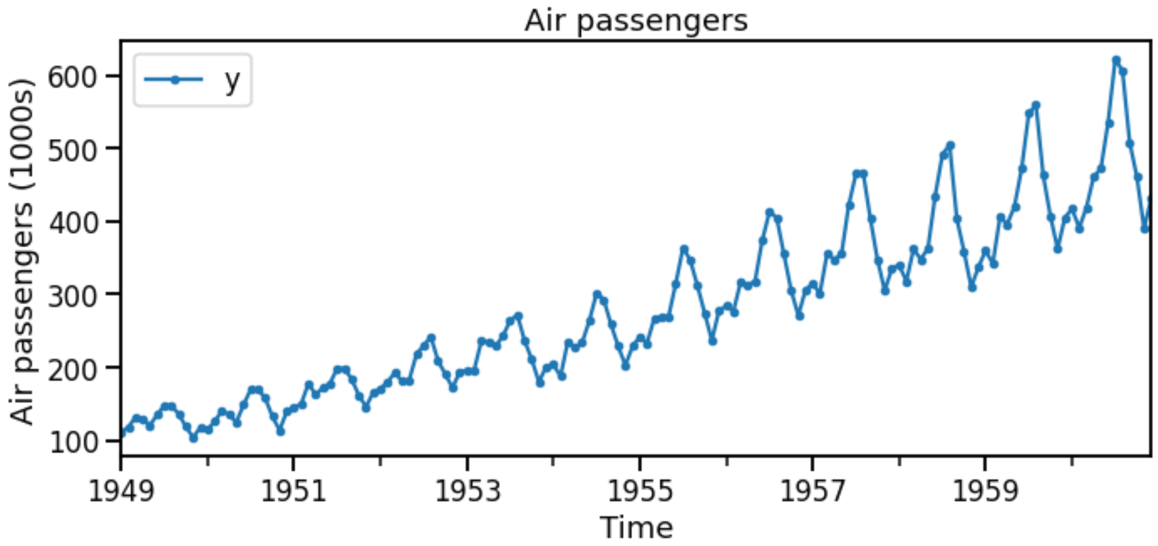
Seasonal lags

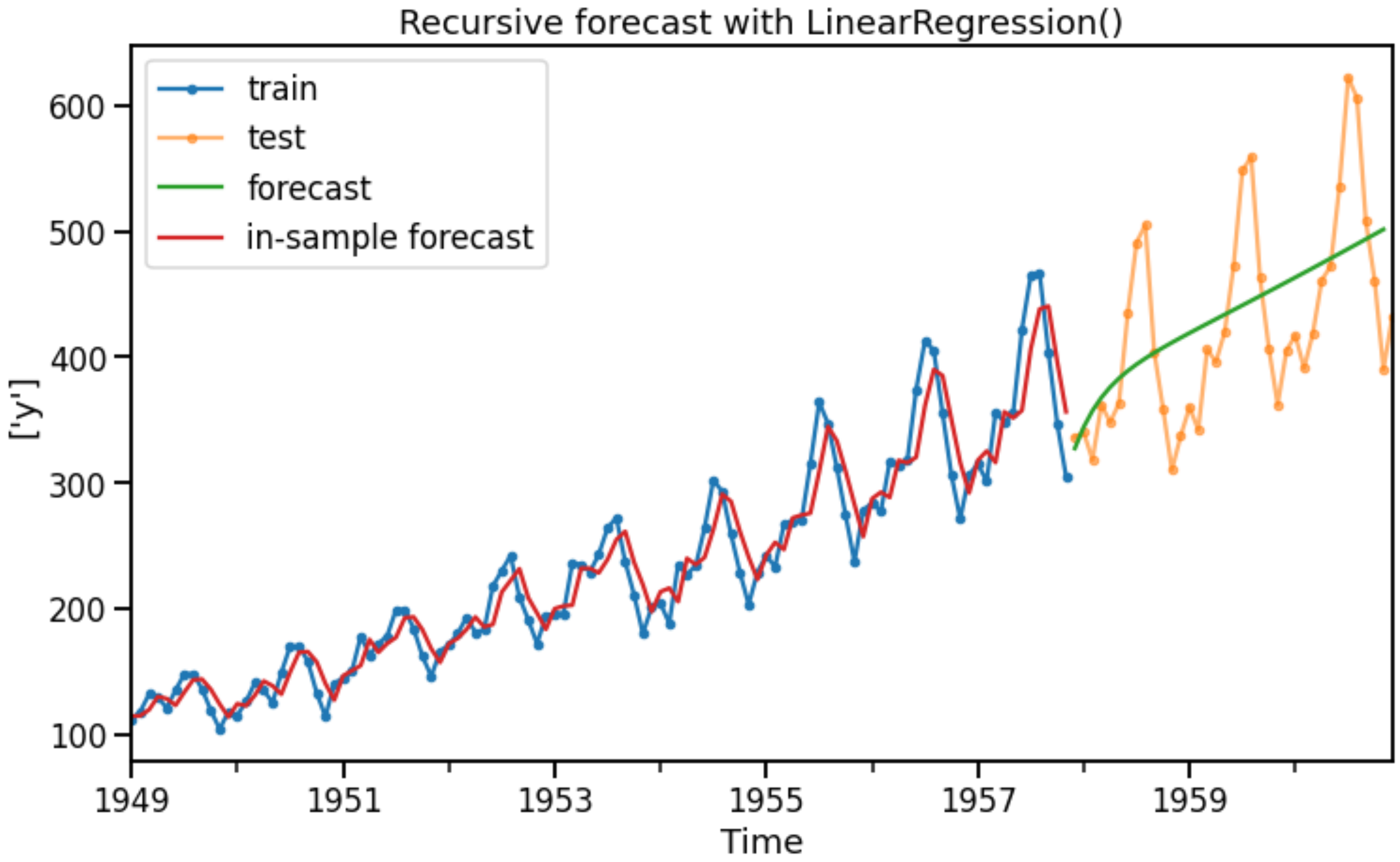
Seasonality features

Lag features can help capture seasonality

- Lag of the seasonal period.
  - E.g: Monthly data, yearly seasonality
  - Feature: y_{t-12}
- If we know the seasonal period then we can use that directly.
- If we don't know the seasonal period then we can use the following to help pick a lag:
  - Domain knowledge
  - Plots
  - ACF and PACF



