DYNAMIC REGRESSION MODELS

Dr. Aric LaBarr
Institute for Advanced Analytics

REGRESSION WITH ARIMA ERRORS

External Variables

- Predictor variables are used for variety of reasons:
  - Account for trend
  - Account for seasonality
  - External information make better forecasts

External Variables

- Predictor variables are used for variety of reasons:
  - Account for trend
  - Account for seasonality
  - External information make better forecasts

Holiday effects, economic variables, changes in policy, etc.

Incorporating Predictor Variables

- Regression with ARIMA errors:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \boxed{Z_t}$$

ARIMA model here!

Incorporating Predictor Variables

- Regression with ARIMA(1,0,1) errors:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \boxed{Z_t}$$

$$Z_t = \omega + \phi_1 Z_{t-1} + e_t + \theta_1 e_{t-1}$$

Incorporating Predictor Variables

- Regression with ARIMA(1,0,1) errors:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + Z_t$$

$$Z_t = \omega + \phi_1 Z_{t-1} + \boxed{e_t} + \theta_1 e_{t-1}$$

White noise

Incorporating Predictor Variables

- Regression with ARIMA errors:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \boxed{Z_t}$$

ARIMA model here!

- Many different names → Dynamic Regression, ARIMAX, Transfer Functions

INTERVENTION VARIABLES

Intervention (Event, Jump, Shift, Bang) Variable

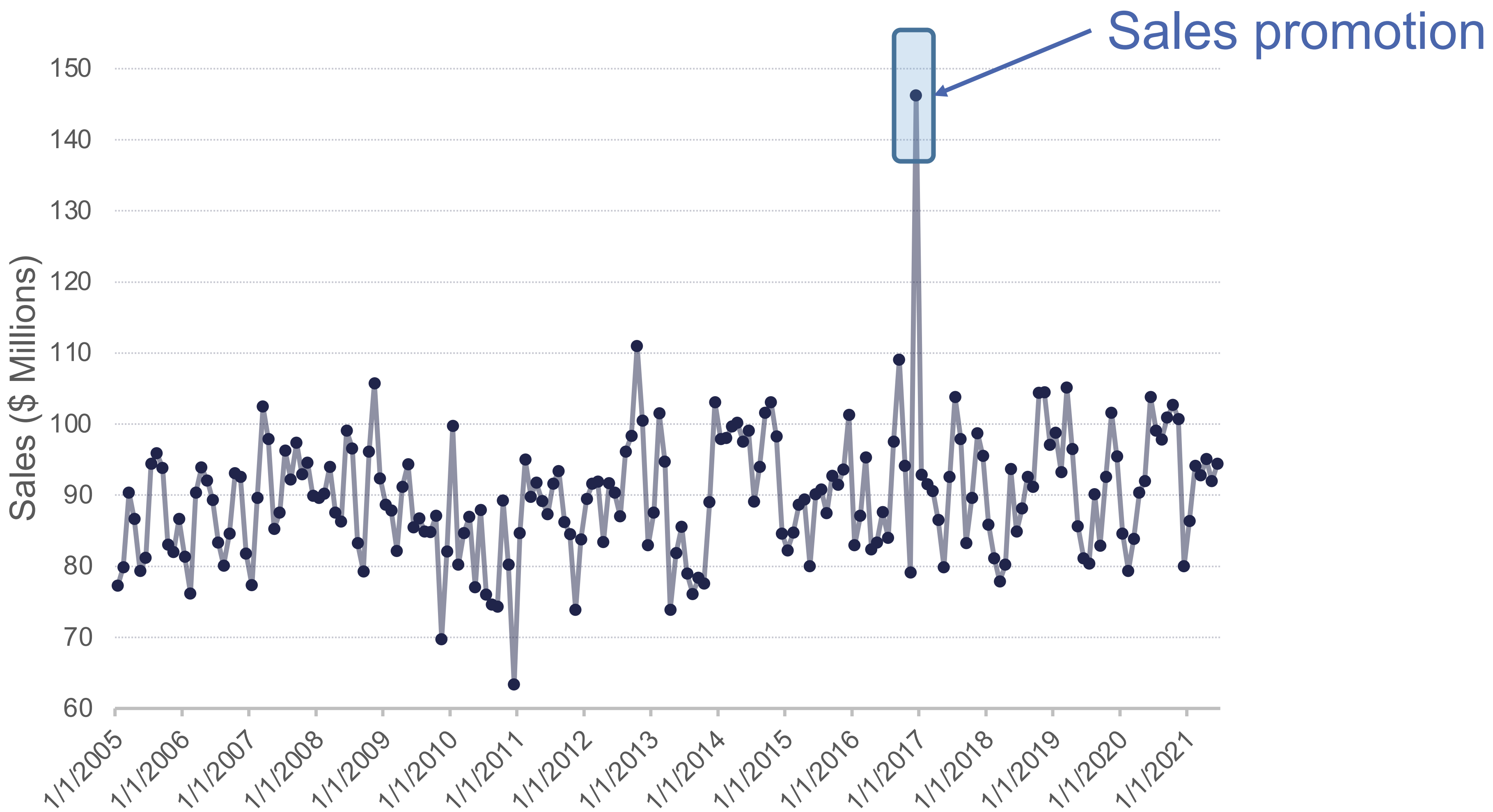
- **Intervention** variable – indicator variable that contains discrete values that flag the occurrence of an event affecting the response series.
- Uses:
  - Model and forecast the response series
  - Analyze the impact of the intervention.
  - Example – monthly revenues from the sale of a product with the implementation of a sales promotion.

- Accommodate **discrete shifts** in time series data through **intercept shifts**.

