Richard Yung
richard.yung@gmail.com
VEE regression analysis spring 2009
(510)9150188

Paid loss triangle: Parameter Stability

Introduction:

As mentioned in the background description of the paid loss triangle, it is mainly affected by three factors: volume of business growth, inflation, and loss payment patterns. Regression analysis can use past or simulated data conveniently to estimate the coefficients of these three factors. Since business growth, inflation and loss payment patterns are very unlikely to remain constant from year to year; one must understand changes that are needed in regression analysis when any or all of these factors changed. In this project, regression analysis is used to generate residual plots of different simulations based on how inflation rate changes and which variables are regressed while volume of business growth and loss payment patterns of the regression equation are constant. Simulated random errors were also introduced to see if the generated residual plots will reflect how the calculated regression equations are affected by stochasticity.

Simulation Setup:

A VBA macro MS Excel was written in a way to simulate paid loss triangle, perform regression analysis using its Data Analysis Toolpak, and generate residual plots against development and calendar years. The macro utilized parameters entered in the worksheet named *parameters* and simulates 5 scenarios:

Simulation 1: constant inflation rate with CY and DY regressed
Simulation 2: discretely changed inflation rate with only CY and DY regressed
Simulation 3: discretely changed inflation rate with all variables regressed
Simulation 4: continuously changed inflation rate with only CY and DY regressed
Simulation 5: continuously changed inflation rate with all variables regressed

The only difference of parameter usage is that β_{2b} was used as the final inflation rate in simulation 4 and 5. Consequently, β_2 continuously shifted towards β_{2b} at a constant rate each consecutive year. The rate of change was calculated from β_2, β_{2b}, and the year when inflation rate started to change. The parameters' names and the method of simulating errors were the same ones listed in illustrative worksheet provided by NEAS. In this project, different levels of stochasticity via different values of σ were tested in all simulations. Also, the F-test method was used to compare simulations 2 and 3 and simulations 4 and 5.

Proposed Equations:

In order to test if a proposed regression equation fits well, sigma is set to 0 in the simulations. Since simulation 1 assumes all parameters (α, β_1, β_2, are constant), the proposed equation is $E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \varepsilon$. Simulation 2 and 3, on the contrary have to use a modified equation from simulation 1 because of the changed inflation rate. By using a dummy variable and introducing another variable *k*, where the end of k^{th} year is when inflation rate changed. Thus, the *E(Y)* should be equations:

$E(Y) = \alpha + \beta_1(DY) + \beta_2(k) + \beta_{2b}(CY - k) + \varepsilon$ when inflation rate changed, and

$E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \varepsilon$ before the inflation rate changed.

Simulation 4 and 5 also use a modified equation from simulation 1 also because of the changed inflation rate. However, the inflation rate is changed continuously instead. Another estimator β_{2a} and variable *k* are introduced here as well. Let *r* be the continuous rate of inflation rate change and k is where the end of k^{th} year is when inflation rate changed. Thus, $\beta_{2a} = \dfrac{\beta_{2b} - \beta_2}{N_{CY} - k}$.Thus the *E(Y)* should be equations:

$E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \beta_{2a}\dfrac{(CY - k)(CY - k + 1)}{2} + \varepsilon$ when inflation rate changed, and $E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \varepsilon$ before the inflation rate changed. (See appendix for the derivations of these equations)

Results of 1st Simulation Run:

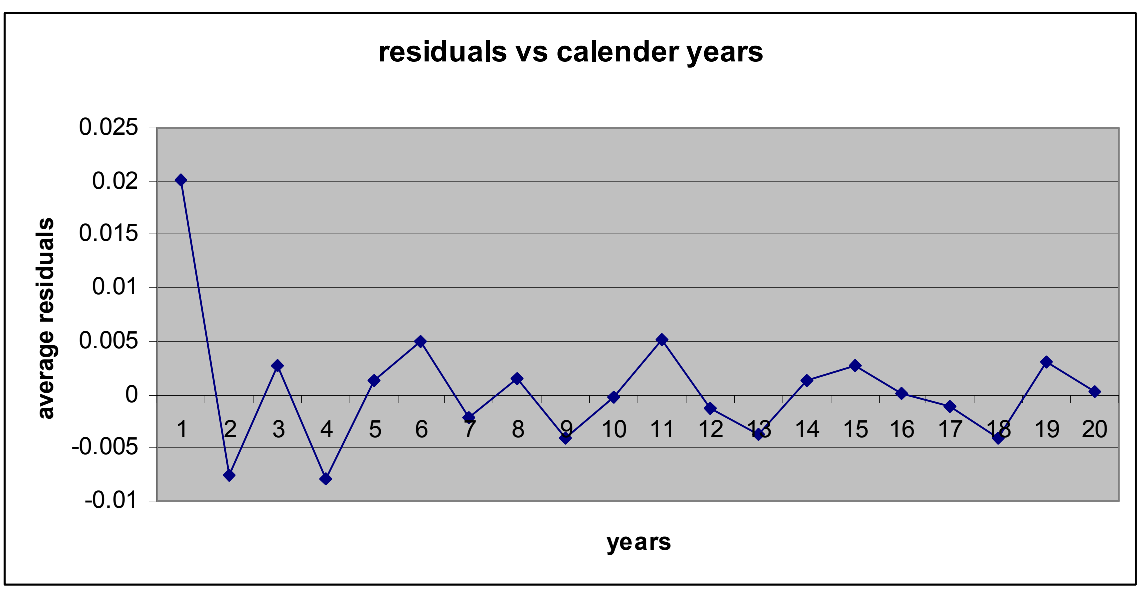
The importance of this run was that all simulations used the very same low sigma of 0.01, thus testing the accuracy of each simulation. The other parameters were also the same in each simulation.

