Swap Regression with US GDP and Public Debt

Viral Chitlangia

Contents

1 Motivation

The idea for SWAP Regression came from a paper published by Mosuk Chow et al, to use alternating predictors to find the relation between human tolerance to temperature and pressure.[1] In such situations, it is not possible to get certain measurements such as pressure, with a set temperature, and vice verse. In such cases, we have to remove the assumption that one variable is regressing the other, as in such cases, for some data points, the role of regressors and predictors switches.

2 Theory for Data Collection

Our idea is that SWAP Regression holds for any Bi Directional causal data points. In this discussion, we limit ourselves to Economic indicators. Glauco De Vita el at concluded that Public Debt and GDP Growth are Bi Directional Causal for the American market.[2] However, in our research, we took US Public Debt and Nominal GDP.

3 Data Collection

The data was found easily on Kaggle, US Public Debt Quarterly Data(in Billion $) and US Nominal GDP Quarterly Data(in Million $), in quarters. Public Debt data was available from 1966-01-01 to 2023-01-01, and Nominal GDP data from 1966-01-01 to 2023-01-01.

4 Analysis of Data

4.1 Testing Bi Directional Causality

To test Bi Directional Causality, we used the **grangertest** function from the **vars** package.
Here, X represents the Nominal GDP, and Y represents Public Debt.

```
Granger causality test

Model 1: X ~ Lags(X, 1:4) + Lags(Y, 1:4)
Model 2: X ~ Lags(X, 1:4)
  Res.Df Df      F    Pr(>F)
1    216
2    220 -4 18.538 4.107e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 1: Granger Test for Y causing X

```
Granger causality test

Model 1: Y ~ Lags(Y, 1:4) + Lags(X, 1:4)
Model 2: Y ~ Lags(Y, 1:4)
  Res.Df Df      F    Pr(>F)
1    216
2    220 -4 4.8931 0.0008519 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2: Granger Test for X causing Y

Fitting the model also showed the significance of each term in the model.

Since for the models trained on GDP, we can find a lag of Public Debt, which is significant, and vice versa, we can conclude Bi Directional Causality.

```
Estimation results for equation X:
==================================
X = X.l1 + Y.l1 + X.l2 + Y.l2 + X.l3 + Y.l3 + X.l4 + Y.l4 + const

        Estimate Std. Error t value Pr(>|t|)
X.l1   1.265e+00  8.644e-02  14.629  < 2e-16 ***
Y.l1   1.235e-01  1.574e-02   7.849 1.93e-13 ***
X.l2  -1.932e-01  1.308e-01  -1.477   0.1410
Y.l2  -1.396e-01  2.613e-02  -5.341 2.34e-07 ***
X.l3   6.275e-02  1.209e-01   0.519   0.6043
Y.l3   2.903e-02  2.765e-02   1.050   0.2949
X.l4  -1.417e-01  7.596e-02  -1.866   0.0634 .
Y.l4  -1.222e-02  1.758e-02  -0.695   0.4878
const  1.031e+04  6.200e+03   1.663   0.0978 .
```

Figure 3: Significance of Y in predicting X

```
Estimation results for equation Y:
==================================
Y = X.l1 + Y.l1 + X.l2 + Y.l2 + X.l3 + Y.l3 + X.l4 + Y.l4 + const

        Estimate Std. Error t value Pr(>|t|)
X.l1  -1.185e+00  4.725e-01  -2.507  0.01290 *
Y.l1   8.038e-01  8.602e-02   9.345  < 2e-16 ***
X.l2  -2.368e-01  7.147e-01  -0.331  0.74074
Y.l2   2.129e-01  1.428e-01   1.491  0.13753
X.l3   1.875e+00  6.610e-01   2.836  0.00500 **
Y.l3   2.913e-01  1.511e-01   1.928  0.05520 .
X.l4  -4.176e-01  4.152e-01  -1.006  0.31566
Y.l4  -2.941e-01  9.606e-02  -3.061  0.00248 **
const  6.922e+03  3.389e+04   0.204  0.83834
```

Figure 4: Significance of X in predicting Y

4.2 Testing Stationarity in Residuals

We got the residuals by fitting the models from the Granger Causality Test, as mentioned above. We applied Augmented Dickey Fuller Test to test stationarity of the residuals. The residuals of both the models gave a p-value of less than 0.01 on Augmented Dickey Fuller test, which we applied using the **adf.test** function in R.