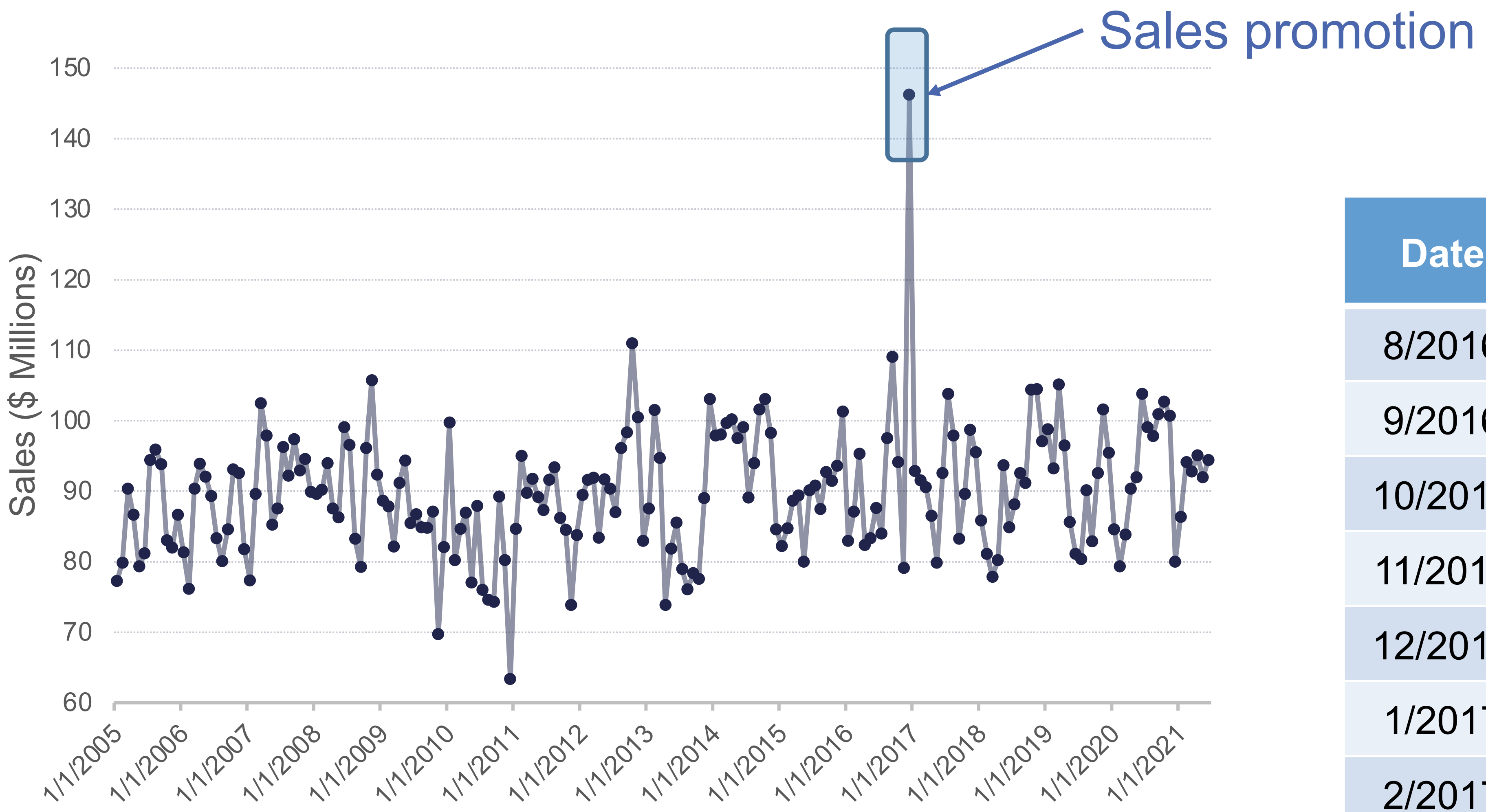
3 Types of Intervention Variables

- There are three common intervention variables:
  1. Point (or Pulse) Interventions
  2. Step Interventions
  3. Ramp Interventions

Point (Pulse) Intervention

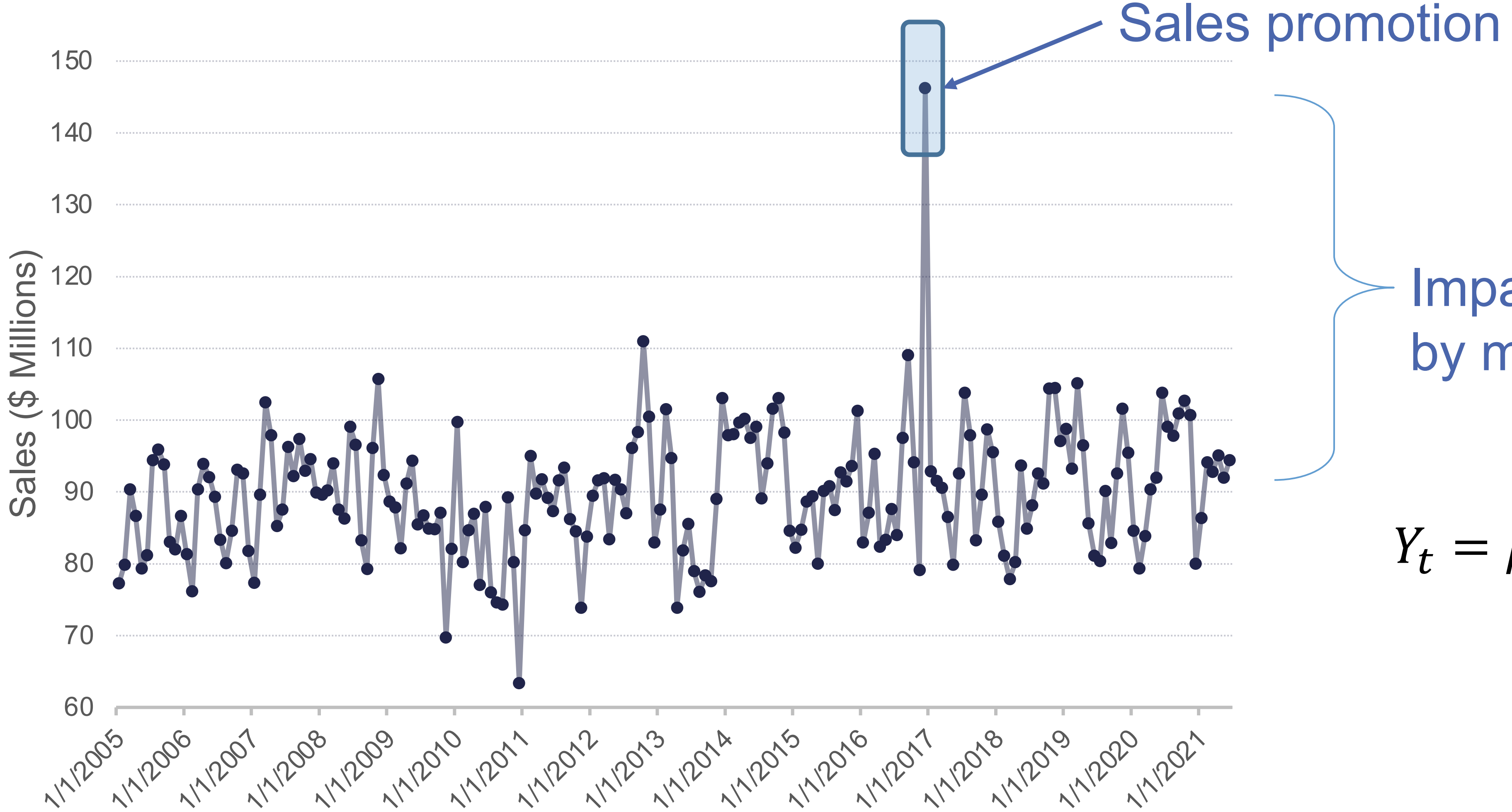


Point (Pulse) Intervention



Date	Intervention Variable I_t
8/2016	0
9/2016	0
10/2016	0
11/2016	0
12/2016	1
1/2017	0
2/2017	0

