Piecewise linear trend: creating the features
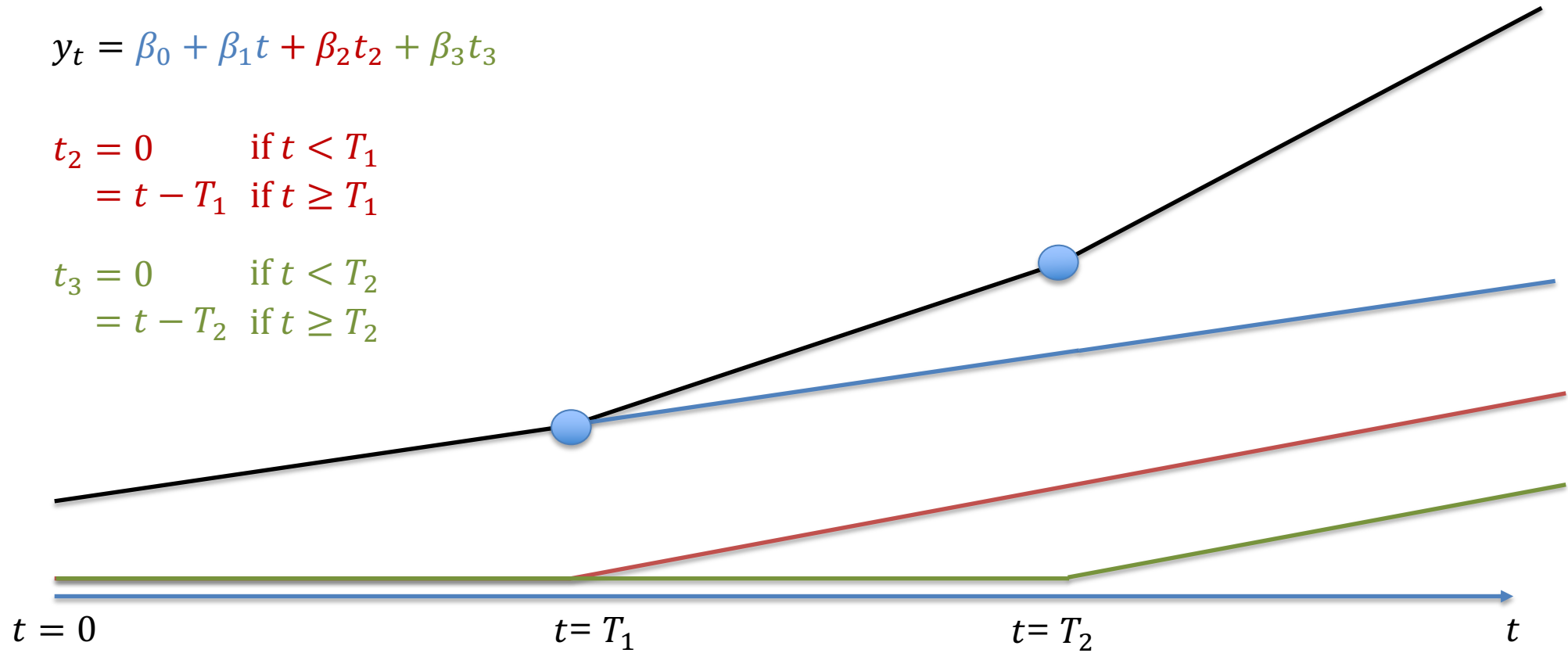
Trend features

Piecewise linear regression

$$y_t = \beta_0 + \beta_1 t + \beta_2 t_2 + \beta_3 t_3$$

$$t_2 = 0 \qquad \text{if } t < T_1$$
$$\quad = t - T_1 \quad \text{if } t \geq T_1$$

$$t_3 = 0 \qquad \text{if } t < T_2$$
$$\quad = t - T_2 \quad \text{if } t \geq T_2$$

$t = 0 \qquad\qquad t = T_1 \qquad\qquad t = T_2 \qquad\qquad t$

Piecewise regression: creating the features

$$y_t = \beta_0 + \beta_1 t + \beta_2 t_2 + \beta_3 t_3$$

$$t_2 = 0 \qquad \text{if } t < T_1$$
$$\quad = t - T_1 \quad \text{if } t \geq T_1$$

$$t_3 = 0 \qquad \text{if } t < T_2$$
$$\quad = t - T_2 \quad \text{if } t \geq T_2$$

Time	y
2020-02-12	23
2020-02-13	30
2020-02-14	35
2020-02-15	30
2020-02-16	20
2020-02-17	34
2020-02-18	12
2020-02-19	?
2020-02-20	?
2020-02-21	?