5 Transforming Data

We have scaled the gdp and public debt data by a factor of 10^{-6}.

6 Modelling the Regression

6.1 Setting up the Variables

Let us denote X_0, Y_0 as the points where Z = 0, and X_1, Y_1 where Z = 1.
For Z = 1, $Y_{1i} = g(X_{1i}) + \epsilon_{1i}$, $\epsilon_{1i} \sim N(0, \sigma_1^2)$
For Z = 0, $X_{0j} = g^{-1}(Y_{0j}) + \epsilon_{0j}$, $\epsilon_{0j} \sim N(0, \sigma_0^2)$
Here, we have the latent variable Z. If we had the variable Z given to us, we could have easily gotten the Likelihood function using
f(x, y | z = 0) = $\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(x-g^{-1}(y))^2}{2\sigma_0^2}} f(y|z=0)$
f(x, y | z = 1) = $\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(y-g(x))^2}{2\sigma_1^2}} f(x|z=1)$

6.2 Relation to GMM

This is not exactly of the form of the GMM, where we have constant means for all the Gaussian Distributions. We also have to consider that the "mean" of the first distribution is not independent of the "mean" of the second distribution.
Our idea is related to the ALT-OPT algorithm, where we solve the problem iteratively. We have a total of 6 parameters to predict - $(\pi_0, \pi_1, g, Z, \sigma_0^2, \sigma_1^2)$.
Here, $Z_i \sim Bernoulli(\pi_1)$ and $\pi_0 = 1 - \pi_1$.

6.3 Proposed Algorithm

Let us initialize $Z^{(0)}$ randomly. We can also initialize $g^{(0)}$ as any suitable model for X on Y. Using that, we can get the vectors $\epsilon_0^{(0)} = X_0^{(0)} - {g^{(0)}}^{-1}(Y_0^{(0)})$ and $\epsilon_1^{(0)} = Y_1^{(0)} - g^{(0)}(X_1^{(0)})$. Getting the variance of those vectors, we can get $\sigma_0^{2(0)}$ and $\sigma_1^{2(0)}$. We also get $\pi_0^{(0)} = \frac{N_0^{(0)}}{N_0^{(0)}+N_1^{(0)}}$ and $\pi_1^{(0)} = 1 - \pi_0^{(0)}$.

6.3.1 Iterations

Here, we assume that the prior distributions of X and Y are independent of Z. Let us denote them by h_X and h_Y for X and Y, respectively.
$P(Z = 1|X, Y) \propto f(X, Y|Z = 1)P(Z = 1) \propto f(Y|X, Z = 1)P(Z = 1)P(X)$
$P(Z = 0|X, Y) \propto f(X, Y|Z = 0)P(Z = 0) \propto f(X|Y, Z = 0)P(Z = 0)P(Y)$
In exact form,

$$P(z^{(t+1)} = 1|x, y) = \frac{\pi_1^{(t)} \frac{1}{\sqrt{2\pi\sigma_1^{2(t)}}} e^{-\frac{(y-g^{(t)}(x))^2}{2\sigma_1^{2(t)}}} h_X(x)}{\pi_1^{(t)} \frac{1}{\sqrt{2\pi\sigma_1^{2(t)}}} e^{-\frac{(y-g^{(t)}(x))^2}{2\sigma_1^{2(t)}}} h_X(x) + \pi_0^{(t)} \frac{1}{\sqrt{2\pi\sigma_0^{2(t)}}} e^{-\frac{(x-{g^{(t)}}^{-1}(y))^2}{2\sigma_0^{2(t)}}} h_Y(y)} \tag{1}$$

$$P(z^{(t+1)} = 0|x, y) = \frac{\pi_0^{(t)} \frac{1}{\sqrt{2\pi\sigma_0^{2(t)}}} e^{-\frac{(x-{g^{(t)}}^{-1}(y))^2}{2\sigma_0^{2(t)}}} h_Y(y)}{\pi_1^{(t)} \frac{1}{\sqrt{2\pi\sigma_1^{2(t)}}} e^{-\frac{(y-g^{(t)}(x))^2}{2\sigma_1^{2(t)}}} h_X(x) + \pi_0^{(t)} \frac{1}{\sqrt{2\pi\sigma_0^{2(t)}}} e^{-\frac{(x-{g^{(t)}}^{-1}(y))^2}{2\sigma_0^{2(t)}}} h_Y(y)} \tag{2}$$