Example: Air passengers



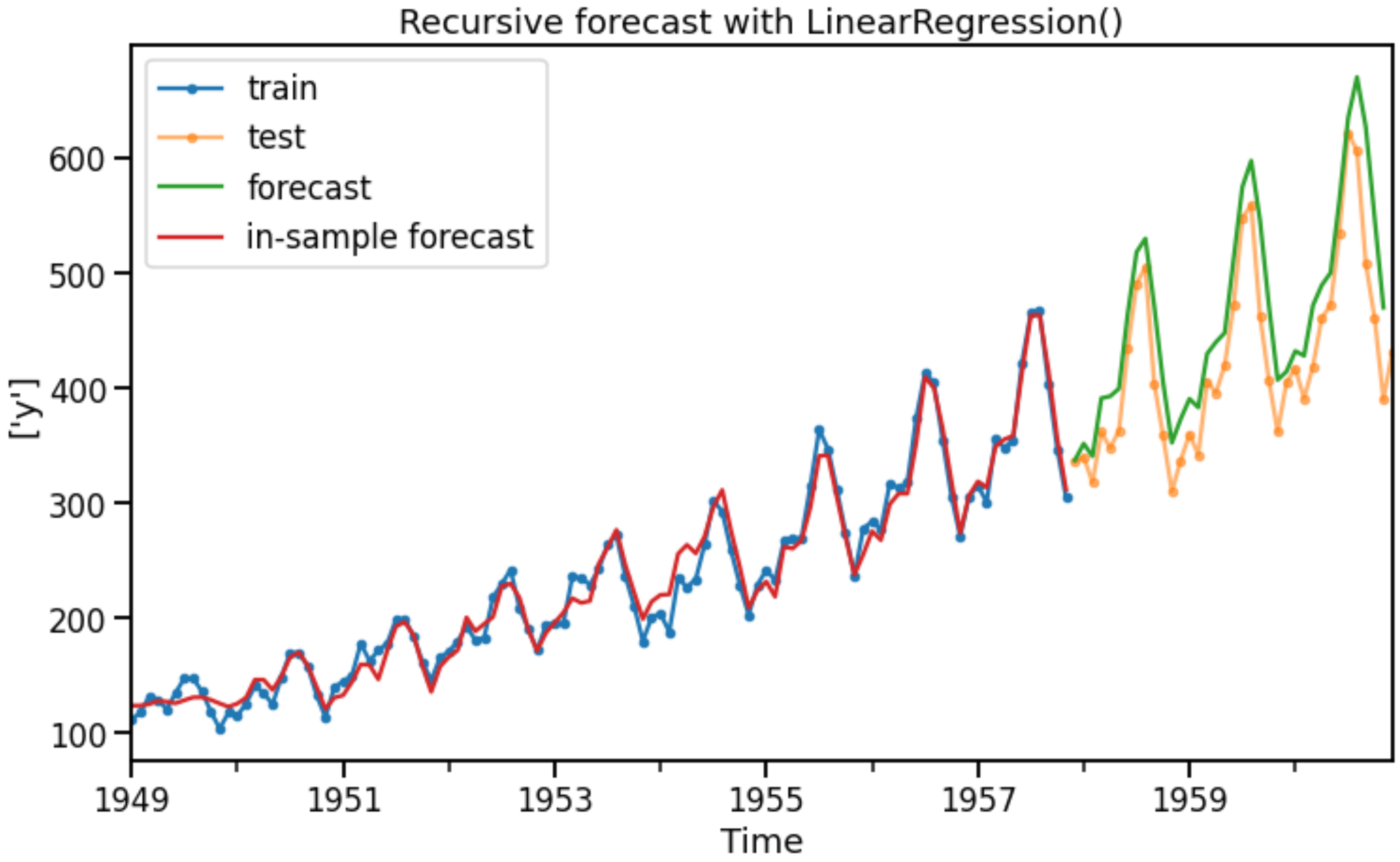
Features

- Trend features: t, t^2
- Lag of 1 month: y_{t-1}

Model

$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 y_{t-1}$$

Example: Air passengers



