ECON 1123 Section 10

Slides at github.com/cjleggett/1123-section

Outline

- Name Circle
- Looking Ahead
- Lecture Recap + Exercises

Name Circle

Name Circle

- Name
- Summer Plans

Looking Ahead

Problem Set 10

- I was wrong, you **can** drop this one if you want
- Like problem set 5, focus more on analysis than on doing everything perfectly

Problem Set 11

- You can drop this one too
- You’re not graded on how well you predict the future, so focus on analysis

Last Section Next Week

- Last week of content **will** probably be on the final
- I’ll go over my tips for the final

Final Exam

- 5/9 at 9am
- Same format as midterm, but longer
- We’ll have:
  - practice exams
  - office hours
  - review sessions

Lecture Recap

Prediction

- Econ PhDs vs. Us
- Very important!
  - Policy-makers make decisions based on causal inference
  - Normal people make decisions based on predictions
    - Expected Earnings
    - Inflation predictions
    - Interest rate forecasts
    - Unemployment / industry predictions

Time Series

- Each data point collected at specific time.
- One set of data points over time called a "series"
- Two models we'll work with:
  - Use past values of Y_t to predict Y_t in the future
  - Use past values of Y_t and X_t to predict Y_t in the future

Vocab

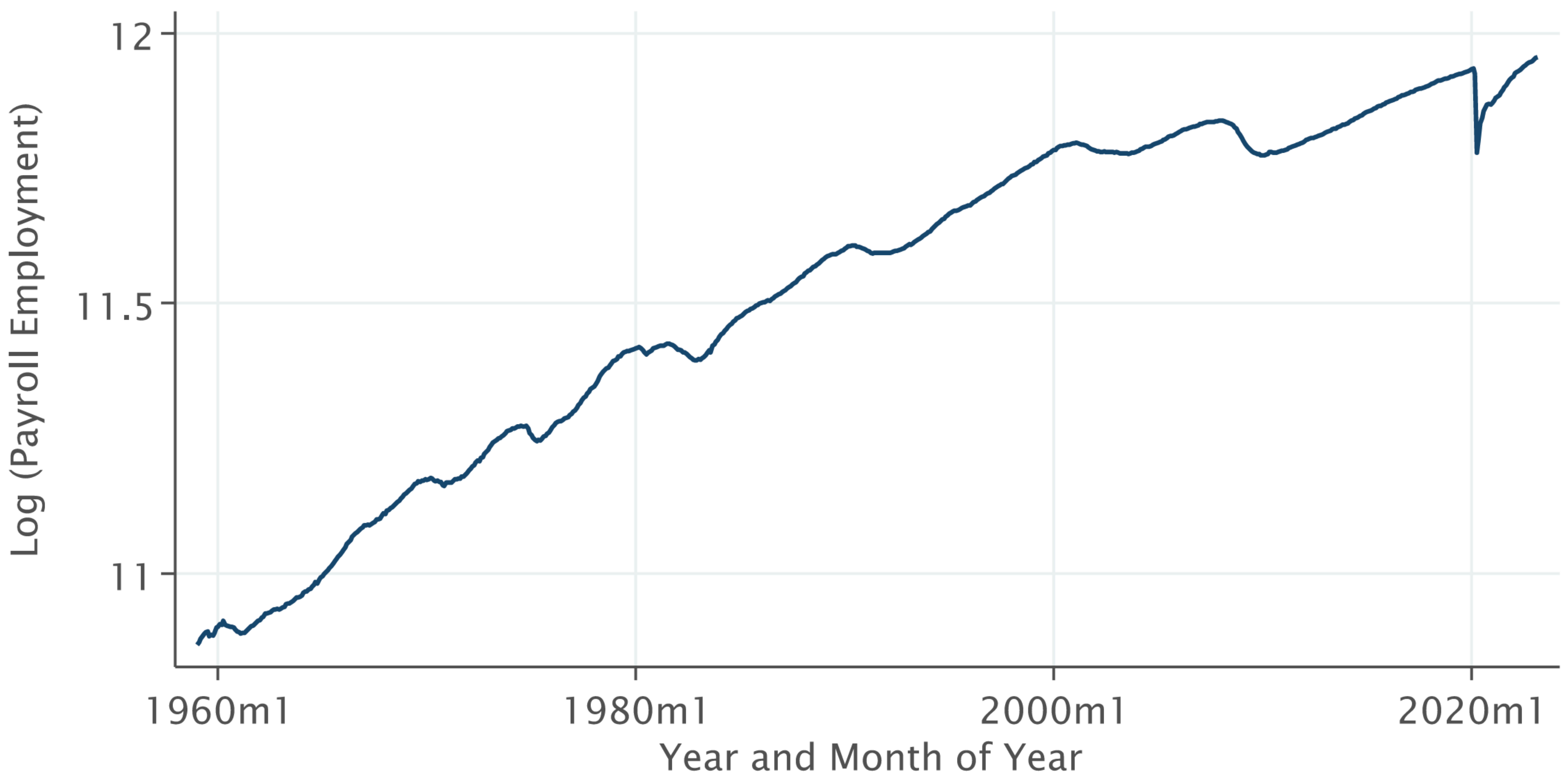
- Time Series: $Y = Y_1, Y_2, \ldots Y_T$
- First Difference: $\Delta Y_t = Y_t - Y_{t-1}$
- Lags: “nth lag of Y_t” = Y_{t-n}
- Autocorrelation: $corr(Y_t, Y_{t-j})$
    - Ranges from -1 to 1
    - large values mean high predictive power