Parameters for Run #1:

Sigma	0.01
Alpha	15
beta1	-0.15
beta2	0.05
beta1b	-0.15
beta2b	0.2

CY	20
DY	20
x^{th} year inflation changed?	12
beta2a	0.01875

Simulation 1:



CY residuals:

1	2	3	4	5
0.020107	-0.00755	0.002731	-0.00782	0.001252
6	7	8	9	10
0.005004	-0.00214	0.001526	-0.00414	-0.00021
11	12	13	14	15
0.005082	-0.00137	-0.00368	0.001311	0.002707
16	17	18	19	20
0.000151	-0.00107	-0.00403	0.003003	0.000232

Simulation 1 statistics:

Regression Statistics	
Multiple R	0.999862
R Square	0.999724
Adjusted R Square	0.999722
Standard Error	0.010675
Observations	210

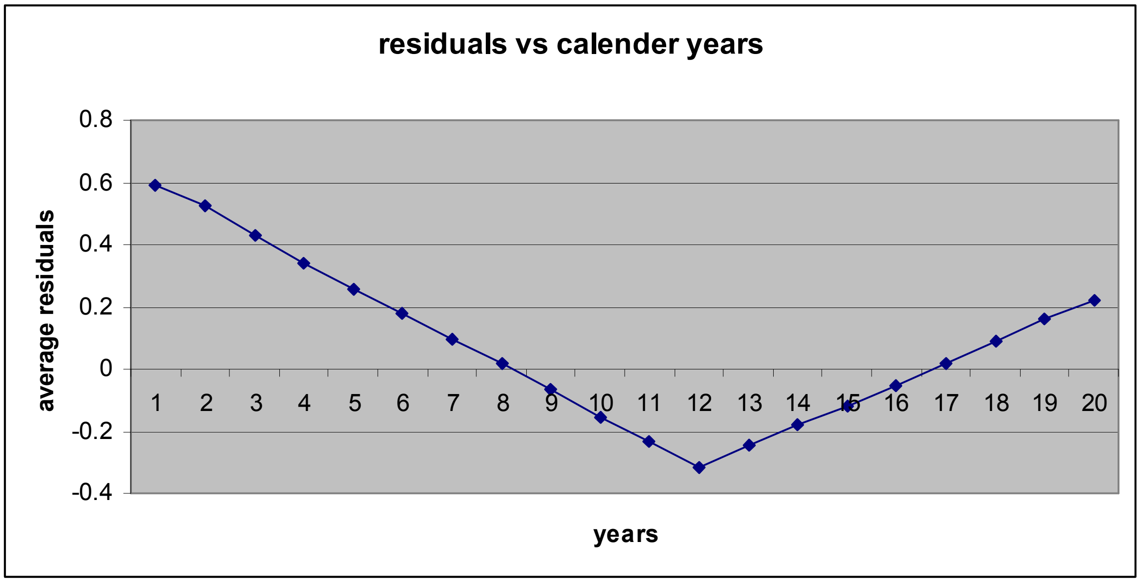
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	85.50894	42.75447	375187.8	0
Residual	207	0.023589	0.000114		
Total	209	85.53253			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Alpha	15.00072	0.002072	7240.894	0	14.99664	15.00481
Beta1	-0.15015	0.000177	-850.673	0	-0.1505	-0.14981
Beta2	0.050085	0.000177	283.7458	3.8E-270	0.049737	0.050433

As expected, the adjusted R^2 is approximately 1, the standard error is close to 0.01, and the coefficient α, β_1, β_2, were also very close to their respective inputted values. Significance of F-stat is 0, indicating that hypothesis of β_1, β_2 are 0 can be rejected. The average residuals plotted against calendar years showed a relatively horizontal line proving that the equation $E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \varepsilon$ is a very good fit if the coefficients are assumed to be constant throughout the development and calendar years.