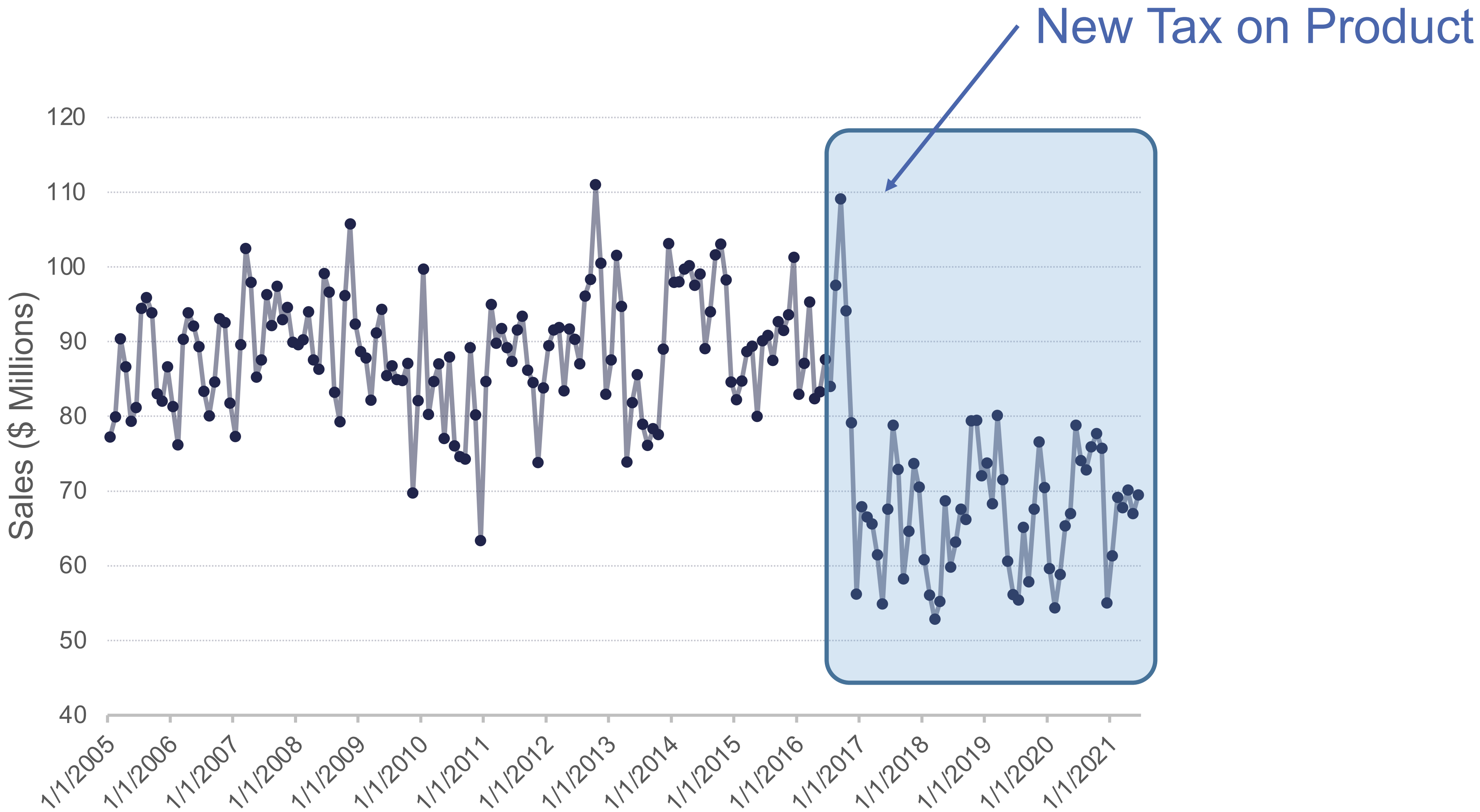
Point (Pulse) Intervention



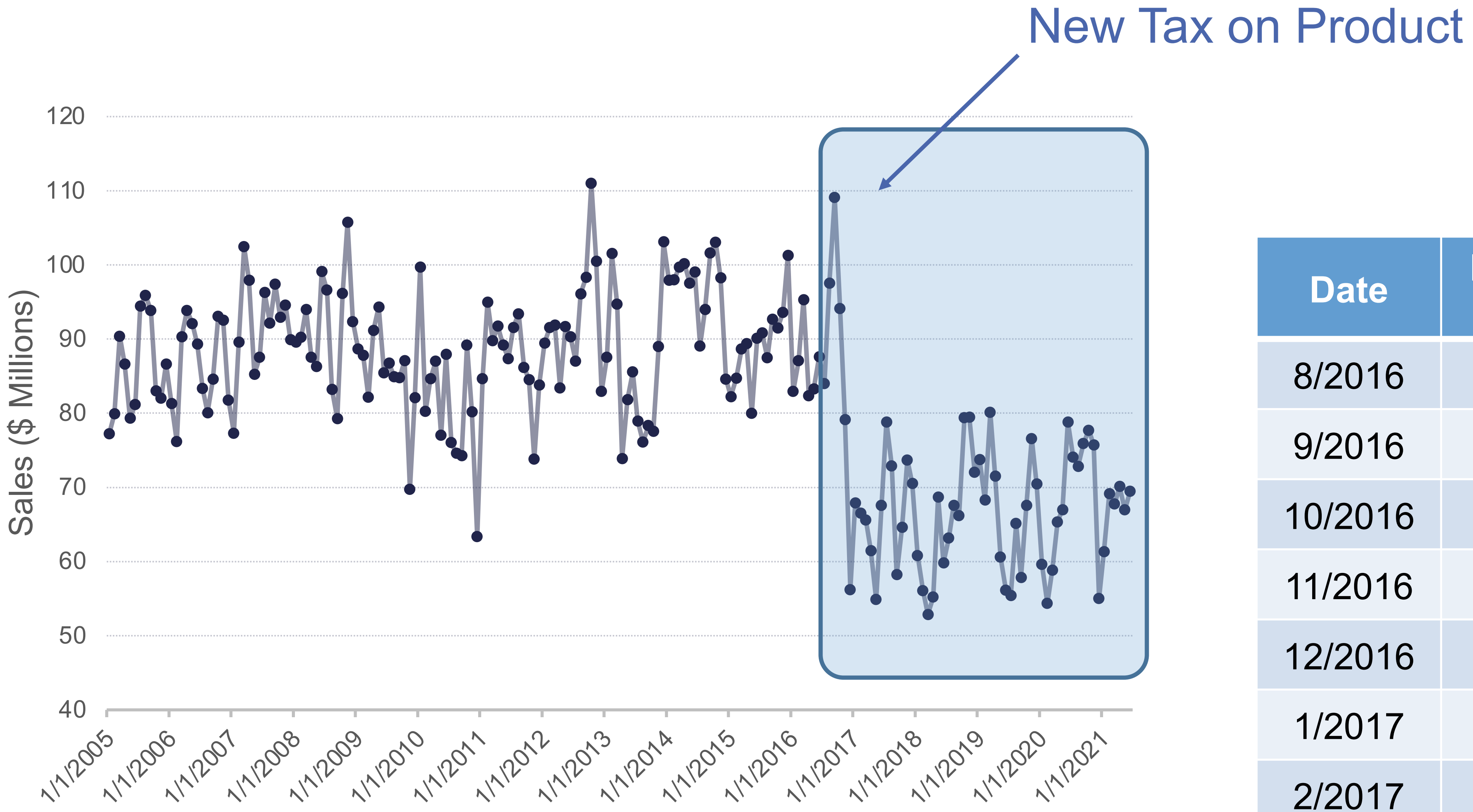
Impact measured by model coefficient

$$Y_t = \beta_0 + \boxed{\beta_1} I_t + Z_t$$