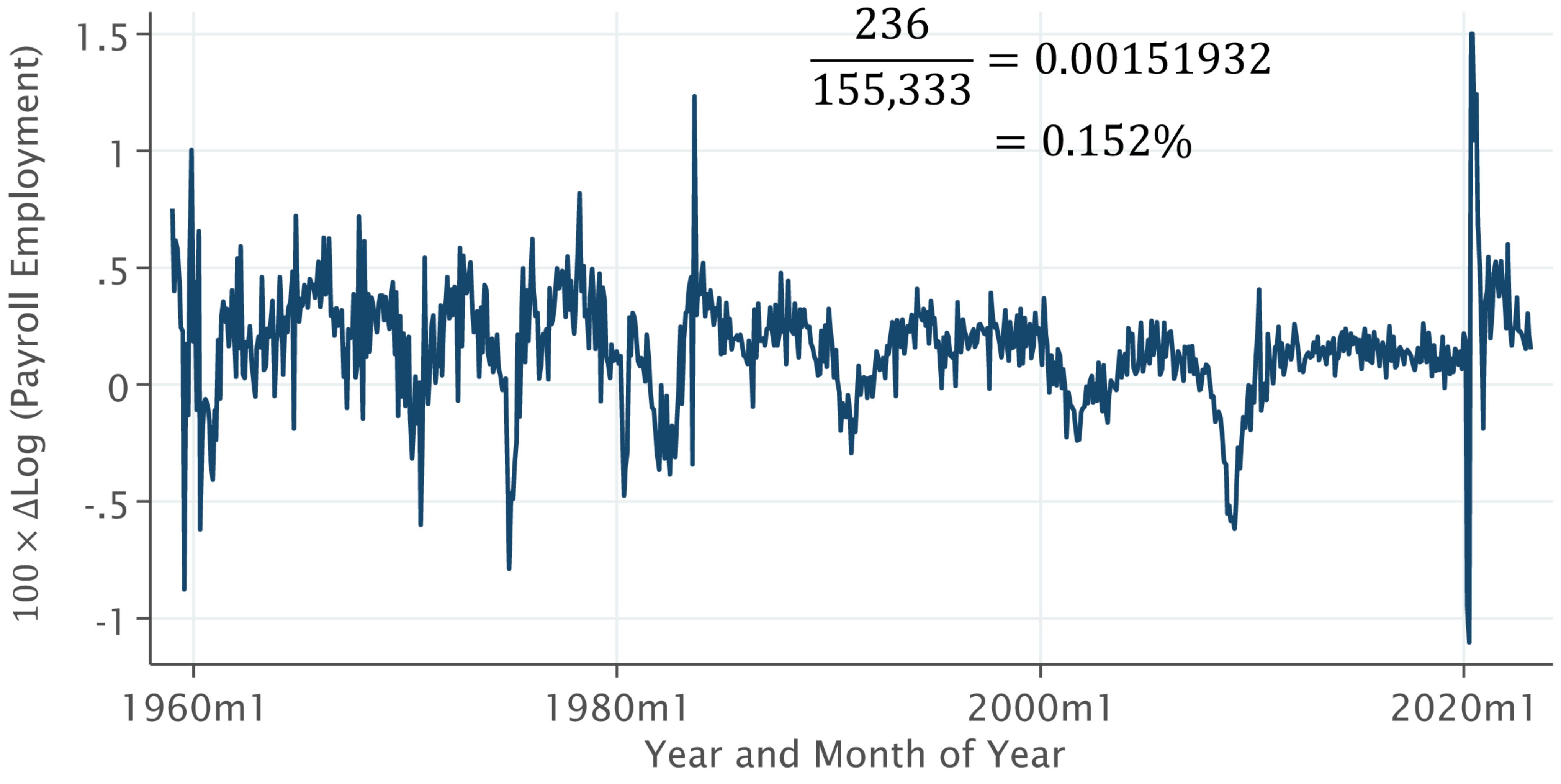
Transformations

- We want to get rid of trends!!!
- So we'll use percentage change:
- We can also use our log trick:
  - $100 \times log(\frac{Y_t}{Y_{t-1}}) = 100 \times \Delta log(Y_t)$
  - So we can use difference in log values for percentage change

Transformations: Why?