We have the conditional likelihood -

$$L(\theta|X, Y, Z^{(t+1)}, \theta^{(t)}) = \prod_{i=1}^{N_0^{(t)}} \frac{1}{\sqrt{2\pi\sigma_0^{2(t)}}} e^{-\frac{(X_{0i}-g^{-1}(Y_{0i}))^2}{2\sigma_0^{2(t)}}} h_Y(Y_{0i}) \prod_{i=1}^{N_1^{(t)}} \frac{1}{\sqrt{2\pi\sigma_1^{2(t)}}} e^{-\frac{(Y_{1i}-g(X_{1i}))^2}{2\sigma_1^{2(t)}}} h_X(X_{1i}) \tag{3}$$

We have $g^{(t+1)} = argmax_g$ L(θ | X, Y, $Z^{(t+1)}$, $\theta^{(t)}$).

6.3.2 Stopping Criterion

Stop when

- $Z^{(t)} = Z^{(t+1)}$.
- $||g^{(t+1)} - g^{(t)}|| < \epsilon$
- $iterations > max\_iters$

Algorithm 1 New GMM Algorithm

1: For all Z_i, we can predict
$Z_i^{(t+1)} = argmax(P(Z_i^{(t+1)} = 0|X_i, Y_i), P(Z_i^{(t+1)} = 1|X_i, Y_i))$, using 1 and 2.

2: $g^{(t+1)} = argmin_g \sum_i^N (\frac{(Y_i - g(X_i))^2}{\sigma_1^{2(t)}} I[Z_i^{(t+1)} = 1] + \frac{(X_i - g^{-1}(Y_i))^2}{\sigma_0^{2(t)}} I[Z_i^{(t+1)} = 0])$, using 3.

3: $\epsilon_0^{(t+1)} = X_0^{(t+1)} - {g^{(t+1)}}^{-1}(Y_0^{(t+1)})$.

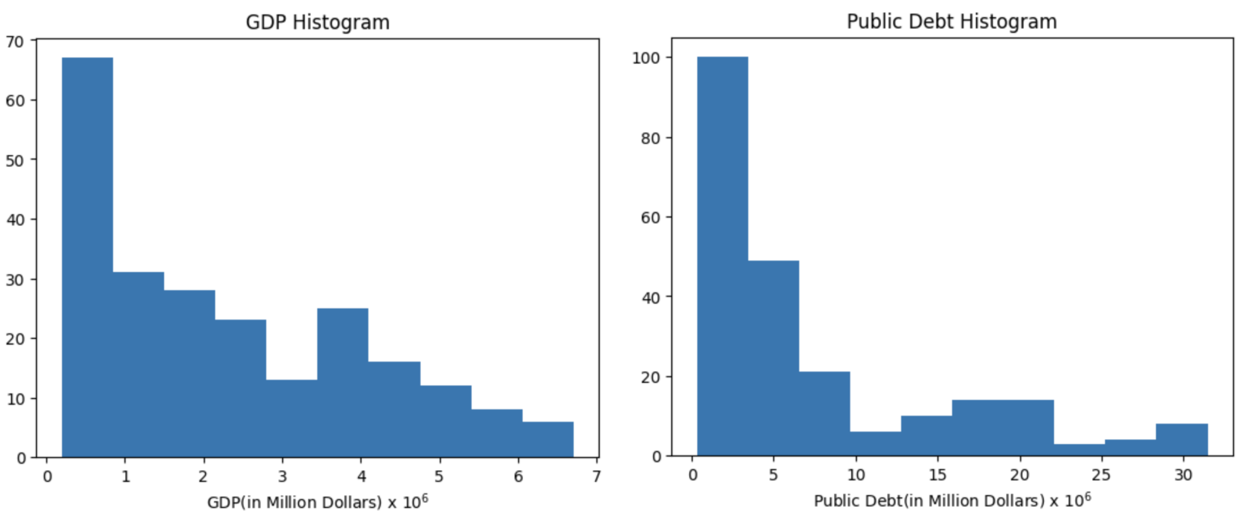
4: $\epsilon_1^{(t+1)} = Y_1^{(t+1)} - g^{(t+1)}(X_1^{(t+1)})$.

5: $\sigma_0^{2(t+1)} = \text{Var}(\epsilon_0^{(t+1)})$.

6: $\sigma_1^{2(t+1)} = \text{Var}(\epsilon_1^{(t+1)})$.

7: $N_0^{(t+1)} = \sum_i^N I[Z_i^{(t+1)} = 0]$.