Simulation 2:



CY residuals:

1	2	3	4	5
0.592351	0.526622	0.428761	0.339294	0.253796
6	7	8	9	10
0.180753	0.094102	0.015025	-0.06804	-0.15225
11	12	13	14	15
-0.23288	-0.31816	-0.24597	-0.18092	-0.11684
16	17	18	19	20
-0.05241	0.020577	0.087825	0.15937	0.222798

Simulation 2 statistics:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.962897
R Square	0.927171
Adjusted R Square	0.926468
Standard Error	0.193122
Observations	210

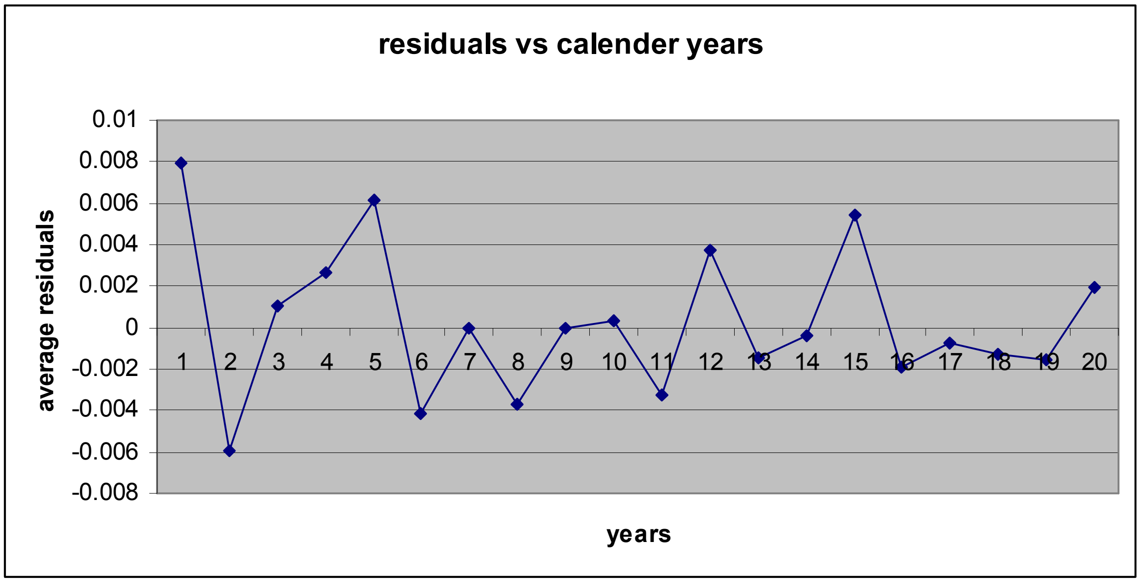
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	98.28578	49.14289	1317.642	1.8E-118
Residual	207	7.720288	0.037296		
Total	209	106.0061			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Alpha	14.41004	0.037479	384.4855	2.1E-297	14.33615	14.48393
Beta1	-0.14993	0.003193	-46.9514	2.5E-112	-0.15623	-0.14363
Beta2	0.132368	0.003193	41.45174	3.4E-102	0.126072	0.138663

The adjusted R^2 and the standard error were calculated to around 0.93 and 0.19 respectively. Furthermore, even though the coefficient α and β_1 are close to the respective inputted values, β_2 was found to be no where near 0.05. The average residuals plotted against calendar years showed a V shape, proving that the equation $E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \varepsilon$ was not a good fit for a linear model.

Simulation 3:



CY residuals:

1	2	3	4	5
0.007984	-0.00591	0.00104	0.002682	0.006154
6	7	8	9	10
-0.00412	-1.1E-06	-0.00373	-6.1E-05	0.000297
11	12	13	14	15
-0.00322	0.003744	-0.00142	-0.00037	0.005462
16	17	18	19	20
-0.00189	-0.00076	-0.00126	-0.00154	0.001939

Simulation 3 statistics:

Regression Statistics	
Multiple R	0.999908
R Square	0.999817
Adjusted R Square	0.999814
Standard Error	0.009724
Observations	210

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	3	106.1614	35.38714	374262.8	0
Residual	206	0.019478	9.46E-05		
Total	209	106.1809			

