Adjustments, Transformations, Features

Handout

Adjustments

Calendar Adjustments

Number of Days Variations

- We may have to look at the difference in quantity etc based on the total number of days in the month

Trading Day Variations

- Due to the changing number of time each day of the week occurs in a month
- **Possible way of adjusting for trading day variation:**

  Let $\{y_t\}$ be a monthly time series that is a total of some variable for each month, e.g. visitor arrivals

  Let $y_t' = \frac{y_t}{\text{no. days of days in month t}} \times 30.45375$

  - Note that this upscales it such that it is upscaled to the average number of days in a month (365/12)

  - **Assumption**: $y_t' = T_t + S_t + C_t + H_t + e_t$
    - T_t - Trend Effect
    - S_t - Seasonal Effect
      - Effect of the different periods, Month etc
    - C_t - Trading Day Effect
      - Effect of it being Mon, Tue, Wed etc
      - **Constraints**: $C_t = \sum_{i=1}^{7} \alpha_i d_{it}$
        - where $\sum_{i=1}^{7} \alpha_i = 0$
        - d_{it} - (Fraction of month t that is day i) $\times 30.45375$
          - This term kind of is the amount of effect that each day will have for that current month after scaling it to be the average number of days in every month of the year
        - Estimations of α_i :
          1. Decompose y_t' into $y_t' = T_t + S_t + e_t'$
          2. Obtain $w_t = y_t' - \hat{T}_t - \hat{S}_t$

3. Regress w_t on $d_{1t}, d_{2t}, \cdots, d_{7t}$
4. Obtain estimates $\hat{\alpha_1}, \cdots, \hat{\alpha_7}$
5. Calendar adjusted time series is $y_t'' = y_t' - \sum_{i=1}^{7} \hat{\alpha}_i d_{it}$

- H_t - Holiday Effect
- e_t - Noise, unexplainable variation in the data
  - What we hope to see is that e_t is white noise

▼ Holiday Variations

- Due to presence / absence of a holiday in a month
- Note that for some the day is always the same, the month could be the same or the day of the week is the only one that is the same. There are many different possible ways in which the holidays can occur

▼ Inflation Adjustments

▼ Price Index

We make use of a standard set of items to compare how much the prices changes from year t_1 to the other year t_2

$$z_t = \frac{\text{price of a set items in year } t_1}{\text{price of a set items in year } t_2} \times 100$$

Note that for this, t_2 is the reference year and t_1 is the year that we want to know how it compares to the year t_2

▼ Adjusted Price

The adjusted price at time t_1 in terms of time t_2 is given by the following

$$x_t = \frac{y_t}{z_t} \times 100$$

- y_t is the unadjusted price for time t_1
- z_t is the Price Index at time t_1

Transformations

We can do the various transformations to stabilize the variance

Box Cox Transformation

In R, the `box_cox()` function can be used to apply the box cox transformation

$$w_t = \begin{cases} \log(y_t), & \lambda = 0 \\ \frac{\text{sign}(y_t)|y_t|^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases}$$

- Note that when $\lambda = 1$, it is a linear transformation
- Note that if λ is too small, it can give extremely large prediction intervals
- Decrease the value of λ → Increase the power of shrinking

Guerrero Function

- In R, `guerrero` function can be used

Tries to find the optimal value of λ

$$\begin{aligned} Var(w) &= Var[g(y)] \\ &\approx [g'(\mu_y)]^2 Var(y) \end{aligned}$$

- We want to find λ that makes the above variance roughly constant

Back-transformation

$$y_t = \begin{cases} \exp(w_t), & \lambda = 0 \\ (\lambda w_t + 1)^{\frac{1}{\lambda}}, & \lambda \neq 0 \end{cases}$$

Proof that using the Naive Version is Biased

First, suppose that we want to estimate $E(y)$, but we want to work with $w = \log y \Rightarrow y = e^w$

Assumption: w has a symmetric pdf. This means that $E(w) = m_2$ where m_2 is the median of w

> 💡 Naive back transform estimates only the Median of y instead of its expectation

$$
\begin{aligned}
P(w \le m_2) &= 0.5 \qquad \text{By Definition of Median} \\
P(e^w \le e^{m2}) &= 0.5 \\
P(y \le e^{m2}) &= 0.5
\end{aligned}
$$

Now, suppose that we use $\hat{w}$ to estimate $E(w) \equiv m_2$, where $\hat{w}$ is just any estimator that we use to estimate the true mean.

The naive back transform $e^{\hat{w}}$ estimates e^{m_2} which we showed to be the median of y above.

