

GDP Advance Estimate, 2016Q4

- Friday, Jan 27
- “Real gross domestic product (GDP) increased at an annual rate of 1.9 percent in the fourth quarter of 2016.”
- “The deceleration in real GDP in the fourth quarter reflected a downturn in exports, an acceleration in imports, a deceleration in PCE, and a downturn in federal government spending ...”
- Real GDP increased 1.6% in 2016

Estimation Uncertainty

- The sample mean

$$b_0 = \frac{1}{T} \sum_{t=1}^T y_{t+h}$$

is an estimate of $\beta_0 = E(y_{t+h})$

- The estimation error is

$$\begin{aligned} b_0 - \beta_0 &= \frac{1}{T} \sum_{t=1}^T y_{t+h} - \beta_0 \\ &= \frac{1}{T} \sum_{t=1}^T (y_{t+h} - \beta_0) \\ &= \frac{1}{T} \sum_{t=1}^T e_t \end{aligned}$$

Estimation Variance

- Under classical conditions,

$$\text{var}(b_0) = \frac{\sigma^2}{T}$$

where $\sigma^2 = \text{var}(e_t)$

- The standard error for b_0 is an estimate of the standard deviation

$$sd(b_0) = \sqrt{\frac{\hat{\sigma}^2}{T}}$$

Forecast Variance

- When the sample mean b_0 is used as the forecast for y_{T+h} then the prediction error is

$$y_{T+h} - b_0 = e_{T+h} + \beta_0 - b_0$$

which is the sum of the forecast error e_{T+h} and the estimation uncertainty $\beta_0 - b_0$.

- The forecast variance is

$$\begin{aligned}\text{var}(y_{T+h} - b_0) &= \text{var}(e_{T+h}) + \text{var}(\beta_0 - b_0) \\ &= \sigma^2 + \frac{\sigma^2}{T} \\ &= \left(1 + \frac{1}{T}\right)\sigma^2\end{aligned}$$

Standard Deviation of Forecast

- The standard deviation of the forecast is the estimate

$$s_{T+h} = \sqrt{\left(1 + \frac{1}{T}\right) \hat{\sigma}^2}$$

- This is slightly larger than the regression standard deviation $\hat{\sigma}$
- Calculated in STATA after a regression using the `stdf` option to the `predict` command:
 - `predict s, stdf`
- This creates variable “s”

Normal Forecast Intervals

- Let \hat{y}_{T+h} be a forecast for y_{T+h}
- The prediction error is $y_{T+h} - \hat{y}_{T+h}$
- Let s_{T+h} be the st. deviation of the forecast
- If the prediction errors are normally distributed, the $(1-\alpha)\%$ forecast interval endpoints are

$$L_{T+h} = \hat{y}_{T+h} + s_{T+h} z_{\alpha/2}$$

$$U_{T+h} = \hat{y}_{T+h} + s_{T+h} z_{1-\alpha/2}$$

where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the normal distribution

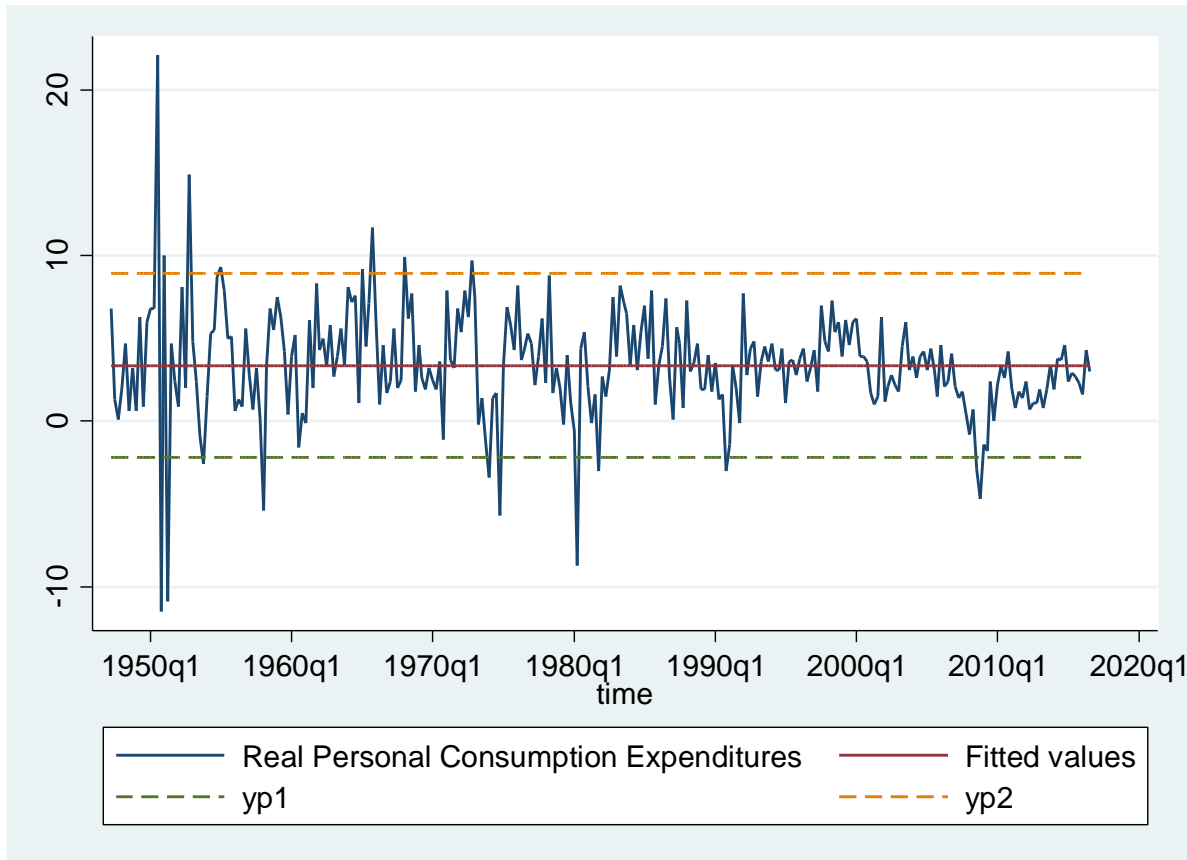
- *e.g.* $\hat{y}_{T+h} \pm 1.64 s_{T+h}$ for a 90% interval

