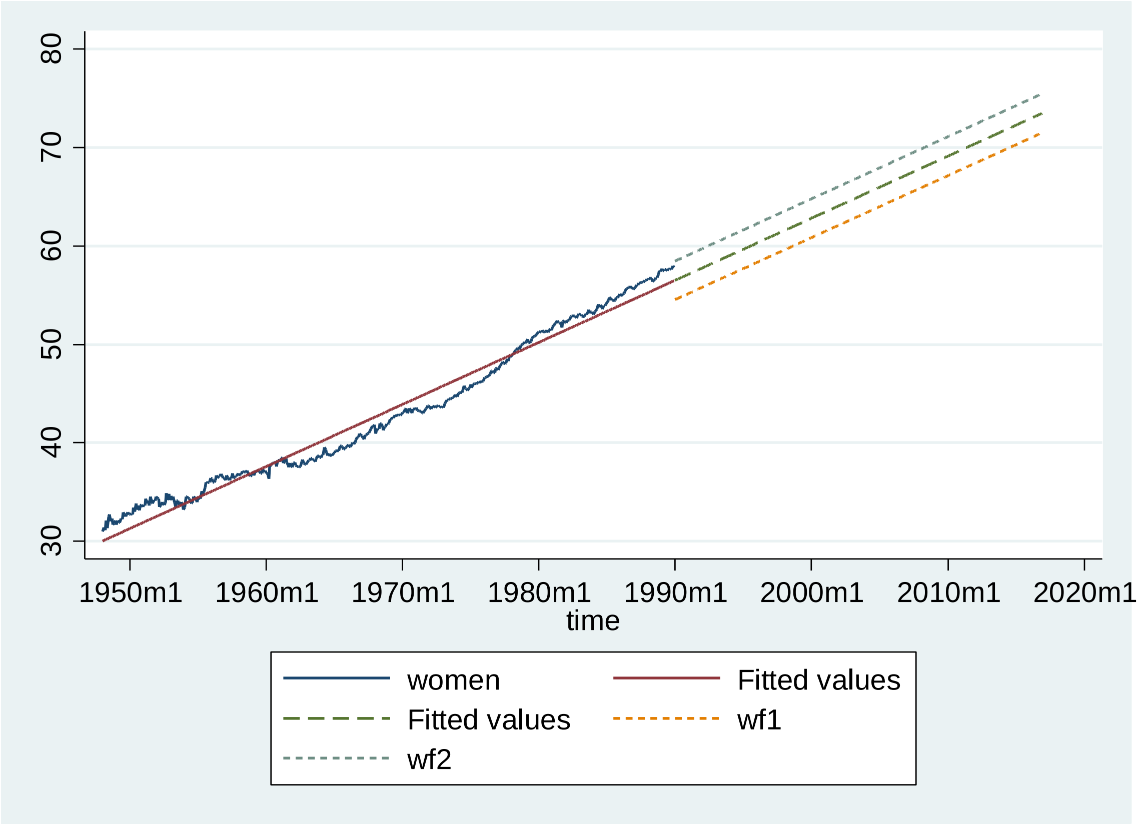


- `predict wp if time<tm(1990m1)`
- `predict sw, stdf`
- `generate wp1 = wp - 1.645*sw`
- `generate wp2 = wp + 1.645*sw`
- `tsline women wp wp1 wp2 if time<tm(1990m1), lpattern (solid solid dash dash)`
  - Does the in-sample fit look good?

Stata do file

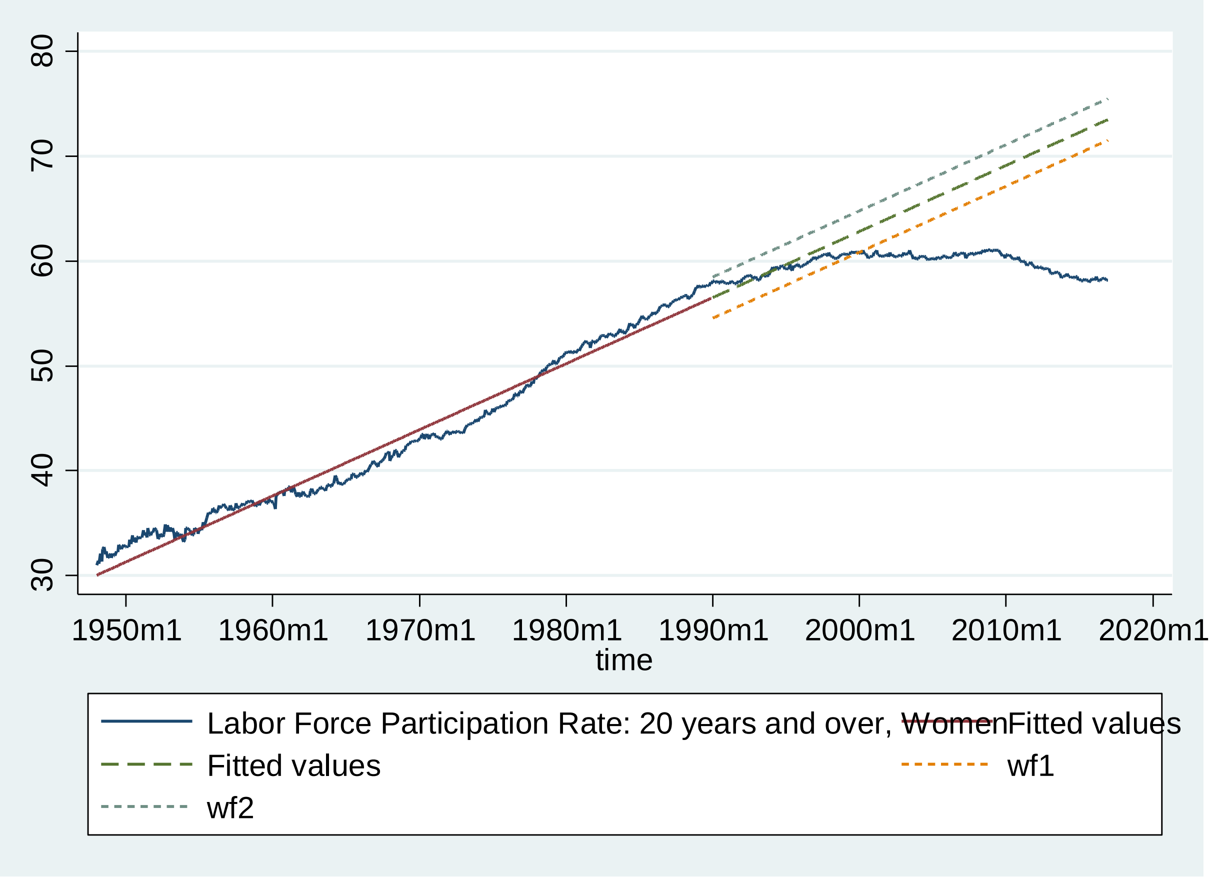
- `regress women time if time<tm(1990m1)`
- `predict wp if time<tm(1990m1)`
- `predict wf if time>=tm(1990m1)`
- `predict sw, stdf`
- `generate wp1 = wf - 1.645*sw`
- `generate wp2 = wf + 1.645*sw`
- `generate women1=women if time<tm(1990m1)`
- `label variable women1 women`
- `tsline women1 wp wf wp1 wp2, lpattern (solid solid dash shortdash shortdash)`
- `tsline women wp wf wp1 wp2, lpattern (solid solid dash shortdash shortdash)`

Out-of-Sample Forecast



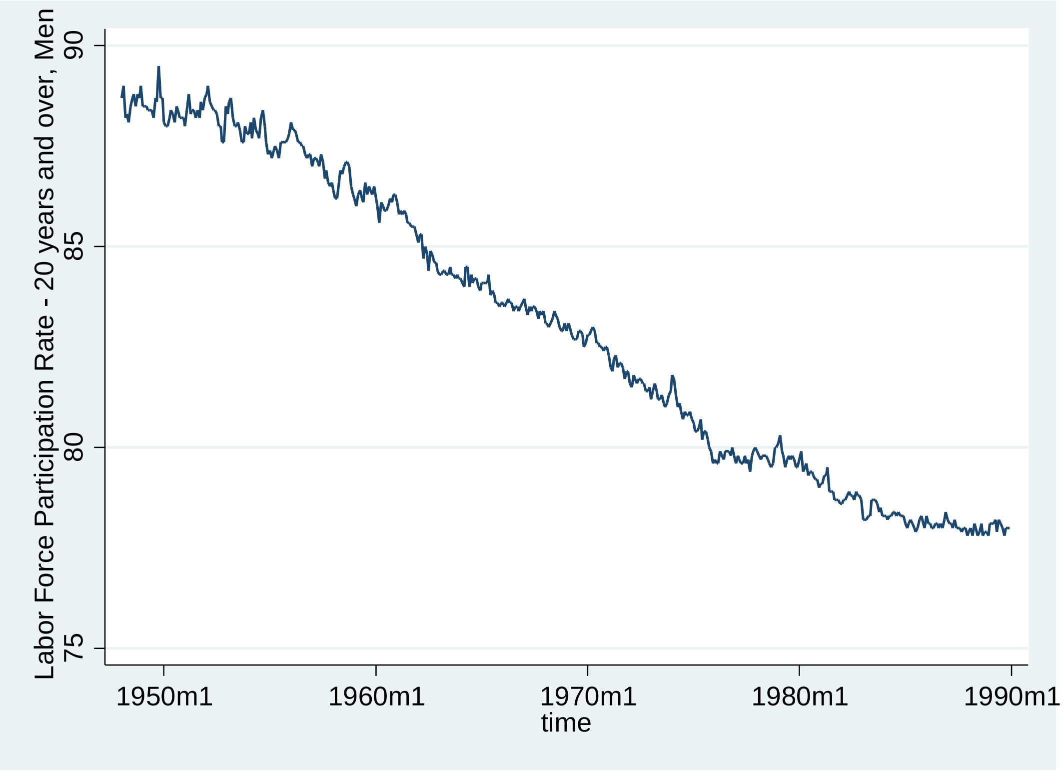
- Out of sample prediction might be too low.

Actual Out-of-Sample



- No: Prediction was way too high!

