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 Real 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
```

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.

```
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
```

```
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 .
```

```
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
```

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 ap value of less than 0.01 on Augmented Dickey Fuller test, which we applied using the **adf.test** function in R.

5 Breaking Point

Here, visually, the breaking point seems to be 3650000, the point at which roles are swapped.

5.1 L2 Norm

Mathematically, Howard Lee proposed a method using the Chow Test to detect the breaking point. [3]
We have $\frac{\frac{e^T e - e_1^T e_1 - e_2^T e_2}{p}}{\frac{e_1^T e_1 + e_2^T e_2}{n+m-2p}} \sim F_{m,n-p}$, where n is number of observations in first dataset, and m in second dataset, and p is number of parameters. e is error of regression on total dataset, e_1 is error of regression on first dataset, and e_2 is error of regression on second dataset. Using this, we have implemented a Linear model, and find the value of x such that the p-value is lowest. In this case, we get x = 2475042.25 and p-value as 1.11 x 10^{-16}.

5.2 L1 Norm

Here, we have used a similar statistic as for L2 Norm.

$$\frac{\frac{\Sigma_i |e_i| - \Sigma_i |e_{1i}| - \Sigma_i |e_{2i}|}{p}}{\frac{\Sigma_i |e_{1i}| + \Sigma_i |e_{2i}|}{n+m-2p}}$$

where n is number of observations in first dataset, and m in second dataset, and p is number of parameters. e is error of regression on total dataset, e_1 is error of regression on first dataset, and e_2 is error of regression on second dataset. Here, we don't find the p-value, because we don't have a distribution, but assuming a distribution G, all we need is a point, we p-value is minimum, or the value of the Statistic is maximum. So, we don't need a distribution to find the Breaking Point. Using this idea, with this data, we get breaking point as GDP = 3509890.0.

6 Transforming Data

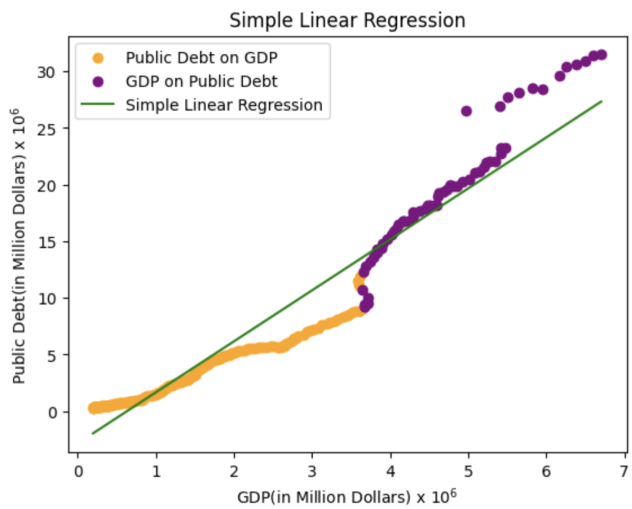
We have scaled the gdp and public debt data by a factor of 10^{-6}.

7 L2 Norm

7.1 Visual Breaking Point

7.1.1 Simple Linear Regression

First, let us treat them as linearly related. So, we can draw a simple linear regression for the model, all together.
With that, we get the line as y = 4.4972477180497235x - 2.840483200739338.



7.1.2 SWAP Linear Regression

We can implement SWAP Regression on the linear model with degree 1.
We use $L(a, b, c) = \Sigma_i (Y_{1i} - mX_{1i} - c)^2 + \Sigma_i (X_{0i} - \frac{Y_{0i} - c}{m})^2$ as the loss function, and minimize it with respect to m and c, using Gradient Descent.
With that, we get the line as y = 4.497247478436949x - 2.8404832396904025.