- Lots of economic data is exponential
    - This means the logarithm grows linearly
- Standard errors approximately proportional to level
    - So we want level to be comparable across the series
- For statistical analysis, we need to assume “stationarity”

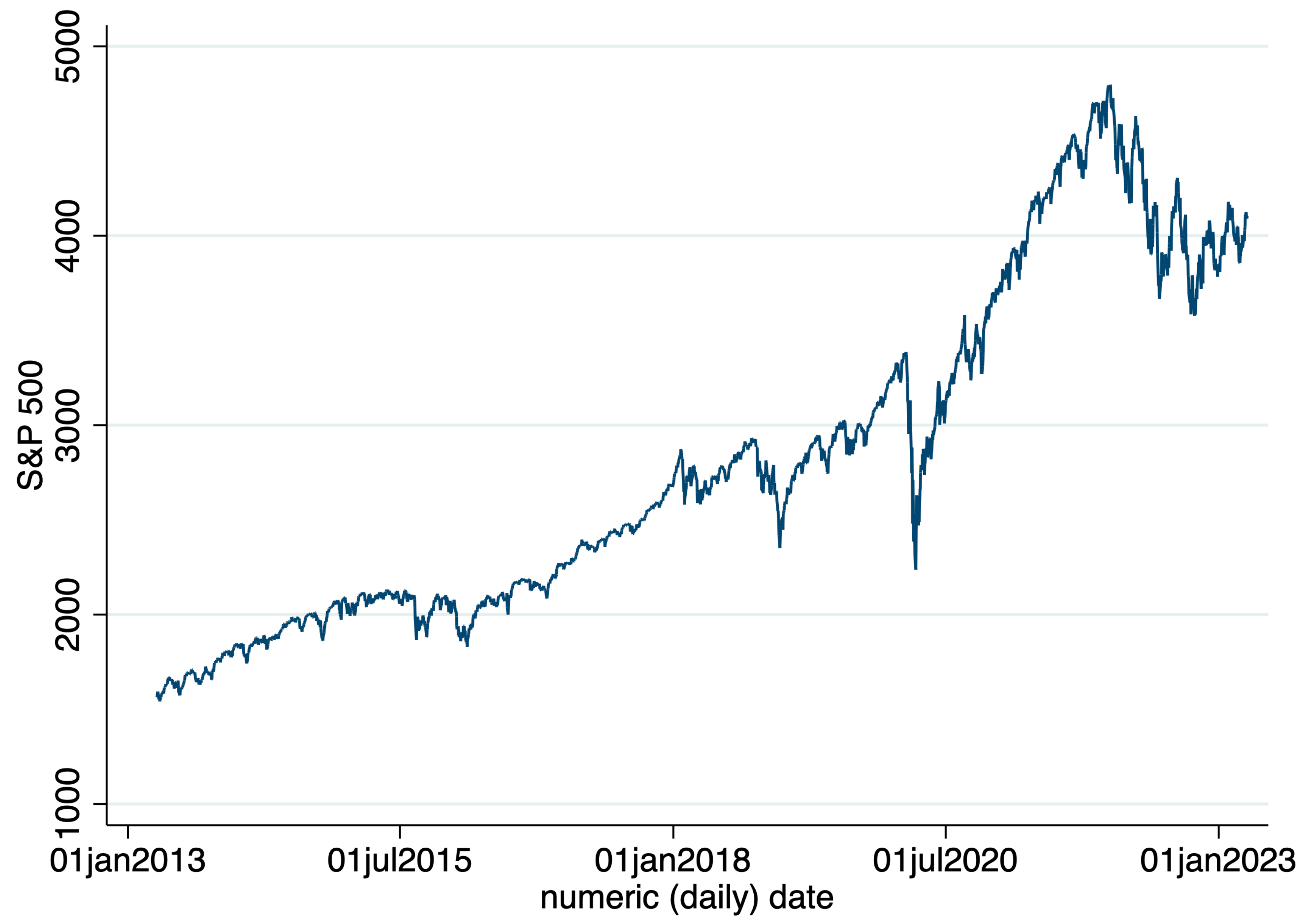

Log (Payroll Employment)
12
11.5
11
1960m1
1980m1
2000m1
2020m1
Year and Month of Year


100 × ΔLog (Payroll Employment)
1.5
1
.5
0
-.5
-1
$\frac{236}{155{,}333} = 0.00151932$
$= 0.152\%$
1960m1
1980m1
2000m1
2020m1
Year and Month of Year

