Business Forecasting

Seasonal Dummy Variables

Regression in Time Series methods

Regression is a useful tool to estimate models where the target variable (Y) can be explained by "causal" independent variables (X)

eg Sales (Y) explained by **Price**, **Promotional Spend**, **Income**, **Interest rates**, **Competitors** (X's)

Regression can also be used to estimate models more indicative of **time series models**

Quasi- explanatory variables (time, seasonal dummies, lagged dependent variables) can be used instead of regular explanatory variables.

Trend Extrapolations

Trend extrapolation based on trend equation

$$Y_t = f\,(\text{time})$$

For a linear trend → $Y_t = \beta_o + \beta_1 * t$

The time index "t" acts as a quasi explanatory variable to help explain/forecast Y_t with regression used to estimate β_o , β_1

Estimated equation **($Y_t = b_o + b_1 * t$)** used to forecast Y_t based on future value of time index

Often used as quick way of **generating forecasts of independent variables** needed in regression forecasts of the target variable.

Seasonality

A quasi –explanatory model can be constructed to extrapolate seasonal time series

Dummy variables included to model seasonal effects

Dummy variables typically modelled as **dichotomous** variables with values of 0 and 1

If there are **p seasons** then **(p-1) dummies** are needed to model seasonality

General equation for Y_t (including trend) is

$$Y_t = f\,(\text{time}, D_1, D_2 \ldots\ldots\ldots D_{p-1})$$

Seasonal Dummy Variables

Each seasonal dummy is modelled as follows;

$D_i = 1$ if observation is season i

$D_i = 0$ otherwise

For **quarterly** time series, there are **3** dummies needed, for **monthly** there are **11** dummies

The season that is omitted is arbitrary. The results will be however, **relative to the omitted season.** Once this is accounted for, the estimated equation will be identical.

Significance tests will be relative to the omitted season. A quarterly example follows

Seasonal Example

Hotel Takings (000's)

MACQUARIE University

Time	Quarter	Sales	D1(Dec)	D2(Mar)	D3(Jun)
1	Dec-11	668532	1	0	0
2	Mar-12	621399	0	1	0
3	Jun-12	553849	0	0	1
4	Sep-12	664512	0	0	0
5	Dec-12	759603	1	0	0
6	Mar-13	691864	0	1	0
7	Jun-13	627765	0	0	1
8	Sep-13	720863	0	0	0
9	Dec-13	749901	1	0	0
10	Mar-14	749365	0	1	0
11	Jun-14	674906	0	0	1
12	Sep-14	777192	0	0	0
13	Dec-14	809598	1	0	0
14	Mar-15	801351	0	1	0
15	Jun-15	732327	0	0	1
16	Sep-15	839229	0	0	0
17	Dec-15	924637	1	0	0

Seasonal Dummy Variables (cont).

In the previous example (quarterly time series) there are 3 dummies constructed (D_1, D_2 ,D_3). The September quarter has been excluded and is the base or reference category

The 3 dummies account for seasonal impacts

Assuming additive seasonality (and linear trend) the following model can be posited;

$$Y_t = \beta_o + \beta_1 * t + \alpha_1 * D_1 + \alpha_2 * D_2 + \alpha_3 * D_3 + \varepsilon_t$$

This model can be estimated using regression

Dummy Variables -Four Models in One

$$Y_t = \beta_o + \beta_1 * t + \alpha_1 * D_1 + \alpha_2 * D_2 + \alpha_3 * D_3 + \varepsilon_t$$

For December ($D_1 = 1$, $D_2 = 0$ $D_3=0$)

Theoretically for **December** the model is

$$Y_t = \beta_o + \beta_1 * t + \alpha_1 * 1 + \alpha_2 * 0 + \alpha_3 * 0 + \varepsilon_t$$

$$Y_t = (\underline{\alpha_1} + \beta_o) + \beta_1 * t + \varepsilon_t$$

For March ($D_1 = 0$, $D_2 = 1$ $D_3=0$)

Theoretically for **March** the model is

$$Y_t = \beta_o + \beta_1 * t + \alpha_1 * 0 + \alpha_2 * 1 + \alpha_3 * 0 + \varepsilon_t$$

$$Y_t = (\underline{\alpha_2} + \beta_o) + \beta_1 * t + \varepsilon_t$$

Four Models in One (cont)

MACQUARIE University

$$Y_t = \beta_o + \beta_1 * t + \alpha_1 * D_1 + \alpha_2 * D_2 + \alpha_3 * D_3 + \varepsilon_t$$

For June (D_1 = 0, D_2 = 0 D_3=1)