8: $N_1^{(t+1)} = \sum_i^N I[Z_i^{(t+1)} = 1]$.



(a) Distribution of GDP

(b) Distribution of Public Debt

Figure 5: Distribution of Data

7 Data Analysis

We run the model with the US GDP vs Public Debt data.
To implement the SWAP Regression model, we would have to assume a prior on X and Y.

7.1 Prior on US GDP

To assume a prior on US GDP, we look at the histogram of US GDP, as shown in 5a. We can see that it resembles an Exponential Distribution, and since US GDP is always Positive, that seems like a fair assumption. We can get the rate parameter $\hat{\lambda}_{MLE} = \frac{1}{\bar{X}} = 0.438780921481478$.

7.2 Prior on Public Debt

To assume a prior on Public Debt, we look at the histogram of Public Debt, as shown in 5b. We can see that it resembles an Exponential Distribution, and since Public Debt is always Positive, that seems like a fair assumption. We can get the rate parameter $\hat{\lambda}_{MLE} = \frac{1}{\bar{Y}} = 0.13497222534768713$.

8 Testing the Model

Now, since we have the prior distribution, we can implement all the models.

8.1 Simple Linear Regression

With Simple Linear Regression, we get the line,
y = 4.49724771804972 x - 2.840483200739337

8.2 Simple Quadratic Regression

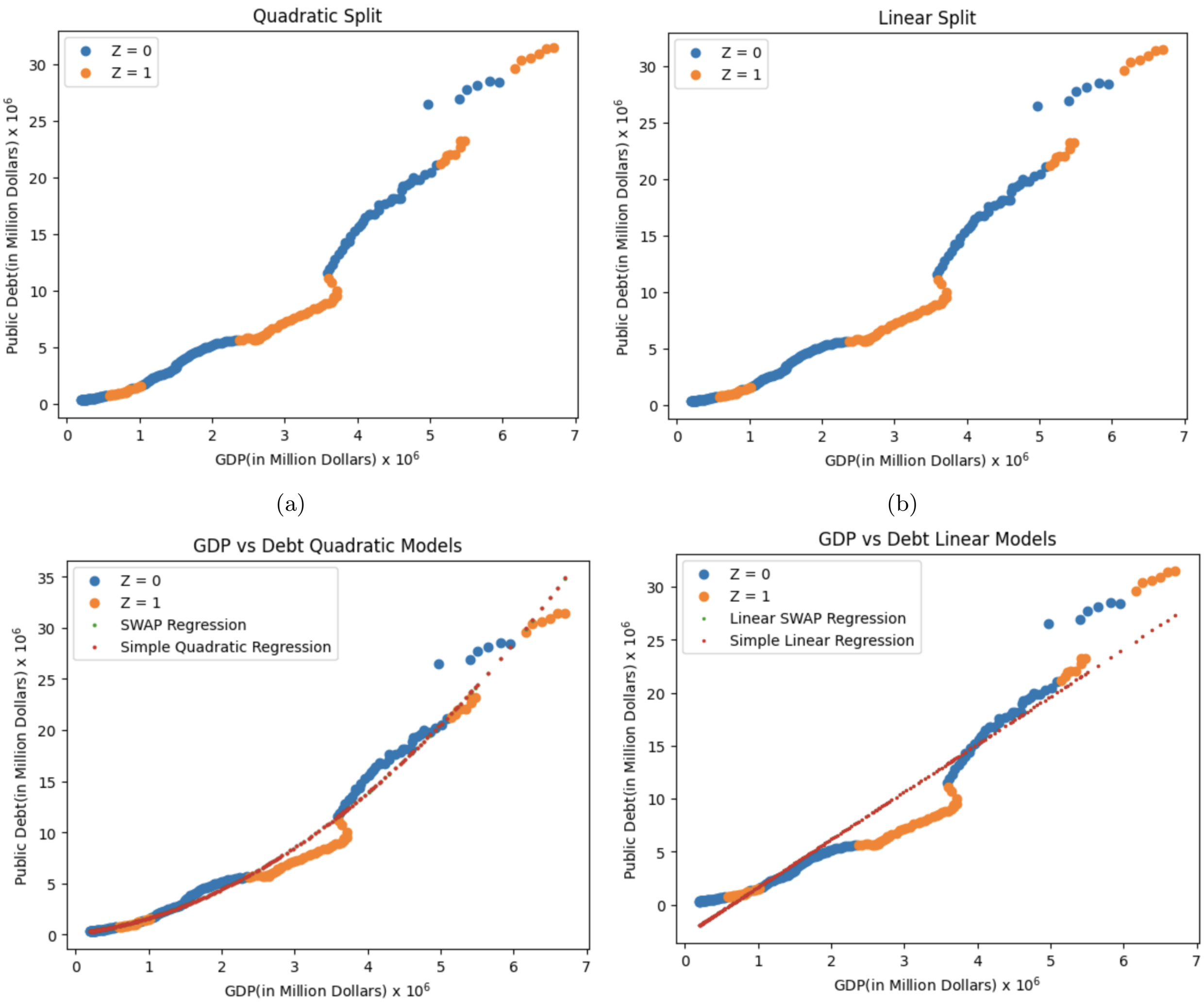
With Simple Quadratic Regression, we get the line,
y = 0.6460656714354706 x^2 + 0.8578272692490471 x + 0.11587014579886798