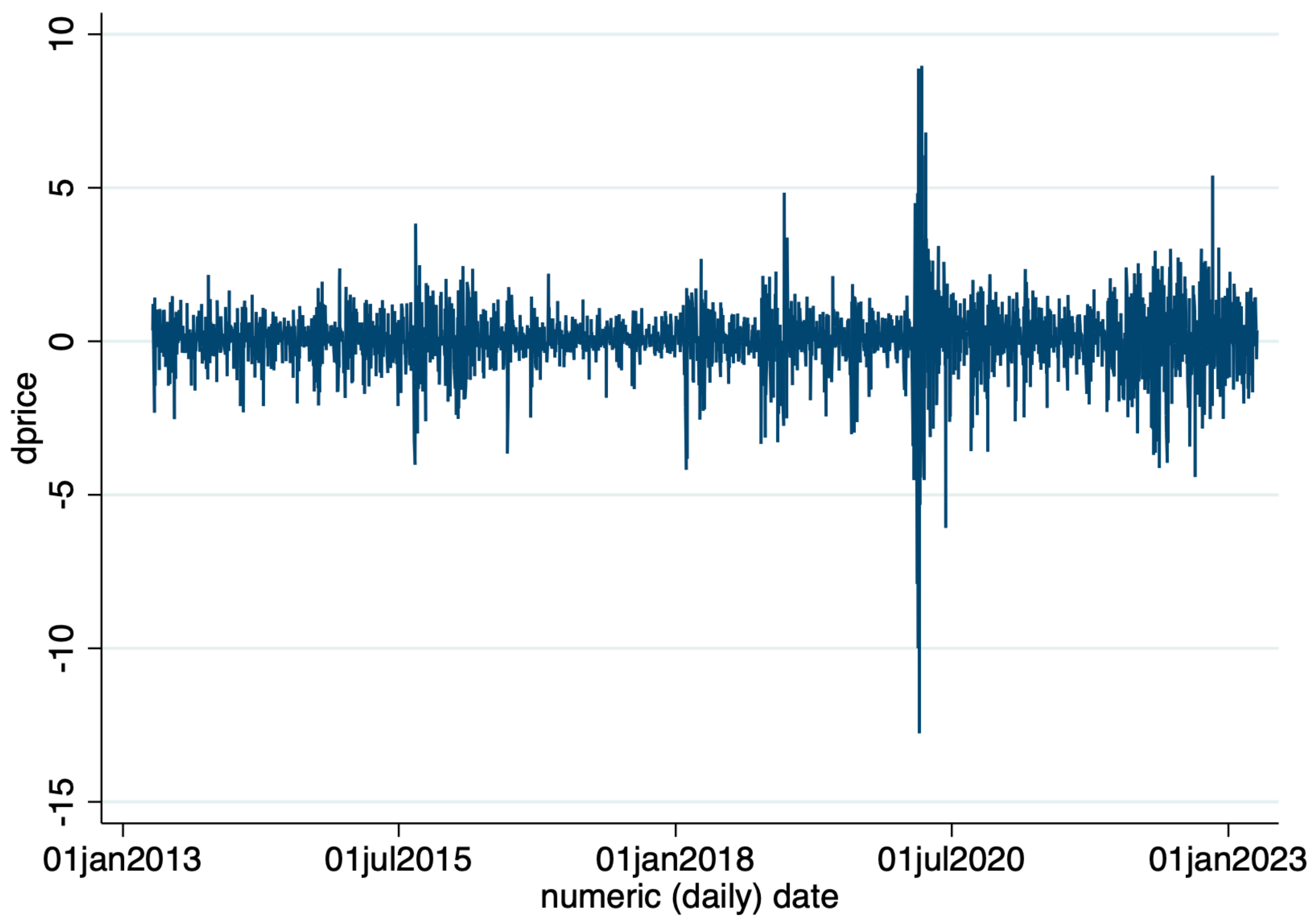
Stationarity

- Assumption: processes do not vary with time
- Definition: Distribution of $(Y_{j+1}, \ldots, Y_{T+j})$ does not depend on j
- If we have stationarity, we can use historical data to forecast future

Exercises: Part 2



S&P 500
1000
2000
3000
4000
5000
01jan2013
01jul2015
01jan2018
01jul2020
01jan2023
numeric (daily) date



dprice
-15
-10
-5
0
5
10
01jan2013
01jul2015
01jan2018
01jul2020
01jan2023
numeric (daily) date

Exercises: Part 3

$AR(p)$

- AR stands for Auto-Regressive
- pth order autoregression
- $Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + u_i$
- Most basic model, but very effective!