Step Intervention

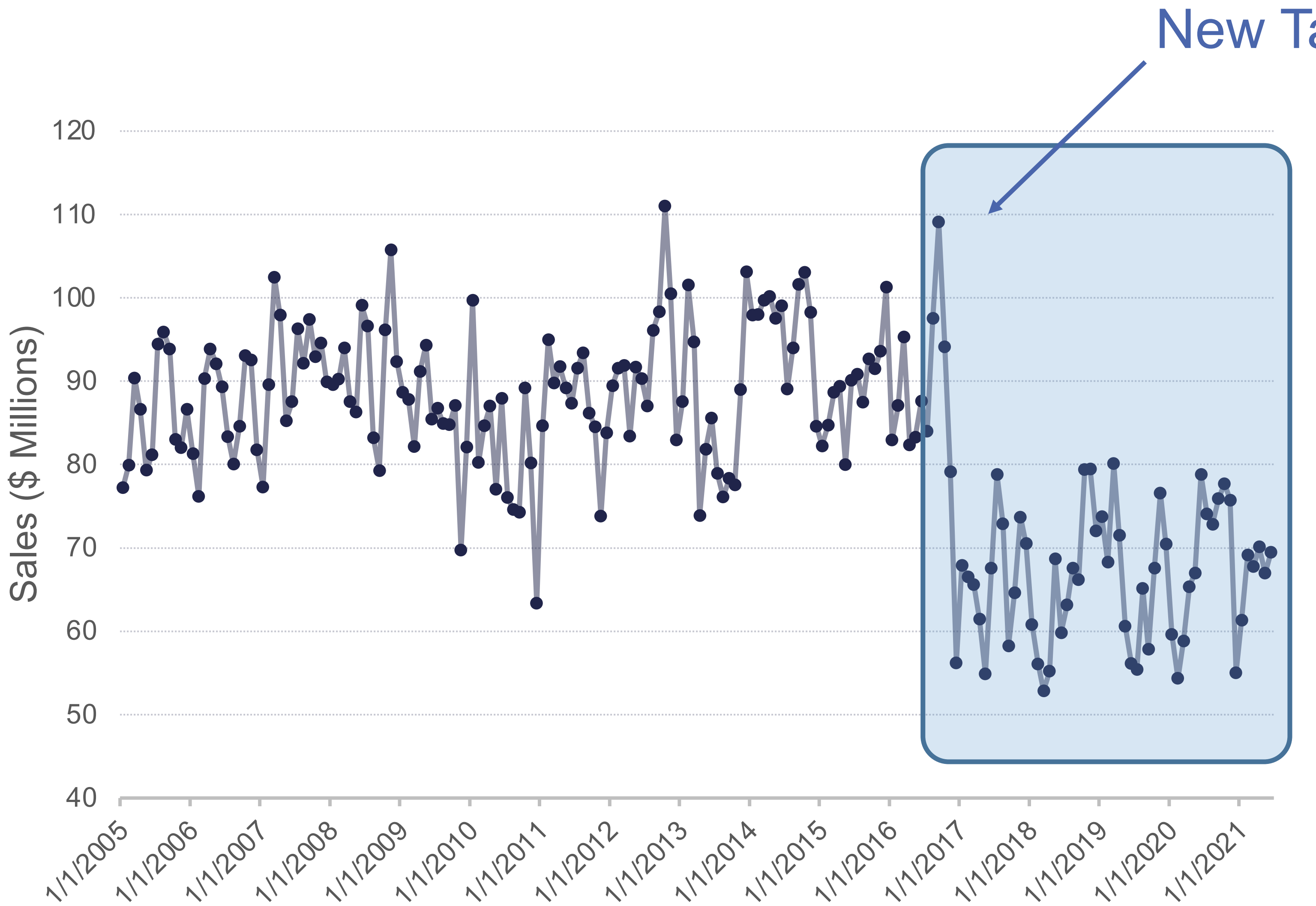


Step Intervention



Date	Intervention Variable I_t
8/2016	0
9/2016	0
10/2016	0
11/2016	0
12/2016	1
1/2017	1
2/2017	1