8.3 SWAP Linear Regression

With SWAP Linear Regression, we get the line,
y = 4.49728064380432 x - 2.8404993840819652

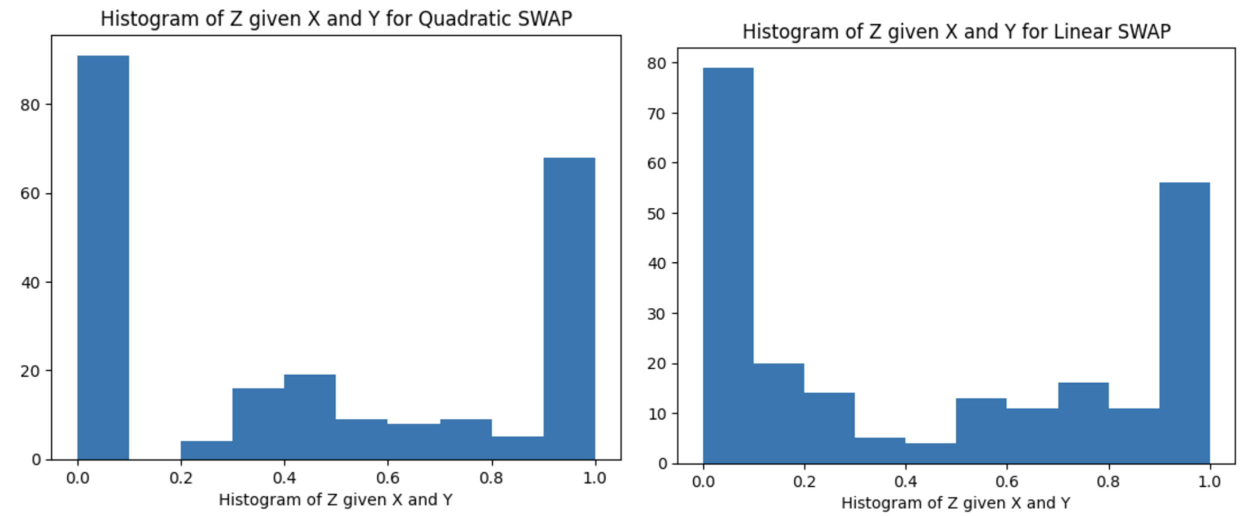
8.4 SWAP Quadratic Regression

With SWAP Quadratic Regression, we get the line,
y = 0.6453359354453013 x^2 + 0.8581953283583627 x + 0.11630141009211563



(a) (b) (c) (d)

Figure 6: Model Fitting



(a) (b)

Figure 7: Conditional Distribution of P(Z = 1)

9 Testing Goodness of SWAP

The test of goodness of the fit can be determined by how well our method can cluster the data into Z = 0 and Z = 1. We can define a good fit as one with more "surety" of Z to be 0 or 1. For the model used, we can look at the probabilities of Z = 1, as done in 1. We can get a Histogram of the probabilities to visualize it. However, since $P_{Z|X,Y}$ is a probability, it is between 0 and 1, so we can assume it to be a Beta Distribution. Since, we want that value to be close to 0 or 1, for more "surety" of whether Z is 0 or 1, we would want α and β to be as less that 1 as possible.

The plots 7a and 7b show that the values of Z we got are generally closer to 0 and 1 than to 0.5. To test it mathematically, we can obtain the Maximum Likelihood Estimates for α and β of both distributions.