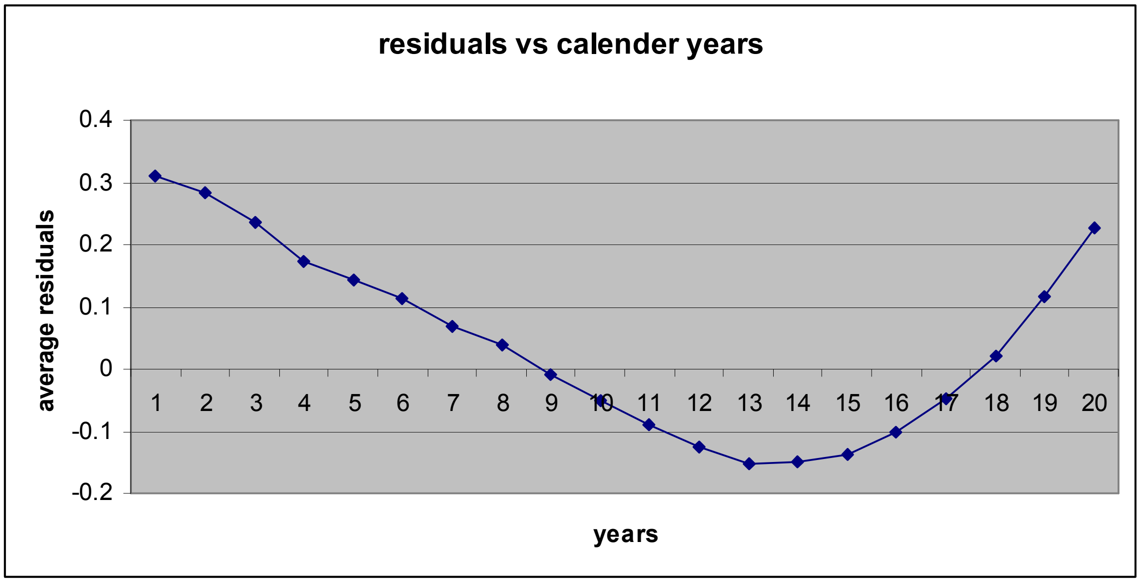
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Alpha	15.00684	0.002806	5347.173	0	15.0013	15.01237
Beta1	-0.15007	0.000161	-933.36	0	-0.15039	-0.14975
Beta2	0.049135	0.000332	148.189	2.9E-211	0.048481	0.049788
Beta2b	0.200295	0.000286	699.5937	0	0.199731	0.20086

By including an extra variable which indicated the year when inflation rate changed in this regression analysis, the adjusted R^2 became very close to 1, and the coefficient α, β_1, β_2, and β_{2b} were also very close to their respective inputted values. F-test would be used to compare simulation 2 and 3. Since simulation 2's regression equation was $E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \varepsilon$ and simulation 3's regression is $E(Y) = \alpha + \beta_1(DY) + \beta_2(k) + \beta_{2b}(CY - k) + \varepsilon$, simulation 2's regression equation would be restricted and simulation 3's regression equation would be unrestricted. So F-test would be:

$$F_{q,N-k} = \frac{(R_{UR}^2 - R_R^2)/q}{(1 - R_{UR}^2)/(N-k)} \to F_{1,210-3} = \frac{(0.999817 - 0.927171)/1}{(1 - 0.999817)/(210-3)} \approx 82173.34$$

The F-test statistic was greater than the critical values for $F_{1,210-3}$ between 3.92 and 3.84 for 5% significance, and between 6.85 and 6.63 for 1% significance. Either way, this showed that variable k introduced in simulation 3 was statistically significant.

Simulation 4:



CY residuals:

1	2	3	4	5
0.310926	0.283901	0.234673	0.172619	0.144736
6	7	8	9	10
0.113365	0.067702	0.037405	-0.00864	-0.04929
11	12	13	14	15
-0.08957	-0.1241	-0.153	-0.14776	-0.13878
16	17	18	19	20
-0.10255	-0.04748	0.021979	0.117356	0.227844

Simulation 4 statistics:

Regression Statistics	
Multiple R	0.97969
R Square	0.959793
Adjusted R Square	0.959404
Standard Error	0.129966
Observations	210