$ADL(p, q)$

- ADL stands for Autoregressive Distributed Lag
- p lags of Y and q lags of X
- $Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + \beta_{p+1} X_{t-1} + \ldots + \beta_{p+q} X_{t-q} + u_i$
- Can be extended to more than one predictor:
  $ADL(p, q_1, q_2 \ldots)$

Exercises: Part 4 (through q12)

Questions 10-12

Date	dprice (price difference)	SP500 (price)
3/30	0.5699	4050.83
3/31	1.4334	4109.31
4/3	0.3692	4124.51
4/4	-0.5814	4100.6
4/5	-0.2496	4090.38
4/6	0.3573	4105.02

AR(1)

Variable	Coefficient
dprice L1	-0.1421
Constant	0.0430

AR(4)

Variable	Coefficient
dprice L1	-0.1326
dprice L2	0.0722
dprice L3	-0.0032
dprice L4	-0.0708
Constant	0.0426

Lag Selection

- How do I choose p and q?
- Why don't I use R^2?
- Methods:
    - Sequential hypothesis testing
    - Information Criterion

Sequential Hypothesis Testing

- Start with small model
- Add a lag. If it’s not statistically significant, stop.
- Repeat

- Start with large model
- If last lag is statistically significant, stop
- Repeat

Sequential Hypothesis Testing

- Start with small model
- Add a lag. If it’s not statistically significant, stop.
- Repeat

- Problem: too-small model

- Start with large model
- If last lag is statistically significant, stop
- Repeat

- Problem: too-large model

Information Criterion

- $\hat{\sigma}_p^2 = \frac{1}{T} \sum_{t=1}^{T} [Y_t - \hat{Y}_t]^2$ is the mean squared error (squared residuals)
- Idea: Try out a bunch of different models, and see which one does the best
- If we try out a ton of different models with lags 1 … n, which one will have lowest MSE?

Information Criterion

- $\hat{\sigma}_p^2 = \frac{1}{T}\sum_{t=1}^{T}[Y_t - \hat{Y}_t]^2$ is the mean squared error (squared residuals)
- Idea: Try out a bunch of different models, and see which one does the best
- If we try out a ton of different models with lags 1 … n, which one will have lowest MSE?
    - n! Because it's largest!
    - So we add a penalty

Penalty Factors

- Bayes Information Criterion (BIC):

$$BIC = log(\hat{\sigma}_p^2) + \frac{log(T)}{T}p$$

- Akaike Information Criterion (BIC):

$$AIC = log(\hat{\sigma}_p^2) + \frac{2}{T}p$$

- Adjusted R^2:

$$log(1 - \bar{R}^2) = log(\hat{\sigma}_p^2) + \frac{1}{T}p + \ldots$$

- Which one chooses smallest model?

Bayes Information Criterion (BIC)

- Cool property: if we're selecting among multiple $AR(p)$ models, and the true model is there, BIC will choose true model as sample size increases
  $P(\hat{p} = TRUE) \to 1$ as $T \to \infty$

- Cool Property: HR Standard Errors will be correct

- Not Cool Property: Very time-intensive

Using BIC

- Come up with range of p and q values to test
- Loop over all combinations and calculate BIC for each model
- Choose model with lowest BIC

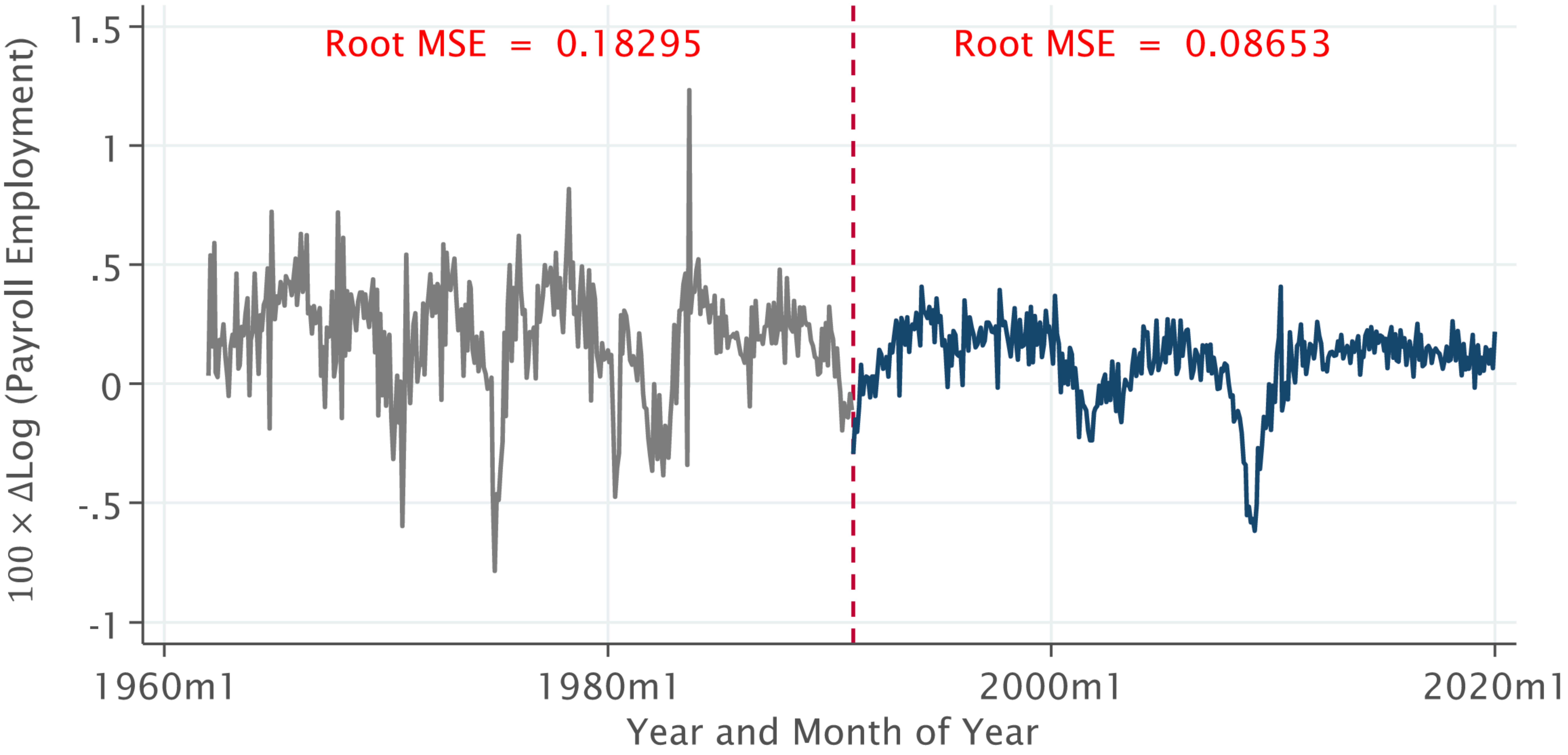
Exercises: Part 4 (the rest)