For the Quadratic Case, we get $\hat{\alpha}_{MLE} = 0.09456035265627243$, $\hat{\beta}_{MLE} = 0.1314989721991328$.

For the Linear Case, we get $\hat{\alpha}_{MLE} = 0.25023439277569504$, $\hat{\beta}_{MLE} = 0.26475833796821263$.

Here, we can see that the estimates for the Quadratic SWAP are smaller than the estimates for the Linear SWAP Model, and are less than 1, so we can say that the Quadratic SWAP Model is a good fit and is better than the Linear SWAP Model.

10 Using the Beta Distribution as Loss Function

As we can see, we would like to see $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$ as low as possible, we can try to model our function as a Beta Distribution. We can do that using the assumption that $p_i \sim Beta(\alpha, \beta)$, where p_i is the probability that $Z_i = 1$. Since, to minimize α and β, we would like to have a closed form expression of $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$. For a Beta distribution, that is not possible. Volodin and Igor Nikolaevich showed that the CDF of a Beta function for a Beta function can be approximated as $J_x(\alpha, \beta)$, where α and β are small.[3] Since, we want α and β to be small, we can use J_x to approximate Beta.

$$J_x(\alpha, \beta) = \begin{cases} (1-\gamma)(\frac{x}{1-x})^{\alpha} & \text{if } 0 \leq x \leq 0.5 \\ 1-\gamma(\frac{1-x}{x})^{\beta} & \text{if } 0.5 \leq x \leq 1 \end{cases} \tag{4}$$

$$\epsilon = \alpha + \beta \tag{5}$$

$$\gamma = \frac{\alpha}{\epsilon} \tag{6}$$

Using this, we can get $j(x;\alpha,\beta) = \frac{d}{dx}J_x(\alpha,\beta)$.

$$j(x;\alpha,\beta) = \begin{cases} (1-\gamma)\alpha(\frac{x}{1-x})^{\alpha-1}\frac{1}{(1-x)^2} & \text{if } 0 \le x \le 0.5 \\ \gamma\beta(\frac{1-x}{x})^{\beta-1}\frac{1}{x^2} & \text{if } 0.5 \le x \le 1 \end{cases} \tag{7}$$

10.1 Maximum Likelihood Estimates for α and β

Suppose we have $X_i \sim J(\alpha,\beta)$, for i = 1 to N. We have X_0 = X, where X $\le$ 0.5, and X_1 where X $\ge$ 0.5. Define $\pi_1 = \frac{N_1}{N_0+N_1} = \frac{1}{N}\sum_i^N I(X_i > 0.5)$.

$$E[\pi_1] = \frac{1}{N}\sum_i^N P(X_i > 0.5) \tag{8}$$

Since, X_i are identically distributed, we can write as -

$$E[\pi_1] = P(X_1 > 0.5) \tag{9}$$

Since, $X_i \ \sim Beta(\alpha,\beta)$, we can approximate it as $J_x(\alpha,\beta)$.

$$E[\pi_1] = \int_{0.5}^{1} j(x;\alpha,\beta)dx \tag{10}$$

$$E[\pi_1] = J_1(\alpha,\beta) - J_{0.5}(\alpha,\beta) \tag{11}$$

$$E[\pi_1] = \gamma \tag{12}$$

Since, π_1 is an unbiased estimator for γ, we can replace γ with π_1. As we will see, this makes our calculations much easier later on.

$$log\ L(\alpha,\beta \mid X) = \sum_{i=1}^{N_0}[log(\frac{1-\gamma}{(1-X_{0i})^2})+log(\alpha)+(\alpha-1)log(\frac{X_{0i}}{1-X_{0i}})]+\sum_{i=1}^{N_1}[log(\frac{\gamma}{X_{1i}^2})+log(\beta)+(\beta-1)log(\frac{1-X_{1i}}{X_{1i}})] \tag{13}$$

Maximizing it wrt α -

$$\hat{\alpha} = \frac{N_0}{\sum_{i=1}^{N_0} log(\frac{1-X_{0i}}{X_{0i}})} \tag{14}$$

Maximizing it wrt β -

$$\hat{\beta} = \frac{N_1}{\sum_{i=1}^{N_1} log(\frac{X_{1i}}{1-X_{1i}})} \tag{15}$$