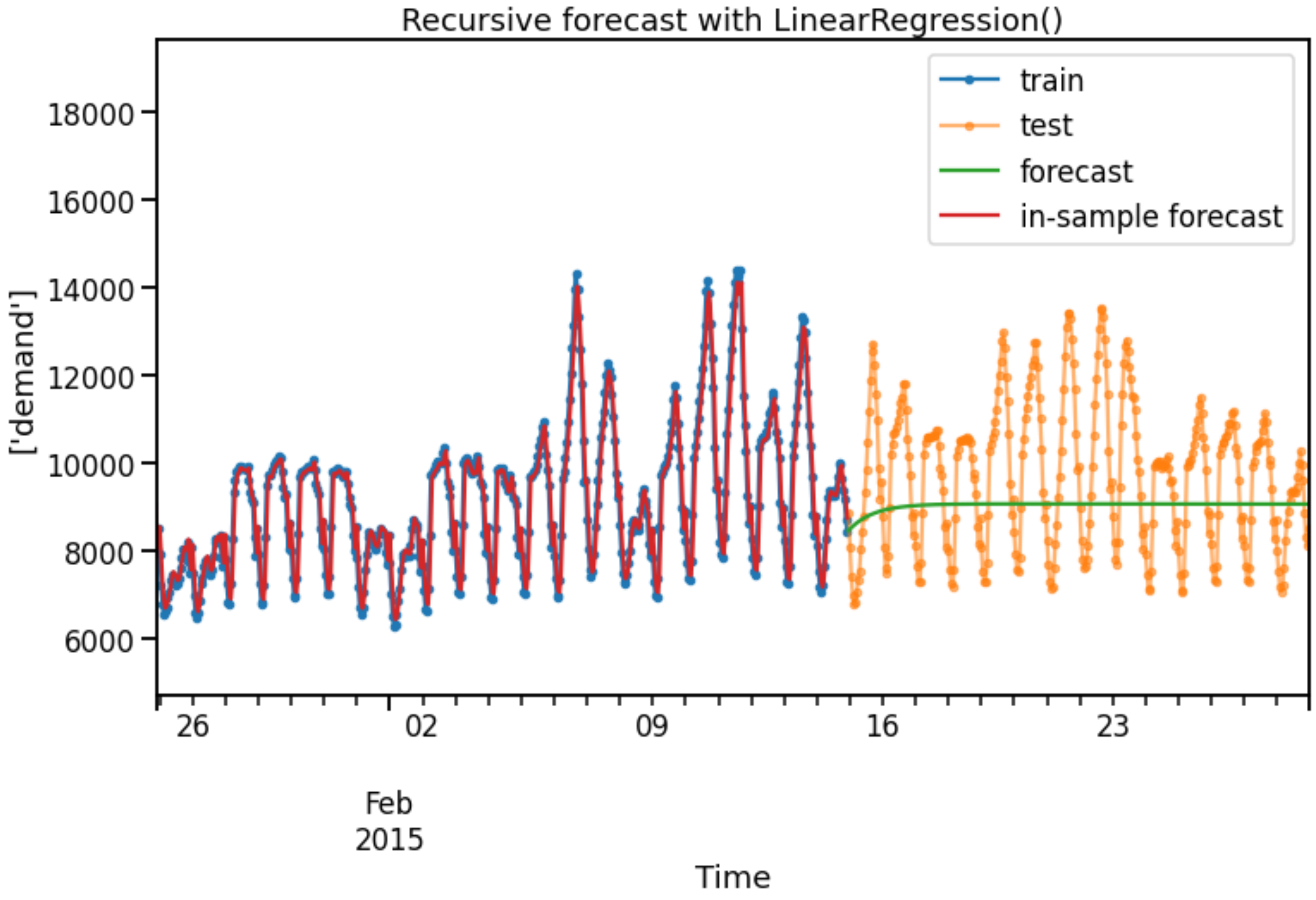
Features

- Trend features: t, t^2
- Lag of 1 month: y_{t-1}
- **Lag of 12 months: $\boldsymbol{y_{t-12}}$**

Model

$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 y_{t-1} + \boldsymbol{\beta_4 y_{t-12}}$$