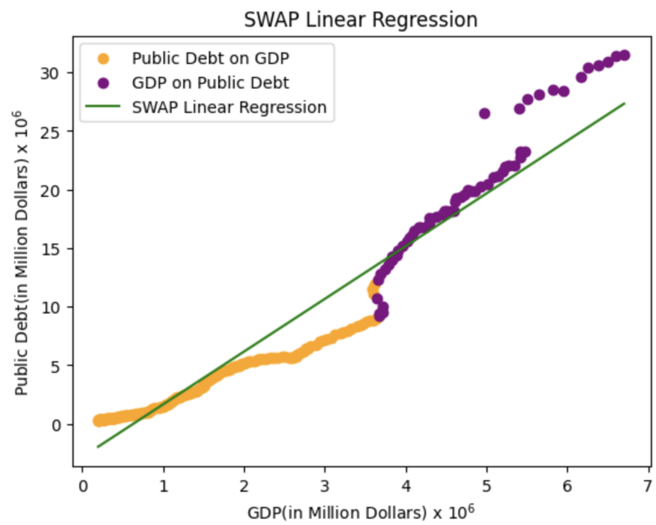
7.1.3 Simple Quadratic Regression

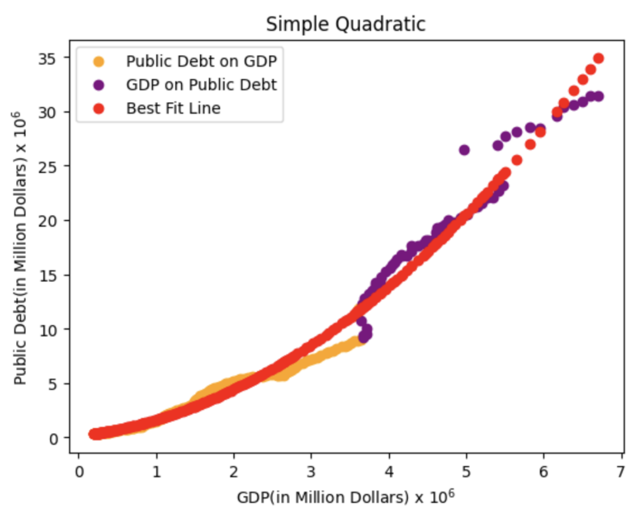
If we try to get this in a Quadratic Model, we will get the quadratic as
y = 0.6460656714354706x^2 + 0.8578272692490471x 0.11587014579886798.

7.1.4 SWAP Regression

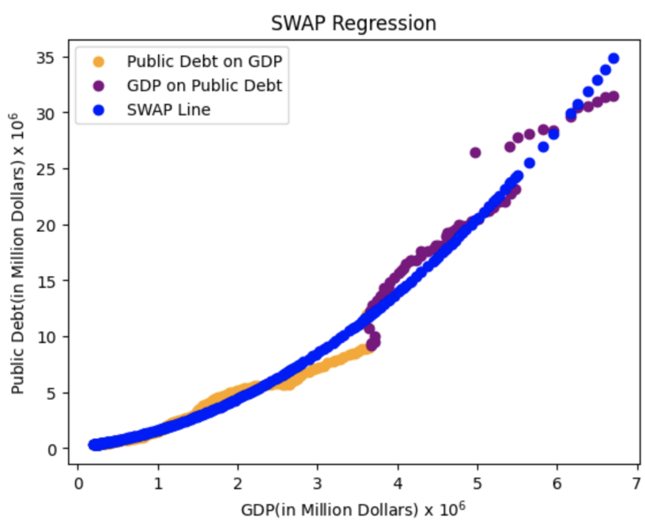
$$L(a, b, c) = \Sigma_i (Y_{1i} - aX_{1i}^2 - bX_{1i} - c)^2 + \Sigma_i (X_{0i} + \frac{b}{2a} - \frac{\sqrt{b^2 + 4ac - 4aY_{0i}}}{2a})^2$$

The inverse of the Quadratic over here is taking as such because we know the nature of the function. We can reverse the sign if the data set is shifted in another direction.