10.2 Loss Function

Since we want to maximize $\hat{\alpha}$ and $\hat{\beta}$ obtained in 14 and 15, we can define the new loss function as -

$$L(\theta|X) = -\frac{N_0}{\hat{\alpha}} - \frac{N_1}{\hat{\beta}} \tag{16}$$

$$L(\theta|X) = -\sum_{i=1}^{N_0} log(\frac{1-X_{0i}}{X_{0i}}) - \sum_{i=1}^{N_1} log(\frac{X_{1i}}{1-X_{1i}}) \tag{17}$$

We can make sense of 16 by the fact that minimizing it will minimize the parameters of the Beta distribution, and we are regularizing the values of α and β using the number of values of observation to determine them. Any loss function could have been chose which minimizes the parameters. The main reason we chose this particular loss function is because of its similarities to the one we got using GMM. Ignoring the constants, we get -

$$L(\theta|X,Y) = \sum_{i=1}^{N_0}[\frac{(X_{0i} - g^{-1}(Y_{0i}))^2}{\sigma_0^2} - \frac{(Y_{0i} - g(X_{0i}))^2}{\sigma_1^2}] + \sum_{i=1}^{N_1}[\frac{(Y_{1i} - g(X_{1i}))^2}{\sigma_1^2} - \frac{(X_{1i} - g^{-1}(Y_{1i}))^2}{\sigma_0^2}] \quad (18)$$

We get the final loss function as 18. We can make sense of it as 3, but with a regularization term, to try to separate Z = 0 from Z = 1, and Z = 1 from Z = 0. We can use 1, to optimize using this loss function.

Algorithm 2 Beta SWAP Algorithm

1: For all Z_i, we can predict $Z_i^{(t+1)} = argmax(P(Z_i^{(t+1)} = 0|X_i, Y_i), P(Z_i^{(t+1)} = 1|X_i, Y_i))$, using 1 and 2.
2: $g^{(t+1)} = argmin_g \sum_i^N [\frac{(Y_i - g(X_i))^2}{\sigma_1^{2(t)}} - \frac{(X_i - g^{-1}(Y_i))^2}{\sigma_0^{2(t)}}]I[Z_i^{(t+1)} = 1] + [\frac{(X_i - g^{-1}(Y_i))^2}{\sigma_0^{2(t)}} - \frac{(Y_i - g(X_i))^2}{\sigma_1^{2(t)}}]I[Z_i^{(t+1)} = 0])$, using 18.
3: $\epsilon_0^{(t+1)} = X_0^{(t+1)} - {g^{(t+1)}}^{-1}(Y_0^{(t+1)})$.
4: $\epsilon_1^{(t+1)} = Y_1^{(t+1)} - g^{(t+1)}(X_1^{(t+1)})$.
5: $\sigma_0^{2(t+1)} = \text{Var}(\epsilon_0^{(t+1)})$.
6: $\sigma_1^{2(t+1)} = \text{Var}(\epsilon_1^{(t+1)})$.
7: $N_0^{(t+1)} = \sum_i^N I[Z_i^{(t+1)} = 0]$.
8: $N_1^{(t+1)} = \sum_i^N I[Z_i^{(t+1)} = 1]$.

11 Other Cool Stuff

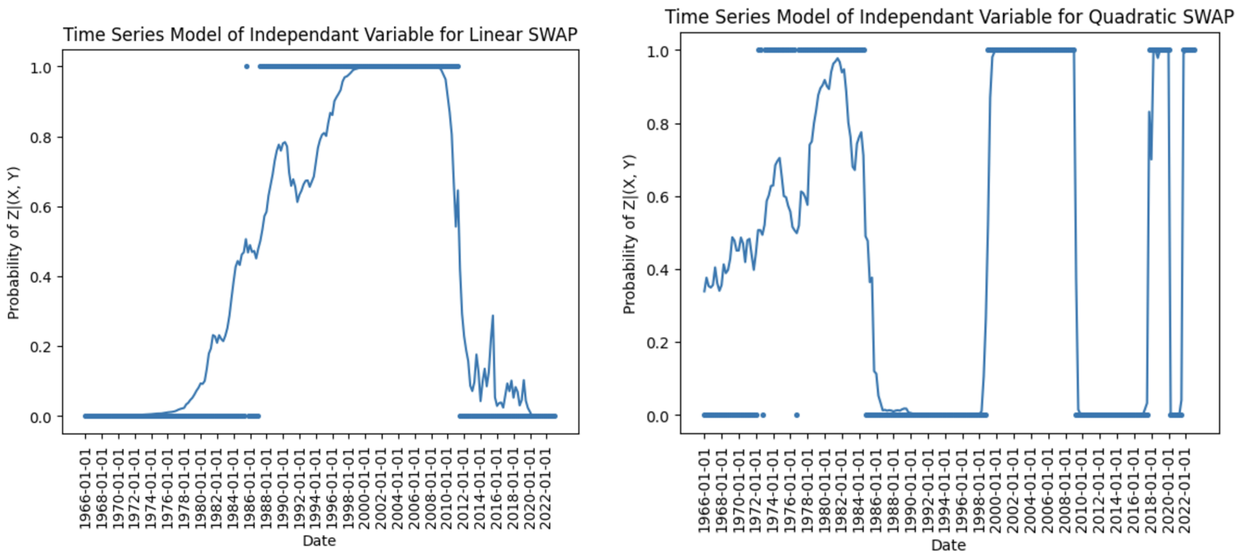
This can be used to determine when X causes Y and when Y causes X. This can be determined by the distribution of Z. Using both models, we can store the final values of $P_{Z^{(t)}|X,Y}$, and $Z^{(t)}$. This can be modeled with the time component to find a relation between X and Y. When Z = 0, X causes Y, and when Z = 1, Y causes X.