Piecewise regression: creating the features

$$y_t = \beta_0 + \boldsymbol{\beta_1 t} + \beta_2 t_2 + \beta_3 t_3$$

$t_2 = 0$ if $t < T_1$
    $= t - T_1$ if $t \geq T_1$

$t_3 = 0$ if $t < T_2$
    $= t - T_2$ if $t \geq T_2$

Time	y	t (days)
2020-02-12	23	0
2020-02-13	30	1
2020-02-14	35	2
2020-02-15	30	3
2020-02-16	20	4
2020-02-17	34	5
2020-02-18	12	6
2020-02-19	?	7
2020-02-20	?	8
2020-02-21	?	9

T_1

Piecewise regression: creating the features

$$y_t = \beta_0 + \beta_1 t + \boldsymbol{\beta_2 t_2} + \beta_3 t_3$$

$$t_2 = 0 \quad \text{if } t < T_1$$
$$\quad = t - T_1 \quad \text{if } t \geq T_1$$

$$t_3 = 0 \quad \text{if } t < T_2$$
$$\quad = t - T_2 \quad \text{if } t \geq T_2$$

	Time	y	t (days)	t_2 (days)
	2020-02-12	23	0	0
	2020-02-13	30	1	0
T_1	2020-02-14	35	2	0
	2020-02-15	30	3	1
	2020-02-16	20	4	2
T_2	2020-02-17	34	5	3
	2020-02-18	12	6	4
	2020-02-19	?	7	5
	2020-02-20	?	8	6
	2020-02-21	?	9	7

Piecewise regression: creating the features

$$y_t = \beta_0 + \beta_1 t + \beta_2 t_2 + \boldsymbol{\beta_3 t_3}$$
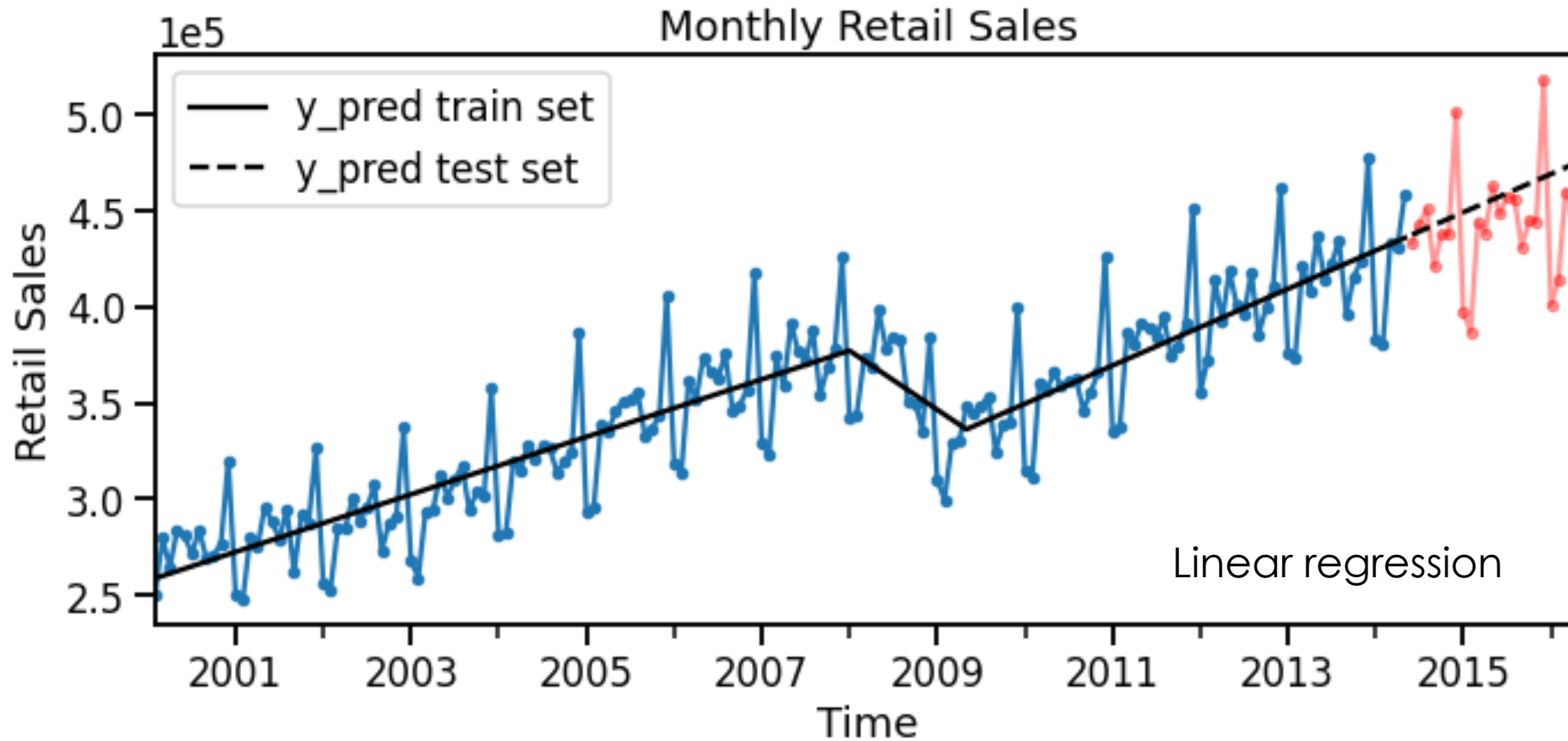
$t_2 = 0 \qquad \text{if } t < T_1$
$\quad = t - T_1 \quad \text{if } t \geq T_1$