SWAP Linear Regression
Public Debt on GDP
GDP on Public Debt
SWAP Linear Regression
Public Debt(in Million Dollars) x 10^6
GDP(in Million Dollars) x 10^6
0
5
10
15
20
25
30
0
1
2
3
4
5
6
7


Simple Quadratic
Public Debt on GDP
GDP on Public Debt
Best Fit Line
Public Debt(in Million Dollars) x 10^6
GDP(in Million Dollars) x 10^6
0
5
10
15
20
25
30
35
0
1
2
3
4
5
6
7

We can optimize this using Gradient Descent. However, the problem arises due to the square root. Over here, we can use the trick of only considering the points where the square root is defined. For considering the final gradient, we can average out the sum with the total number of points which were defined. We get the final quadratic as
y = $0.6447196501501261x^2 + 0.8574630615840181x + 0.11578189044160224$



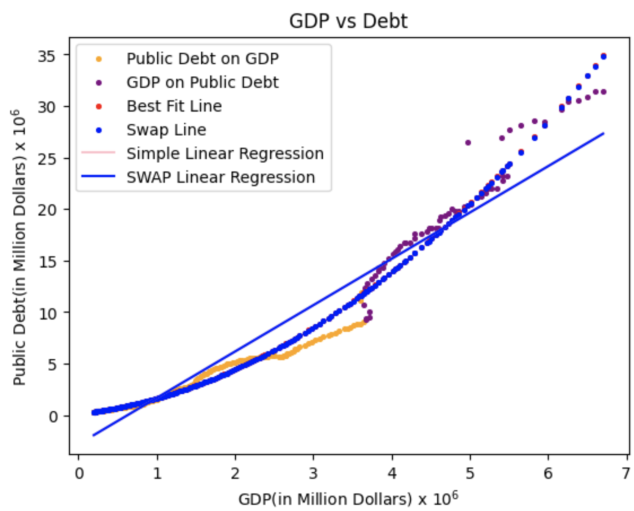
7.1.5 Error Analysis

The Gradient of the error, i.e. $(\frac{\partial L}{\partial a})^2 + (\frac{\partial L}{\partial b})^2 + (\frac{\partial L}{\partial c})^2$ is decreasing, and appears to tend to 0, as iterations increases.

7.1.6 Comparison and Analysis

Like in the paper, we have defined the error as
$L(a, b, c) = \Sigma_i(Y_{1i} - aX_{1i}^2 - bX_{1i} - c)^2 + \Sigma_i(X_{0i} + \frac{b}{2a} - \frac{\sqrt{b^2+4ac-4aY_{0i}}}{2a})^2$