Example: Electricity demand



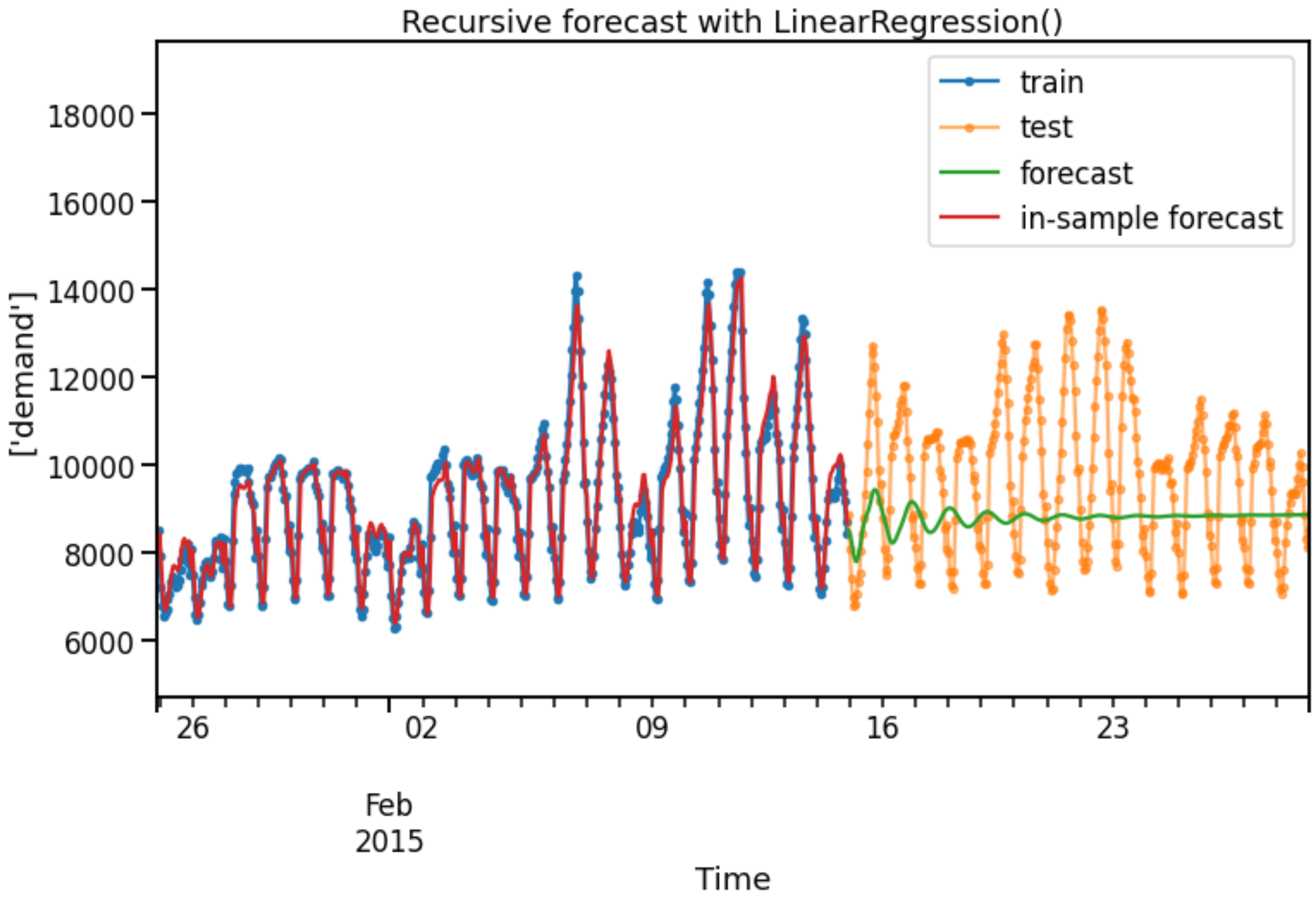
Features

- Trend features: t, t^2
- Lag of 1 hour: y_{t-1}

Model

$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 y_{t-1}$$

Example: Electricity demand



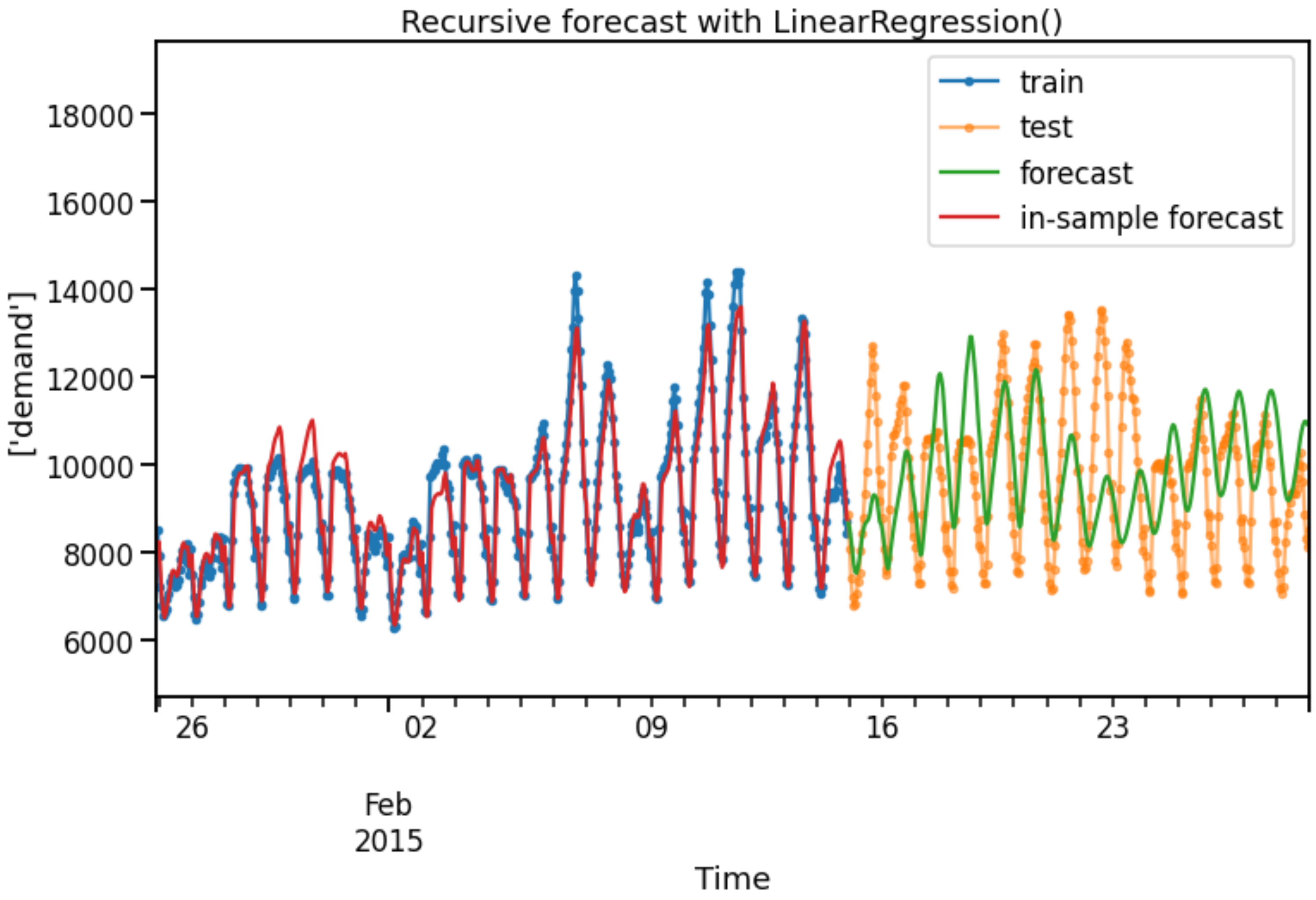
Features

- Trend features: t, t^2
- Lag of 1 hour: y_{t-1}
- **Lag of 24 hours: $\boldsymbol{y_{t-24}}$**

Model

$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 y_{t-1} + \boldsymbol{\beta_4 y_{t-24}}$$