$t_3 = 0 \qquad \text{if } t < T_2$
$\quad = t - T_2 \quad \text{if } t \geq T_2$

|	Time	y	t (days)	t_2 (days)	t_3 (days)
	2020-02-12	23	0	0	0
	2020-02-13	30	1	0	0
T_1	2020-02-14	35	2	0	0
	2020-02-15	30	3	1	0
	2020-02-16	20	4	2	0
T_2	2020-02-17	34	5	3	0
	2020-02-18	12	6	4	1
	2020-02-19	?	7	5	2
	2020-02-20	?	8	6	3
	2020-02-21	?	9	7	4

Example: Monthly retail sales

Features:

- Time since start (t)
- Changepoint 1
  - 2008-01-01
- Changepoint 2
  - 2009-05-01

Example: Monthly retail sales

Monthly Retail Sales

Features:
- Time since start (t)
- Changepoint 1
  - 2008-01-01
- Changepoint 2
  - 2009-05-01
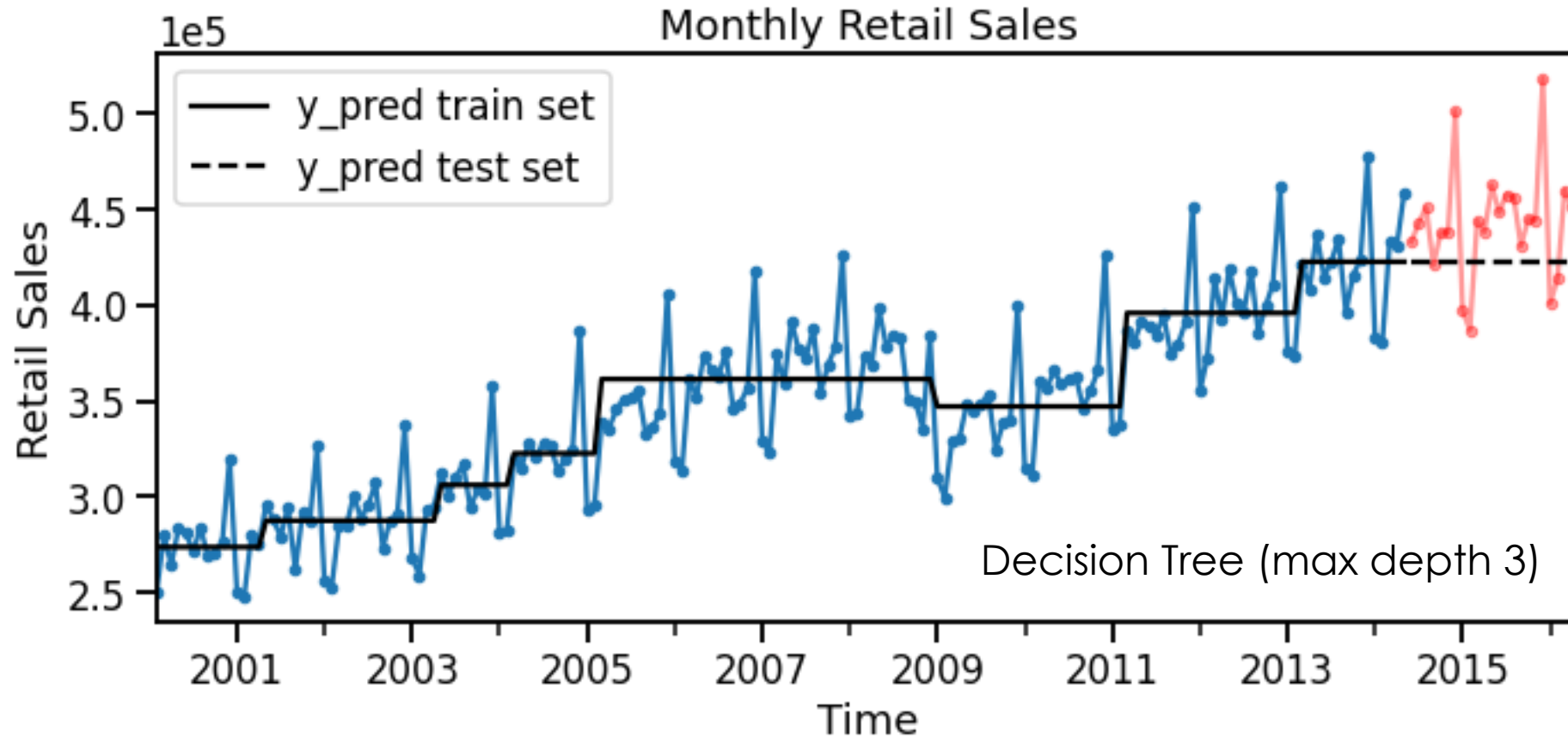
Example: Monthly retail sales