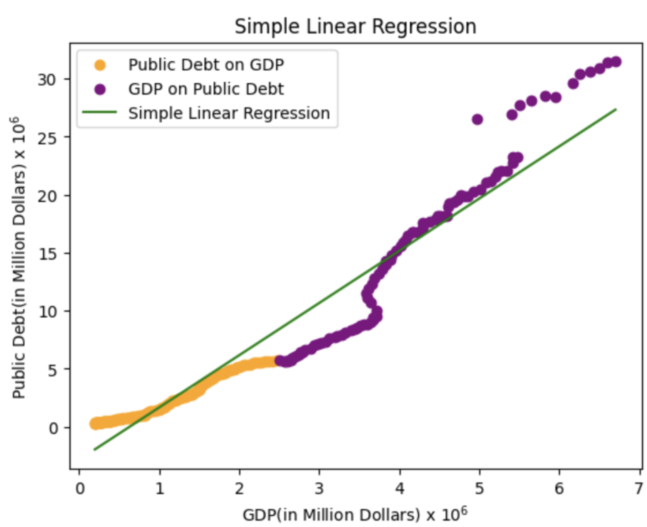
Model	Error
Simple Linear Regression	2.9538907319184533
SWAP Linear Regression	2.9538898722529736
Simple Quadratic Regression	0.4644160500769963
SWAP Regression	0.4595191446188699



7.2 Chows Test Breaking Point

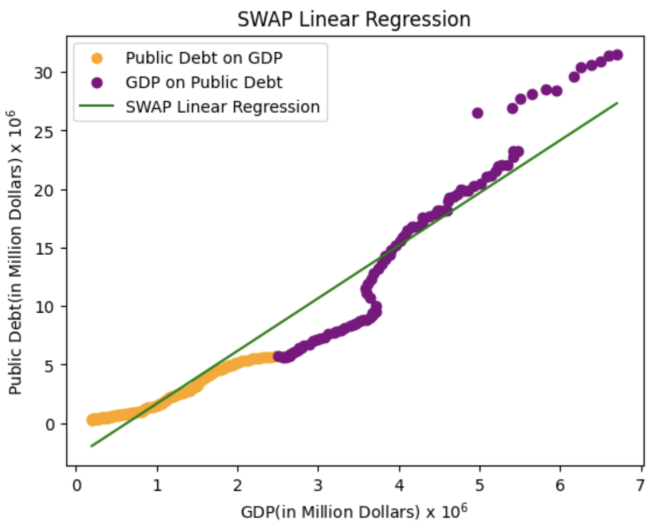
7.2.1 Simple Linear Regression

First, let us treat them as linearly related. So, we can draw a simple linear regression for the model, all together.
With that, we get the line as y = 4.4972477180497235x - 2.840483200739338.



7.2.2 SWAP Linear Regression

We can implement SWAP Regression on the linear model with degree 1.
We use $L(a,b,c) = \Sigma_i(Y_{1i} - mX_{1i} - c)^2 + \Sigma_i(X_{0i} - \frac{Y_{0i}-c}{m})^2$ as the loss function, and minimize it with respect to m and c, using Gradient Descent.
With that, we get the line as y = 4.497247689574723x - 2.8404831719534753.



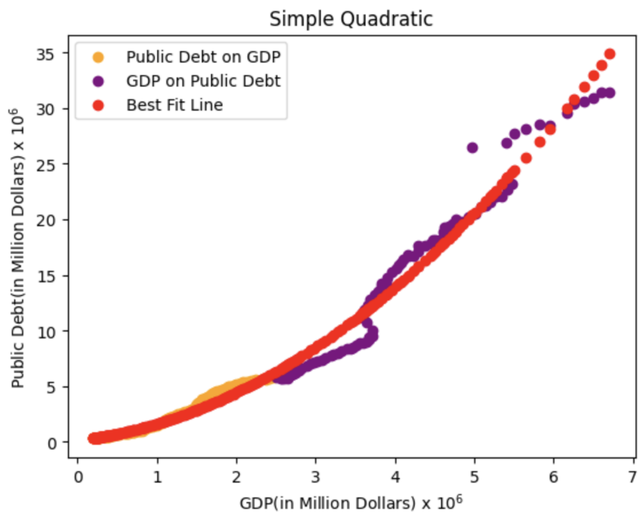
7.2.3 Simple Quadratic Regression

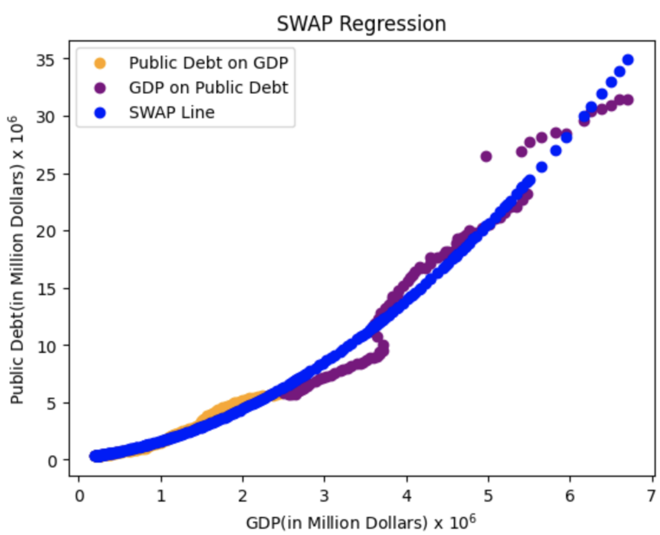
If we try to get this in a Quadratic Model, we will get the quadratic as
y = $0.6460656714354706x^2 + 0.8578272692490471x + 0.11587014579886798$.