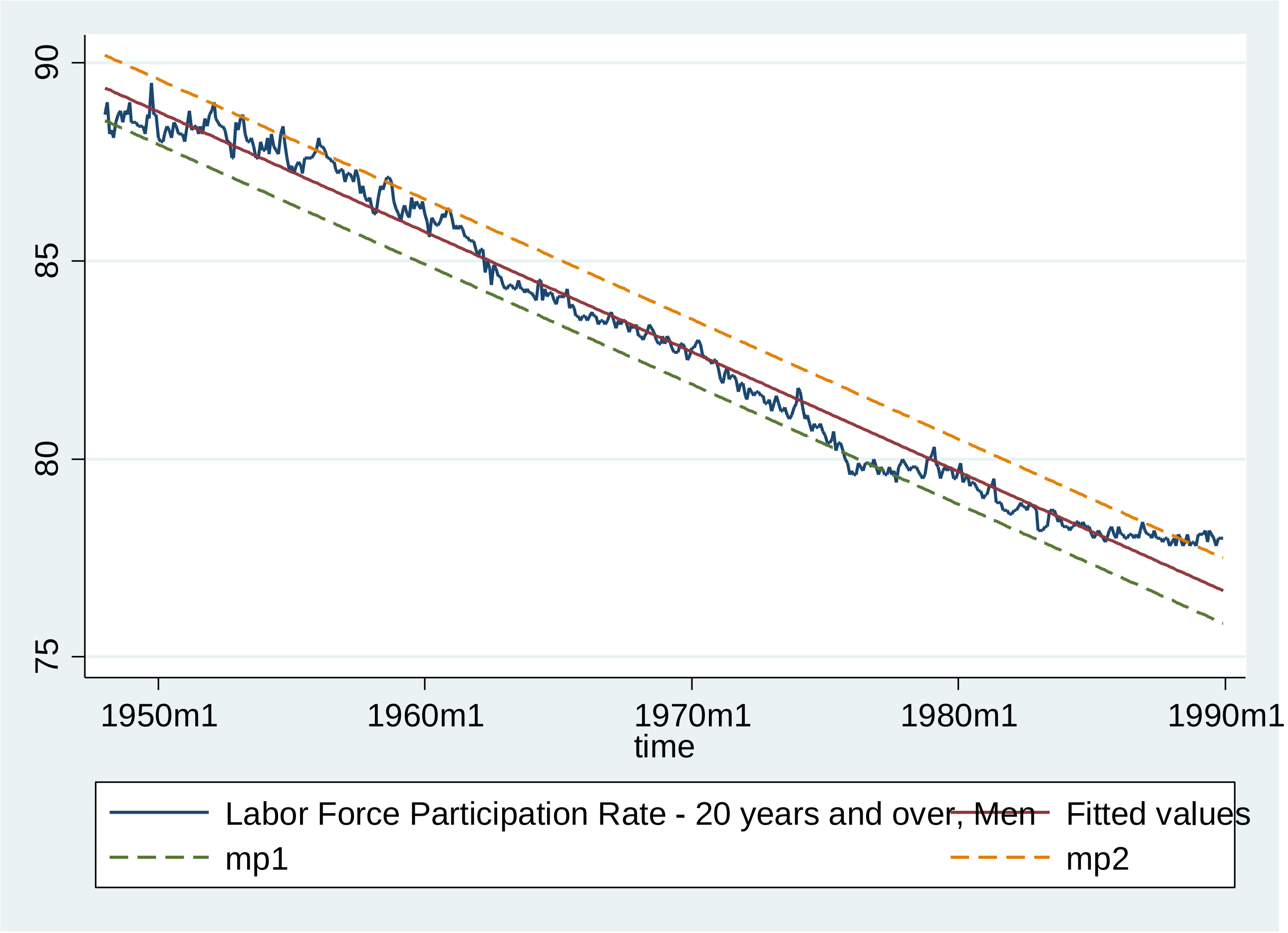
Men’s Labor Force Participation Rate



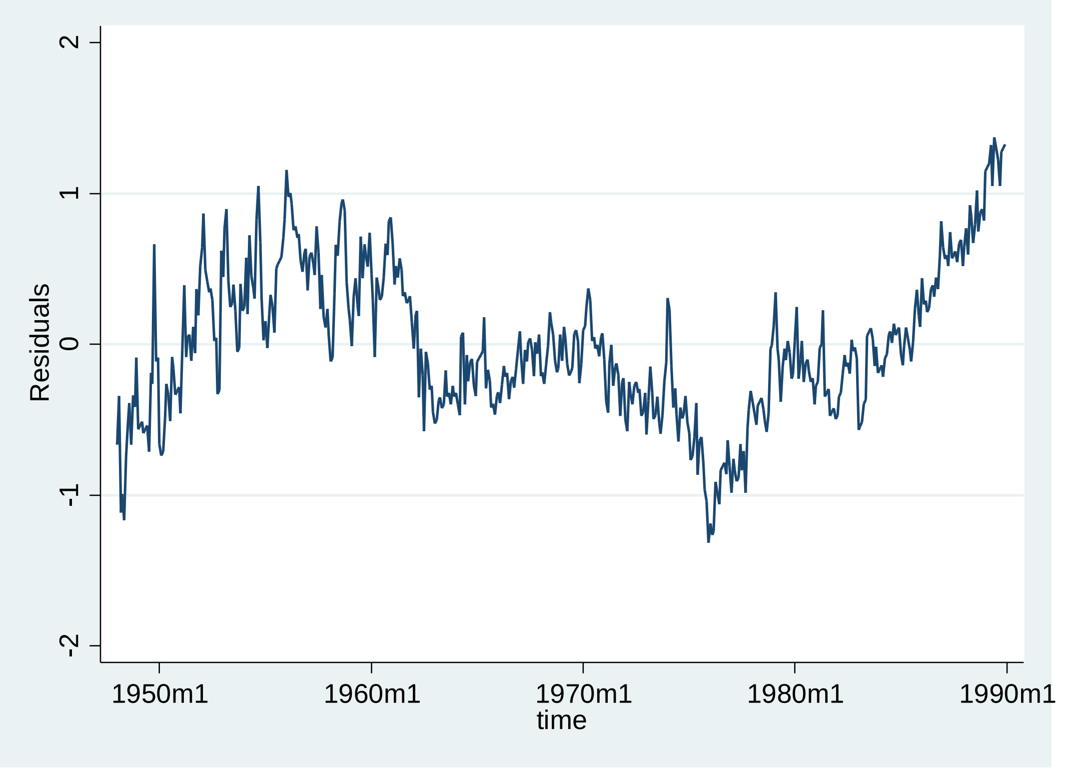
Estimation

```
. regress men time if time<tm(1990m1)

      Source |       SS       df       MS              Number of obs =     504
-------------+------------------------------           F(  1,   502) =27289.02
       Model |  6793.87072     1  6793.87072           Prob > F      =  0.0000
    Residual |  124.977854   502  .248959868           R-squared     =  0.9819
-------------+------------------------------           Adj R-squared =  0.9819
       Total |  6918.84857   503  13.7551661           Root MSE      =  .49896

------------------------------------------------------------------------------
         men |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |  -.0252351   .0001528  -165.19   0.000    -.0255352   -.0249349
       _cons |    85.7342   .0276341  3102.48   0.000      85.6799    85.78849
------------------------------------------------------------------------------
```