Example: Electricity demand



Features

- Trend features: t, t^2
- Lag of 1 hour: y_{t-1}
- **Lag of 24 hours: $\boldsymbol{y_{t-24}}$**
- **Lag of 24*7 hours: $\boldsymbol{y_{t-24*7}}$**

Model

$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 y_{t-1} + \boldsymbol{\beta_4 y_{t-24}} + \boldsymbol{\beta_5 y_{t-24*7}}$$

Pros and cons

Pros

- Simple to create.

Pros and cons

Cons

- Creates missing data at the start of the time series.
  - Big problem for long seasonality (e.g., yearly) with high frequency time series (e.g., hourly).

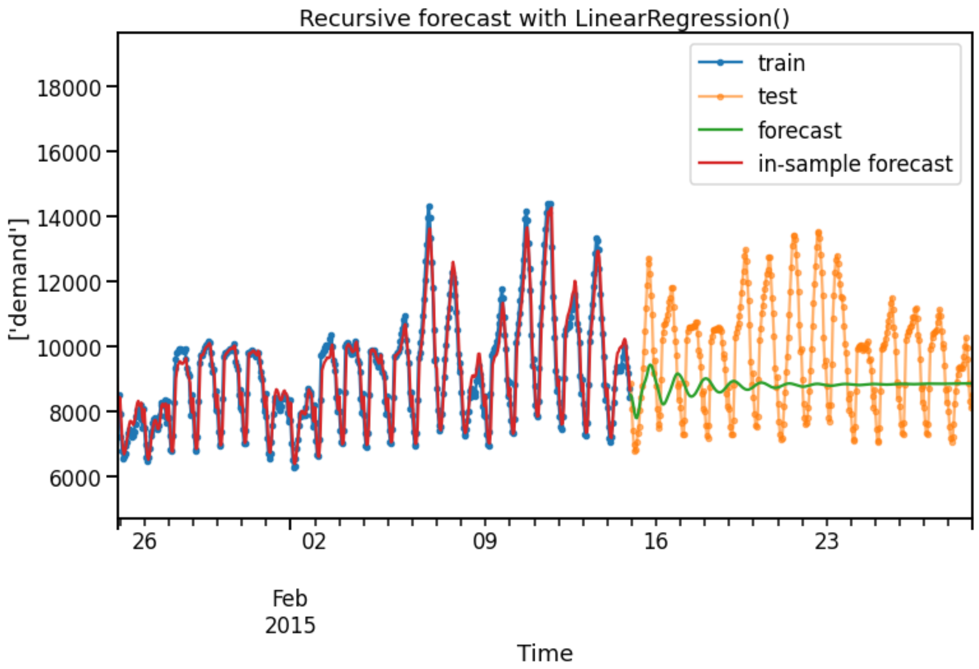
Date	Sales	Sales Lag 2
2020-02-12	23	NaN
2020-02-13	30	NaN
2020-02-14	35	23
2020-02-15	30	30
2020-02-16	?	35

Pros and cons

Cons

- Creates missing data at the start of the time series.
    - Big problem for long seasonality (e.g., yearly) with high frequency time series (e.g., hourly).