7.2.4 SWAP Regression

$L(a,b,c) = \Sigma_i(Y_{1i} - aX_{1i}^2 - bX_{1i} - c)^2 + \Sigma_i(X_{0i} + \frac{b}{2a} - \frac{\sqrt{b^2+4ac-4aY_{0i}}}{2a})^2$

The inverse of the Quadratic over here is taking as such because we know the nature of the function. We can reverse the sign if the data set is shifted in another direction.
We can optimize this using Gradient Descent. However, the problem arises due to the square root. Over here, we can use the trick of only considering the points where the square root is defined. For considering the final gradient, we can average out the sum with the total number of points which were defined.
We get the final quadratic as
y = $0.6463311007780554x^2 + 0.8579547627153442x + 0.11592895773043216$

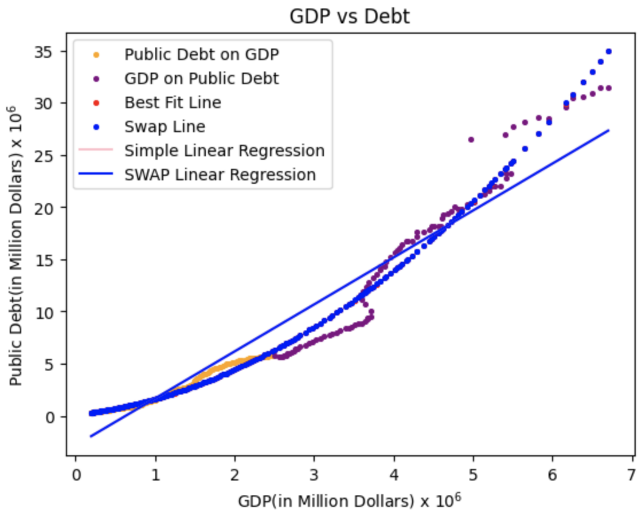




7.2.5 Error Analysis

The Gradient of the error, i.e. $(\frac{\partial L}{\partial a})^2 + (\frac{\partial L}{\partial b})^2 + (\frac{\partial L}{\partial c})^2$ is decreasing, and appears to tend to 0, as iterations increases.

7.2.6 Comparison and Analysis



Like in the paper, we have defined the error as
$L(a,b,c) = \Sigma_i(Y_{1i} - aX_{1i}^2 - bX_{1i} - c)^2 + \Sigma_i(X_{0i} + \frac{b}{2a} - \frac{\sqrt{b^2+4ac-4aY_{0i}}}{2a})^2$ for Quadratic Case and
$L(a,b,c) = \Sigma_i(Y_{1i} - mX_{1i} - c)^2 + \Sigma_i(X_{0i} - \frac{Y_{0i}-c}{m})^2$ for Linear Case.

Model	Error
Simple Linear Regression	1.19428932806907l4
SWAP Linear Regression	1.194289304430548
Simple Quadratic Regression	0.11524650520418465
SWAP Regression	0.11503248109406963

8 L1 Norm

It is difficult to implement Gradient Descent on an L1 Norm, so we have implement different techniques to solve it.