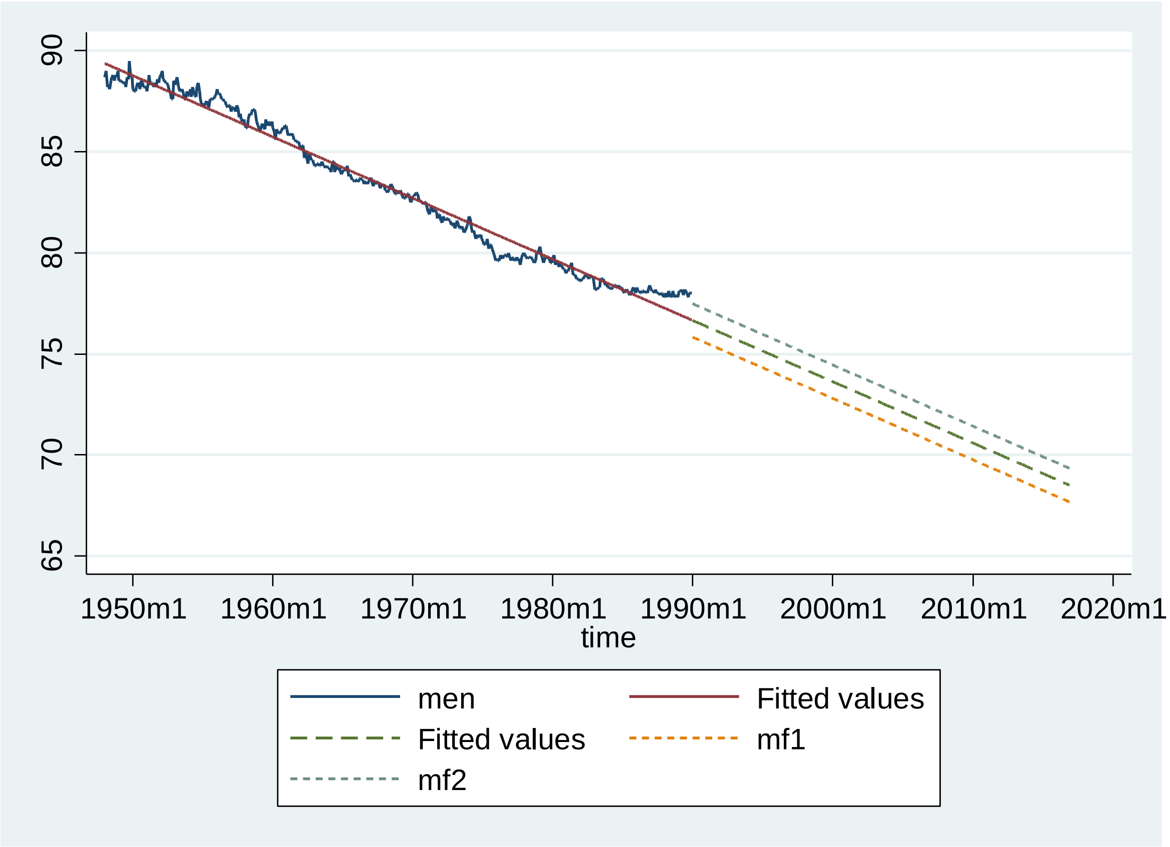
In-Sample Fit



Residuals

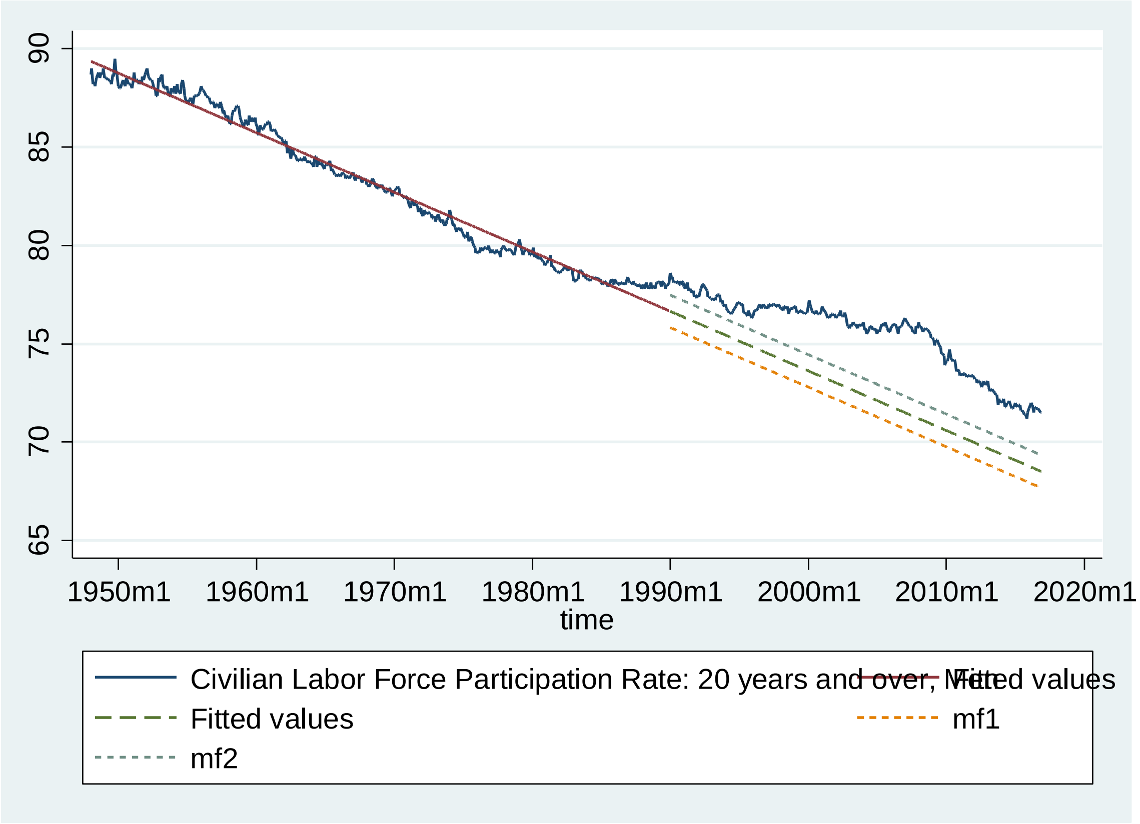


Forecast



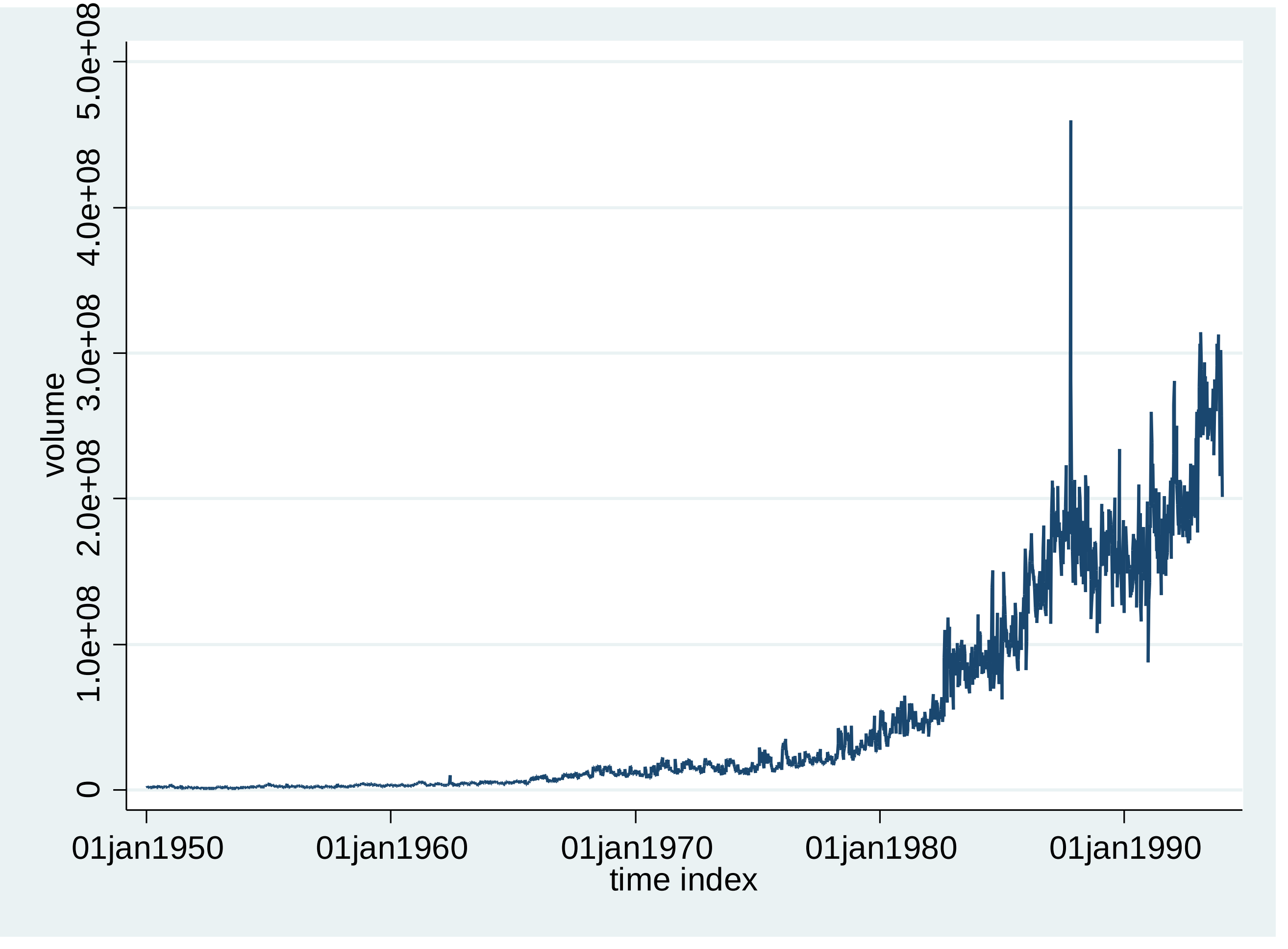
- End of Sample looks worrying

Actual Out-of-Sample



- Linear Trend Terrible

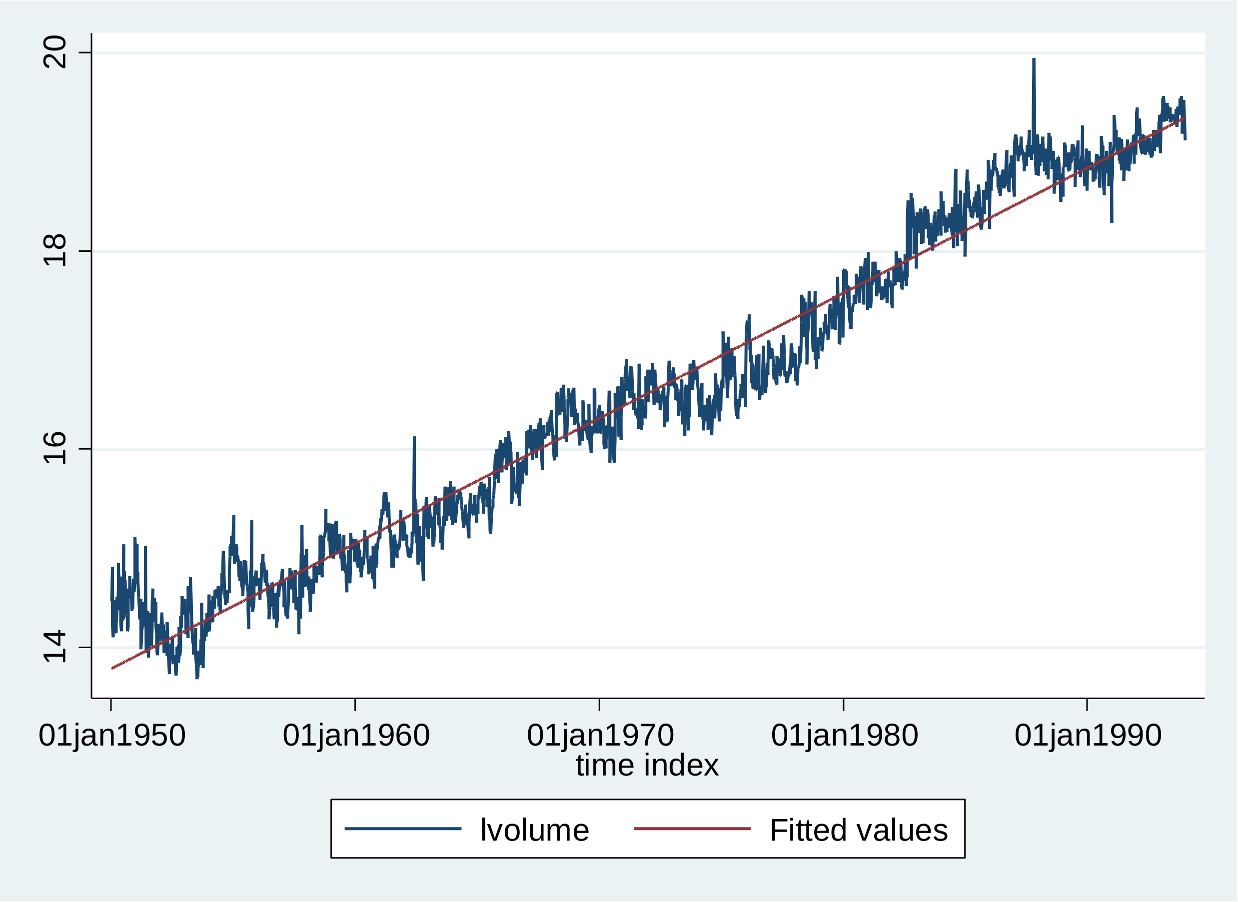
Example 2: Transaction Volume



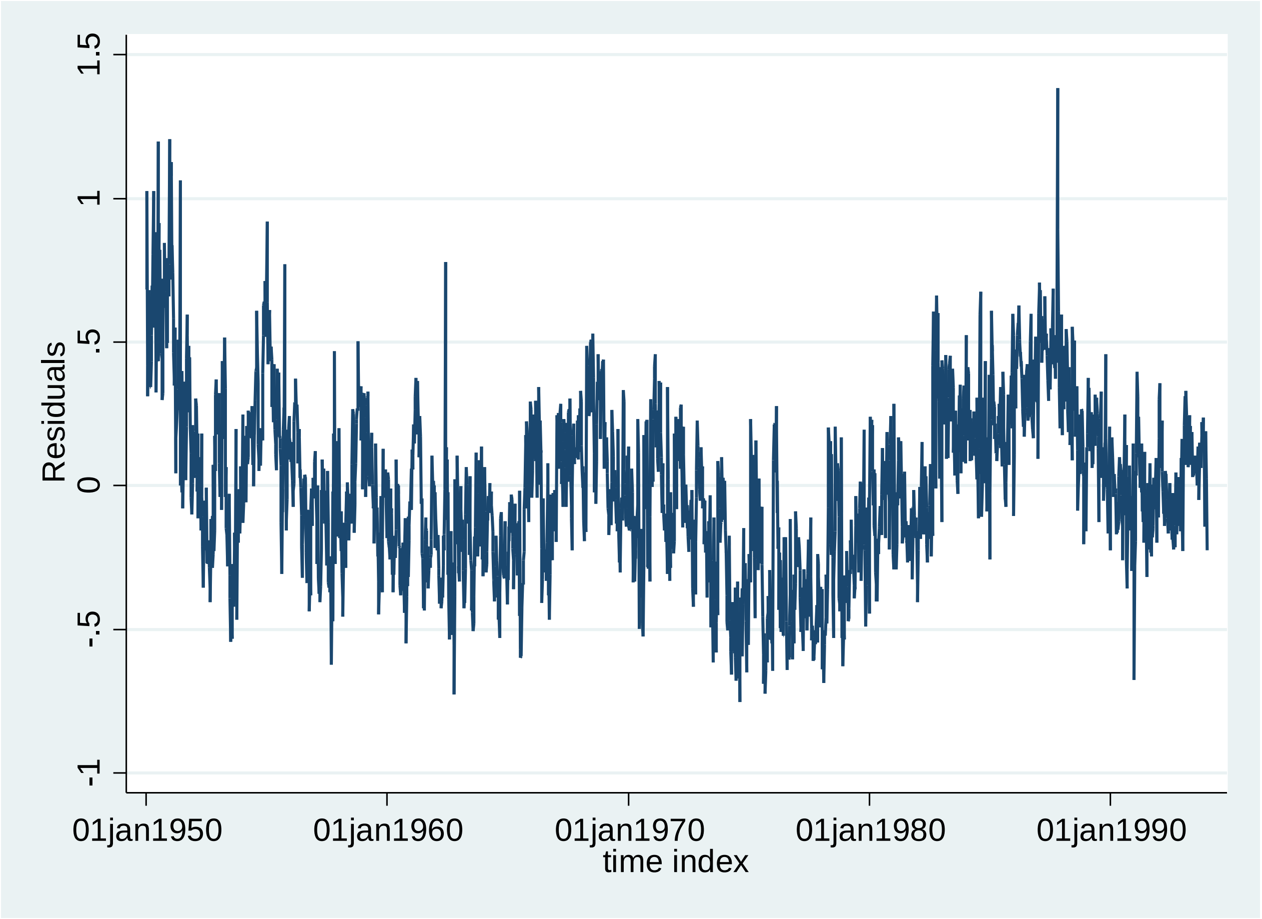
Estimating Logarithmic Trend

```
. regress lvolume time if time<td(07jan1994)

      Source |       SS           df       MS      Number of obs   =     2,296
-------------+----------------------------------   F(1, 2294)      =  67135.58
       Model |  5910.07977         1  5910.07977   Prob > F        =    0.0000
    Residual |  201.945409     2,294     .088032   R-squared       =    0.9670
-------------+----------------------------------   Adj R-squared   =    0.9669
       Total |  6112.02518     2,295   2.6631918   Root MSE        =     .2967

------------------------------------------------------------------------------
     lvolume |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0003458   1.33e-06   259.11   0.000     .0003432    .0003484
       _cons |   15.04939   .0085205  1766.26   0.000     15.03268     15.0661
------------------------------------------------------------------------------
```

Fitted Trend

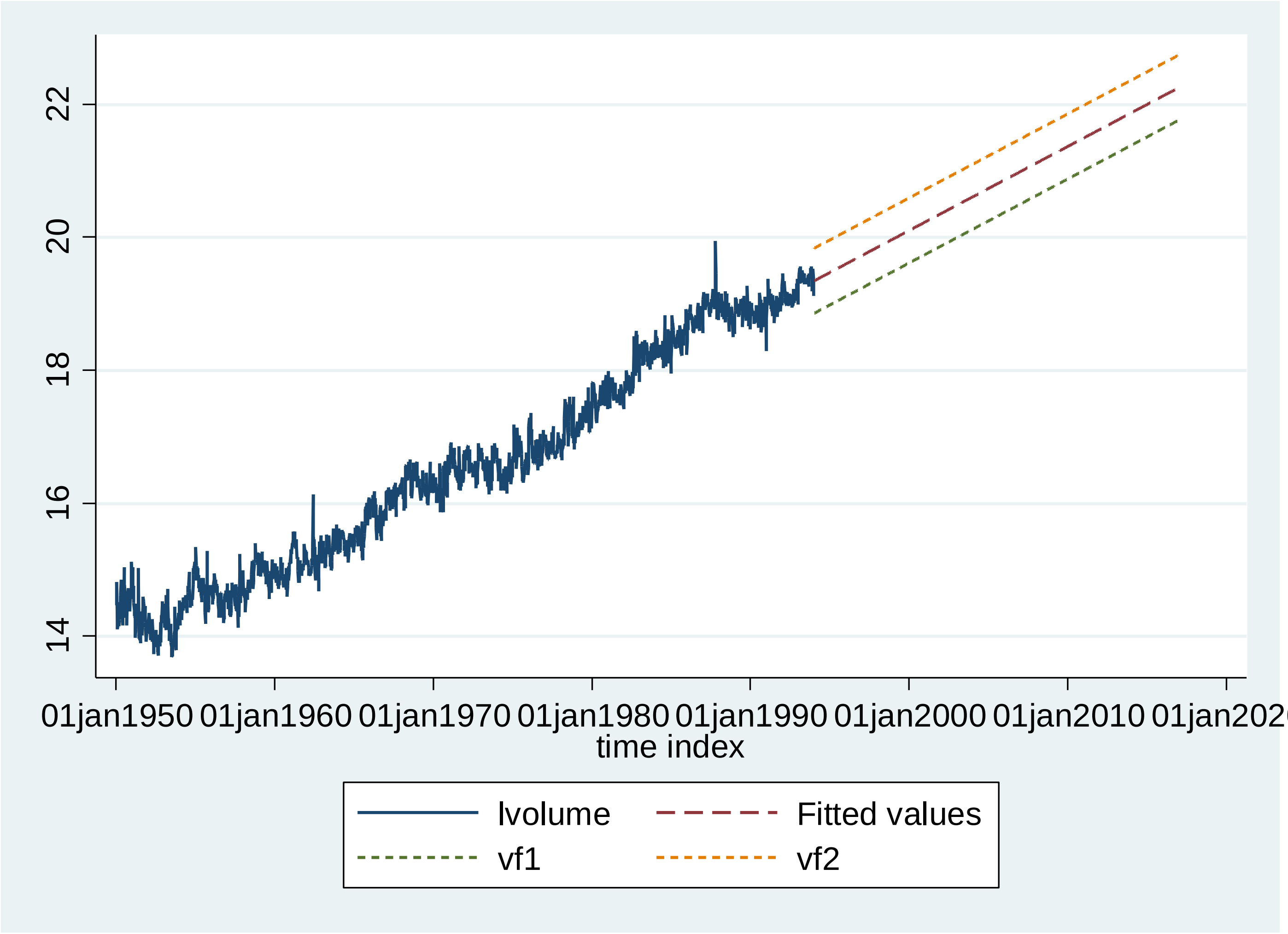


Residuals



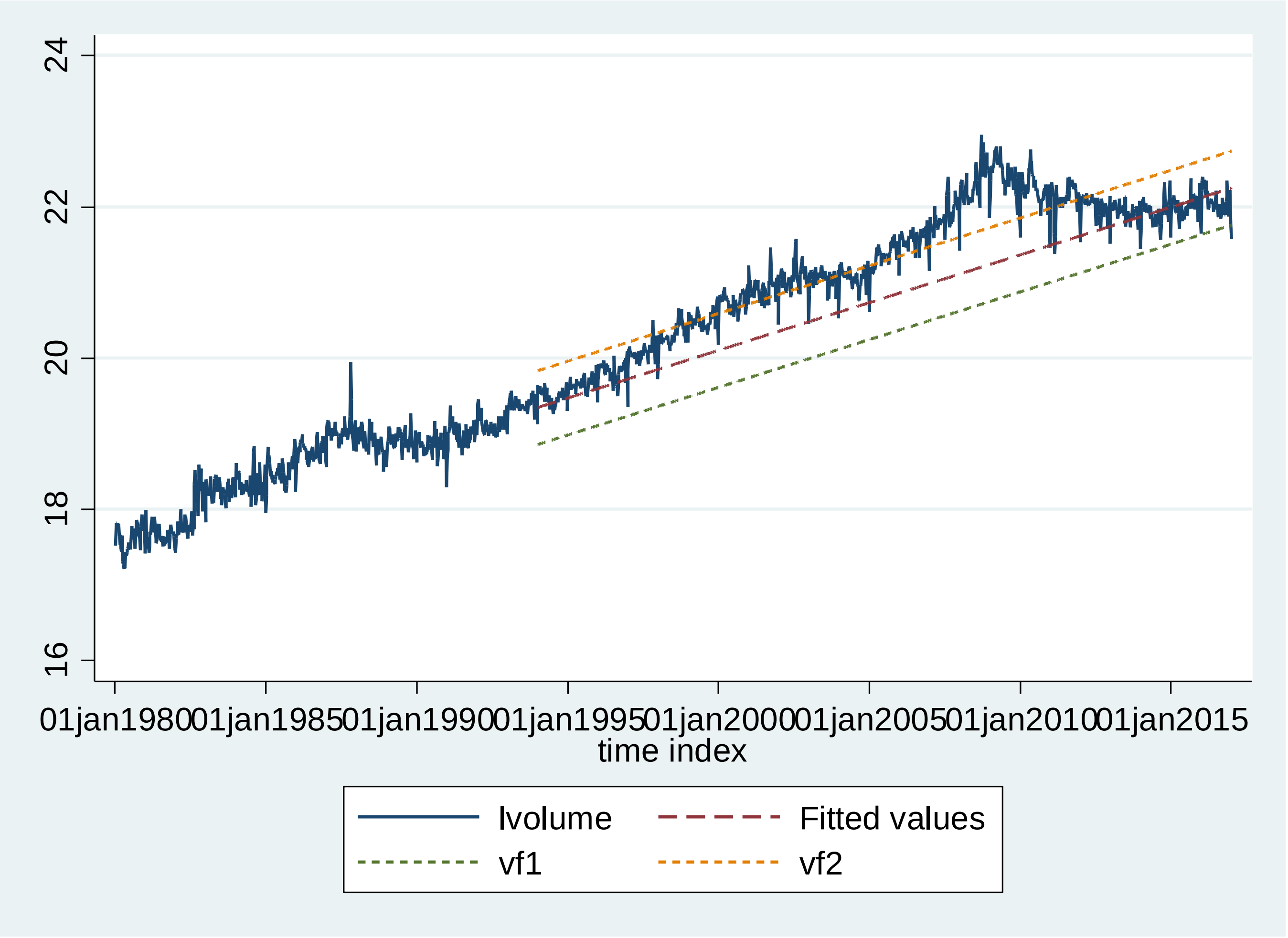
Residuals
1.5
1
.5
0
-.5
-1
01jan1950
01jan1960
01jan1970
01jan1980
01jan1990
time index

Forecast



22
20
18
16
14
01jan1950
01jan1960
01jan1970
01jan1980
01jan1990
01jan2000
01jan2010
01jan2020
time index
lvolume
Fitted values
vf1
vf2