Example Revisited: PCE

- U.S. Real Personal Consumption Expenditure
- From National Income Accounts
- Quarterly, percentage change from previous period, annualized
- 1947q2 to 2016q4
 - use `realgdpgrowth`
 - regress `pce`
 - predict `yp`
 - predict `s, stdf`
 - gen `yp1 = yp - 1.645*s`
 - gen `gp2 = yp + 1.645*s`
 - `tsline pce yp yp1 yp2, lpattern (solid solid dash dash)`

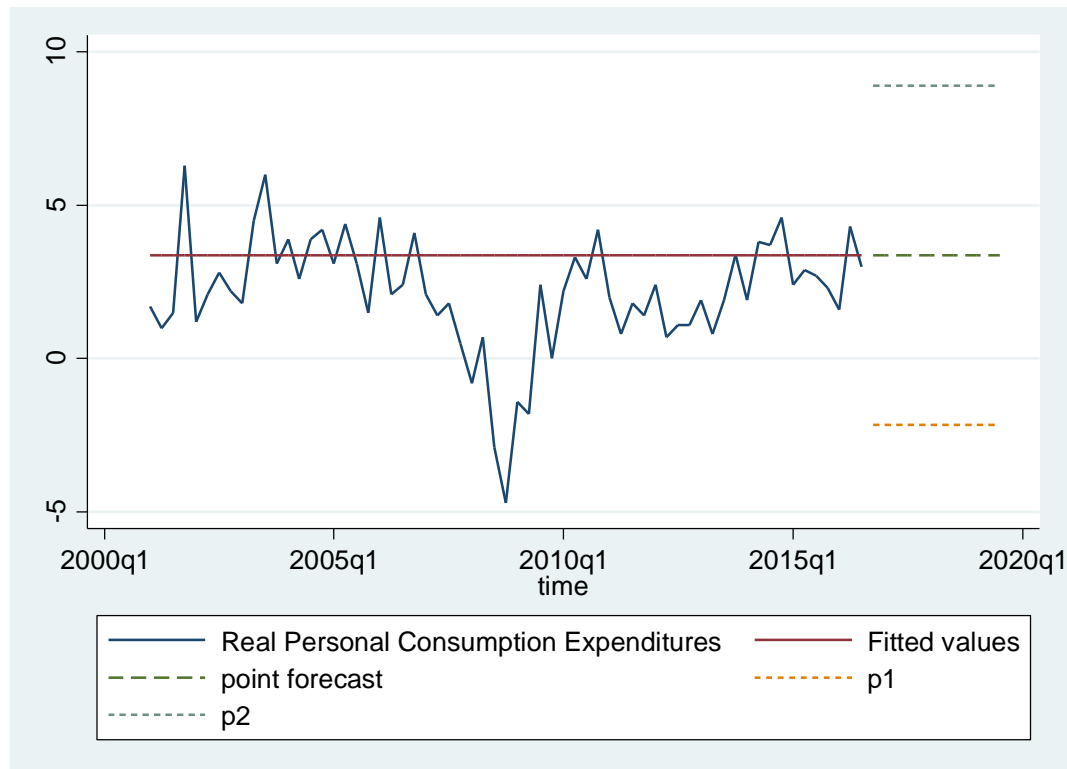
In-Sample Prediction Intervals

- PCE, fitted values, and intervals



Out of Sample

- `tsappend, add(12)`
- `predict p if time>tq(2016q4)`
- `predict s, stdf`
- `gen p1 = p - 1.645*s`
- `gen p2 = p + 1.645*s`
- `tsline pce yp p p1 p2 if time>tq(2000q4),
lpattern (solid solid dash shortdash shortdash)`