8.1 Revisiting the MM Algorithm

The MM Algorithm [4] is a method used to minimize a Loss Function, f(θ). It is an iterative algorithm. The idea is to find a function g($\theta|\theta_{(k)}$) such that for each iteration k,
1) f(θ) $\leq$ g($\theta|\theta_{(k)}$)
2) f($\theta_{(k)}$) = g($\theta_{(k)}|\theta_{(k)}$)
Then, the algorithm proceeds as -
$\theta_{(k+1)}$ = argmin$_\theta$ [g($\theta|\theta_{(k)}$)]
We stop once we get $||\theta_{(k+1)} - \theta_{(k)}||_2$ small enough.

8.2 Solving an L1 Norm System

Let us take the function f(u) = $u^{\frac{1}{2}}$. Since this is a concave function, by rooftop theorem, we get
$u^{\frac{1}{2}} \leq {u^*}^{\frac{1}{2}} + (u - u^*)(\frac{1}{2}{u^*}^{-\frac{1}{2}})$
$u^{\frac{1}{2}} \leq \frac{1}{2}{u^*}^{\frac{1}{2}} + \frac{1}{2}u{u^*}^{-\frac{1}{2}}$
In this case, we can see that
$g(u|u^*) = \frac{1}{2}{u^*}^{\frac{1}{2}} + \frac{1}{2}u{u^*}^{-\frac{1}{2}}$
For L1 Norm, we have
$L(\beta) = \Sigma_i|y_i - f(x_i|\beta)|$
Using the rooftop equation, we can get
$u_i^{\frac{1}{2}} \leq \frac{1}{2}{u_i^*}^{\frac{1}{2}} + \frac{1}{2}u_i{u_i^*}^{-\frac{1}{2}}$
Define : $u_i = (y_i - f(x_i|\beta))^2$ and $u_i^* = (y_i - f(x_i|\beta_{(k)}))^2$
$L(\beta) \leq \frac{1}{2}\Sigma_i(|y_i - f(x_i|\beta_{(k)})| + \frac{(y_i - f(x_i|\beta))^2}{|y_i - f(x_i|\beta_{(k)})|})$
Define : $g(\beta|\beta_{(k)}) = \frac{1}{2}\Sigma_i(|y_i - f(x_i|\beta_{(k)})| + \frac{(y_i - f(x_i|\beta))^2}{|y_i - f(x_i|\beta_{(k)})|})$
we can clearly see how it follows 1) and 2) of the MM Algorithm taking f = L function.
$\text{argmin}_\beta g(\beta|\beta_{(k)}) = \text{argmin}_\beta \Sigma_i \frac{(y_i - f(x_i|\beta))^2}{|y_i - f(x_i|\beta_{(k)})|}$
The L1 Norm has effectively been converted to an L2 Norm with Weights, as the denominator is constant with respect to β. If a direct solution is not possible, Numerical Methods such as Gradient Descent or Newton Raphson is possible on this new function.

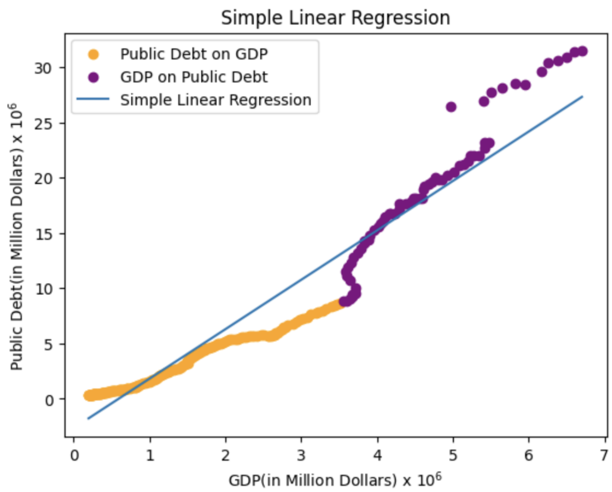
8.3 L1 Norm SWAP Algorithm

For the SWAP Quadratic case, we have
$f(x_i|a, b, c) = ax_i^2 + bx_i + c$
Define : $w_{i,(k),1} = |y_{1i} - a_{(k)}x_{1i}^2 - b_{(k)}x_{1i} - c_{(k)}|$ and $w_{i,(k),0} = |x_{0i} + \frac{b_{(k)}}{2a_{(k)}} - \frac{\sqrt{b_{(k)}^2 + 4a_{(k)}c_{(k)} - 4a_{(k)}y_{0i}}}{2a_{(k)}}|$
Effectively, we have to minimize
$\Sigma_i \frac{(y_{1i} - ax_{1i}^2 - bx_{1i} - c)^2}{w_{i,(k),1}} + \Sigma_i \frac{(x_{0i} + \frac{b}{2a} - \frac{\sqrt{b^2 + 4ac - 4ay_{0i}}}{2a})^2}{w_{i,(k),0}}$
We can see how this is equivalent to the OLS with SWAP Regression. We can directly implement Gradient Descent on it, like with SWAP Regression, with an additional factor of the weights, which are constant with respect to a, b, c. This directly gives us the iteration.