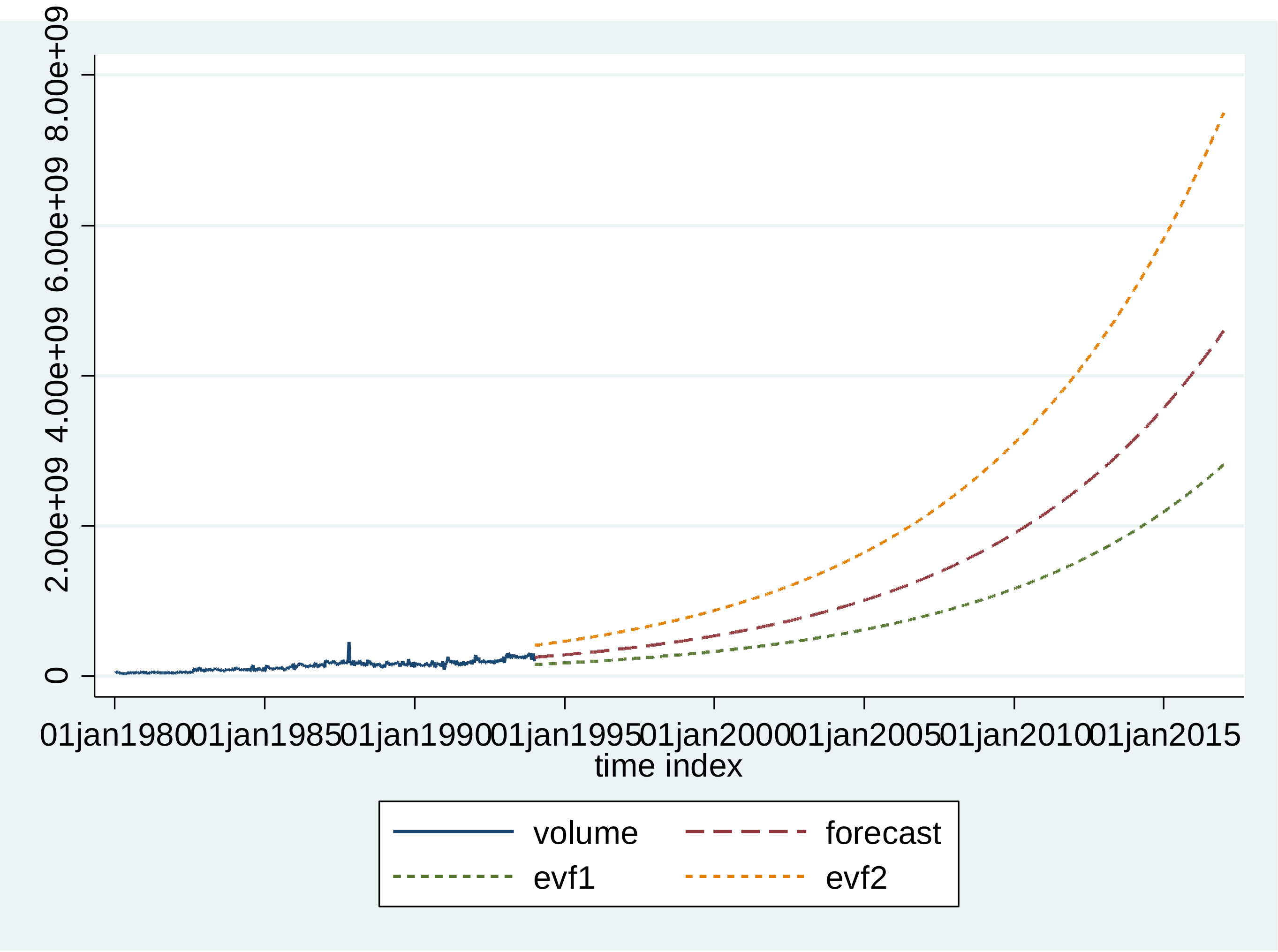
Actual Out-of-Sample



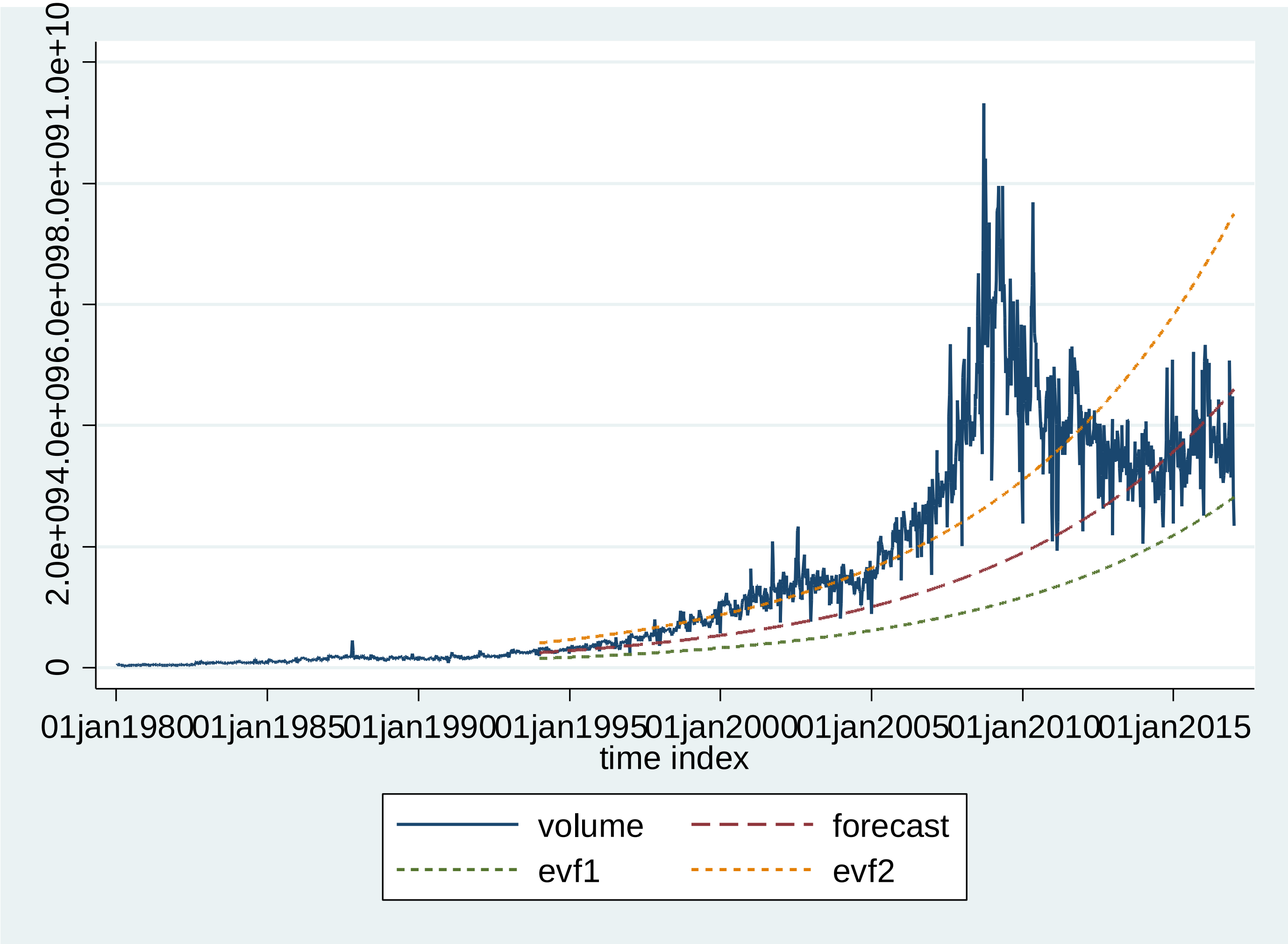
Forecasting Levels from a Forecast of Logs

- 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



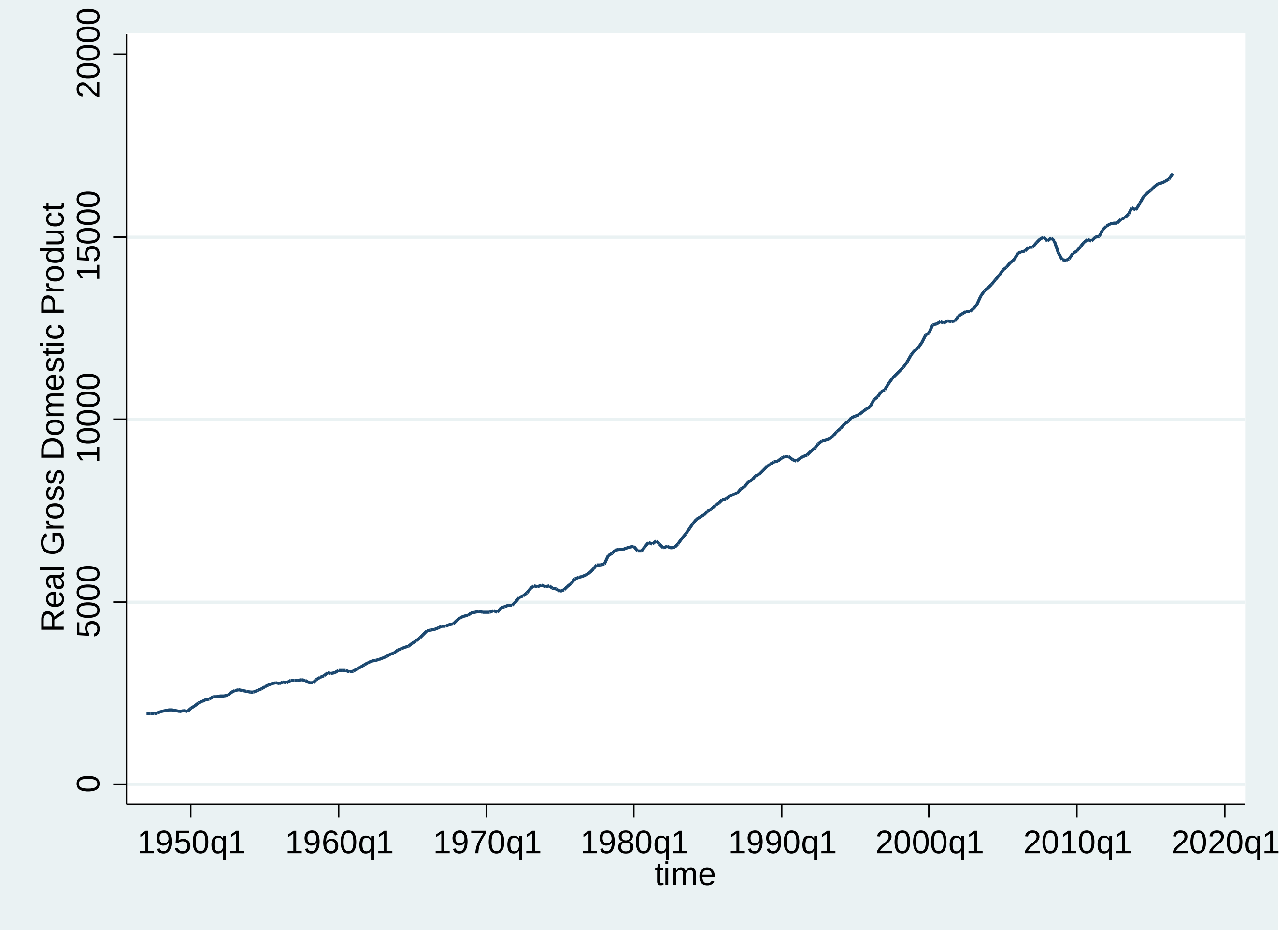
Actual Out-of-Sample



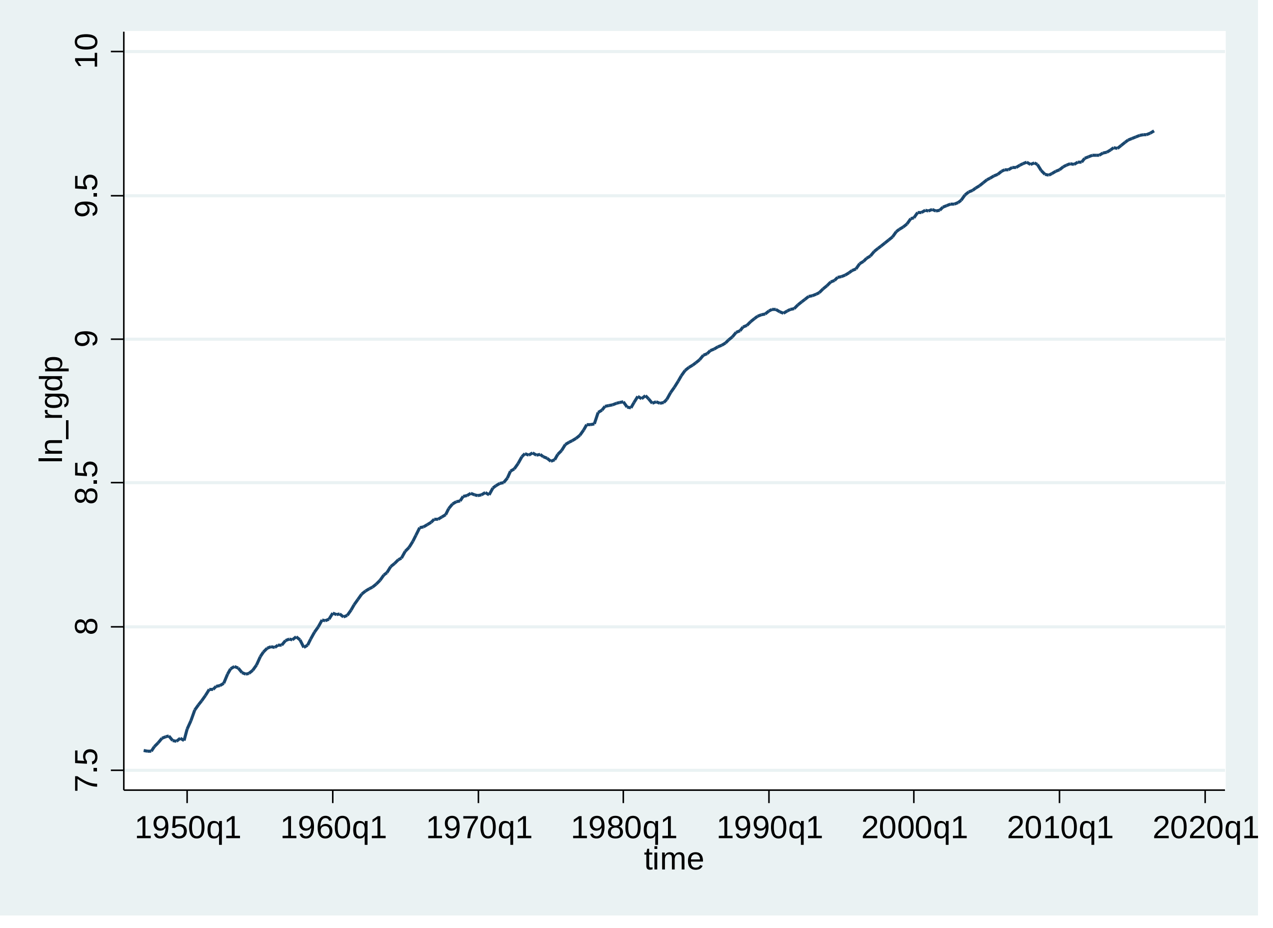
Stata do file

- use s&p
- gen lvolume = ln(volume)
- regress lvolume time if time<td(07jan1994)
- predict vf if time>=td(07jan1994)
- predict sv, stdf
- generate vf1 = vf - 1.645*sv
- generate vf2 = vf + 1.645*sv
- generate lvolume1=lvolume if time<td(07jan1994)
- label variable lvolume1 lvolume
- tsline lvolume1 vf vf1 vf2, lpattern (solid dash shortdash shortdash)
- tsline lvolume vf vf1 vf2 if time>=td(01jan1980), lpattern (solid dash shortdash shortdash)
- generate volume1=volume if time<td(07jan1994)
- label variable volume1 volume
- generate evf = exp(vf)
- label variable evf forecast
- generate evf1 = exp(vf1)
- generate evf2 = exp(vf2)
- tsline volume1 evf evf1 evf2 if time>=td(01jan1980), lpattern (solid dash shortdash shortdash)
- tsline volume evf evf1 evf2 if time>=td(01jan1980), lpattern (solid dash shortdash shortdash)