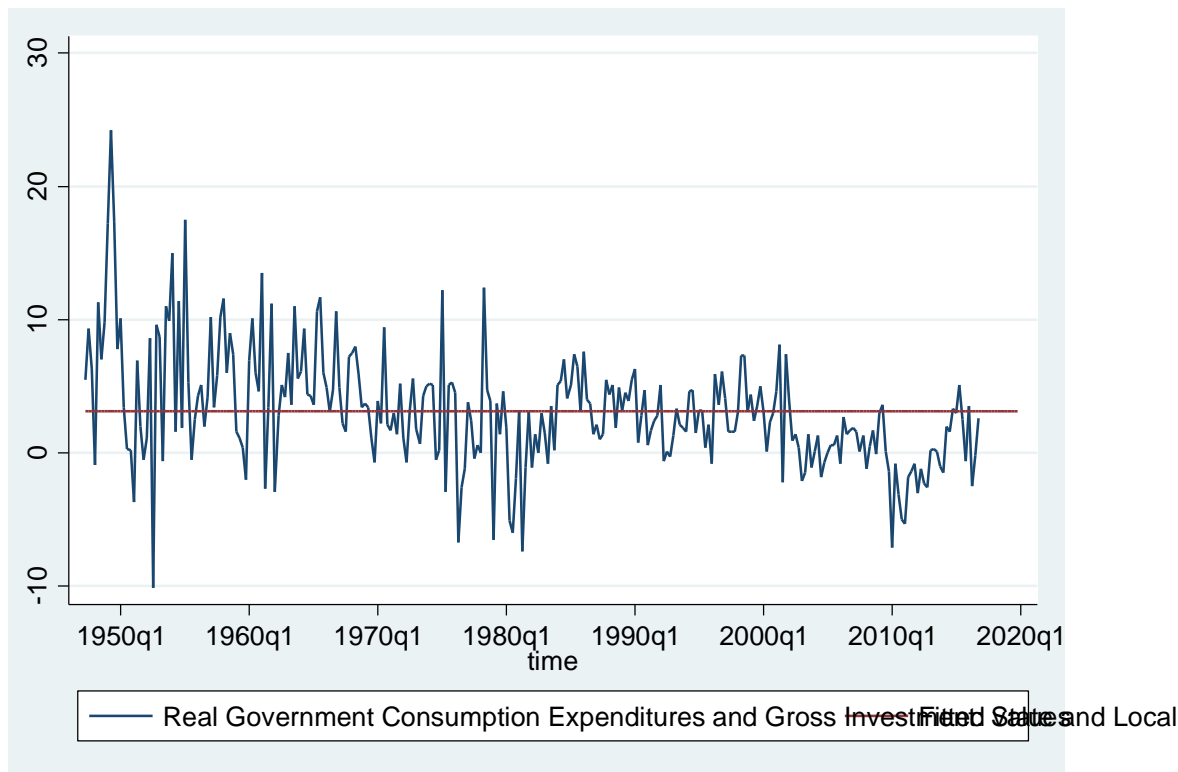


Mean Shifts

- Sometimes the mean of a series changes over time
- It can drift slowly, or change quickly
 - Possibly due to a policy change
- In this case, forecasting based on a constant mean model can be misleading

State and Local Government Spending Percentage Growth Rate (Quarterly)

- Average for 1947-2016: 3.1%
- But this has not been typical in recent years.



Alternatives

- Subsample estimation
 - Estimate the mean on subsamples
 - Forecasts are based on the most recent
- Dummy Variable formulation

$$E(y_{t+h} | \Omega_t) = \beta_0 + \beta_1 d_t$$
$$d_t = 1(t \geq \tau)$$

- τ is the breakdate
 - The date when the mean shifts
 - The coefficient β_0 is the mean before $t=\tau$
 - The coefficient β_1 is the shift at $t=\tau$
 - The sum $\beta_0+\beta_1$ is the mean after $t=\tau$