Theoretically for **June** the model is

$$Y_t = \beta_o + \beta_1 * t + \alpha_1 * 0 + \alpha_2 * 0 + \alpha_3 * 1 + \varepsilon_t$$

$$Y_t = (\alpha_3 + \beta_o) + \beta_1 * t + \varepsilon_t$$

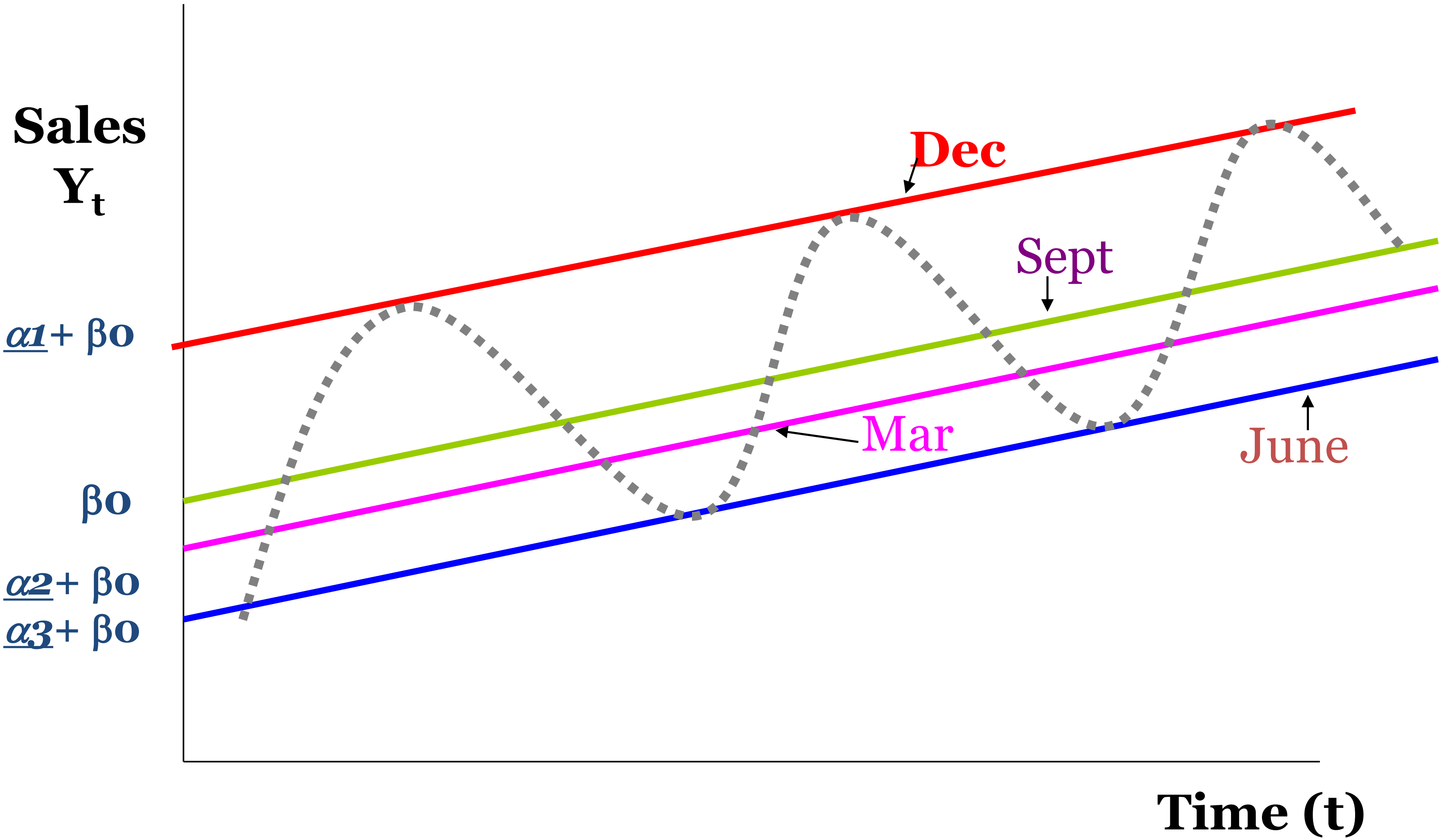
For September (D_1 = 0, D_2 = 0 D_3=0)

Theoretically for **September** the model is

$$Y_t = \beta_o + \beta_1 * t + \alpha_1 * 0 + \alpha_2 * 0 + \alpha_3 * 0 + \varepsilon_t$$

$$Y_t = \beta_o + \beta_1 * t + \varepsilon_t \text{ (base model)}$$

Additive Seasonality with Dummies



Assessing Seasonality

Once the model is formulated and dummies constructed, regression can be used to estimate β_0 , β_1 and α_1 , α_2 and α_3 (estimates $\mathbf{b_0}$, $\mathbf{b_1}$ and $\mathbf{a_1}$, $\mathbf{a_2}$ **and** $\mathbf{a_3}$)