Example 3: Real GDP



Ln(Real GDP)

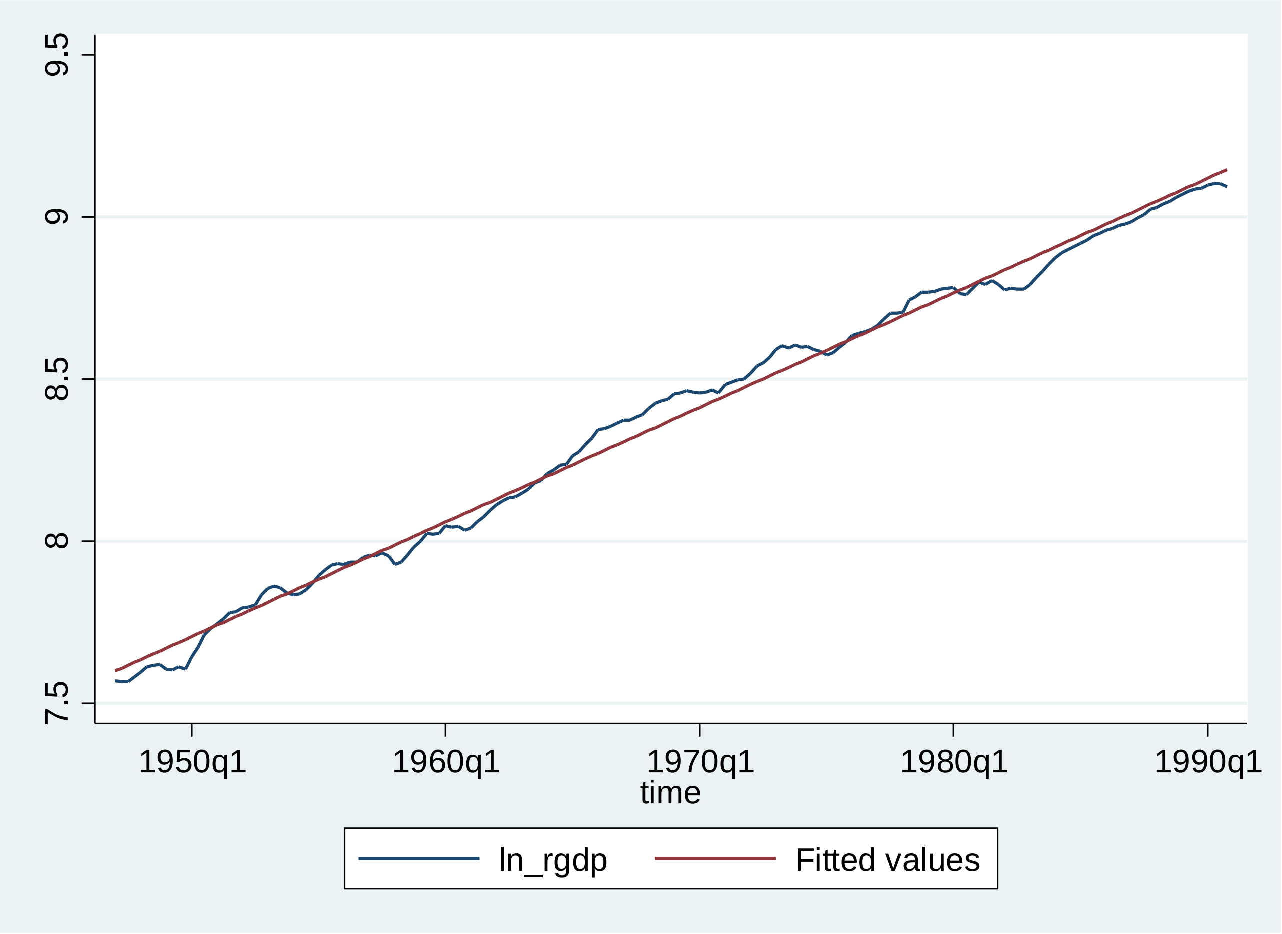


10
9.5
9
8.5
8
7.5
ln_rgdp
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
2020q1
time

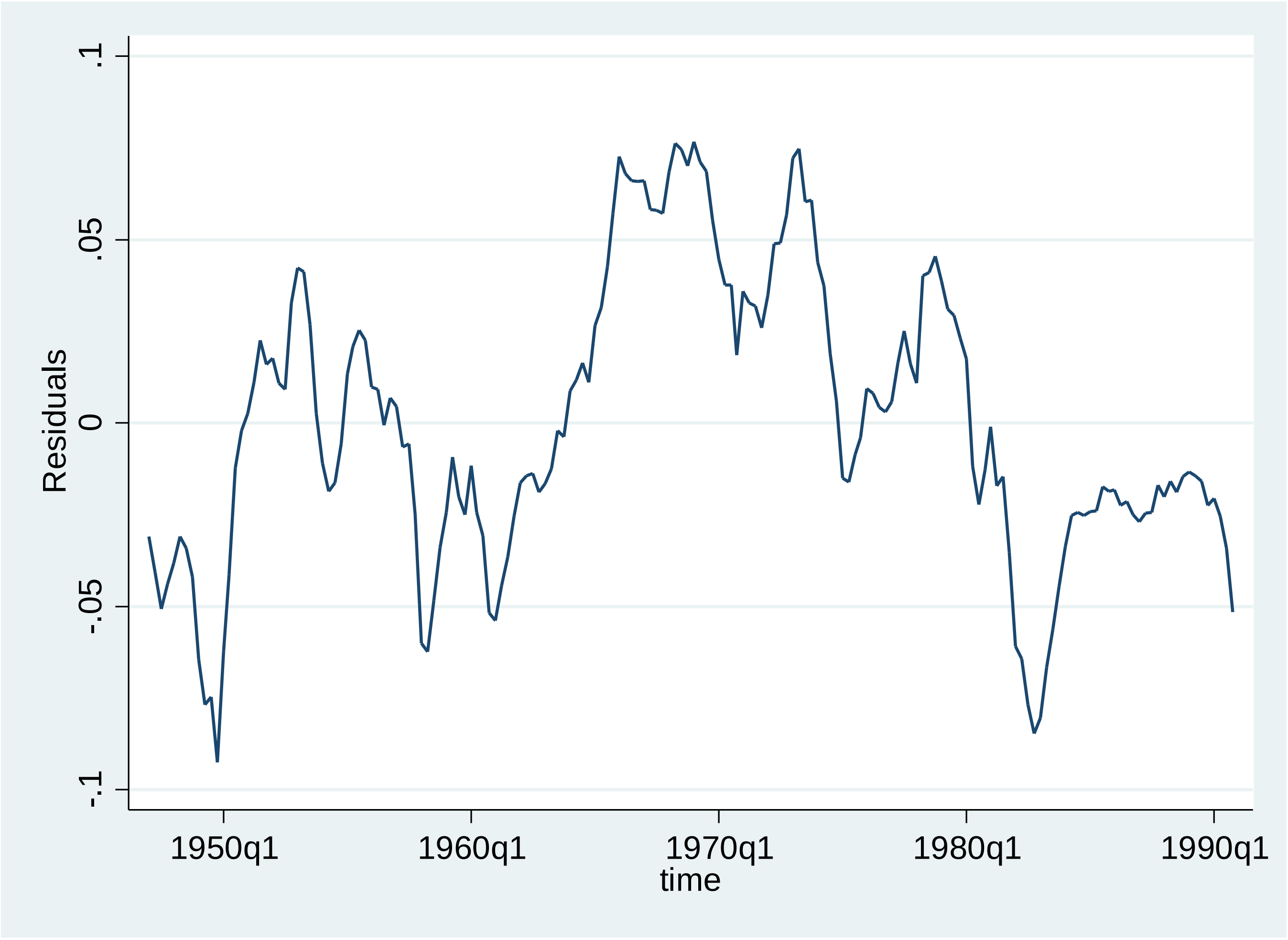
Estimation

```
. regress ln_rgdp time if time<=tq(1990q4)

      Source |       SS       df       MS              Number of obs =     176
-------------+------------------------------           F(  1,   174) =23264.59
       Model |  35.5325367     1  35.5325367           Prob > F      =  0.0000
    Residual |  .265754166   174  .001527323           R-squared     =  0.9926
-------------+------------------------------           Adj R-squared =  0.9925
       Total |  35.7982909   175  .204561662           Root MSE      =  .03908

------------------------------------------------------------------------------
     ln_rgdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0088439    .000058   152.53   0.000     .0087294    .0089583
       _cons |   8.058359   .0035937  2242.34   0.000     8.051266    8.065452
------------------------------------------------------------------------------
```