▼ **Bias Adjustment**

▼ **Proof for $\lambda = 0$**

$$E(y) = E[e^w] = \int e^w f(w) dw \tag{*}$$

We will denote the following:

$h(w) = e^w$

$a = \mu_w = \int w f(w) dw = E(w)$

💡 Taylor's Series Expansion of $h(x)$ about a

$$h(x) = h(a) + h'(a)(x-a) + h''(a)\frac{(x-a)^2}{2!} + \cdots$$

From (*), we can have the following:

$$
\begin{aligned}
E(y) &= \int [e^{\mu_w} + e^{\mu_w}(w-\mu_w) + e^{\mu_w}\frac{(w-\mu_w)^2}{2!} + \cdots]f(w) \quad dw \\
&\approx \int e^{\mu_w} f(w)\, dw + \int e^{\mu_w}(w-\mu_w)f(w)\, dw + \int e^{\mu_w}\frac{(w-\mu_w)^2}{2!}]f(w)\, dw \\
&= e^{\mu_w}(1) + 0 + \frac{1}{2}e^{\mu_w}\int (w-\mu_w)^2 f(w) dw \\
&= e^{\mu_w} + \frac{1}{2}e^{\mu_w}\sigma_w^2 \\
&= e^{\mu_w}(1 + \frac{1}{2}\sigma_w^2)
\end{aligned}
$$

Note: Here are some of things that are used in the above derivation

1. We expanded e^w about μ_w using Taylor's Series
2. We also make a second-order Taylor Series approximation since as the terms get bigger they should tend to 0 and most of the time, a second order approximation is good enough
3. $\int f(w)\,dw = 1$ - Since this is the pdf of w and we assume it is continuous so the integral of the pdf will be 1
4. Derivation for the second term

$$\begin{gathered}\int e^{\mu_w}(w-\mu_w)f(w)\,dw \\ = e^{\mu_w}[\int wf(w)\,dw - \int \mu_w f(w)\,dw] \\ = e^{\mu_w}[\mu_w - \mu_w \int f(w)\,dw] \\ = e^{\mu_w}[\mu_w - \mu_w] \\ = 0\end{gathered}$$

5. Derivation of σ_w^2

$$\begin{gathered}\int (w-\mu_w)^2 f(w)dw = E[(w-\mu)^2] \\ \text{The above is using the below identities} \\ E[g(x)] = \int g(x)f(x)\,dx \\ Var(X) = E[(X - E[X])]\end{gathered}$$

Conclusion

If $\hat{w}$ estimates $E(w) = \mu_w$,

then $e^{\hat{w}}[1 + \frac{1}{2}\sigma_w^2]$ estimates $E(y)$

Unbiased Estimator of Back-Transform

We can make use of the following back transform instead so that it is unbiased

$$y_t = \begin{cases}\exp(w_t)[1 + \frac{\sigma^2}{2}], & \lambda = 0 \\ (\lambda w_t + 1)^{\frac{1}{\lambda}}[1 + \frac{\sigma^2(1-\lambda}{2(\lambda w_t + 1)^2}], & \lambda \neq 0\end{cases}$$

Note that σ^2 is the variance of w_t. The larger this value, the bigger the adjustment required

▼ **Proof**

$$\begin{gathered}Let\ g(\lambda) = \frac{(y_t^\lambda - 1)}{\lambda} \\ \lim_{\lambda \to 0} g(\lambda) = \log y_t\end{gathered}$$

- **Note**
    - Transformations can have very large effect on Prediction Intervals (PI)
    - If some data are zero or negative, then use $\lambda > 0$
    - Choosing $\log$ is a simple way to force the forecasts to be positive
    - Transformations must be reversed to obtain forecasts on the original scale.

Features

Summary Statistics

Note that this provides a good understand of the spread of the various but note that there may not be a notion of time series in those statistics.

- For instance, `quantile`. We may only know that only 50% of the points are above a certain value but there is no intuition of the time

Tiled Statistics

Divides our data into non-overlapping windows

- Note that if there are periods then it breaks them up into those periods

We can have the following statistics:

1. **Variance of the window means**: Measures how "stable" the series is, i.e. how much variation there is across the means of windows
    a. Can be used to determine which series is the most stable
2. **Variance of the window variances**: Measures how "lumpy" the series is, i.e. how much variation there is across the variations of the dataset.
    a. If the variation of the variation is small, then the data will look pretty lumpy since the dataset will not have much variation in terms of the variation across time
    b. If the variation of the variation is large, we will see that there are changes in variation across time
    c. If the lumpiness of the time series is related to the level, i.e. if it is more lumpy at the start and increases over time and there is such a correlation → Box Cox can then be useful for stabilising the variance

Roll Statistics

Computed using overlapping windows

It will identify those locations whereby there is a large change in the mean / variance for instance.

Usage

- Can be used to identify places where there are sudden changes in the levels, variability and distribution