Testing Stationarity

- Test for structural break at specific time
- Test for structural break at any time
- End-of-sample stability check

Chow Test (Structural Break)

- Decide on a time you want to test for a break (covid? 2008?)
    - Indicator D_t is 0 if $t < r$, and 1 if $t \geq r$
- Fully interact AR(1) model with indicator for after this time
    - $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 D_t + \beta_3 D_t \times T_{t-1} + u_t$
- Do normal F test for whether slope/intercept are equal before and after date



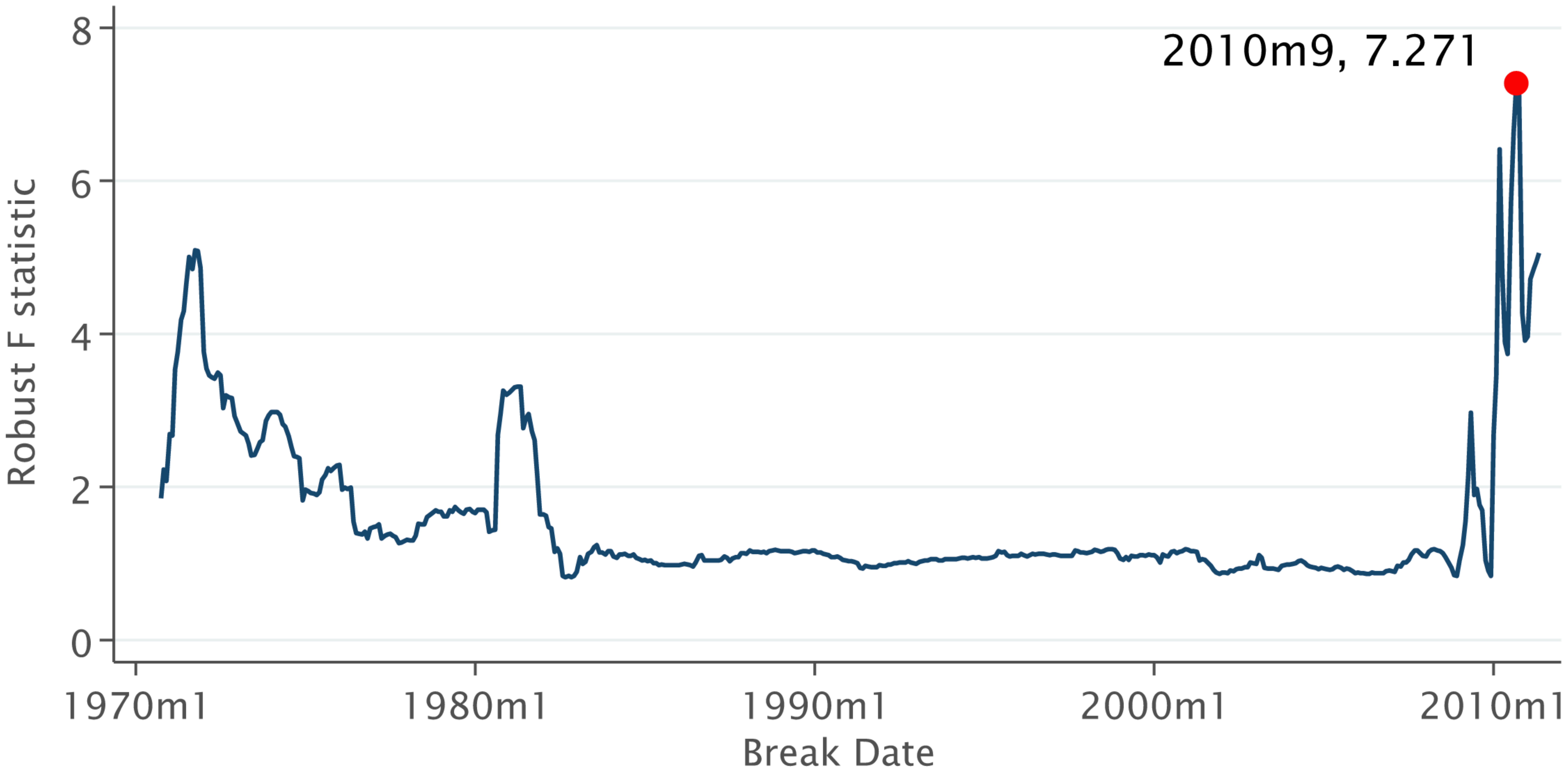
Root MSE = 0.18295
Root MSE = 0.08653
1.5
1
.5
0
-.5
-1
100 × ΔLog (Payroll Employment)
1960m1
1980m1
2000m1
2020m1
Year and Month of Year