8.4 Fitting the Models

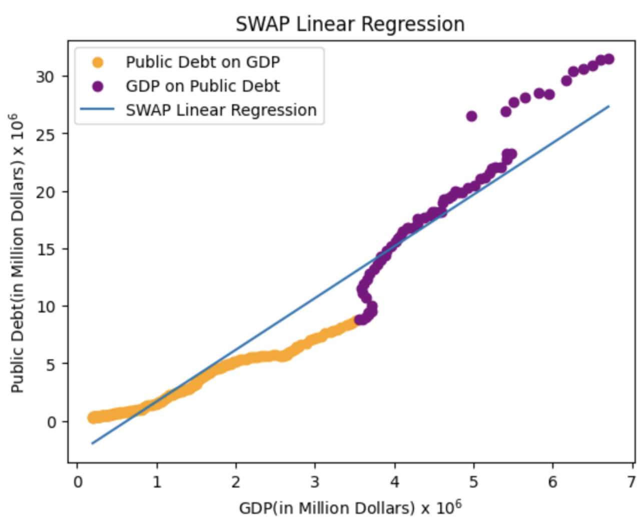
8.4.1 Simple Linear Regression

First, let us treat them as linearly related. So, we can draw a simple linear regression for the model, all together.
With that, we get the line as y = 4.47297417x - 2.68125833.



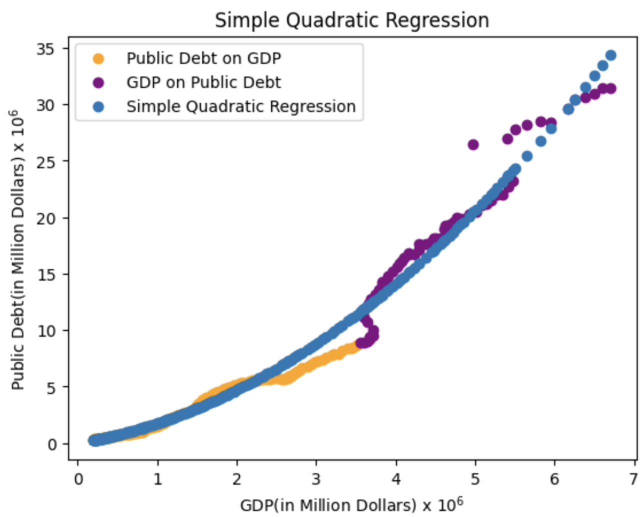
8.4.2 SWAP Linear Regression

We can implement SWAP Regression on the linear model with degree 1.
We use $L(a,b,c) = \Sigma_i (Y_{1i} - mX_{1i} - c)^2 + \Sigma_i (X_{0i} - \frac{Y_{0i} - c}{m})^2$ as the loss function, and minimize it with respect to m and c, using Gradient Descent.