Fitted Trend

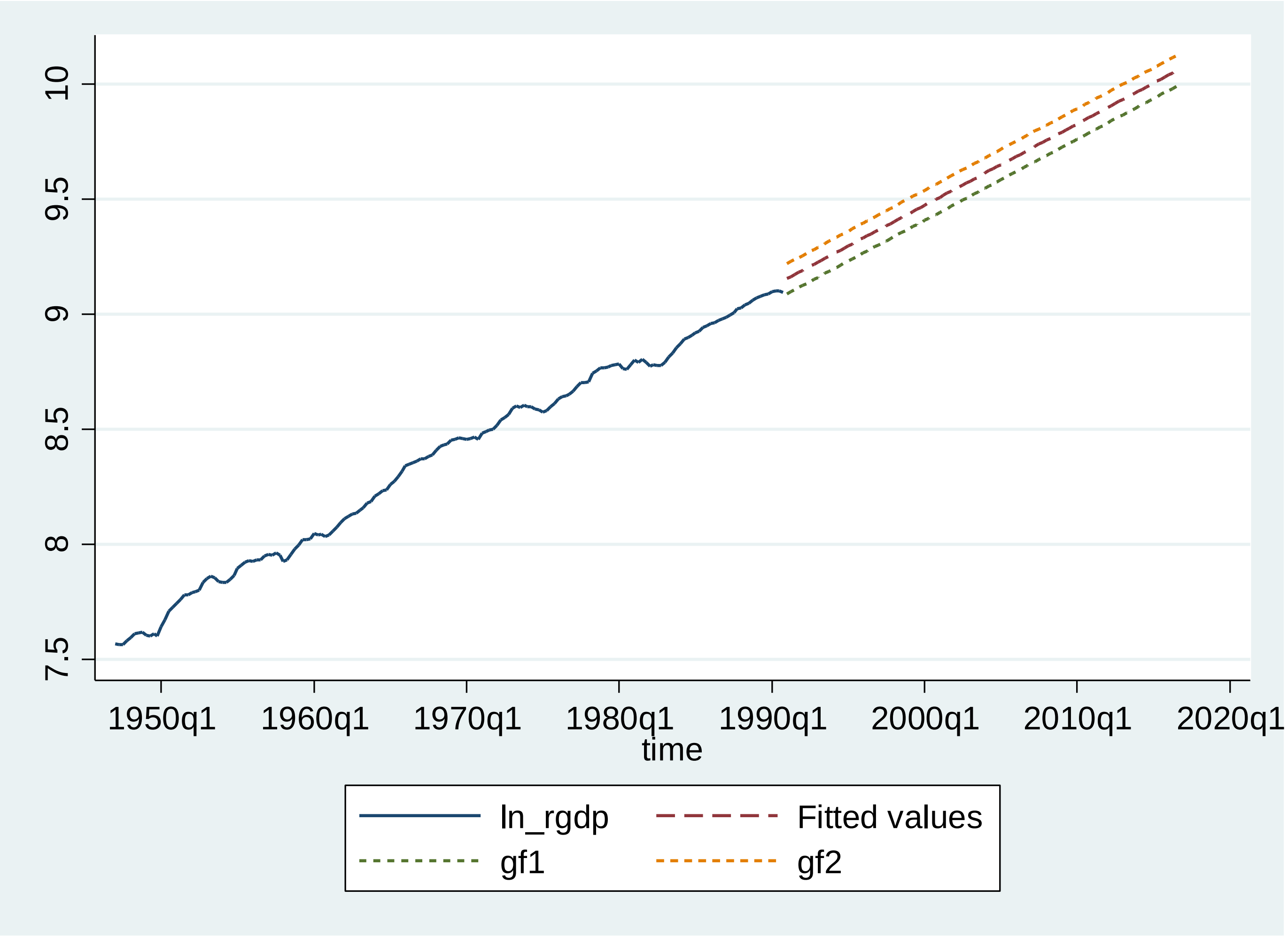


Residuals

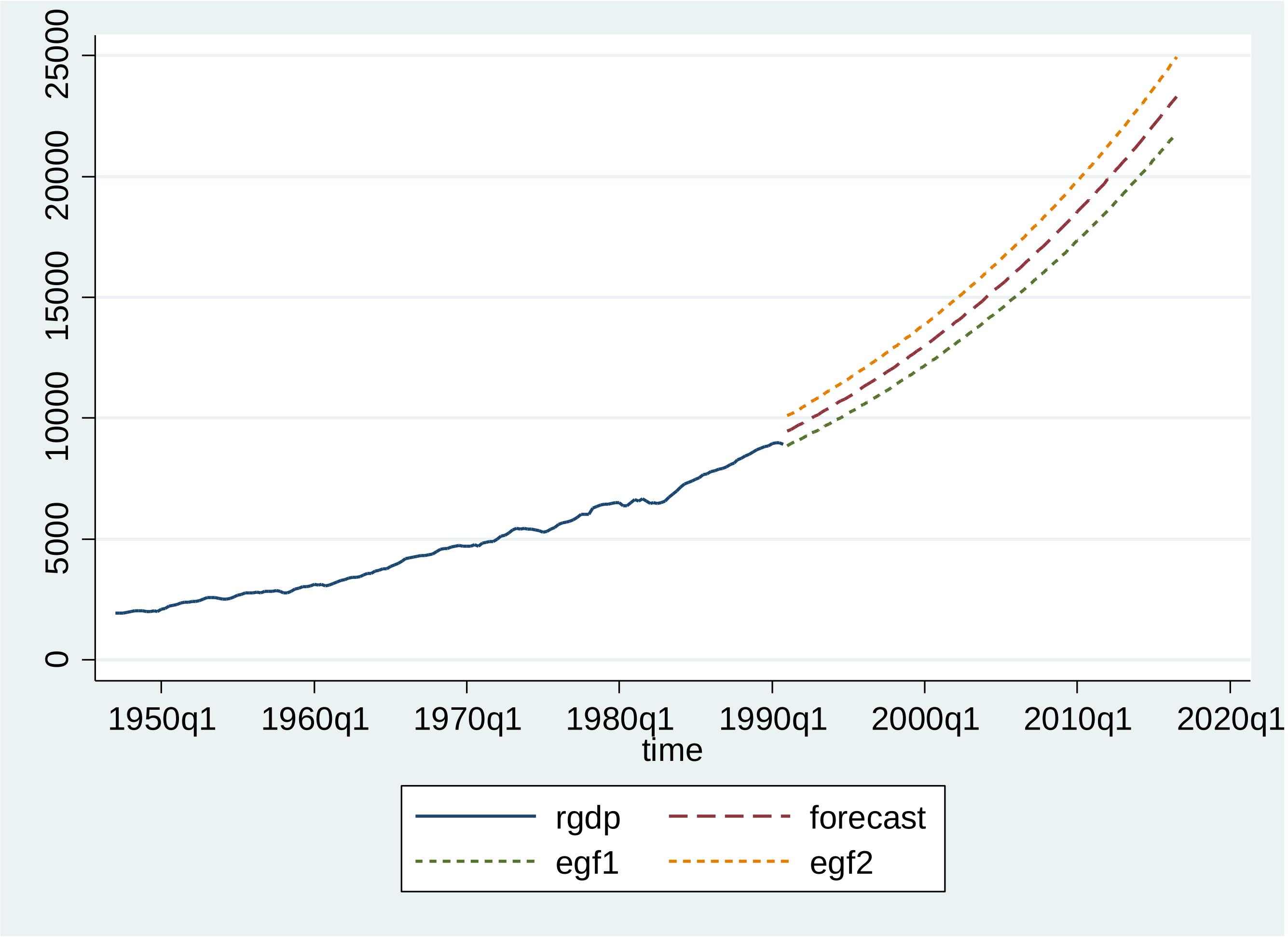


Residuals
.1
.05
0
-.05
-.1
1950q1
1960q1
1970q1
1980q1
1990q1
time

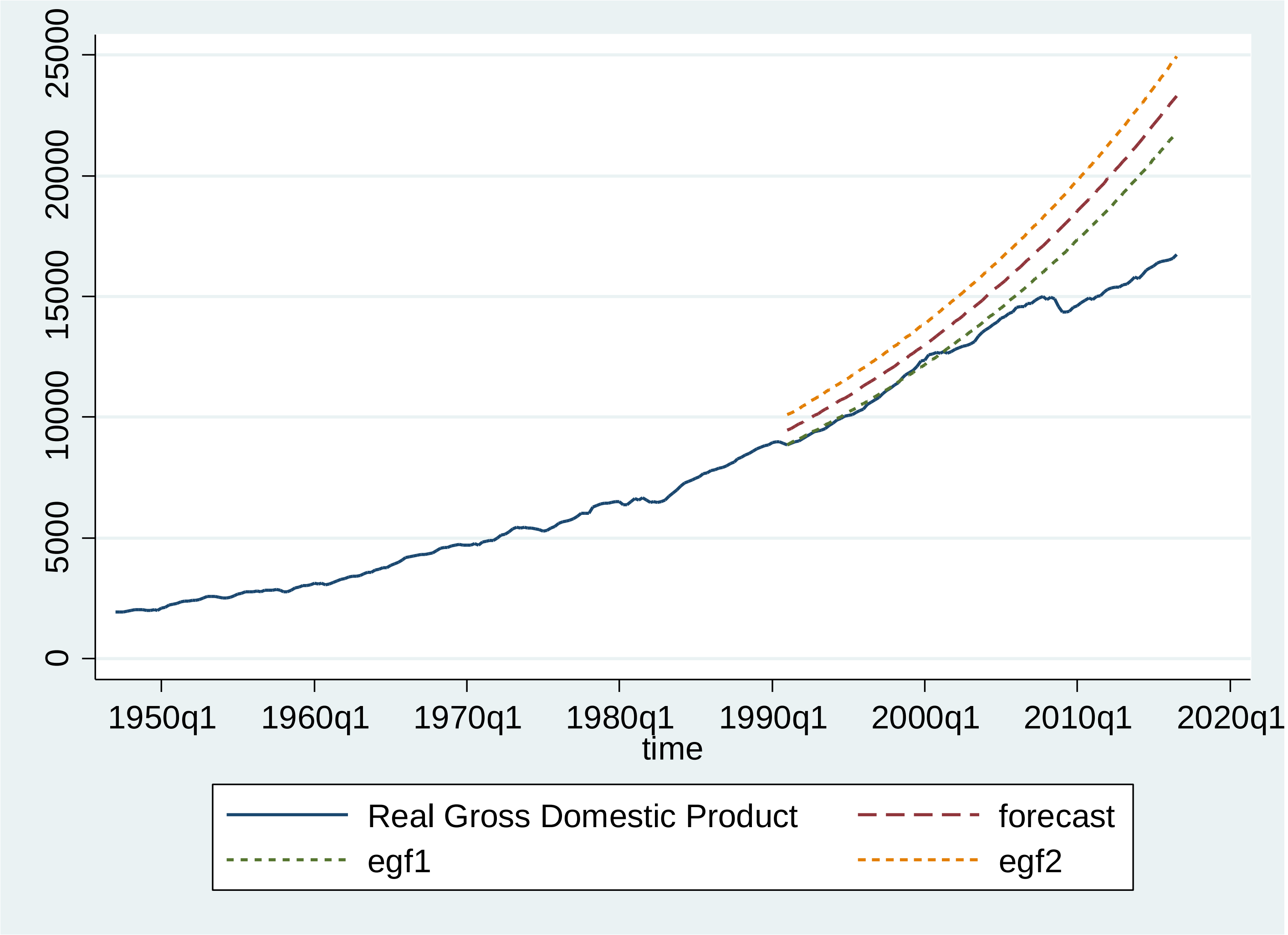
Forecast of ln(RGDP)



Forecast of RGDP (in levels)



Actual Out-of-Sample



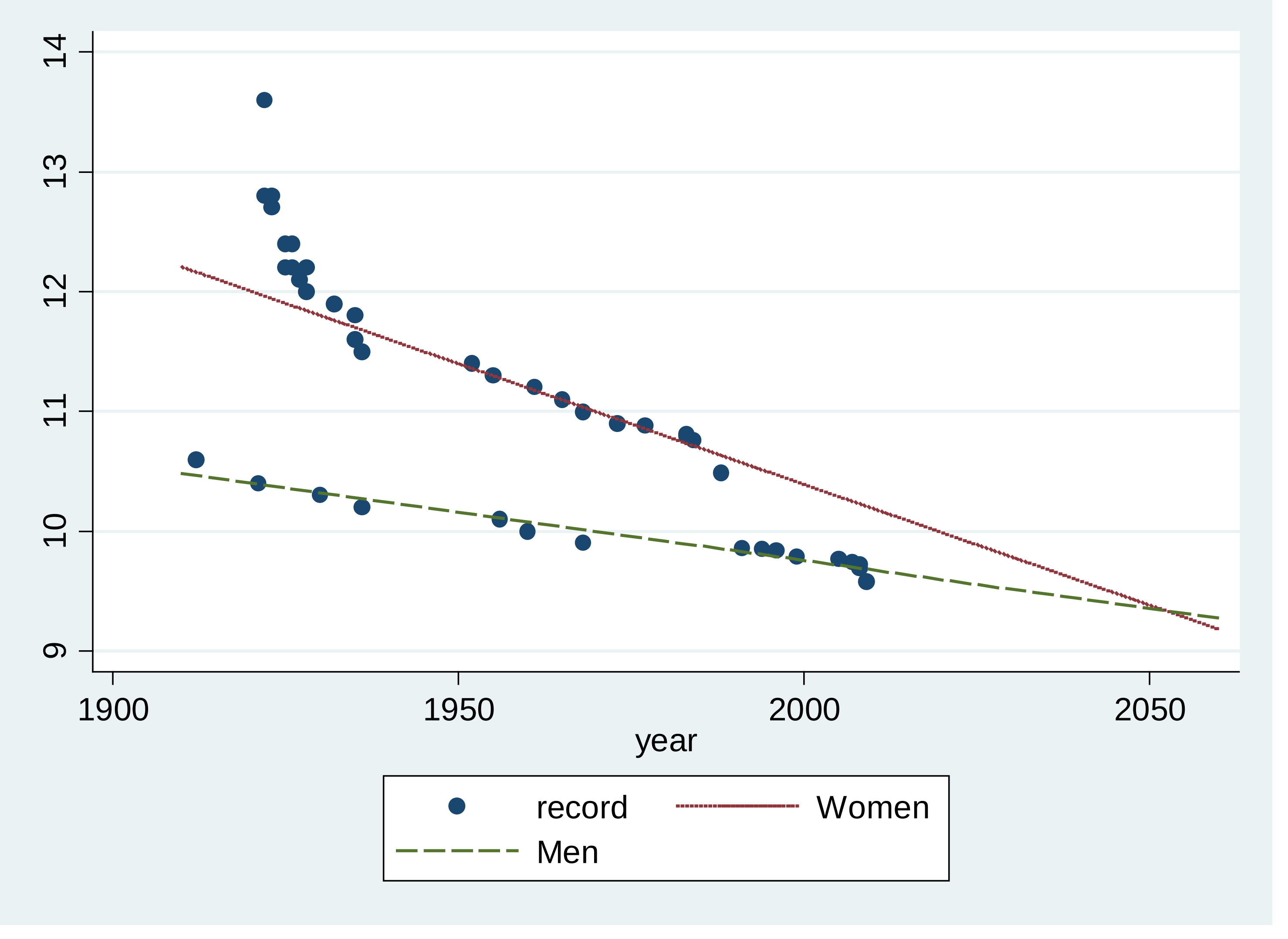
Problems with Pure Trend Forecasts

- **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).

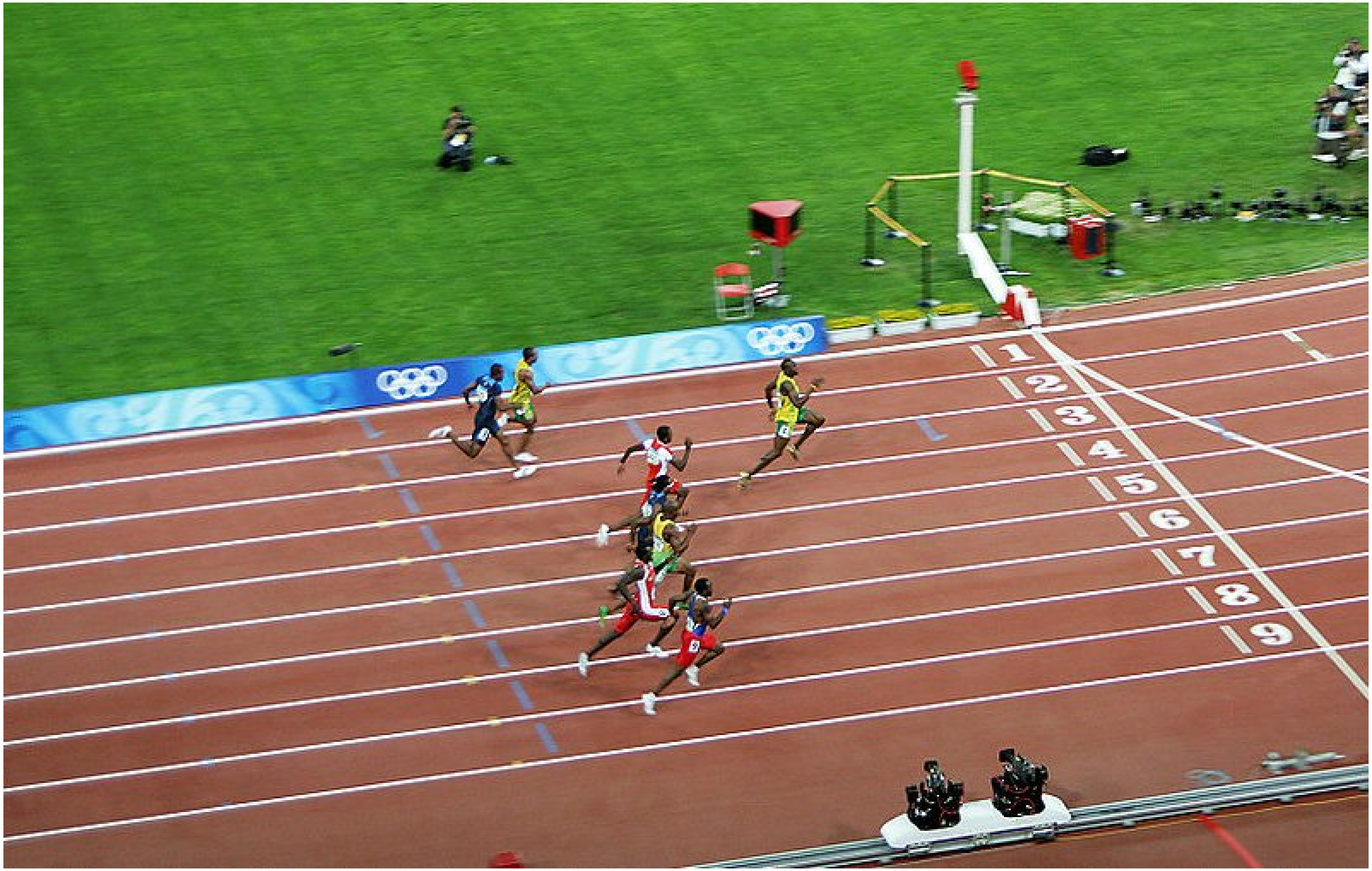
Trend Models

- 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



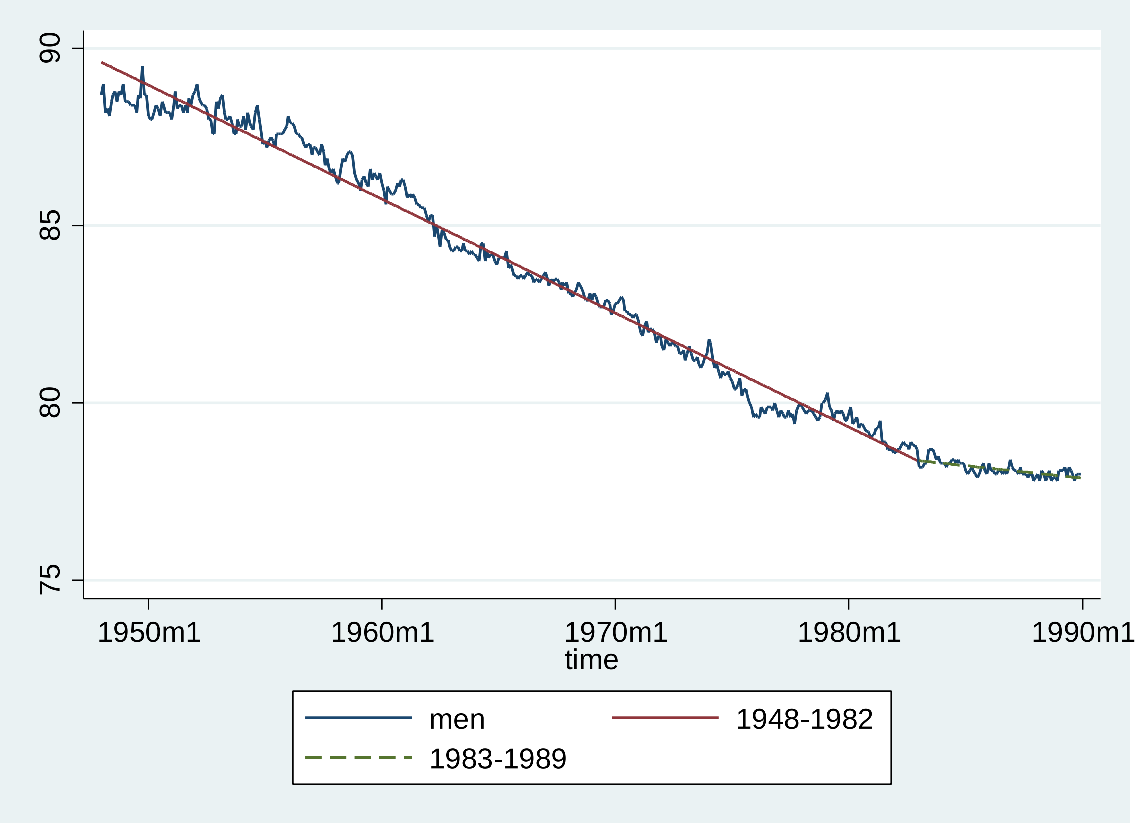
Usain Bolt



Changing Trends

- We have seen in some cases that it appears that the trend slope has changed at some point.
- This is a type of structural change, sometimes called a **changing trend** or **breaking trend**.
- We can model this using the interaction of dummy variables with the trend.