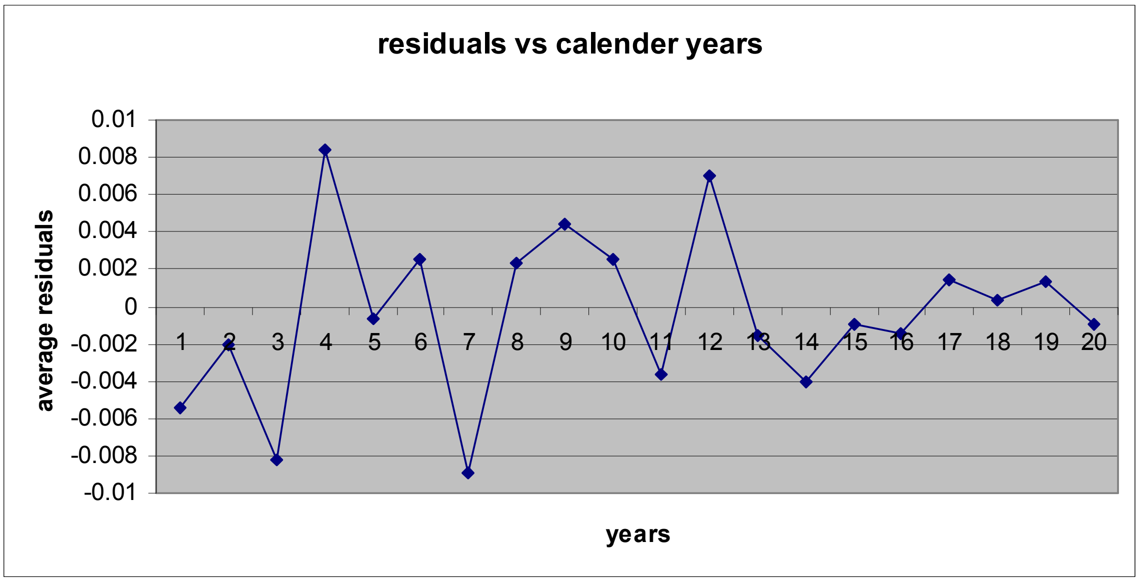
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	83.4647	41.73235	2470.66	3.5E-145
Residual	207	3.496472	0.016891		
Total	209	86.96117			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Alpha	14.68983	0.025222	582.4159	0	14.6401	14.73955
Beta1	-0.15008	0.002149	-69.8372	7.3E-146	-0.15432	-0.14584
Beta2	0.08994	0.002149	41.85183	5.7E-103	0.085703	0.094177

The adjusted R^2 and the standard error were calculated to around 0.95 and 0.13 respectively. Also, only the coefficient β_1 are close to the respective inputted value. α was found to be 14.69 and β_2 was found to be near 0.09. The average residuals plotted against calendar years showed a rounded v shape, proving that the equation $E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \varepsilon$ was not a good fit for a linear regression model.

Simulation 5:



CY residuals:

1	2	3	4	5
-0.00539	-0.002000603	-0.00816	0.008431	-0.00062
6	7	8	9	10
0.002575	-0.008891699	0.002328	0.00441	0.002493
11	12	13	14	15
-0.00362	0.007003282	-0.00157	-0.00405	-0.00093
16	17	18	19	20
-0.00142	0.001417959	0.000372	0.001297	-0.00096

Simulation 5 statistics:

Regression Statistics	
Multiple R	0.999881
R Square	0.999762
Adjusted R Square	0.999758
Standard Error	0.010024
Observations	210

ANOVA

	df	SS	MS	F	Significance F
Regression	3	86.79378	28.93126	287941	0
Residual	206	0.020698	0.0001		
Total	209	86.81448			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Alpha	14.99691	0.002561	5855.644	0	14.99187	15.00196
Beta1	-0.14999	0.000166	-904.96	0	-0.15032	-0.14967
Beta2	0.050328	0.000271	185.702	2.7E-231	0.049793	0.050862
Beta2a	0.018589	0.000101	184.5962	9.2E-231	0.01839	0.018787

Just like simulation 3, by including an extra variable which indicated the year when inflation rate changed in this regression analysis, the adjusted R^2 became very close to 1, and the coefficient α, β_1, β_2, and β_{2b} were also very close to their respective inputted values. F-test would also be used to compare simulation 4 and 5. Since simulation 4's regression equation was $E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \varepsilon$ and simulation 5's regression equation was $E(Y) = \alpha + \beta_1(DY) + \beta_2(CY) + \beta_{2a}\dfrac{(CY-k)(CY-k+1)}{2} + \varepsilon$, simulation 4's regression equation would be restricted and simulation 5's regression equation would be unrestricted. So F-test would be:

$$F_{q,N-k} = \frac{(R_{UR}^2 - R_R^2)/q}{(1-R_{UR}^2)/(N-k)} \rightarrow F_{1,210-3} = \frac{(0.999762 - 0.959793)/1}{(1-0.999762)/(210-3)} \approx 34762.95$$

The F-test statistic was greater than the critical values for $F_{1,210-3}$ between 3.92 and 3.84 for 5% significance, and between 6.85 and 6.63 for 1% significance. Either way, this showed that variable k introduced in simulation 5 was statistically significant.