Step Intervention

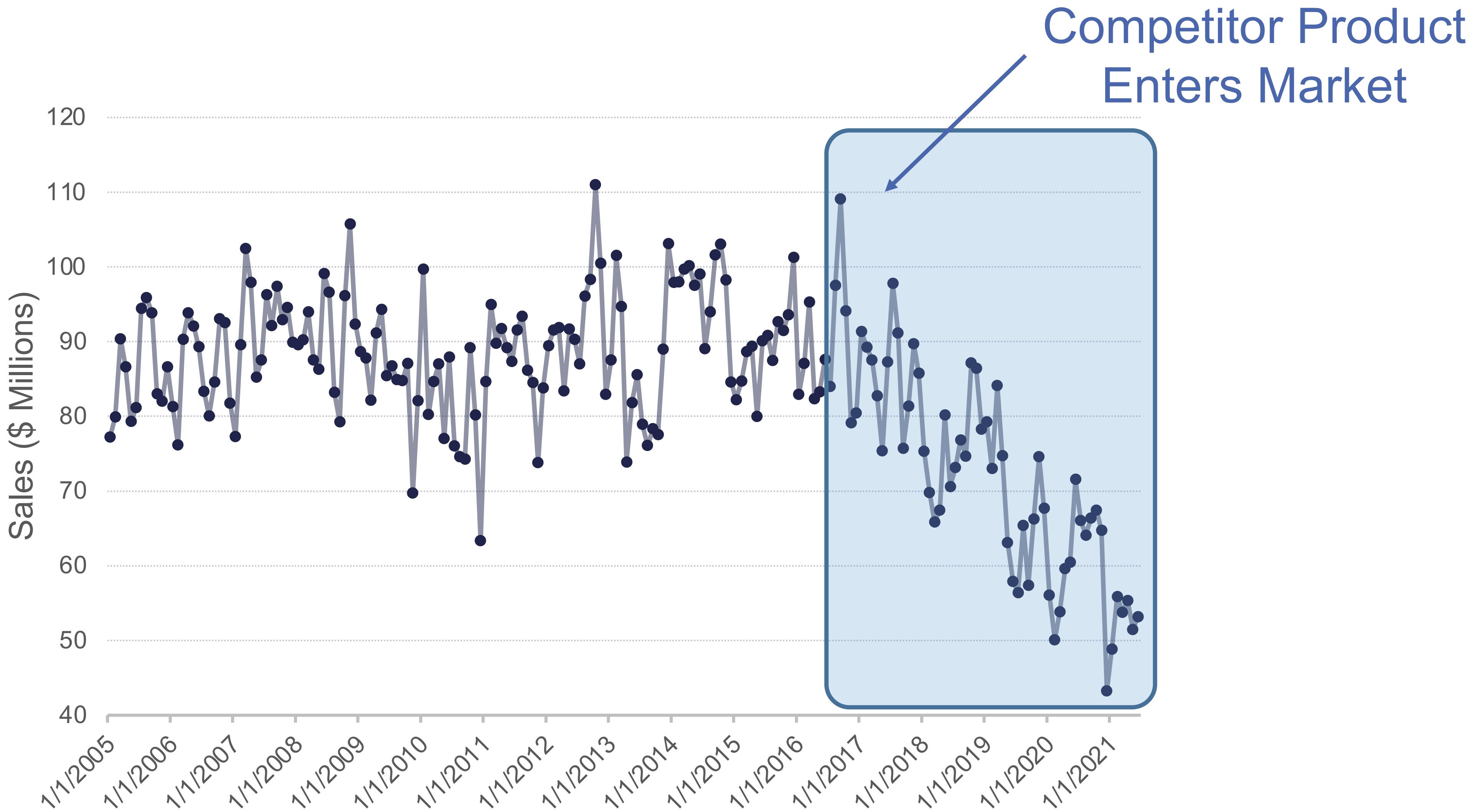


Impact measured by model coefficient

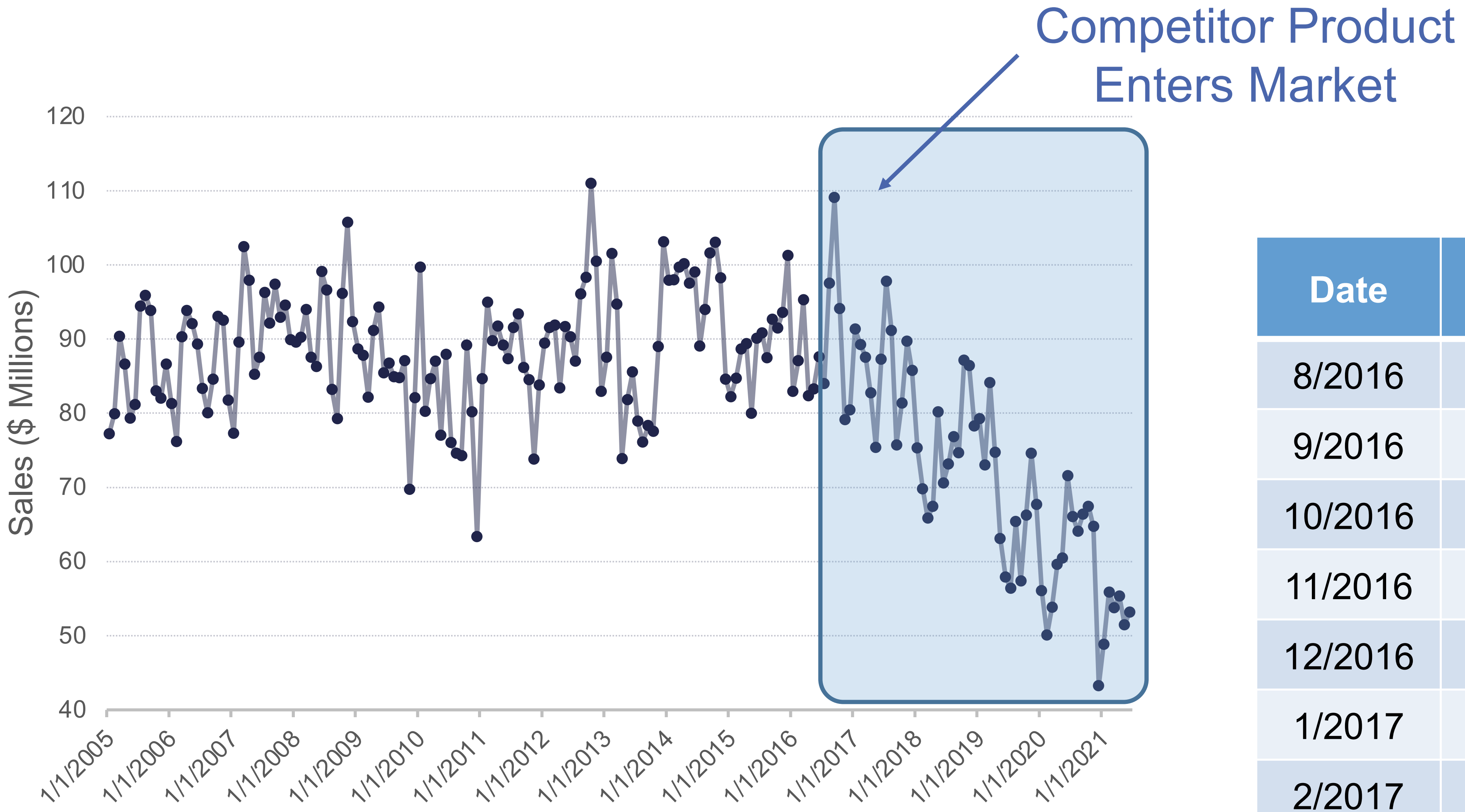
$$Y_t = \beta_0 + \boxed{\beta_1} I_t + Z_t$$

