# FORECASTING REGIONAL INFLATION IN THE PHILIPPINES USING SUPPORT VECTOR MACHINES OPERATIONS MANUAL

PRODUCED BY: VIDAL MARVIN C. GABRIEL

VMGMARV@GMAIL.COM

THIS USER MANUAL IS THE APPLICATION OF THE PAPER *Forecasting regional inflation in the Philippines using machine learning techniques: A new approach.*

# Table of Contents

SECTION 1

# Forecasting regional inflation in the Philippines

A study was conducted by Gabriel, et al. which develops the first regional inflation forecasting models for the Philippines employing non-linear machine learning approaches for a few representative regions of the country. The result of their paper found that support vector regression models outperformed ARIMA, artificial neural networks (ANN) and long-short term memory (LSTM) in the month-ahead and dynamic forecasting exercises in the Philippine setting. The paper can be found in BSP working paper series entitled *Forecasting regional inflation in the Philippines using machine learning techniques: A new approach*. This manual briefly discusses the published paper and guides the users on how to use the developed algorithms.

## 1.1 Support vector regression (SVR)

SVR is an application of support vector machines (SVM). In an ordinary least squares regression (OLS) with one predictor, the goal is to minimize the objective function:

$$min \sum_{t=1}^{n} (y_t - w_t x_t)^2 \tag{1}$$

where $y_t$ is the target, $w_t$ is the coefficient, and $x_t$ is the predictor (feature). SVR takes this a step further by allowing the definition of the acceptable level of error. The objective of SVR is to minimize the coefficient $w$ while the error term is handled in constraint, i.e., error should be less than the specified threshold, $\epsilon$. This modifies **Equation 1** to:

$$min \frac{1}{2} ||w||^2 \tag{2}$$

and the constraint is given by:

$$\left| y_t - w_t x_t \right| \leq \epsilon \tag{3}$$

However, there will be cases wherein data will violate **Equation 3**. Data can exceed $\epsilon$ with some deviation. To accommodate this instance, $\xi$ is added in **Equation 2** as regularization parameter which then yields:

$$min\frac{1}{2}||w||^2 + c\sum_{t=1}^{n}|\xi_t| \tag{4}$$

with a new constraint

$$\left| y_t - w_t x_t \right| \leq \epsilon + |\xi| \tag{5}$$

In this configuration, there is an additional hyper-parameter, $C$ that should be tuned. Furthermore, it can be noted that $C$ and $\epsilon$ are directly proportional. Ths means that if $C$ approaches 0, **Equation 4** reverts back to **Equation 2**.

## 1.2 SVR architecture

Two models were developed using SVRs. The first model is a univariate process with lag. The second model is a multivariate process wherein the inputs are inflation and shocks (i.e. 0 or 1) as determined by:

$$\text{shocks} = \begin{cases} 1, & \text{if } \pi_n - \pi_n - 1 > 0 \\ 0, & \text{otherwise} \end{cases}$$

Both models will used a lag value of 1 ($m = 1$), $\epsilon$ of 0.1 and a radial base function (RBF) kernel. RBF is a type of algorithm wherein it reconstructs unknown functions from a known data. Further, the functions are multivariate and may be solutions to partial differential equations satisfying certain additional conditions. Since the kerel searches for a function to describe the data, greed search was further conducted.

## 1.3 Data preparation

The data are taken from the Philipine Statistics Authority (PSA) and covers the period January 1994 to present. Inflation is defined as:

$$\pi_{year-on-year} = \left(\frac{CPI_t - CPI_{t-1}}{CPI_{t-1}}\right) * 100 \tag{6}$$

where $CPI_t$ is the CPI of the current month while $CPI_{t-1}$ is the same month's CPI from the previous year.
The data is structured to follow the requirements of the SVR models. This involves preparing the input to the network as the feature vector $\left[y_t, y_{t-\tau}, ..., y_{t-(m-1)_\tau}\right]$ and the corresponding output vector as $y_{t-\tau}$

SECTION 2

# How to run the models

The following provides the steps on how to run the models. The python scripts discussed in this manual are optimized and automated to some extent. Thus, if you wish to know more about the algorithm, please coordinate with the authors of the paper.

## 2.1 Python 3.0

The scripts were written in Python 3.6 using an Ubuntu based computer. Thus, to successfully run the models, you should make sure your desktop runs a Python version 3+ programming language. Further, it is recommended to use Spyder as the integrated development environment (IDE).

## 2.2 main.py module

This module contains both models. To run main. py:

1. Open *'main.py'* with Spyder
2. Click run
3. In the terminal(console) type **svm_uni** to forecast inflation using univariate SVR. Or, type **svm_mul** to forecast inflation using multivariate SVR. Then, press enter
4. Type the region/s that you want to forecast. Please follow the codes in the **Table** 4. In the case of multiple regions, use <space> as the divider between regions (eg. to forecast regions 1 and 2, type: r1 r2). Then, press enter
5. To view the results, open *'results.csv'*.

| Regions | Code |
|---|---|
| NCR | ncr |
| CAR | car |
| Region 1 | r1 |
| Region 2 | r2 |
| Region 3 | r3 |
| Region 4 | r4 |
| Region 5 | r5 |
| Region 6 | r6 |
| Region 7 | r7 |
| Region 8 | r8 |
| Region 9 | r9 |
| Region 10 | r10 |
| Region 11 | r11 |
| Region 12 | r12 |
| Region 13 | r13 |
| ARMM | armm |

Table 1: Codes for each region

## 2.3 error.py module

Although the mean absolute error (MAE) is already presented in the *'main.py'* module, it can be calculated separately. To run error.py:

1. Open *'error.py'* with Spyder
2. Click run
3. Type the region/s that you want to check for MAE. If you wish to check for the MAE in the aggregate level, type phil. Then, click enter

3

SECTION

# Updating the data

In order to forecast the following month, the data in the *'main_data.csv'* and *'summary.xlsx'* should be updated to the most recent CPI and inflation data. This section discusses how to update the files which are used by the forecasting modules.

## 3.1 main_data.csv

The CPI data from January 1994 is collected and stored in *'main_data.csv'*. To update *'main_data.csv'*:

1. Open *'main_data.csv'*.
2. Copy CPI data from Philippine Statistics Authority (PSA) - Open stat
3. Paste to the corresponding format of *'main_data.csv'*

## 3.2 summary.xlsx

This excel file is needed to compute for the mean absolute error of the regions and the aggregate inflation. The actual and predicted inflation is stored in this file. Thus, it must be updated every time a new inflation is released. To update *'summary.xlsx'*:

1. Open *'summary.xlsx'*
2. Copy actual inflation data from Philippine Statistics Authority (PSA) - Open stat
3. Paste to the corresponding format of *'summary.xlsx'*

4. Copy predicted inflation data from *'results.csv'*

5. Paste to the corresponding format of *'summary.xlsx'*