Results of 2nd Simulation Run:

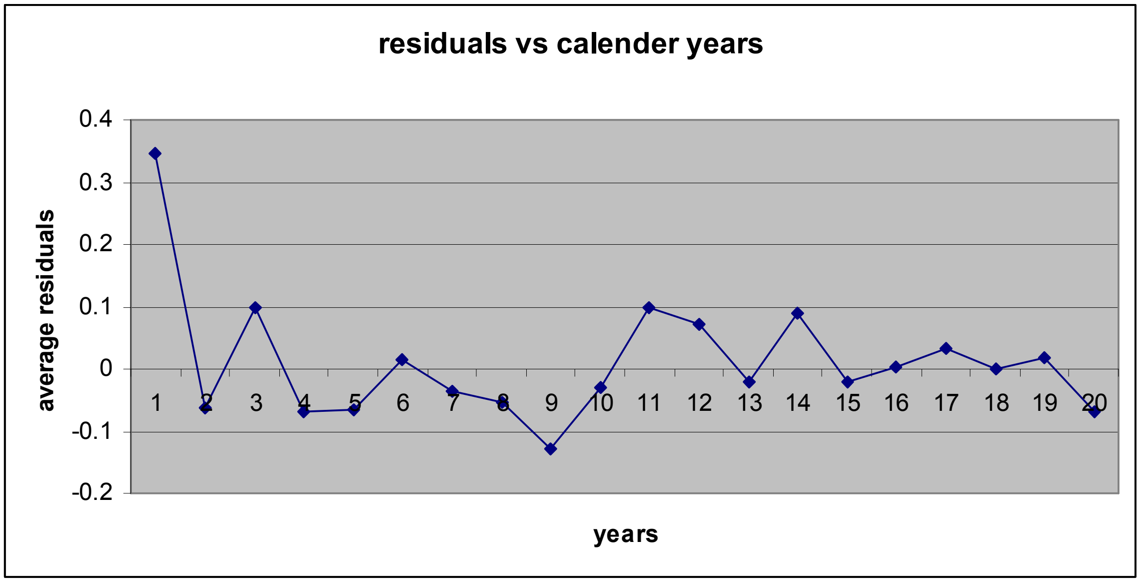
After using low stochasticity to find out the accuracies and patterns of the five simulations, a larger value of sigma could be used to see if the same patterns would be masked by stochasticity in each of the simulations. This run was that all simulations used a higher sigma of 0.25. The other parameters were also the same as the ones in previous each simulation run.

Parameters for Run #2:

Sigma	0.25
Alpha	15
beta1	-0.15
beta2	0.05
beta1b	-0.15
beta2b	0.2

CY	20
DY	20
x^{th} year inflation changed?	12
beta2a	0.01875

Simulation 1:



CY residuals:

1	2	3	4	5
0.347449	-0.06357	0.098299	-0.06772	-0.06485
6	7	8	9	10
0.013869	-0.03644	-0.05482	-0.12779	-0.02934
11	12	13	14	15
0.099659	0.071725	-0.02224	0.090337	-0.02116
16	17	18	19	20
0.001793	0.03405	-0.00128	0.01757	-0.06999

Simulation 1 statistics:

Regression Statistics	
Multiple R	0.941807
R Square	0.887001
Adjusted R Square	0.885909
Standard Error	0.240661
Observations	210

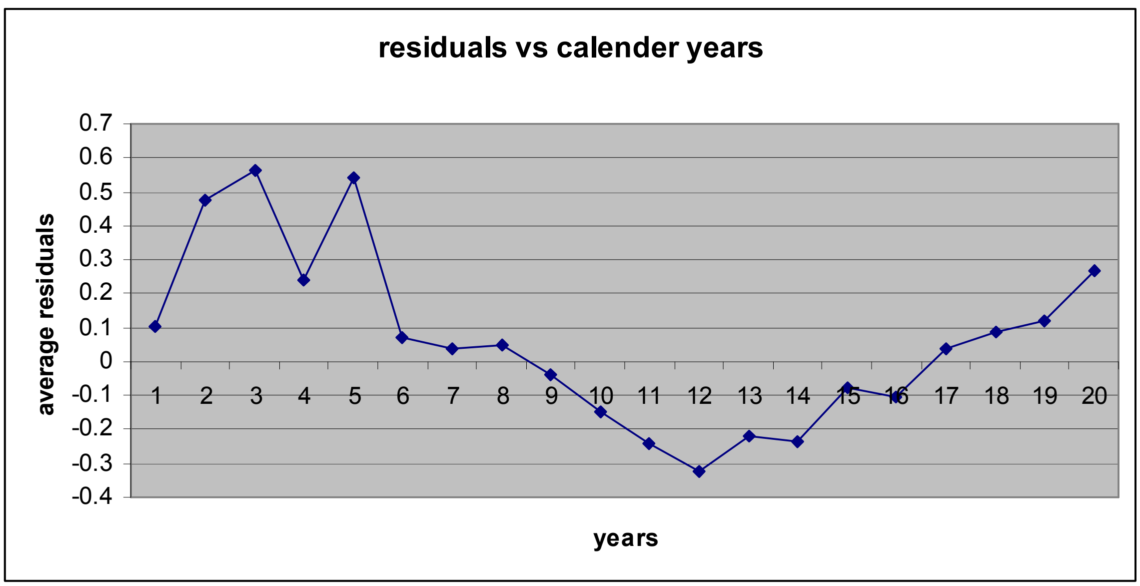
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	94.10888	47.05444	812.4351	9.8E-99
Residual	207	11.98898	0.057918		
Total	209	106.0979			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Alpha	14.90675	0.046705	319.171	1.1E-280	14.81467	14.99883
Beta1	-0.15882	0.003979	-39.91	3.6E-99	-0.16666	-0.15097
Beta2	0.059894	0.003979	15.05106	4.5E-35	0.052048	0.067739