Step Intervention – Example

Ramp Intervention

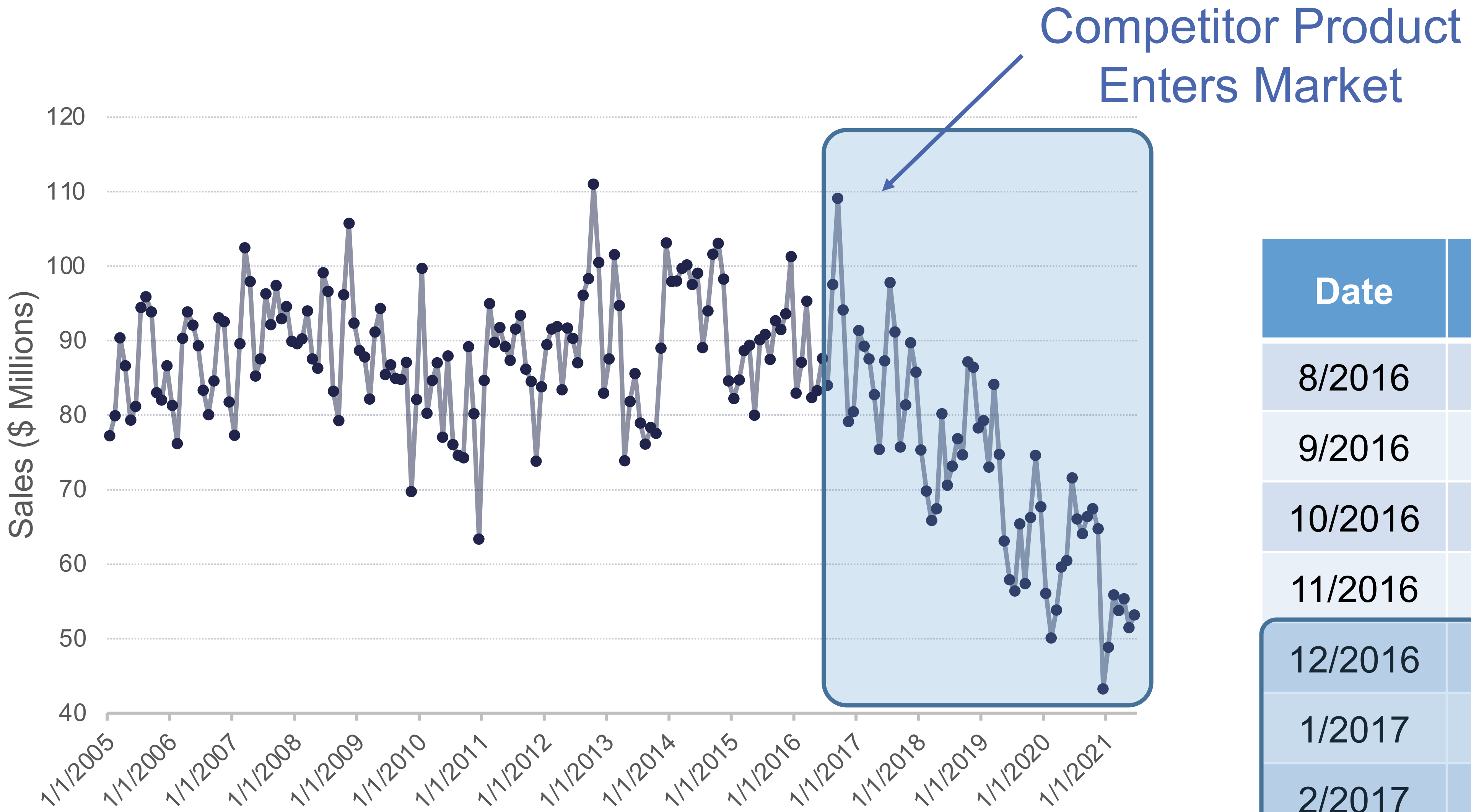


Ramp Intervention



Date	Intervention Variable I_t
8/2016	0
9/2016	0
10/2016	0
11/2016	0
12/2016	1
1/2017	2
2/2017	3

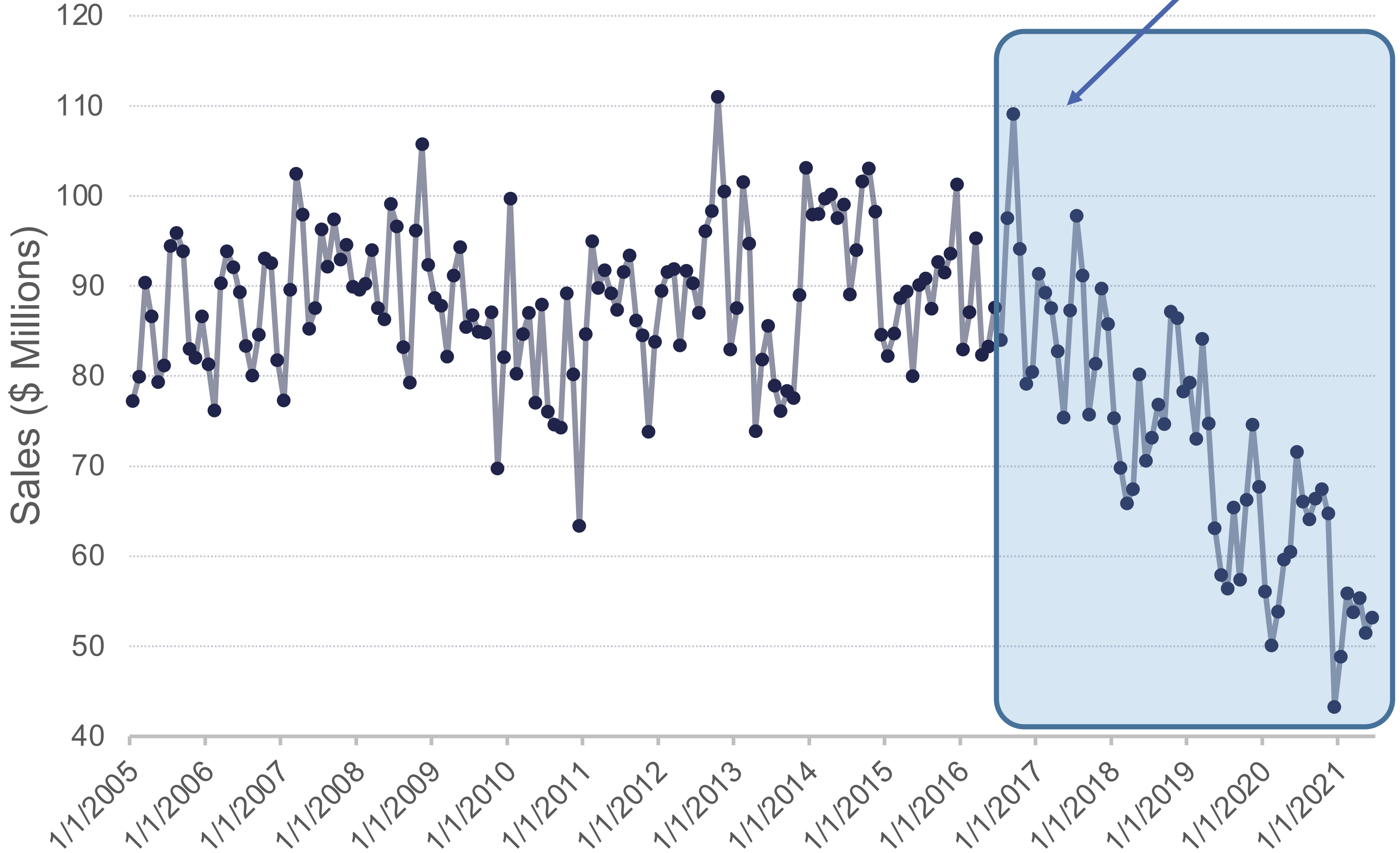
Ramp Intervention



Date	Intervention Variable I_t
8/2016	0
9/2016	0
10/2016	0
11/2016	0
12/2016	1
1/2017	2
2/2017	3

Distance from previous state

Ramp Intervention



Competitor Product Enters Market

Impact (new slope) measured by model coefficient

$$Y_t = \beta_0 + \boxed{\beta_1} I_t + Z_t$$

Point Intervention

```
Sep11 <- rep(0, 207)
Sep11[141] <- 1
```

```
Full.ARIMA <- auto.arima(training, seasonal = TRUE, xreg = Sep11, method = "ML")
```

PREDICTOR VARIABLES

Including External Variables

- Most forecasting models also need to account for explanatory variables such as price, advertising, or income.
- These models have many names – ARIMAX, dynamic regression models, transfer functions, etc.