Normal regression diagnostic checks apply to determine model adequacy

Seasonality assessed **individually** (**relative to omitted season**) and **collectively (relative to excluding all seasonal dummies)**

Individual season – **t tests**; Overall.- **F test**

Regression Results -Example

Regression Statistics	
Multiple R	0.985018076
R Square	0.97026061
Adjusted R Square	0.962330106
Standard Error	19720.11211
Observations	20

ANOVA

	df	SS	MS	F
Regression	4	1.90312E+11	47578020646	122.3453905
				Sig F :
Residual	15	5833242324	388882821.6	2.93670957E-11
Total	19	1.96145E+11		

	Coefficients	Standard Error	t Stat	P-value
Intercept	600807.8125	12855.94023	46.73386791	1.15315E-17
D1(Dec)	42981.27688	12689.43629	3.387169918	0.004063358
D2(Mar)	-1570.92875	12569.15456	-0.124982849	0.902196759
D3(Jun)	-93361.0744	12496.42981	-7.471019786	1.97878E-06
Time	15407.22563	779.5058747	19.76537461	3.7433E-12

Assessing the Results

R^2 is high (0.97) (but typically will be for time series data with trends)

Model overall is significant (**Sig F stat < 0.05**)

Time (t) appears a significant explanatory variable (**p value for t stat is < 0.05**)

Normal diagnostic checks to determine model adequacy employed before making further inferences

In addition to the normal checks, residuals should be **categorised and examined by season**

Assessing Seasonality

Assuming all the normal diagnostic checks suggest the model is adequate, test seasonality by the following:

Individual seasonal variables- December, June appear significant explanatory variables but **March** not (**p-values on respective t stats**)

However, these are **relative to omitted season (Sep) not absolute.** Assess seasonal variable significance **collectively**

Collective test of seasonality using a **modified F test**; **compare R^2** from **separate regressions** with and without seasonal dummies

Interpreting Results

Assuming all the normal diagnostic checks suggest the model is adequate;

Estimated Equation

$$E(Y_t) = 600{,}807.81 + 15407.22 * t + 42981.27 * D_1 + -1570.928 * D_2 + -93361.07 * D_3$$

Base Sales: 600,807.81 (Average Sales in period 1)

For every one unit increase in time (t) (ie progression from one quarter to the next) there will be an increase of **15407.22** to base sales

This is the slope of an estimated trend line for this time series (**the trendline that applies to all seasons is identical**)

Interpreting Results (cont)

Interpreting Seasonal coefficients;

December: (**relative to September**) Sales in Dec will be **42981.27 higher**

March: (**relative to September**) Sales in March will be **1570.928 lower**

June: (**relative to September**) Sales in Dec will be **93361.07 lower**

Coefficients can be considered as **additive seasonality indexes**
However, they are again **relative to the omitted season** (Sept)

Overall Seasonality

Compare R^2 **with (wi) seasonal dummies** to R^2 **without (wo) seasonal dummies**

In this example R^2 (with) = **0.9702**, R^2 without = **0.7204** (Difference ≈ 0.25)

F test is $\{(R^2_{(wi)} - R^2_{(wo)})/j\} / \{(1 - R^2_{(wi)})/(n-k-1)_{(wi)}\}$

J = number of dummies in model; k = number variables including dummies in full model

In this case, $F = \{0.25/3\}/\{0.0298/15\} \approx 41.94$

p-value ≈ 0 (Critical F (0.05, 3, 15) = 3.287)

Since **p-value <0.05** >>**Reject null** (Null = all coefficients (α_i) on seasonal dummy variables = 0). Hence, **seasonality appears present**

Forecasting

$E(Y_t) = 600{,}807.81 + 15407.22 * t + 42981.27 * D_1 + -1570.928 * D_2 + -93361.07 * D_3$

Forecasting with the model; Suppose we wish to forecast for Dec 2016 (period 21) and March 2017 (period 22)

<u>Dec 2007</u> $= 600{,}807.81 + 15407.22 * 21 + 42981.27 * 1 + -1570.928 * 0 + -93361.07 * 0$ $= \underline{967{,}340.7}$

$(D_1 = 1,\ D_2 = 0,\ D_3 = 0 \quad t = 21)$

<u>Mar 2008</u> $= 600{,}807.81 + 15407.22 * 22 + 42981.27 * 0 + -1570.928 * 1 + -93361.07 * 0$ $= 9\underline{38{,}195.\ 72}$

$(D_1 = 0,\ D_2 = 1,\ D_3 = 0 \quad t = 22)$