Because of the larger error values due to higher sigma, the adjusted R^2 is now approximately 0.89 and the standard error is about 0.24. Similarly, the coefficient α, β_1, β_2 also deviated slightly from their respective inputted values. The average residuals plotted against calendar years also showed a noisier horizontal line, proving that it was affected by higher value of sigma.

Simulation 2:



CY residuals:

1	2	3	4	5
0.104601	0.473054	0.562853	0.242271	0.540844
6	7	8	9	10
0.06916	0.040542	0.051026	-0.03673	-0.15068
11	12	13	14	15
-0.24336	-0.32143	-0.21684	-0.23547	-0.0746
16	17	18	19	20
-0.1049	0.037017	0.086841	0.119568	0.265031

Simulation 2 statistics:

Regression Statistics	
Multiple R	0.921567
R Square	0.849287
Adjusted R Square	0.84783
Standard Error	0.293812
Observations	210

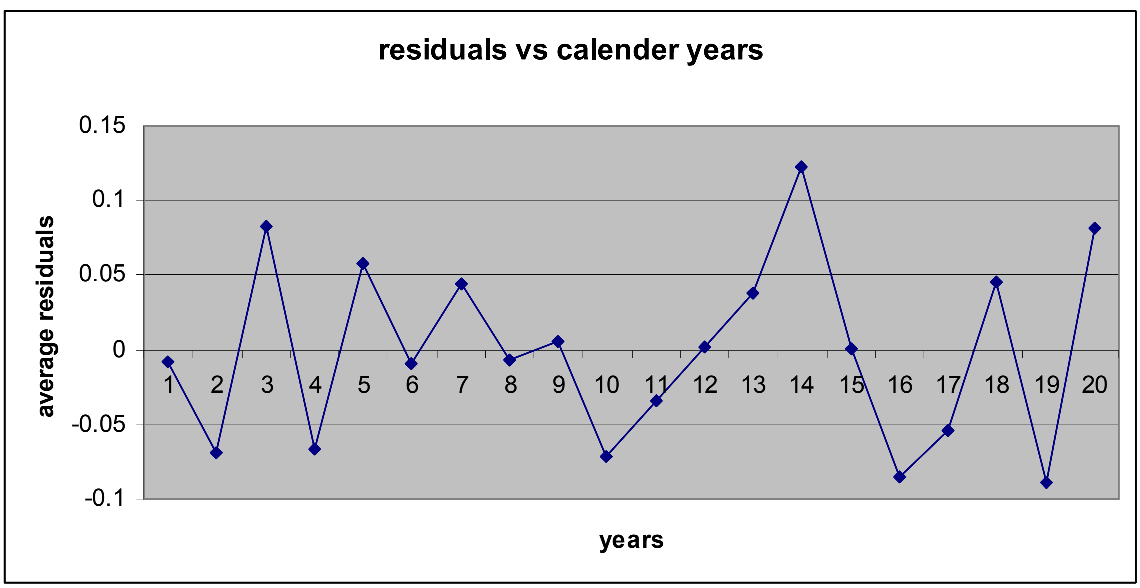
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	100.6961	50.34806	583.2337	8.68E-86
Residual	207	17.86942	0.086326		
Total	209	118.5655			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Alpha	14.44135	0.05702	253.2701	5.8E-260	14.32893	14.55376
Beta1	-0.15281	0.004858	-31.4545	8.46E-81	-0.16239	-0.14324
Beta2	0.132394	0.004858	27.25159	2.1E-70	0.122816	0.141972

Unlike previous simulation 2, the adjusted R^2 dropped to 0.85 and the standard error increased to 0.29. Not only that, the coefficients α, and β_1 also drifted slightly away from their respective inputted values. The average residuals plotted against calendar years showed a V shape,, but it was not a perfect V shape as seen in run 1 simulation 2. Thus, this also showed that the regression was affected by higher value of sigma.

