OSS Project

The University of Auckland | 825159894

How does the number of issues change over time

Ayato Tanemura

2022

# Introduction

**What is OSS?**

<https://www.sdxcentral.com/networking/nfv/definitions/whats-network-functions-virtualization-nfv/nfv-software/operations-support-system-oss-definition/>

An operations support system (OSS) is a software component that enables a service provider to monitor, control, analyse, and manage the services on its network.

<https://www.openlogic.com/blog/top-5-benefits-open-source-software>

When you have a product that someone is working on because they want to work on it, (not just because they’re getting paid) their personal drive to add creativity and contribute their best ideas tends to be a lot higher. This is often what inspires enthusiast open source communities to develop features that are new and disruptive, and why we see open source projects leading the way in terms of this disruption.

**What this paper is for?**

By applying longitudinal data models, you need to provide an insight about the projects evolution. Clearly define your research questions and test your hypotheses using this dataset.

**What is the research design?**

You have been provided with a panel dataset of projects hosted on GitHub. It contains two years (8 quarters) of data for each project. The dataset contains …

**What this paper is going to focus on?**

Issues are pointed out by people who are interested in the projects. Issues column shows the total number of problem/bugs raised or request for new features and those are pointed put by members, watchers or people who requested pull for review. This means that if the number of those people increase, then the number of issues pointed out would be increase as well since the total number of reviews would increase. Also, if the total number of commits increase, possibility of causing issues is supposed to increase. This can tell the number of commits increase, then the number of total issues increase.

**Why the focus is worth?**

Issue is not the bad thing. Finding issue is necessary to improve systems or algorithms. There is no perfect code in the world and there must be some issues or points which can be improve. It’s good if you could figure out if there is any effect from other factors to increase the number of problems or bugs raised or requested for new features. If you find that more watchers more issues, then you can work on to increase the number of watchers by marketing to increase the review opportunities for the new features.

# Research Question and Hypothesis

Do the trajectories of the number of issues differ by the number of members, commits, watchers, and pull requests?

* Predictor: The number of members:
  + Null hypothesis: There is no effect from the number of members.
  + Alternative hypothesis: There is an effect from the number of members.
* Predictor: The number of commits:
  + Null hypothesis: There is no effect from the number of commits.
  + Alternative hypothesis: There is an effect from the number of commits.
* Predictor: The number of watchers:
  + Null hypothesis: There is no effect from the number of watchers.
  + Alternative hypothesis: There is an effect from the number of watchers.
* Predictor: The number of pull requests:
  + Null hypothesis: There is no effect from the number of pull requests.
  + Alternative hypothesis: There is an effect from the number of pull requests.

# Data cleaning and preparation

Do I have to explain Dataset?

Who is the audience? Only the researcher or public as well.

For the preparation of data, I created Time\_1 column to add another time factor. The Time column in original data shows a sequence for time of observation with starting from 1 to 8. However, in my case study, I would like to set the start point as 0. Thus, I created the Time\_1 column to show the sequence for time observation starting with 0. Other than the Time factor, I used Members, Commits, Issues, Watchers, and PullReq for the predictors which were continuous variables without issues.

# Definition of main variables, visualisation exploration

Definition of main variables

* Issues
* Members
* Commits
* Watchers
* Pull Request
* Time\_1

Visualisation exploration

* Sampling visualisation
* Overall visualisation

Overview of the data

# Methodology

## Model justification

Unconditional Means Model

\* The composite model is:

Issues = 24.84 + e

\* Estimate of fixed effect:

The initial status of Issues at occasion 0 is 24.84 (p = 0) at 0.01 level of significance.

Unconditional Growth model

\* The composite model:

Level1: Issues = a + b \* Time\_1 + j

Level2: a = 1.00 + y\_0i, b = 6.81 + y\_1i

Issues = 1.00 + 6.81 \* Time\_1 + e (With composite residual: e = y\_0i + y\_1i \* Time\_1 + j)

\* Estimate of fixed effect

1. The estimated initial Issues is 2.60 (p-vale = 0.35). However, it is non-significant at 0.01 level of significance.

2. The rate of change at the Time\_1 0 is 0.11 (p-vale = 0) at 0.01 level of significance.

Members, Commits, Watchers, and Pull Requests

\* Estimate of Fixed effect

1. The estimated initial Issues controlling for members, commits, watchers, and pull requests is -1.64 (p > 0.01). However, it is non-significant at 0.01 level of significance.

2. The estimated differential in initial Issues for one member difference in Members controlling for other predictors at the initial stage is 0.28 (p<0.01) at 0.01 level of significance.

3. The estimate rate of change in Issues controlling for Members, Commits, Watchers, and Pull Requests is 2.66 (p<0.05) at 0.05 level of significance.

4. The estimated differential in initial Issues for one commit difference in Commits controlling for other predictors at the initial stage is 0.049 (p<0.01) at 0.01 level of significance.

5. The estimated differential in initial Issues for one watcher difference in Watchers controlling for other predictors at the initial stage is 0.0098 (p<0.05) at 0.05 level of significance.

6. The estimated differential in initial Issues for one pull request difference in Pull Request controlling for other predictors at the initial stage is -0.091 (p<0.01) at 0.01 level of significance.

7. The estimated differential in the rate of change in Issues of Members is -0.019 (p > 0.05). However, it is indistinguishable since the p value is 0.60.

8. The estimated differential in the rate of change in Issues of Commits is 0.0059 (p < 0.01) at 0.01 level of significance.

9. The estimated differential in the rate of change in Issues of Watchers is 0.0070 (p < 0.01) at 0.01 level of significance.