11.1 GMM Model

Refer 8.

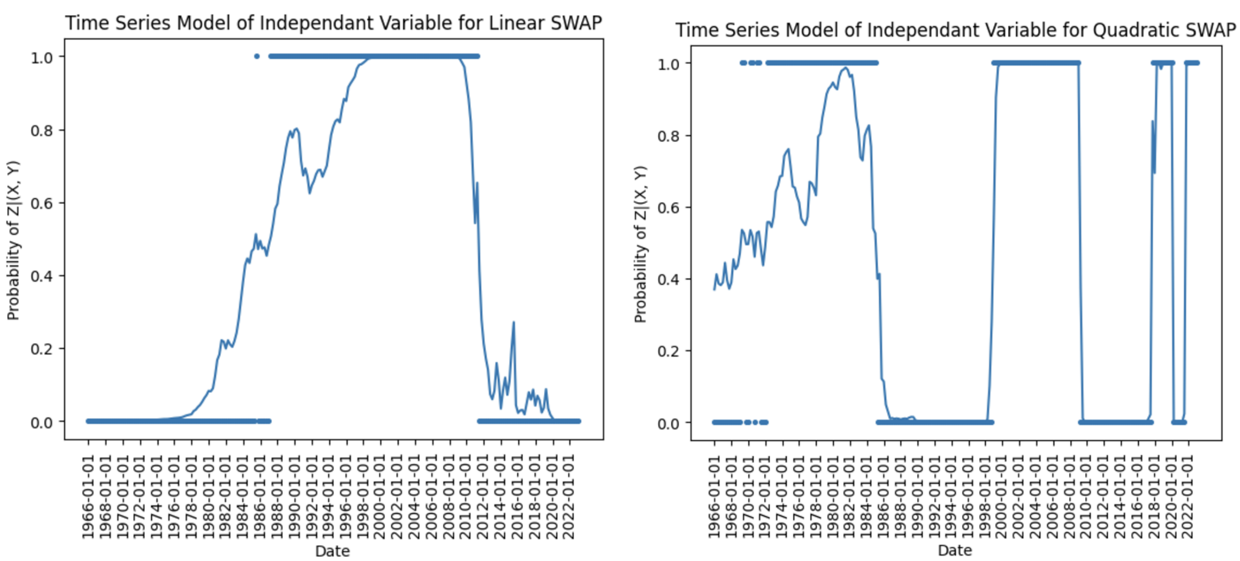
11.2 Beta Model

Refer 9.



(a) (b)

Figure 8: Swapping Variables wrt Time for GMM Model



(a) Beta Linear (b) Beta Quadratic

Figure 9: Swapping Variables wrt Time for Beta Model

References

[1] M. Chow, B. Li, and J. Q. Xue, "On regression for samples with alternating predictors and its application to psychrometric charts," *Statistica Sinica*, vol. 25, no. 3, pp. 1045–1064, 2015.

[2] G. De Vita, E. Trachanas, and Y. Luo, "Revisiting the bi-directional causality between debt and growth: Evidence from linear and nonlinear tests," *Journal of International Money and Finance*, vol. 83, pp. 55–74, 2018.

[3] I. N. Volodin, "Beta-distribution for small parameters," *Theory of Probability & Its Applications*, vol. 15, no. 3, pp. 546–549, 1970.