Labor Force Participation - Men



- Separate trends fit to 1948-1982 and 1983-1989

Sub-Sample Trend Lines

- If you fit a trend for observations before and after a breakdate τ, then for $t \leq \tau$

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

  and for $t > \tau$

$$T_t = \alpha_0 + \alpha_1 Time_t$$

- Notice that both the intercept and slope change

Estimation

- You can simply estimate on each sub-sample separately, and then forecast using the second set of estimates.
- Or, you can use dummy variable interactions.
- Define the dummy variable for observations after time τ

$$d_t = 1(t \geq \tau)$$

Dummy Equation

$$
\begin{aligned}
T_t &= (\beta_0 + \beta_1 Time_t) 1(t < \tau) + (\alpha_0 + \alpha_1 Time_t) 1(t \geq \tau) \\
&= (\beta_0 + \beta_1 Time_t) + ((\alpha_0 - \beta_0) + (\alpha_1 - \beta_1) Time_t) 1(t \geq \tau) \\
&= \beta_0 + \beta_1 Time_t + \beta_2 d_t + \beta_3 Time_t d_t
\end{aligned}
$$

where

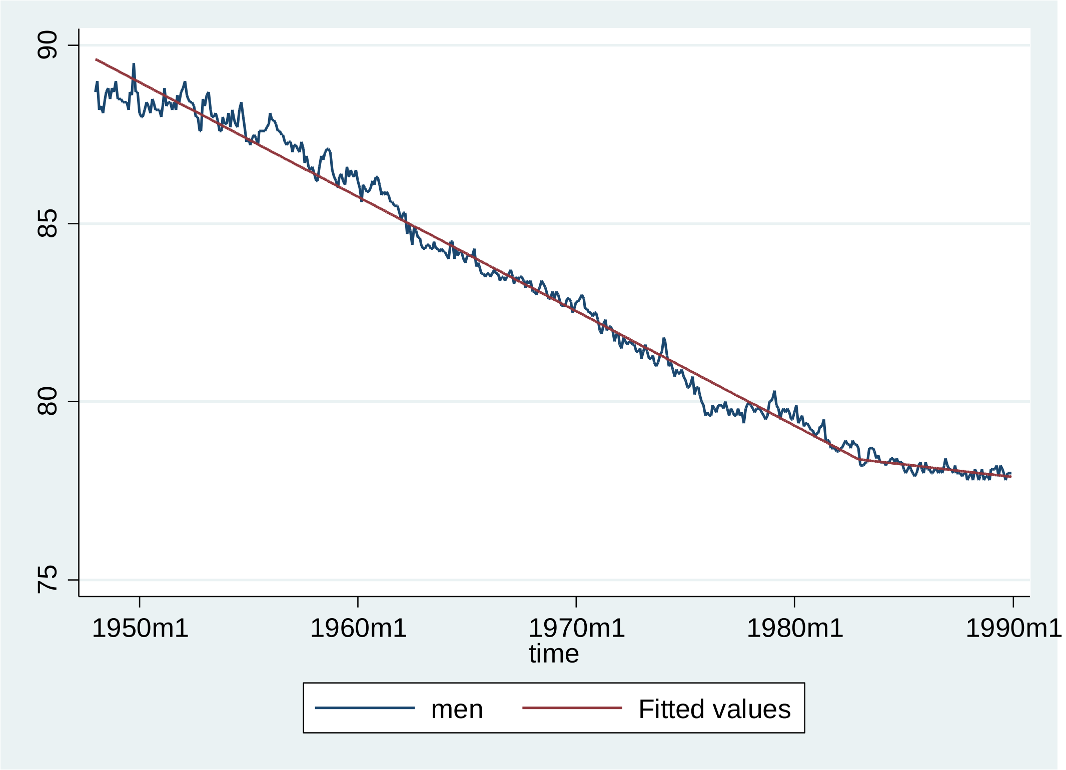
$$\beta_2 = \alpha_0 - \beta_0$$

$$\beta_3 = \alpha_1 - \beta_1$$

- This is a linear regression, with regressors $Time_t$, d_t and $Time_t d_t$

Estimation

- gen d = (time>=tm(1983m1))
- gen dtime = d*time
- regress men time d time if time<tm(1990m1)
- predict p3 if time<tm(1990m1)
- tsline men p3 if time<tm(1990m1)



Discontinuity

- One problem with this method is that the estimated trend function can be discontinuous
  - At the breakdate τ there might be a jump in the trend function
  - This might not be sensible
  - We may wish to impose continuity
- In the model, this requires

$$\beta_0 + \beta_1 \tau = \alpha_0 + \alpha_1 \tau$$

or

$$\beta_2 + \beta_3 \tau = 0$$

Continuous Break

- You can impose a continuous trend by using a technique known as a **spline**

$$
\begin{aligned}
T_t &= \beta_0 + \beta_1 Time_t + \beta_2 (Time_t - \tau) 1(t \geq \tau) \\
&= \beta_0 + \beta_1 Time_t + \beta_2 Time_t^*
\end{aligned}
$$

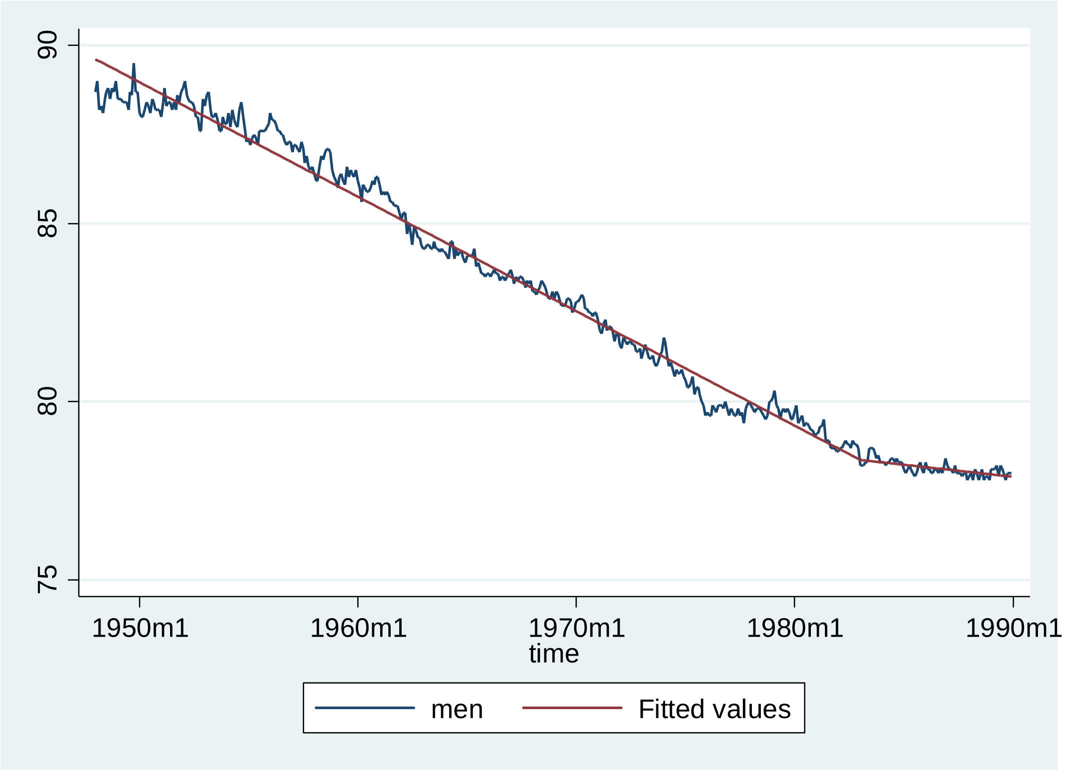
where

$$Time_t^* = (Time_t - \tau) 1(t \geq \tau)$$

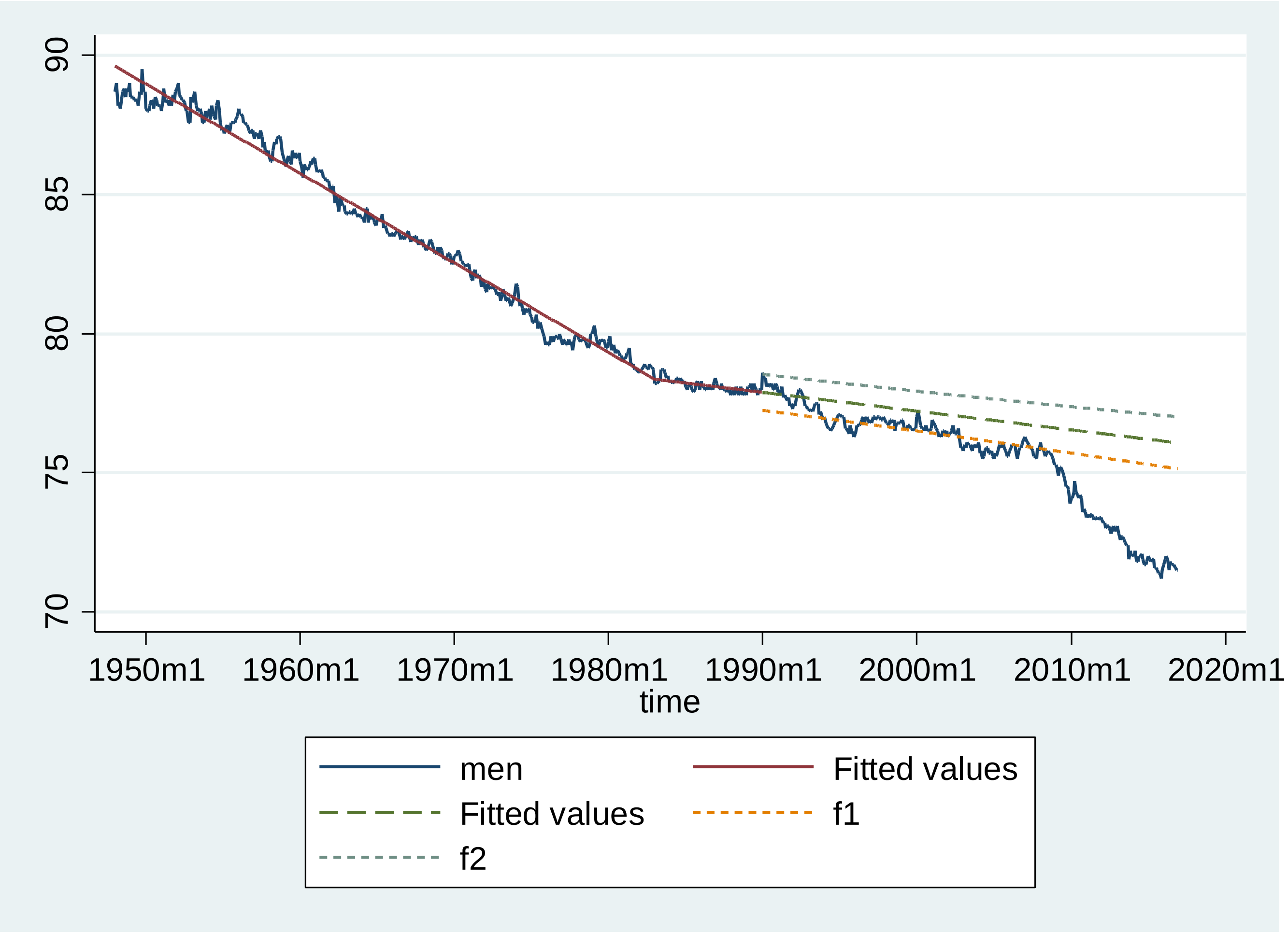
- The variable $Time_t^*$ is 0 before the breakdate, and is a smoothly increasing trend afterwards.