Monthly Retail Sales
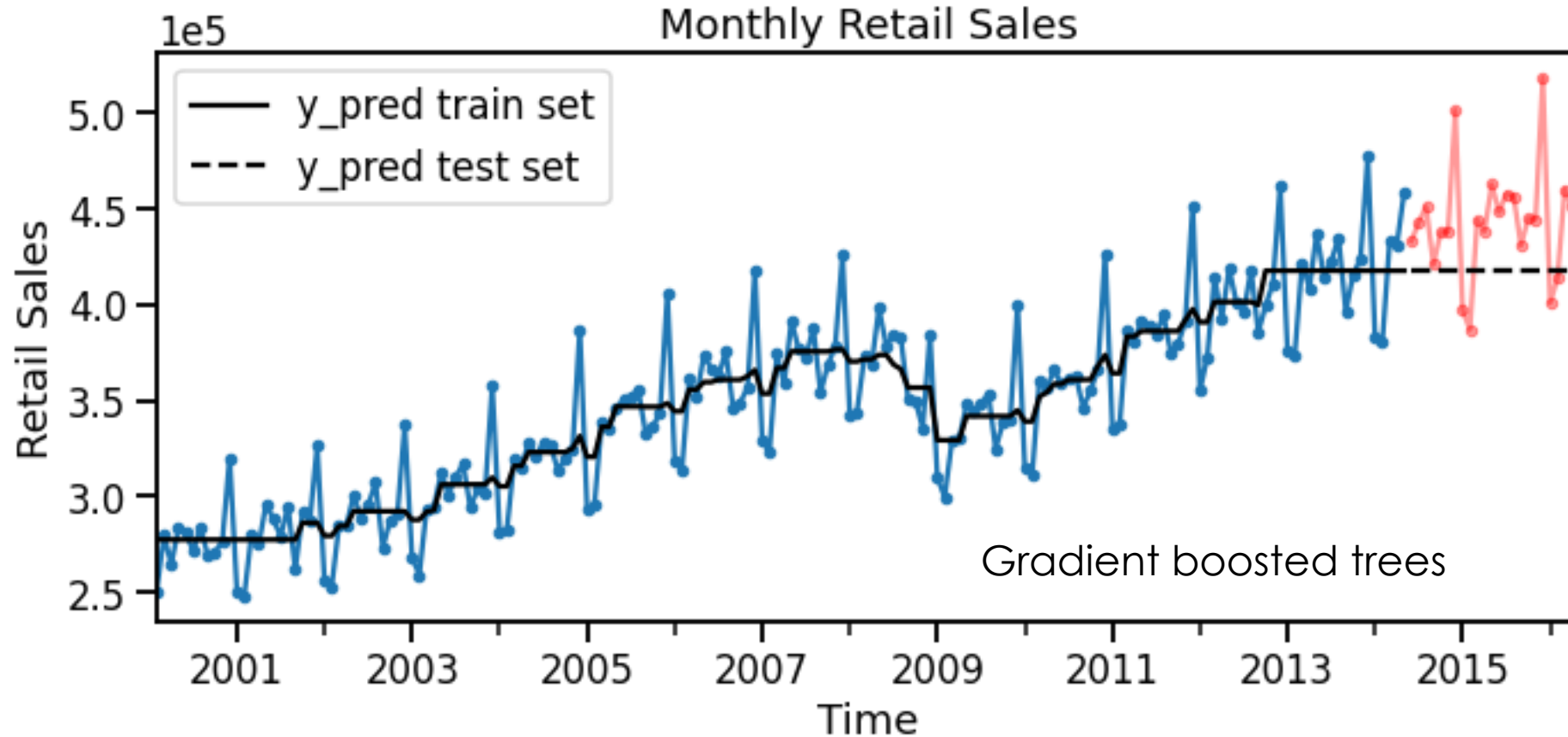
Features:

- Time since start (t)
- Changepoint 1
  - 2008-01-01
- Changepoint 2
  - 2009-05-01

Implementation - sktime

```python
from sktime.transformations.series.time_since import TimeSince
```

Manually specify the changepoints.

```python
# Specify changepoints in monthly period
changepoints = [
                "1992-01", # start of time series
                "2008-01", # changepoint
                "2009-04", # changepoint
                ]
```

Implementation - sktime

```python
time_since_transformer = TimeSince(start=changepoints,
                                   positive_only=True,
                                   keep_original_columns=True)

time_since_transformer.fit_transform(df)
```

Implementation - sktime

```python
time_since_transformer = TimeSince(start=changepoints,
                                   positive_only=True,
                                   keep_original_columns=True)

time_since_transformer.fit_transform(df)
```

Implementation - sktime

ds	y	time_since_1992-01	time_since_2008-01	time_since_2009-04
1992-01	146376	0	0	0
1992-02	147079	1	0	0