10. The estimated differential in the rate of change in Issues of Pull Request is 0.0024 (p < 0.01) at 0.01 level of significance.

Members (only initial status), Commits, Watchers, and Pull Requests

1. The estimated initial Issues controlling for members, commits, watchers, and pull requests is -1.58 (p > 0.01). However, it is non-significant at 0.01 level of significance.

2. The estimated differential in initial Issues for one member difference in Members controlling for other predictors at the initial stage is 0.28 (p<0.01) at 0.01 level of significance.

3. The estimated differential in initial Issues for one commit difference in Commits controlling for other predictors at the initial stage is 0.049 (p<0.01) at 0.01 level of significance.

4. The estimate rate of change in Issues controlling for Members, Commits, Watchers, and Pull Requests is 2.41 (p<0.05) at 0.05 level of significance.

5. The estimated differential in initial Issues for one watcher difference in Watchers controlling for other predictors at the initial stage is 0.0097 (p<0.05) at 0.05 level of significance.

6. The estimated differential in initial Issues for one pull request difference in Pull Request controlling for other predictors at the initial stage is -0.092 (p<0.01) at 0.01 level of significance.

7. The estimated differential in the rate of change in Issues of Commits is 0.0059 (p < 0.01) at 0.01 level of significance.

8. The estimated differential in the rate of change in Issues of Watchers is 0.0070 (p < 0.01) at 0.01 level of significance.

9. The estimated differential in the rate of change in Issues of Pull Request is 0.0024 (p < 0.01) at 0.01 level of significance.

Members (only initial status), Commits, Watchers, and Log (Pull Requests)

1. The estimated initial Issues controlling for members, commits, watchers, and pull requests is -1.96 (p > 0.05). However, it is non-significant at 0.05 level of significance.

2. The estimated differential in initial Issues for one member difference in Members controlling for other predictors at the initial stage is 0.19 (p<0.05) at 0.05 level of significance.

3. The estimated differential in initial Issues for one commit difference in Commits controlling for other predictors at the initial stage is 0.035 (p<0.01) at 0.01 level of significance.

4. The estimate rate of change in Issues controlling for Members, Commits, Watchers, and Log (Pull Requests) is 0.51 (p>0.05). However, it is non-significant at 0.05 level of significance.

5. The estimated differential in initial Issues for one watcher difference in Watchers controlling for other predictors at the initial stage is 0.0060 (p>0.05). However, it is non-significant at 0.05 level of significance.

6. The estimated differential in initial Issues for one pull request difference in Log (Pull Request) controlling for other predictors at the initial stage is 7.82 (p<0.01) at 0.01 level of significance.

7. The estimated differential in the rate of change in Issues of Commits is 0.0099 (p < 0.01) at 0.01 level of significance.

8. The estimated differential in the rate of change in Issues of Watchers is 0.0091 (p < 0.01) at 0.01 level of significance.

9. The estimated differential in the rate of change in Issues of Pull Request is 0.25 (p>0.05). However, it is non-significant at 0.05 level of significance.

Members (only initial status), Commits, Watchers, and Log (Pull Requests) (Only initial status)

1. The estimated initial Issues controlling for members, commits, watchers, and pull requests is -2.16 (p > 0.05). However, it is non-significant at 0.05 level of significance.

2. The estimated differential in initial Issues for one member difference in Members controlling for other predictors at the initial stage is 0.19 (p<0.05) at 0.05 level of significance.

3. The estimated differential in initial Issues for one commit difference in Commits controlling for other predictors at the initial stage is 0.035 (p<0.01) at 0.01 level of significance.

4. The estimate rate of change in Issues controlling for Members, Commits, Watchers, and Log (Pull Requests) is 0.72 (p>0.05). However, it is non-significant at 0.05 level of significance.

5. The estimated differential in initial Issues for one watcher difference in Watchers controlling for other predictors at the initial stage is 0.0058 (p>0.05). However, it is non-significant at 0.05 level of significance.

6. The estimated differential in initial Issues for one pull request difference in Log (Pull Request) controlling for other predictors at the initial stage is 8.18 (p<0.01) at 0.01 level of significance.

7. The estimated differential in the rate of change in Issues of Commits is 0.010 (p < 0.01) at 0.01 level of significance.

8. The estimated differential in the rate of change in Issues of Watchers is 0.0093 (p < 0.01) at 0.01 level of significance.

Members (only initial status), Commits, and Log (Pull Requests) (Only initial status)

1. The estimated initial Issues controlling for members, commits, watchers, and pull requests is -1.99 (p > 0.05). However, it is non-significant at 0.05 level of significance.

2. The estimated differential in initial Issues for one member difference in Members controlling for other predictors at the initial stage is 0.19 (p<0.05) at 0.05 level of significance.

3. The estimated differential in initial Issues for one commit difference in Commits controlling for other predictors at the initial stage is 0.033 (p<0.01) at 0.01 level of significance.

4. The estimate rate of change in Issues controlling for Members, Commits, Watchers, and Log (Pull Requests) is 1.34 (p>0.05). However, it is non-significant at 0.05 level of significance.

5. The estimated differential in initial Issues for one pull request difference in Log (Pull Request) controlling for other predictors at the initial stage is 8.18 (p<0.01) at 0.01 level of significance.

6. The estimated differential in the rate of change in Issues of Commits is 0.011 (p < 0.01) at 0.01 level of significance.

## Model selection

Based on the outcomes of the models above, the table below shows the AICs and BICs for each model.

|  |  |  |
| --- | --- | --- |
| Model | AIC | BIC |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Model Interpretation

# Results and discussions

# References