Fitted Continuous Breaking Trend

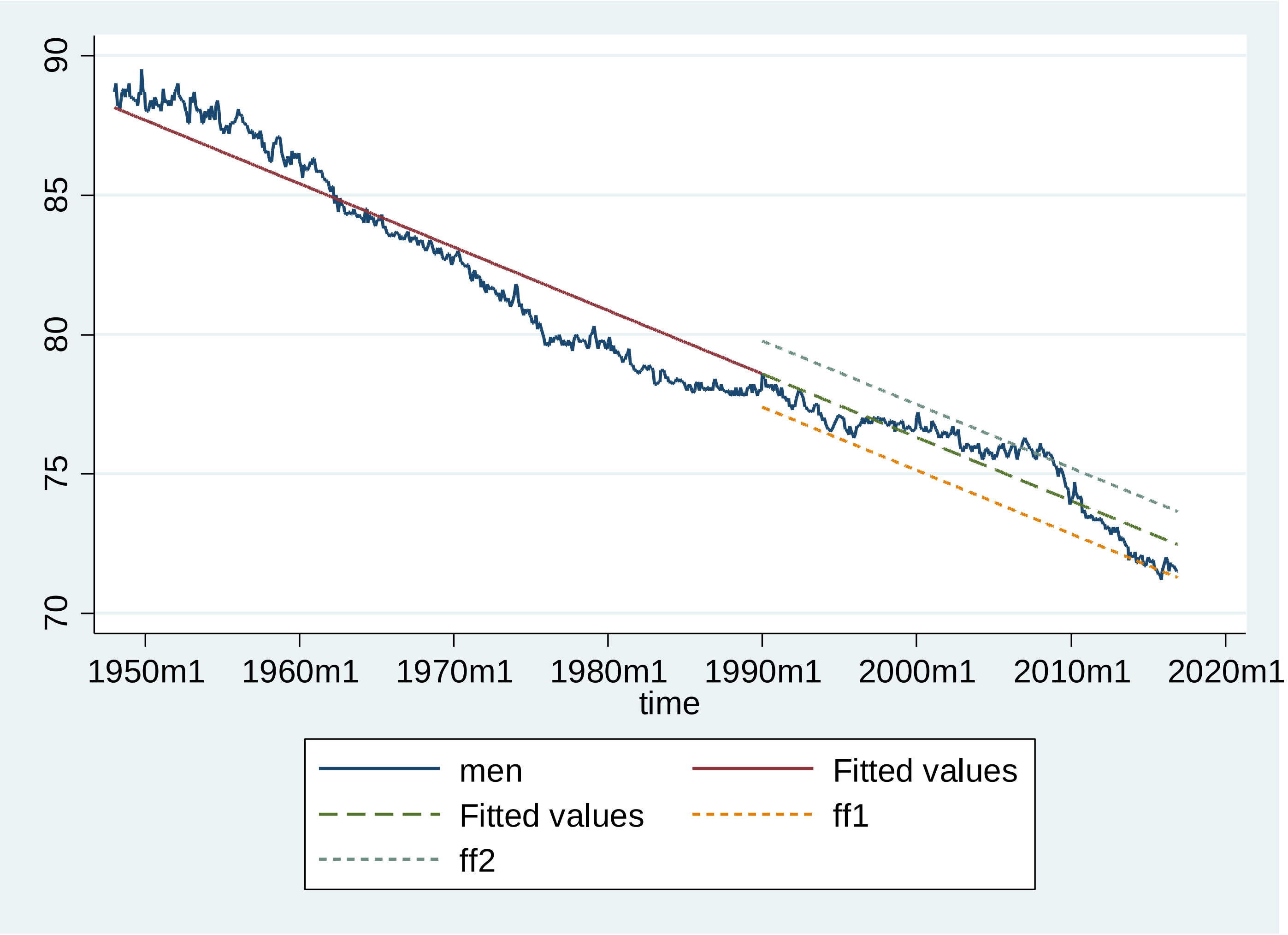
- `gen stime = (time-tm(1983m1))*d`
- `regress men time stime if time<tm(1990m1)`
- `predict p4 if time<tm(1990m1)`
- `tsline men p4 if time<tm(1990m1)`



Continuous Breaking Trend Forecast



Contrast with Linear Trend Forecast



Real GDP

```
. generate tstar=(t-tq(1974q1))*(t>=tq(1974q1))

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

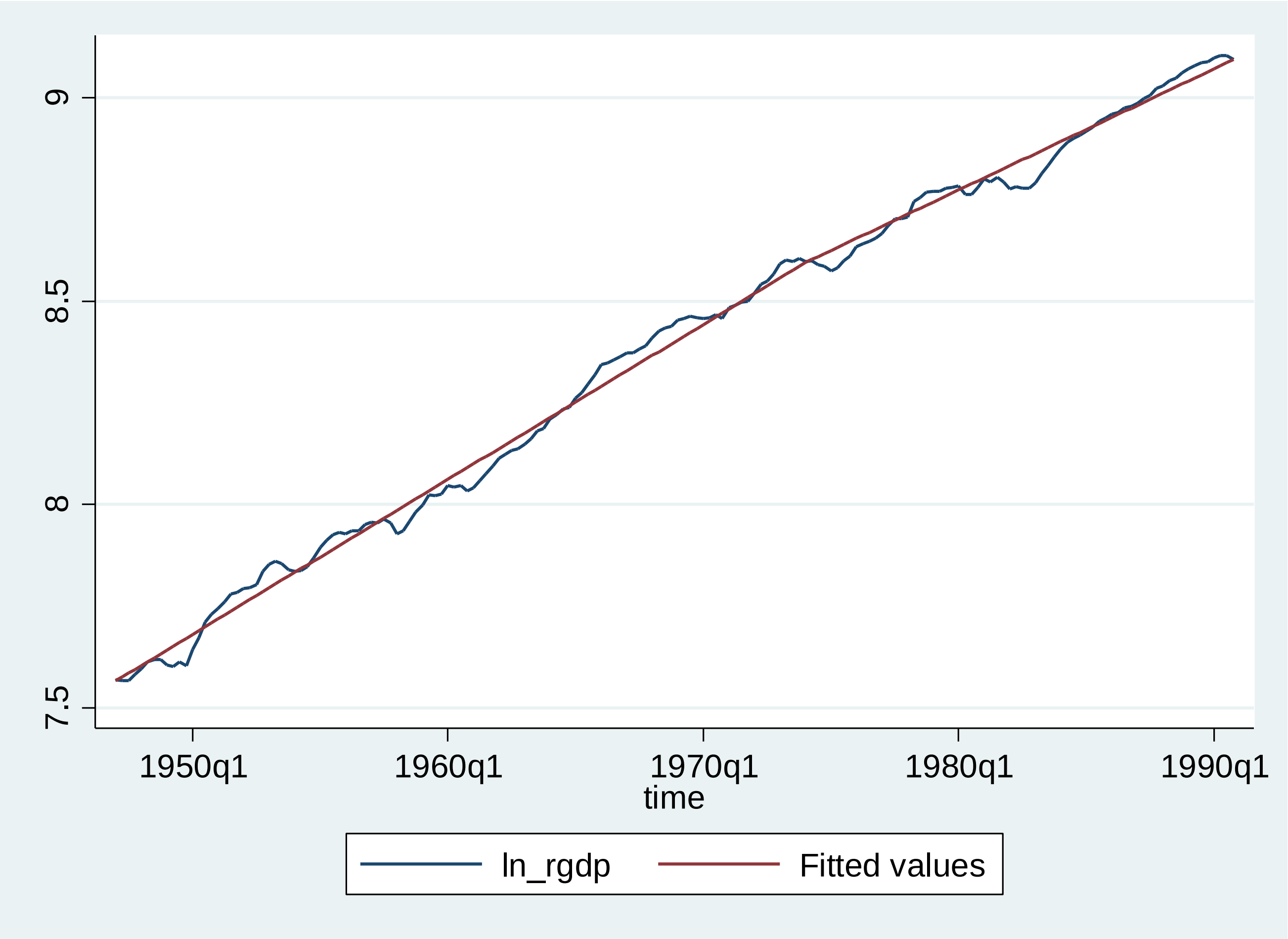
Source	SS	df	MS
Model	34.7825634	2	17.3912817
Residual	.155349843	173	.000897976
Total	34.9379132	175	.199645218

Number of obs	= 176
F(2, 173)	=19367.20
Prob > F	= 0.0000
R-squared	= 0.9956
Adj R-squared	= 0.9955
Root MSE	= .02997

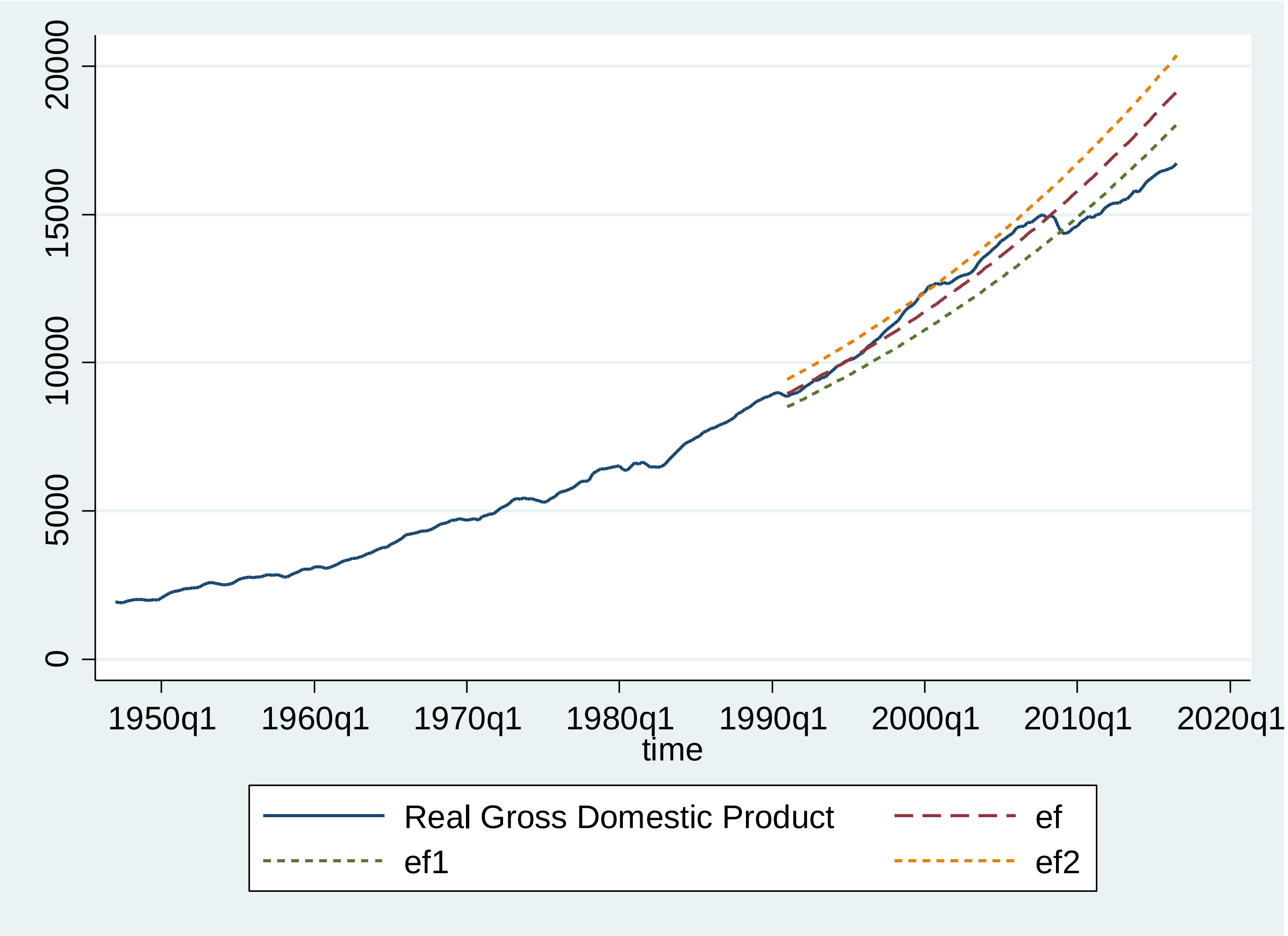
y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
t	.0094132	.0000775	121.43	0.000	.0092602	.0095662
tstar	-.0020582	.0001934	-10.64	0.000	-.0024398	-.0016766
_cons	7.96762	.0027667	2879.80	0.000	7.962159	7.973081

- Break in 1974q1

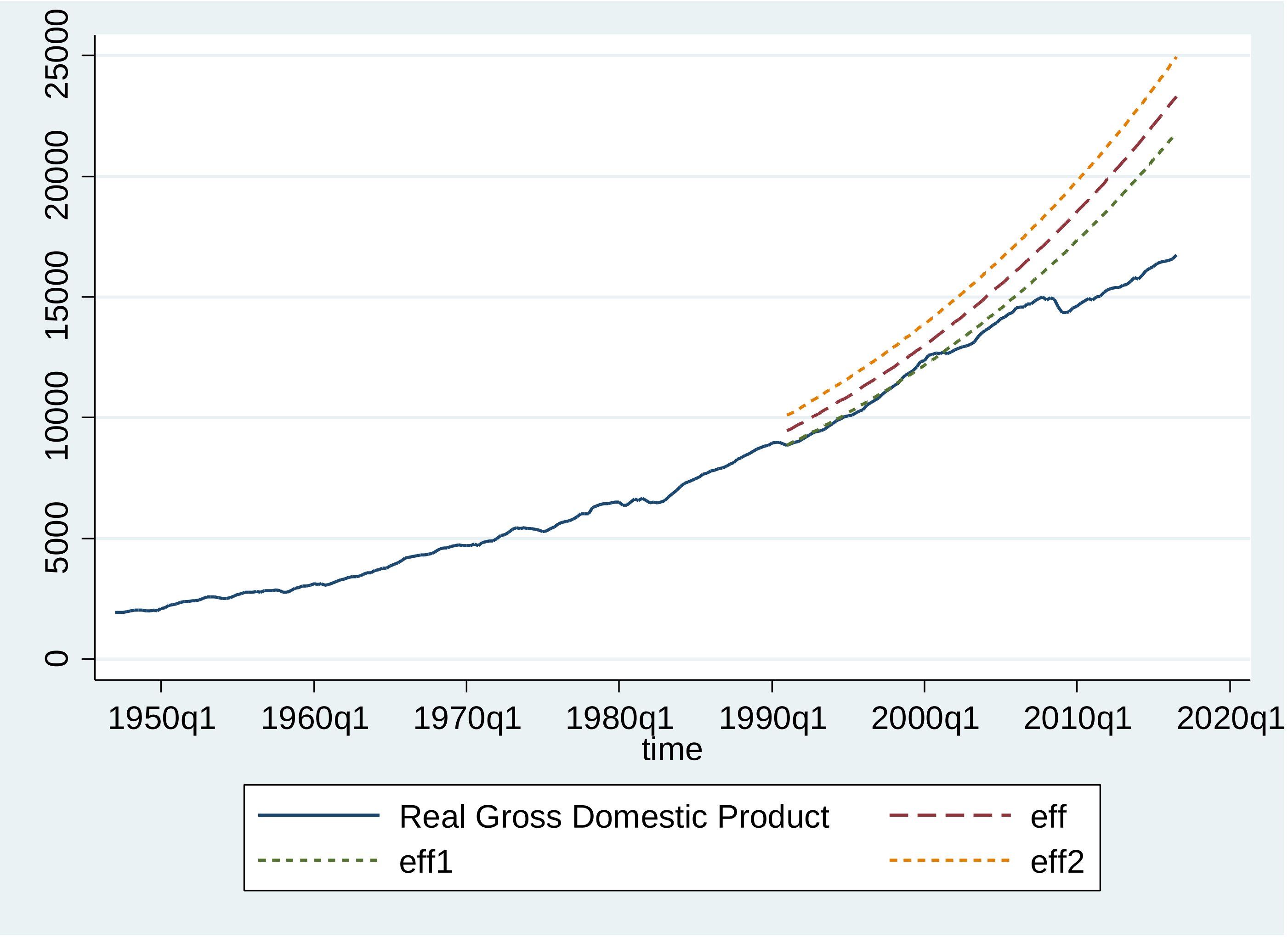
Real GDP - fitted



Forecast – Breaking Trend Model



Contrast – Forecast from Linear Trend



How to pick Breaks/Breakdates

- With caution, and skeptically
- Always have plenty of data (at least 10 years) after the breakdate
- Look for economic explanations
- Formally, the breakdate can be selected by minimizing the sum of squared errors

Assignments

- Read Diebold through Chapter 5
- Problem Set # 3
  - Due Tuesday (2/7)
- Read Chapter 3 from *The Signal and the Noise*
  - Reading Reflection
  - Due Thursday (2/9)