- Does not reliably capture seasonality.

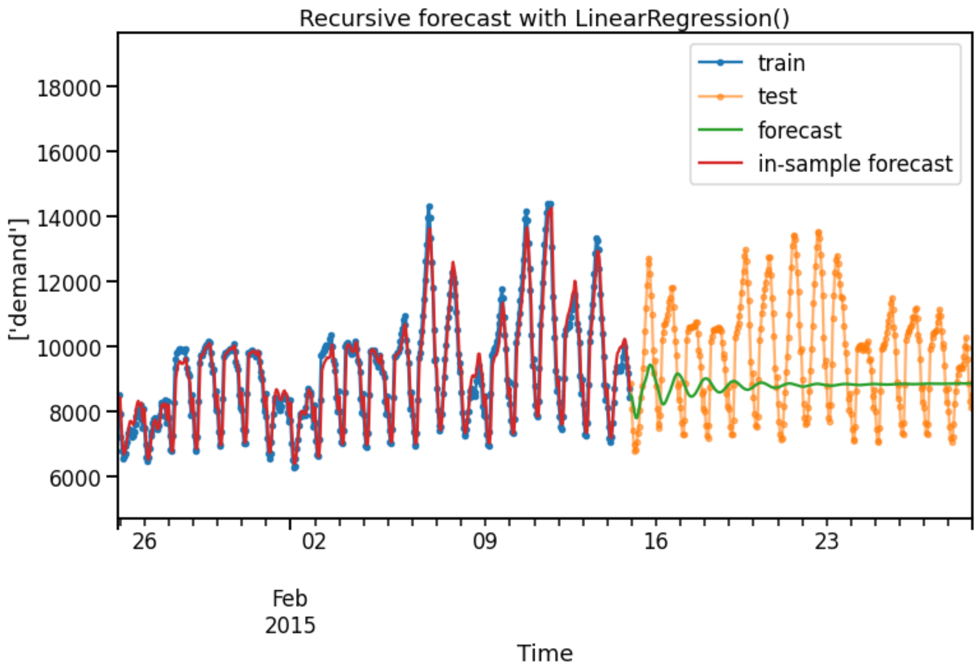


$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 y_{t-1} + \beta_4 y_{t-24}$$

Pros and cons

Cons

- Creates missing data at the start of the time series.
  - Big problem for long seasonality (e.g., yearly) with high frequency time series (e.g., hourly).
- Does not reliably capture seasonality.
- Lag features ignore the calendar. If seasonality is driven by the calendar then use datetime features or seasonal dummies.