QLR Test (Structural Break)

- QLR is maximum Chow F statistic over all possible breaks in middle 70% of time span
- Why restrict this?

QLR Test (Structural Break)

- QLR is maximum Chow F statistic over all possible breaks in middle 70% of time span
- Why restrict this?
  - So we have enough data on either side



2010m9, 7.271
Robust F statistic
0
2
4
6
8
1970m1
1980m1
1990m1
2000m1
2010m1
Break Date

QLR Test Critical Values

- QLR = max(F tests)
- This is a distribution itself!
- Critical values of this are difficult, and were derived semi-recently (1993)
- We calculate this with a computer, but need to know # of restrictions:
    - 1 restriction for dummy variable
    - p restrictions for lags of Y
    - q restrictions for lags of X
    - total of 1 + p + q

What to do when we detect a break?

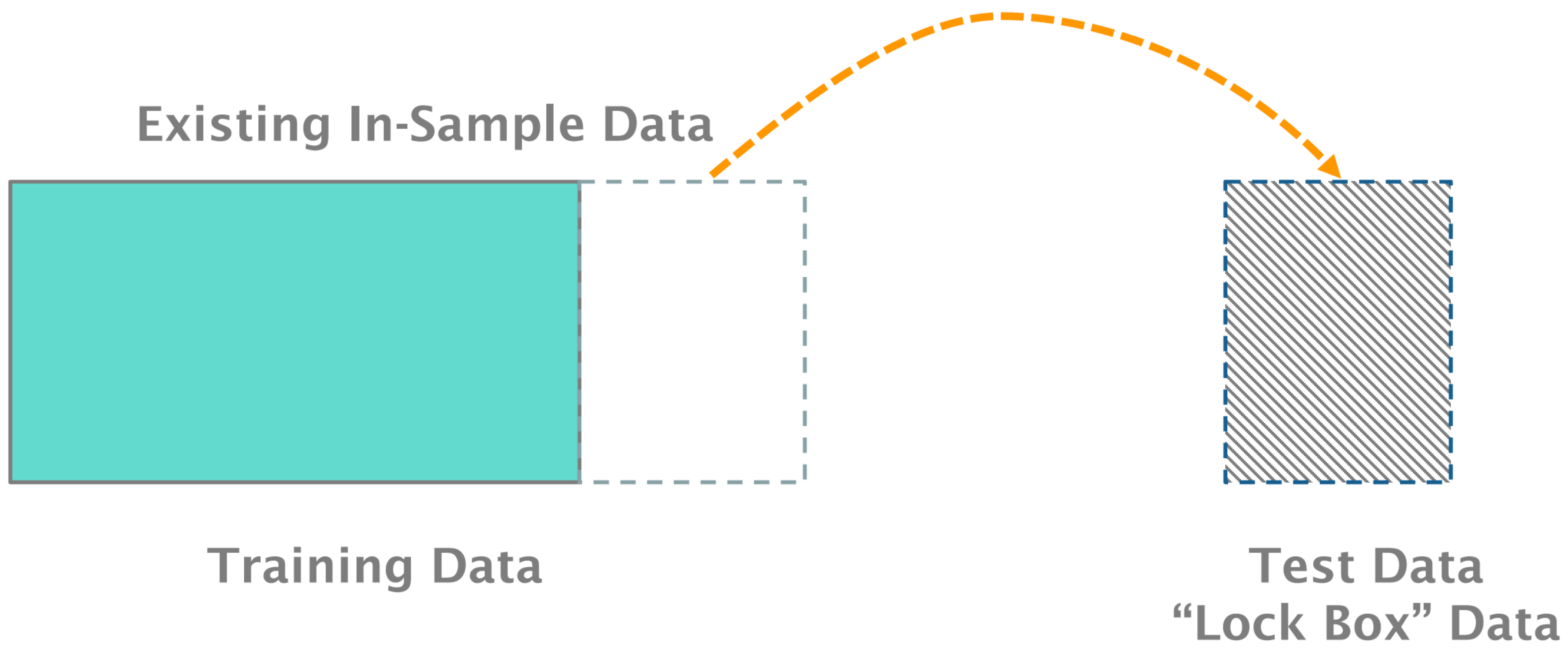
- Split data at the break
- Use only second-half data

