1 Aligning the dates for the three datasets

```
head:
    year  num_mortage  num_possession  Unemployment rate  IR rate
0  1975    5076000.0          4870.0                4.5  11.0000
1  1976    5322000.0          4950.0                5.4  11.1137
2  1977    5582000.0          4680.0                5.6   8.8772
tail:
     year  num_mortage  num_possession  Unemployment rate  IR rate
36  2011   11384000.0         37300.0                8.1      0.5
37  2012   11284000.0         33900.0                8.0      0.5
38  2013   11186000.0         28900.0                7.6      0.5
```

Figure 1: head and tails of the aligned dataset

	num_moratage	num_possession	Unemployment rate
IR_rate	-0.760988	-0.285785	0.242834

Table 1: coefficients of correlation of IR_rate with independent variables

From table 1 we observed a strong negative correlation between IR_rate and "num_mortage". In following sections, the independent variables are "num_mortage", "num_ possession", "Unemployment rate", the dependent variable is "IR rate", we hope to model the relationship between independent and dependent variables using simple regression models, polynomial regression models and piecewise polynomial regression models.

2 Simple Regression Models

Model name	R^2	Log-likelihood	AIC	BIC
SR	0.598	-93.853	195.7	202.4
SR_updated	0.598	-93.853	193.7	198.7

Table 2: Model quality quantities table for SR and SR_updated

For SR_updated model, we removed the "Unemployment rate" columns from the predictors as SR model summary suggests $P > |t|$ of this coefficient is 0.988 hence this column is highly likely irrelevant. From the AIC and BIC value in table 1, the SR_updated is better.

3 Polynomial Regression Models

We observed from table 3 that the best model is adding every predictor raised to power of 2, and orthogonalising them.

4 Piecewise Polynomial Regression Models

We picked 4 knots for each predictor for the piecewise polynomial regression by using quantile method. From Table 4 and Figure 4 we can see the this PPR outperforms all previous models in every quantity chosen, though the BIC of PPR is still lower than the of PR_2_ortho, we slightly worry about the problem of overfitting.

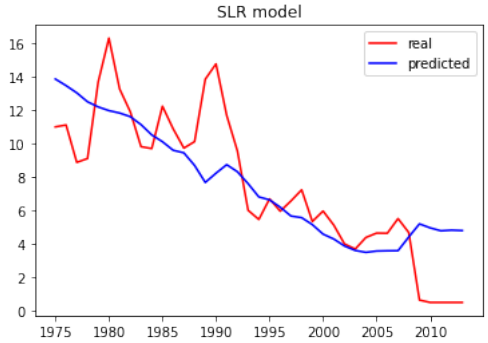
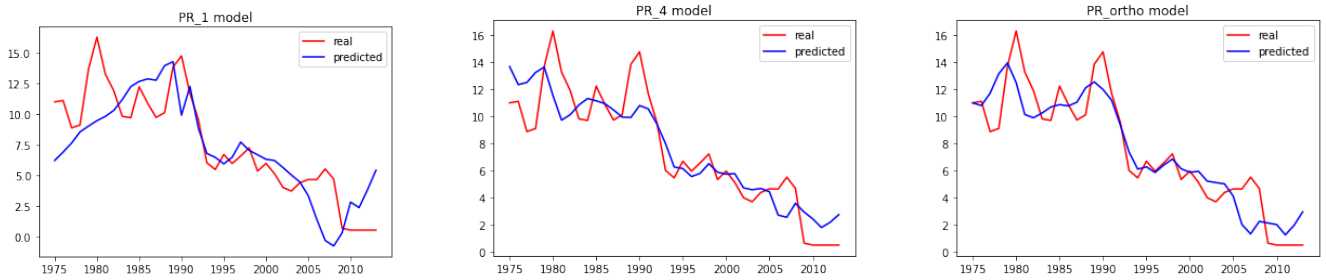


Figure 2: Plotting of the real and predicted IR_rate using SR_updated model against year

Model name	raised power	orthogonalised?	R^2	Log-likelihood	AIC	BIC
PR_4	4	No	0.544	-96.306	198.6	203.6
PR_3	3	No	0.726	-86.386	180.8	187.4
PR_2	2	No	0.777	-82.351	176.7	186.7
PR_2_ortho	2	Yes	0.824	-77.733	169.5	181.1

Table 3: Model quality quantities table for various polynomial regression models



(a) PR_4 (b) PR_2 (c) PR_2_orthogonal

Figure 3: plottings with different polynomial regression models

Model name	R^2	Log-likelihood	AIC	BIC
PPR	0.926	-60.774	153.5	180.2
SR_updated	0.598	-93.853	193.7	198.7
PR_2_ortho	0.824	-77.733	169.5	181.1

Table 4: Model quality quantities table for PPR and best SR and PR models

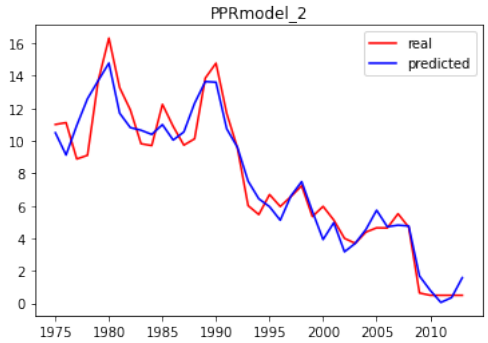


Figure 4: Plotting of the real and predicted IR_rate using PR model against year