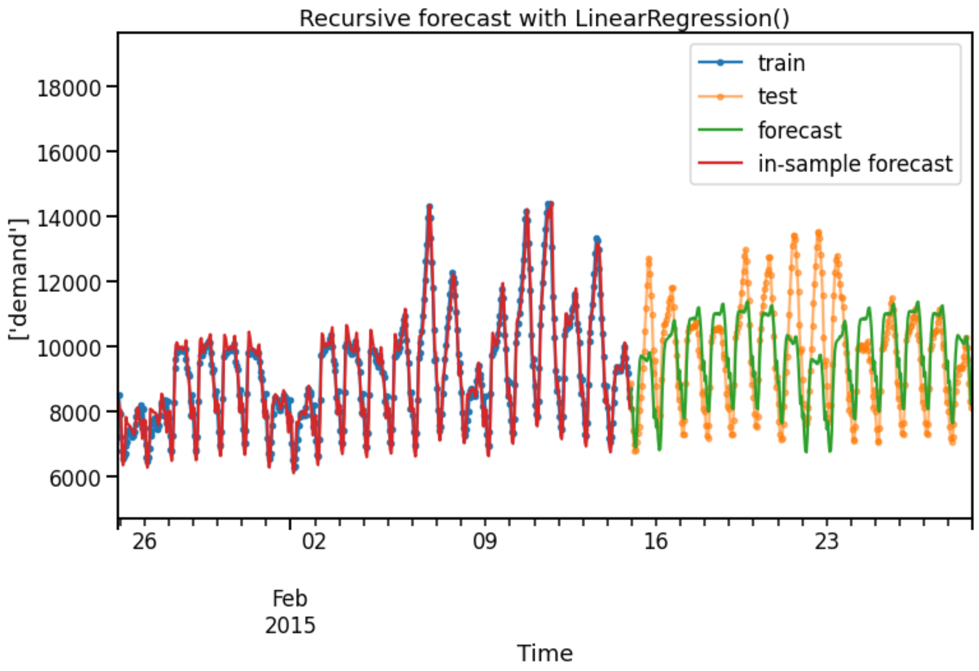


$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 y_{t-1} + \beta_4 y_{t-24}$$

Pros and cons

Cons

- Creates missing data at the start of the time series.
  - Big problem for long seasonality (e.g., yearly) with high frequency time series (e.g., hourly).
- Does not reliably capture seasonality.
- Lag features ignore the calendar. If seasonality is driven by the calendar then use datetime features or seasonal dummies.



$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 y_{t-1} +$$
$$\beta_5 is\_hour\_1 + \ldots$$
$$\beta_{28} is\_monday + \ldots$$

Pros and cons

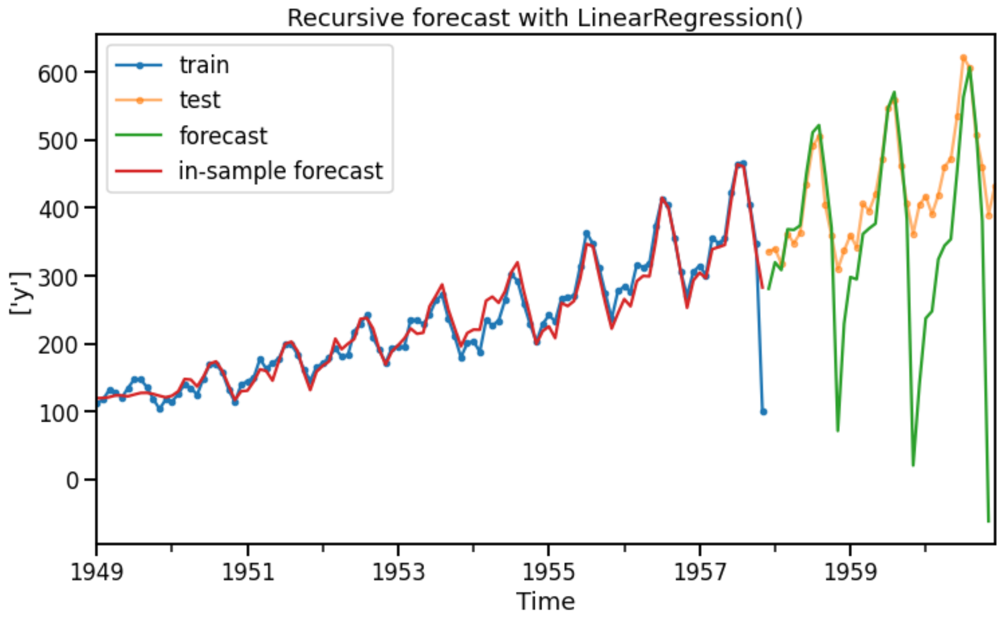
Cons

- Creates missing data at the start of the time series.
  - Big problem for long seasonality (e.g., yearly) with high frequency time series (e.g., hourly).

- Does not reliably capture seasonality.

- Lag features ignore the calendar. If seasonality is driven by the calendar then use datetime features or seasonal dummies.

- Creates a strong dependence on what happened exactly one seasonal period ago. So outliers or other unexpected behavior impacts predictions. Can cause unstable forecasts.



$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 y_{t-1} + \beta_4 y_{t-12}$$

Summary

Lag features can be used to capture seasonality.

The lag of the seasonal period is used.

Seasonal lags are simple but come with many cons.

It can be better to combine seasonal lags with other features or use other methods altogether to capture seasonality.