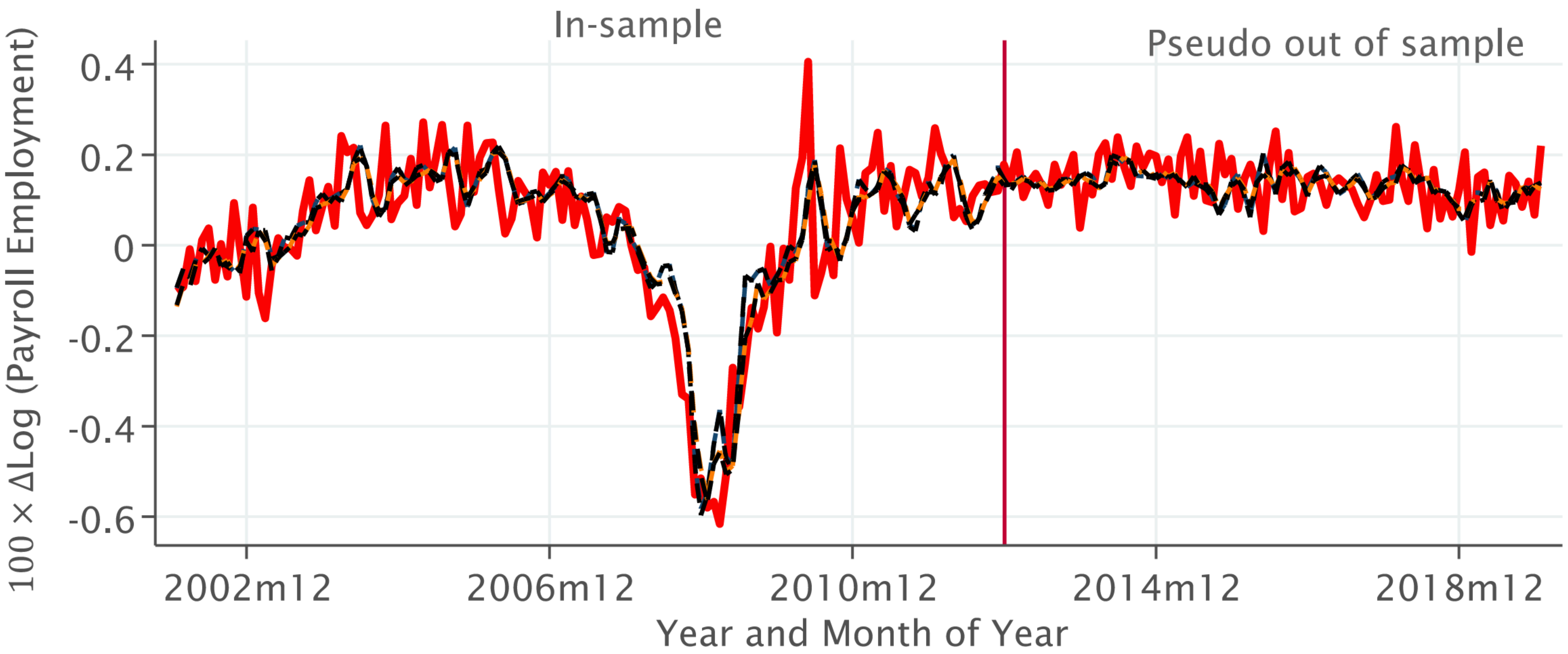
End-of-Sample Stability Check

- Worst break is right at end of sample: think covid
- Use pseudo out of sample forecasting (POOS) as check for end stability
- This is an **informal** test

POOS Test

- Train data: used to build a model
- Test data (POOS data): used to evaluate model
- In time series, we take a chunk out of the end and use it to test our predictions




In-sample
Pseudo out of sample
100 × ΔLog (Payroll Employment)
0.4
0.2
0
-0.2
-0.4
-0.6
2002m12
2006m12
2010m12
2014m12
2018m12
Year and Month of Year