Including External Variables

- Most forecasting models also need to account for explanatory variables such as price, advertising, or income.
- These models have many names – ARIMAX, dynamic regression models, transfer functions, etc.
- Already have done this…
    - Trend models
    - Seasonal dummy variables
    - Harmonic regression
    - Intervention variables

Including External Variables

- Most forecasting models also need to account for explanatory variables such as price, advertising, or income.
- These models have many names – ARIMAX, dynamic regression models, transfer functions, etc.

- Often, there are **lagged impacts** as well as (or instead of) immediate impacts - that is past values of explanatory variables can be important.

How Many Lags?

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{1,t-1} + \cdots + \beta_k X_{1,t-k} + Z_t$$

- Multiple ways to evaluate how many lags of a predictor variable you need in a model
  - Cross-correlation functions and pre-whitening of series
  - Evaluate many different lag combination models with AIC/BIC on validation set.

How Many Lags?

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{1,t-1} + \cdots + \beta_k X_{1,t-k} + Z_t$$

- Multiple ways to evaluate how many lags of a predictor variable you need in a model
  - Cross-correlation functions and pre-whitening of series
    - Time consuming
    - Requires modeling of the predictor variables
    - Best used for small number of predictors
  - Evaluate many different lag combination models with AIC/BIC on validation set.

How Many Lags?

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{1,t-1} + \cdots + \beta_k X_{1,t-k} + Z_t$$

- Multiple ways to evaluate how many lags of a predictor variable you need in a model
  - Cross-correlation functions and pre-whitening of series
  - Evaluate many different lag combination models with AIC/BIC on validation set.
    - More efficient
    - Handles many variables much easier
    - Similar in accuracy of the “elegant” first approach

Adding Lags to Model

```
Sep11 <- rep(0, 207)
Sep11[141] <- 1

Sep11.L1 <- rep(0, 207)
Sep11.L1[142] <- 1

Sep11.L2 <- rep(0, 207)
Sep11.L2[143] <- 1

...

Sep11.L6 <- rep(0, 207)
Sep11.L6[147] <- 1

Anniv <- rep(0, 207)
Anniv[153] <- 1

Full.ARIMA <- auto.arima(training, seasonal = TRUE, xreg = cbind(Sep11, Sep11.L1, Sep11.L2, Sep11.L3,
                                                                 Sep11.L4, Sep11.L5, Sep11.L6, Anniv),
                         method = "ML")
summary(Full.ARIMA)
```

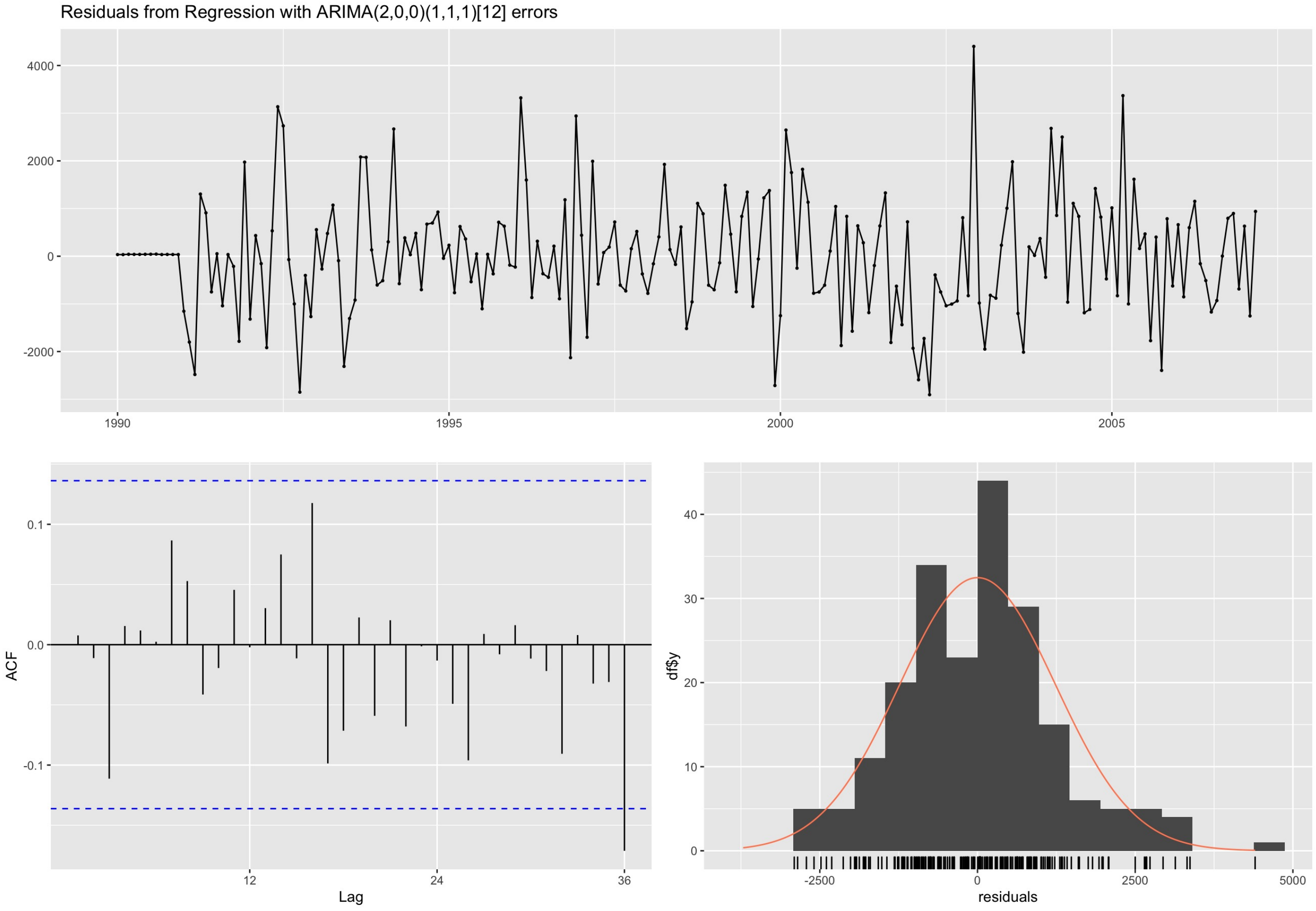
Adding Lags to Model

```
## Series: training
## Regression with ARIMA(2,0,0)(1,1,1)[12] errors
##
## Coefficients:
##          ar1     ar2    sar1    sma1     drift       Sep11     Sep11.L1
##       0.6298  0.2207  0.1926  -0.696  124.7562  -17400.420  -12116.115
## s.e.  0.0714  0.0726  0.1143   0.081   21.1622    1162.401    1271.324
##         Sep11.L2   Sep11.L3   Sep11.L4   Sep11.L5   Sep11.L6       Anniv
##        -8076.014  -7670.030  -4344.649  -2173.140  -749.6299  -2306.1784
## s.e.    1387.179   1427.366   1403.914   1271.271  1105.3247    998.2399
##
## sigma^2 estimated as 1736410:  log likelihood=-1673.71
## AIC=3375.42   AICc=3377.75   BIC=3421.24
##
## Training set error measures:
##                    ME     RMSE      MAE        MPE     MAPE      MASE
## Training set 1.076103 1235.596 944.9564 -0.0820269 1.937634 0.3509825
##                     ACF1
## Training set 0.007704655
```