With that, we get the line as y = 4.49724772x - 2.8404832.



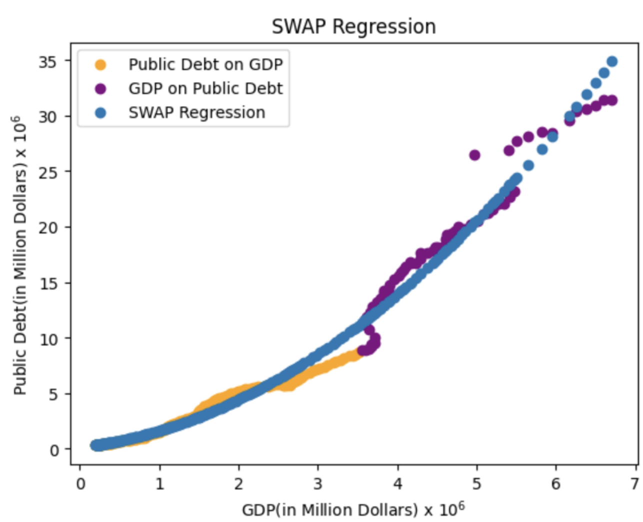
8.4.3 Simple Quadratic Regression

If we try to get this in a Quadratic Model, we will get the quadratic as
y = 0.58982848x^2 + 1.17061799x - 0.01448315.

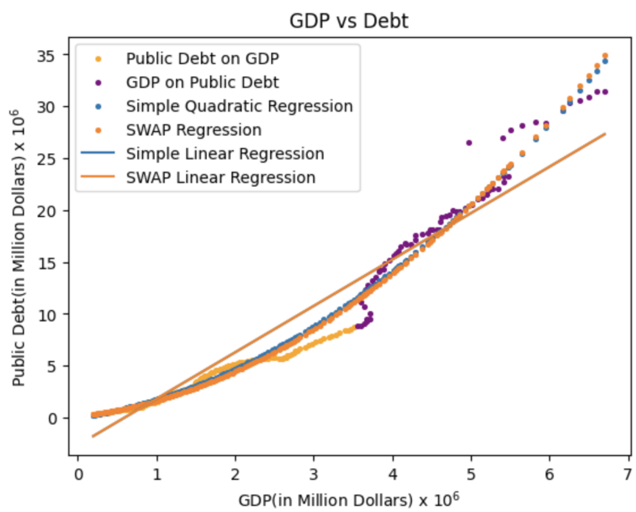


8.4.4 SWAP Regression

Finally, implementing the SWAP Regression with L1 Norm, we get the final quadratic as
y = $0.64606567x^2 + 0.85782727x + 0.11587015$.



8.4.5 Comparison and Analysis



Like in the paper, we have defined the error as
$L(a,b,c) = \Sigma_i|Y_{1i} - aX_{1i}^2 - bX_{1i} - c| + \Sigma_i|X_{0i} + \frac{b}{2a} - \frac{\sqrt{b^2+4ac-4aY_{0i}}}{2a}|$ for Quadratic Case
and
$L(a,b,c) = \Sigma_i|Y_{1i} - mX_{1i} - c| + \Sigma_i|X_{0i} - \frac{Y_{0i}-c}{m}|$ for Linear Case.

Model	Error
Simple Linear Regression	1.20614599662417
SWAP Linear Regression	1.2185641198876926
Simple Quadratic Regression	0.34869045972817997
SWAP Regression	0.3371849609170245

9 Final Analysis

Here, we can see, in this type of data, we can see that SWAP is working better than the other Linear Models for this Bi Directional Causal Dataset.

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] H. Lee, "Using the chow test to analyze regression discontinuities," *Tutorials in Quantitative Methods for Psychology*, vol. 4, 09 2008.

[4] K. Lange, *The MM Algorithm*, pp. 185–219. New York, NY: Springer New York, 2013.