Forecast

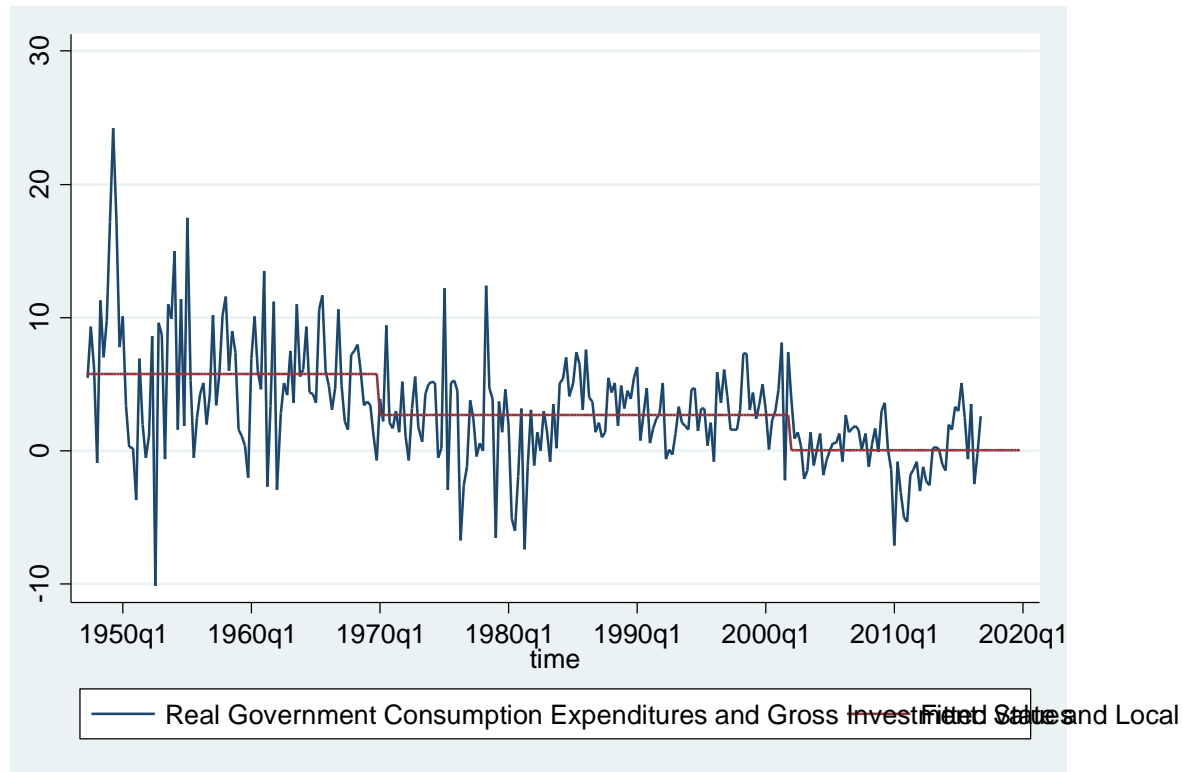
- Linear Regression y_{t+h} on d_t
- Example: State and Local Government Percentage Growth
 - Mean breaks in 1970q1 and 2002q1
 - generate d1=(time>=tq(1970q1))
 - generate d2=(time>=tq(2002q1))
 - regress state d1 d2
 - predict db
 - tsline state sb

Source	SS	df	MS	Number of obs	=	279
Model	1212.06992	2	606.034962	F(2, 276)	=	40.89
Residual	4090.17975	276	14.8194919	Prob > F	=	0.0000
				R-squared	=	0.2286
				Adj R-squared	=	0.2230
Total	5302.24968	278	19.0728406	Root MSE	=	3.8496

state	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d1	-3.065908	.527853	-5.81	0.000	-4.105038	-2.026779
d2	-2.614531	.6023029	-4.34	0.000	-3.800223	-1.42884
_cons	5.76044	.4035487	14.27	0.000	4.966015	6.554864

Fitted

- Out-of-sample forecast falls from 3.2% to 0%!



Should you use Mean Shifts?

- Only after great hesitation and consideration.
- Should use shifts and breaks reluctantly and with care.
- Do you have a model or explanation?
- What is the forecasting power of a mean shift?
 - If they have happened in the past, will there be more in the future?
- Yet, if there has been an obvious shift, a simple constant mean model will forecast terribly.

How to Select Breakdates

- Judgmental
 - Dates of known policy shifts
 - Important events
 - Economic crises
- Informal data-based
 - Visual inspection
- Formal data-based
 - Estimate regression for many possible breakdates
 - Select one which minimizes sum of squared error
 - This is the least-squares breakdate estimator

Trend Models

- A trend model is

$$T_t = g(\textit{Time}_t)$$

where \textit{Time}_t is the time index.

- In STATA, \textit{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 examples from another textbook (Diebold: Elements of Forecasting)
- They did not forecast well out of sample.
- A constructive alternative is to forecast growth rates, as we did for consumption expenditure.