Seasonal ARIMA

```
checkresiduals(Full.ARIMA)
```

```
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,0,0)(1,1,1)[12]
## Q* = 16.046, df = 11, p-value = 0.1394
##
## Model df: 13.   Total lags used: 24
```



FORECASTING

Forecasting with External Variables

- Forecasting in time series with only lagged values of the target variable is easy – recursive formula that just feeds into itself.
- Forecasting in time series with external variables is much trickier.
  - What are the future values of the external variables?

Forecasting with External Variables

- Forecasting in time series with only lagged values of the target variable is easy – recursive formula that just feeds into itself.
- Forecasting in time series with external variables is much trickier.
  - What are the future values of the external variables?

- Known future values (interventions)
- External estimates of future values
- Need to forecast future values ourselves

Forecasting

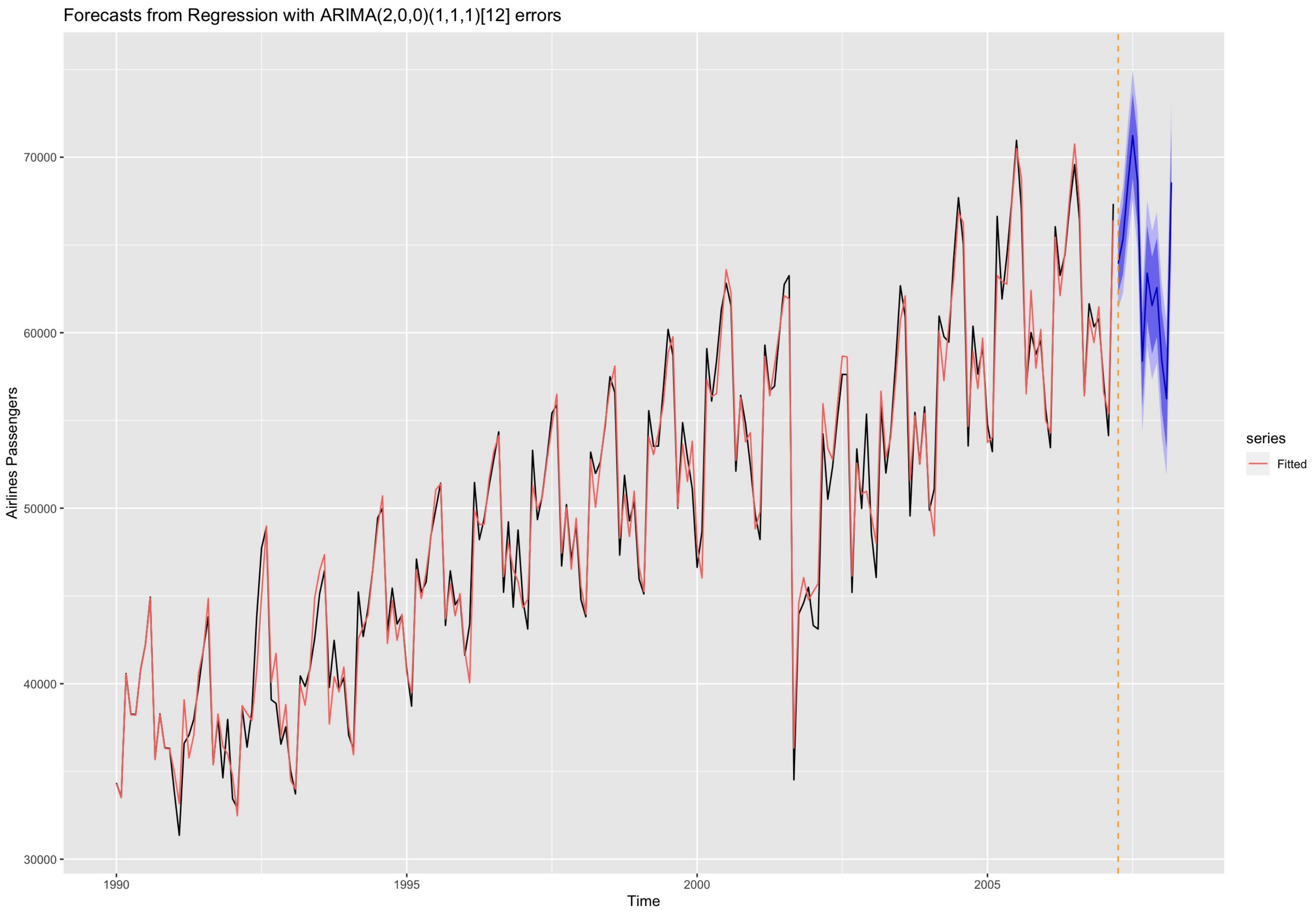
```
Sep11 <- rep(0, 12)
Sep11.L1 <- rep(0, 12)
Sep11.L2 <- rep(0, 12)
Sep11.L3 <- rep(0, 12)
Sep11.L4 <- rep(0, 12)
Sep11.L5 <- rep(0, 12)
Sep11.L6 <- rep(0, 12)
Anniv <- rep(0, 12)

forecast::forecast(Full.ARIMA, xreg = cbind(Sep11, Sep11.L1, Sep11.L2, Sep11.L3, Sep11.L4, Sep11.L5,
                                            Sep11.L6, Anniv),
                   h = 12)

Full.ARIMA.error <- test - forecast::forecast(Full.ARIMA, xreg = cbind(Sep11, Sep11.L1, Sep11.L2, Sep11.L3, S
ep11.L4, Sep11.L5, Sep11.L6, Anniv), h = 12)$mean

Full.ARIMA.MAE <- mean(abs(Full.ARIMA.error))
Full.ARIMA.MAPE <- mean(abs(Full.ARIMA.error)/abs(test))*100
```

Forecasting



Model Evaluation on Test Data

Model	MAE	MAPE
HW Exponential Smoothing	1134.58	1.76%
Seasonal ARIMA	1229.21	1.89%
Dynamic Regression ARIMA	1180.99	1.80%

?