1992-03	159336	2	0	0
1992-04	163669	3	0	0
1992-05	170068	4	0	0

Implementation - sktime

ds	y	time_since_1992-01	time_since_2008-01	time_since_2009-04
1992-01	146376	0	0	0
1992-02	147079	1	0	0
1992-03	159336	2	0	0
1992-04	163669	3	0	0
1992-05	170068	4	0	0

Implementation - sktime

ds	y	time_since_1992-01	time_since_2008-01	time_since_2009-04
2007-10	368681	189	0	0
2007-11	377802	190	0	0
2007-12	426077	191	0	0
2008-01	342697	192	0	0
2008-02	343937	193	1	0
2008-03	372923	194	2	0
2008-04	368923	195	3	0
2008-05	397969	196	4	0
2008-06	378490	197	5	0
2008-07	383686	198	6	0

Pros and cons

Pros

Easy to implement.

Provides a method of handling simple non-linear trends.

Cons

Need to choose change points.

Mostly useful for linear models.

If the time series is highly non-linear it would be tedious to break it down into change points.

Summary

Simple changes in trend can be modelled using piecewise regression.

Piecewise regression amounts to creating new features, one for each changepoint.

Enables linear models to handle non-linear trends in the time series.

References

Piecewise linear trends:

https://otexts.com/fpp3/nonlinear-regression.html