Example 1

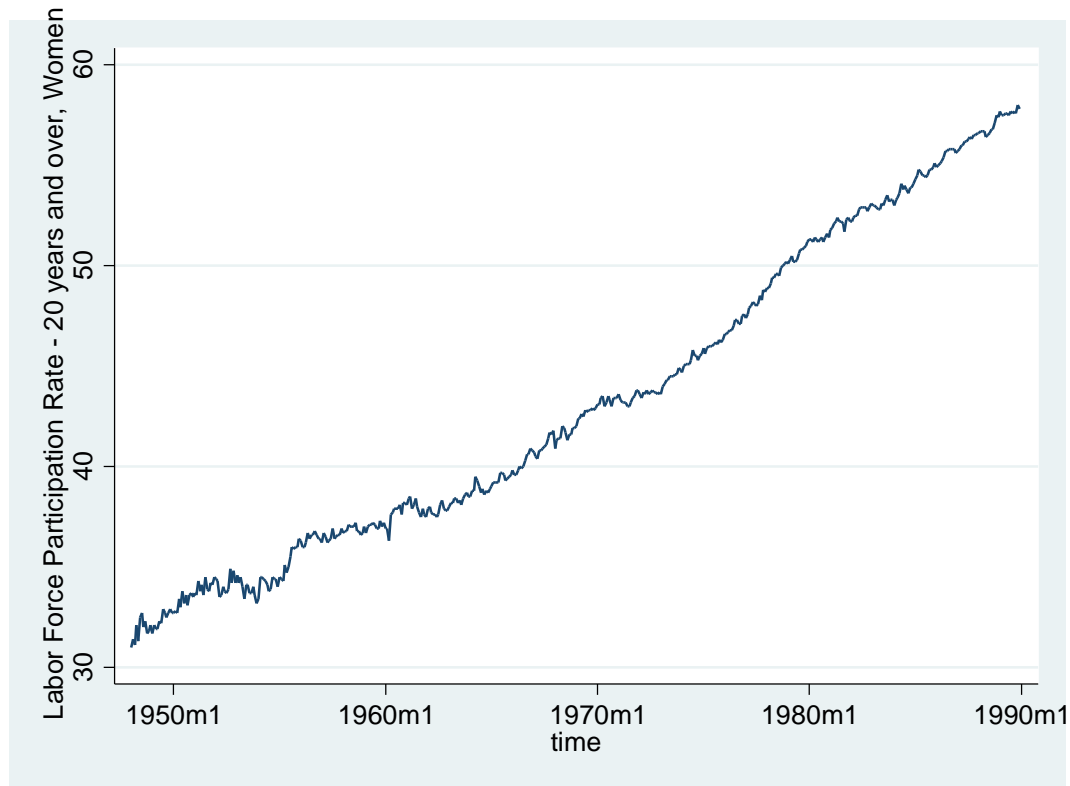
Labor Force Participation Rate

- From BLS
- Monthly, 1948-2016, Seasonally adjusted
- Men and Women, ages 20+
- Percentage of population in labor force (employed plus unemployed divided by population)
- We will estimate on 1948-1989
- Forecast 1990-2016

Example

- `freduse LNS11300025 LNS11300026`
- `rename LNS11300025 men`
- `rename LNS11300026 women`
- `tsmktime time, start(1948m1)`
- `tsset time`
- `tsline women if time<tm(1990m1)`
- `tsline men if time<tm(1990m1)`

Women's Labor Participation Rate 1948-1989



Men's Labor Participation Rate 1948-1989



Linear Trend Model

- The labor force participation rates have been smoothly and linearly increasing (for women) and smoothly and linearly decreasing (for men) over 1948-1989
- 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

Quadratic Trends

- An alternative model is a quadratic trend

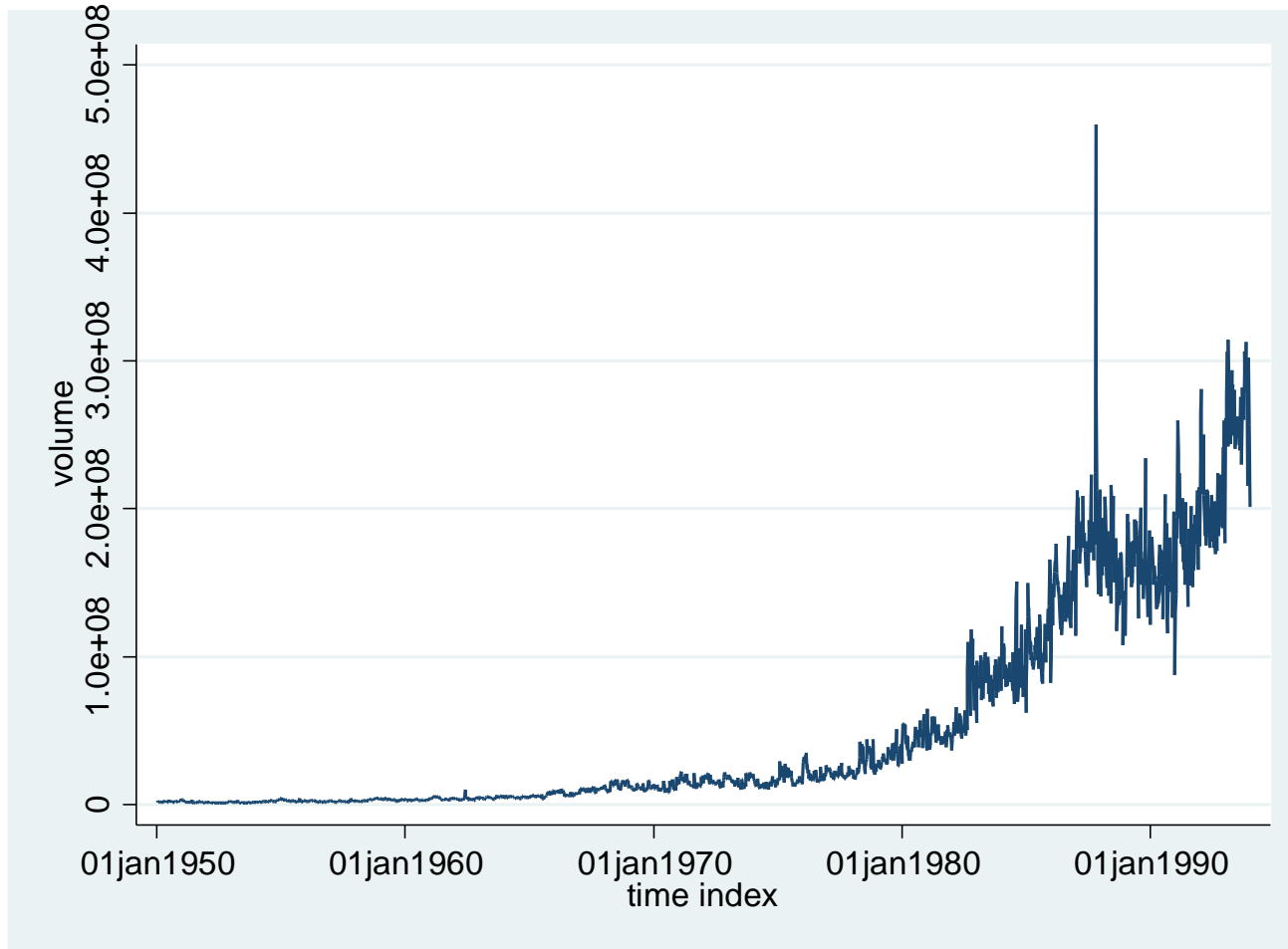
$$T_t = \beta_0 + \beta_1 Time_t + \beta_2 Time_t^2$$

Example 2

Transaction Volume, S&P Index

- From Yahoo Finance
 - GSPC, closing price and average volume
- Weekly, 1950-current
- Stata file `s&p.dta` on course webpage
- We estimate on 1955-1993
- Forecast 1994-current

Transaction Volume



Exponential Trend

- To model this we will use an exponential trend

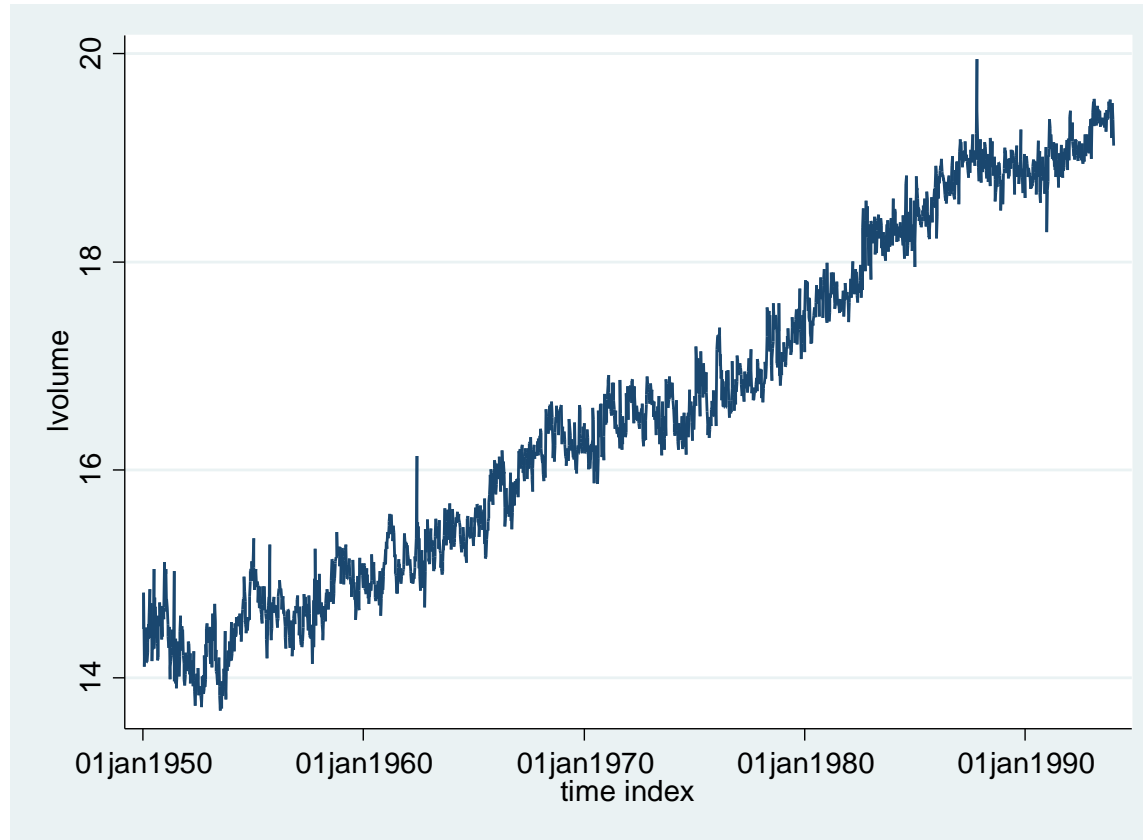
$$T_t = e^{\beta_0 + \beta_1 Time_t}$$

- 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

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

or

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

where

$$- x_{1t} = \textit{Time}_t$$

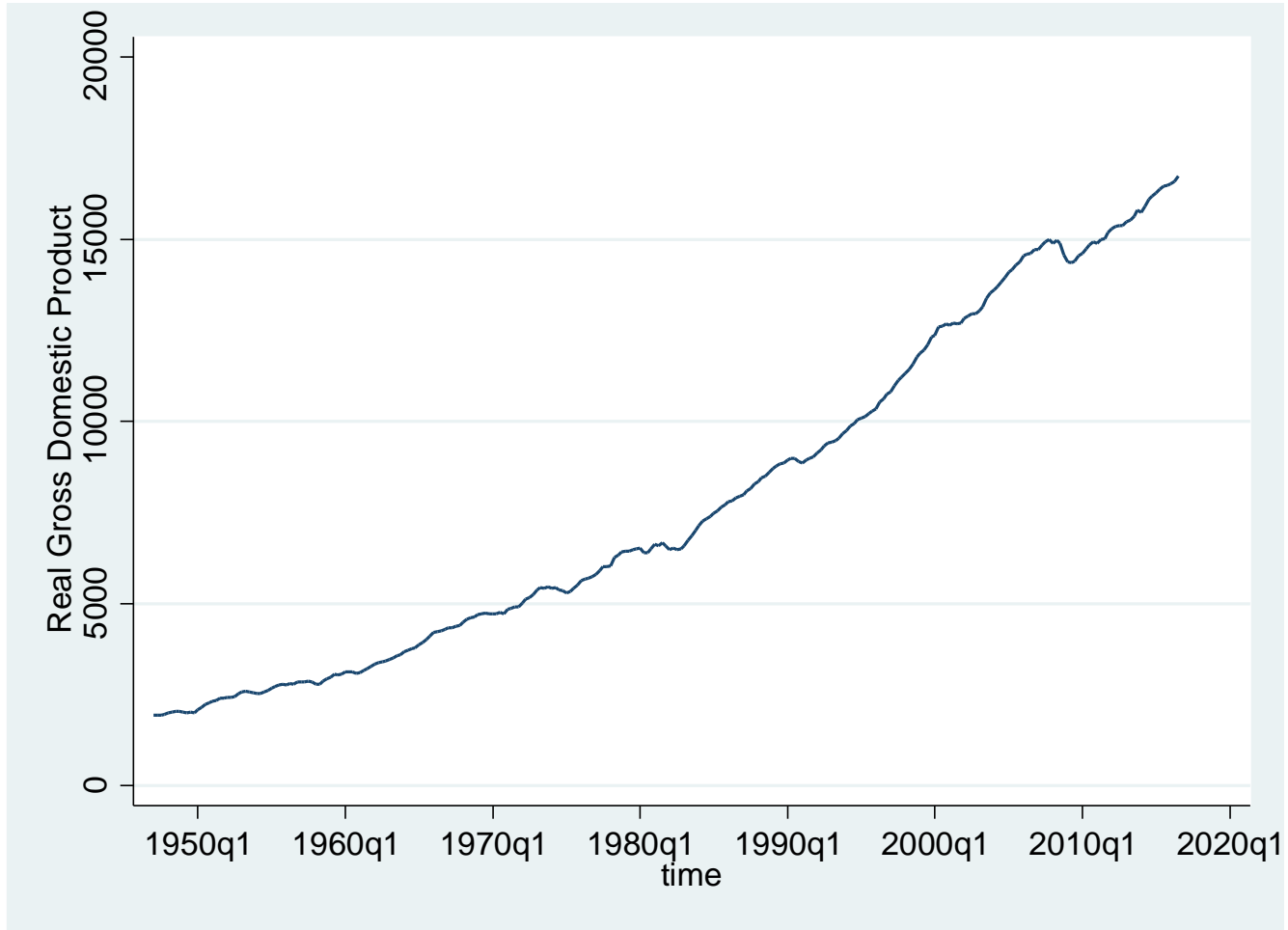
$$- x_{2t} = \textit{Time}_t^2$$

Example 3

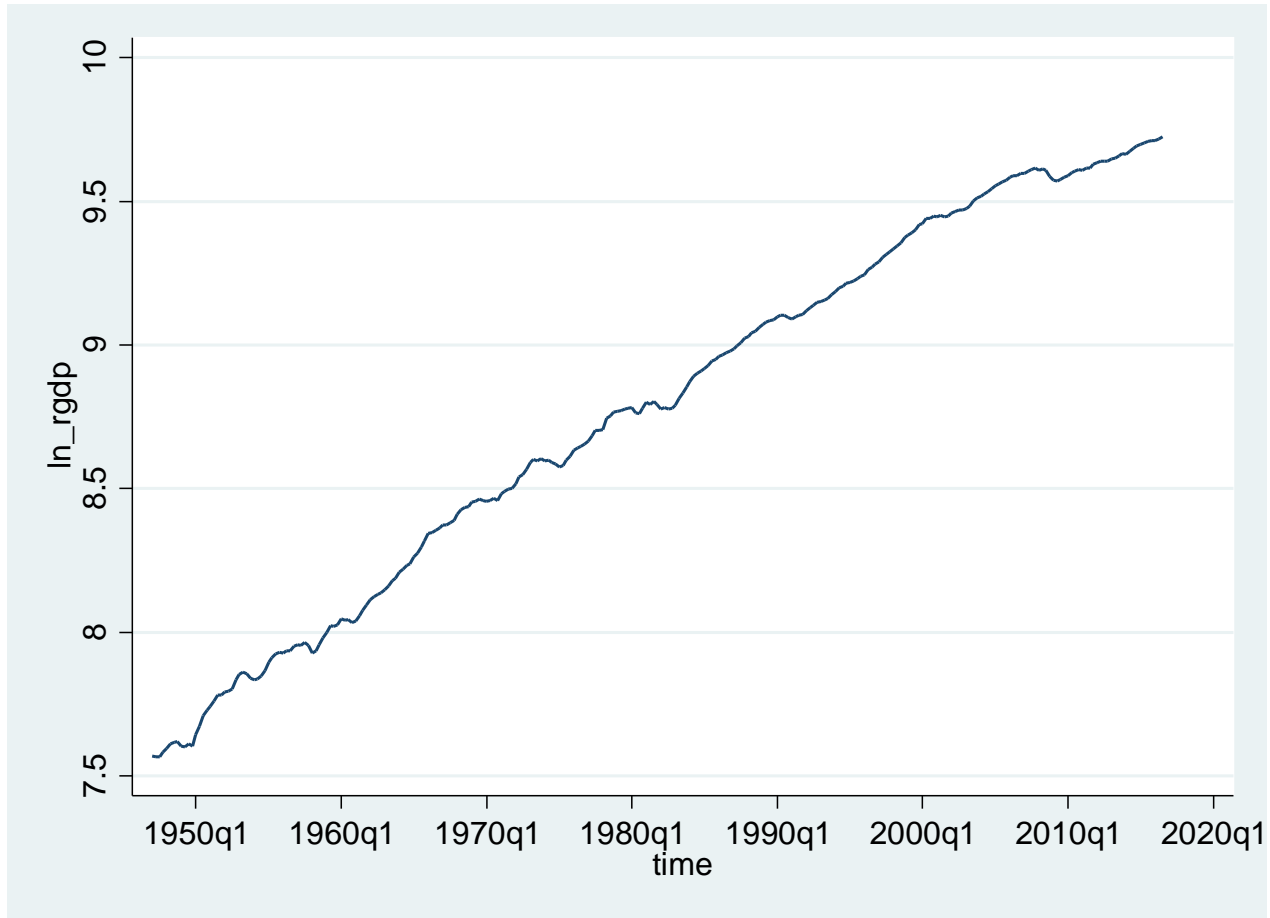
Real GDP

- From BEA
- Quarterly, 1947-2016
- We will estimate on 1947-1990, forecast 1991-2016
- Also use an exponential trend

Real GDP



Ln(Real GDP)



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 $T_t = \beta_0 + \beta_1 x_t$
the forecast for y_{t+h} is $\hat{y}_{t+h} = b_0 + b_1 x_t$ 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

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

- 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

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

- 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

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

- 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

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

- For example, the quadratic trend

$$T_t = \beta_0 + \beta_1 \text{Time}_t + \beta_2 \text{Time}_t^2$$

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

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

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

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

- In STATA, use **theregress** command

Assignments

- Read Diebold through Chapters 5.1-5.2
- Problem Set # 3
 - Due Tuesday (2/7)
- Read Chapter 2 from *The Signal and the Noise*
 - Reading Reflection
 - Due Thursday (2/2)