Simulation 3:



CY residuals:

1	2	3	4	5
-0.00779	-0.06883	0.082706	-0.06591	0.057834
6	7	8	9	10
-0.00951	0.043908	-0.00692	0.005484	-0.07156
11	12	13	14	15
-0.03414	0.001647	0.038092	0.122258	0.000892
16	17	18	19	20
-0.08452	-0.05381	0.045447	-0.08917	0.0811

Simulation 3 statistics:

Regression Statistics	
Multiple R	0.938237
R Square	0.880288
Adjusted R Square	0.878545
Standard Error	0.254188
Observations	210

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	97.87334	32.62445	504.9332	1.21E-94
Residual	206	13.30995	0.064611		
Total	209	111.1833			

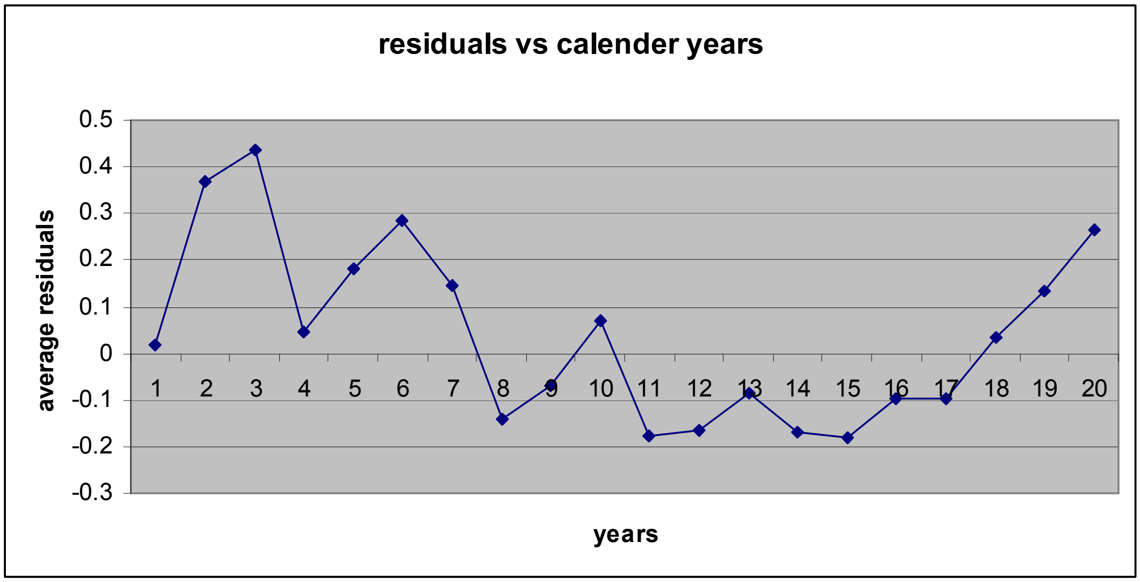
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Alpha	14.97439	0.073364	204.1099	1.1E-239	14.82975	15.11903
Beta1	-0.14879	0.004203	-35.4005	1.59E-89	-0.15708	-0.1405
Beta2	0.056737	0.008667	6.546034	4.61E-10	0.039649	0.073826
Beta2b	0.17842	0.007484	23.83952	3.8E-61	0.163664	0.193175

As noted earlier, this was the more accurate regression of discretely changed inflation rate. The adjusted R^2 was 0.88 opposed to 0.85 from previous simulation. Just like the previous run, F-test would be used to compare simulation 2 and 3, which is:

$$F_{q,N-k} = \frac{(R_{UR}^2 - R_R^2)/q}{(1-R_{UR}^2)/(N-k)} \rightarrow F_{1,210-3} = \frac{(0.880288 - 0.849287)/1}{(1-0.880288)/(210-3)} \approx 53.605$$

The F-test statistic was a lot smaller than the one calculated previously, but it was still greater than the critical values for $F_{1,210-3}$ between 3.92 and 3.84 for 5% significance, and between 6.85 and 6.63 for 1% significance. Either way, this showed that variable k introduced in simulation 3 was still statistically significant. Also, just like F-test statistic, the residuals plot against calendar years showed that the seemingly horizontal line was affected more by higher value of sigma.

Simulation 4:



CY residuals:

1	2	3	4	5
0.019313	0.367315	0.437882	0.04536	0.181621
6	7	8	9	10
0.286408	0.146636	-0.14176	-0.06763	0.070059
11	12	13	14	15
-0.17608	-0.16458	-0.08689	-0.17054	-0.18172
16	17	18	19	20
-0.09856	-0.09633	0.033012	0.135462	0.267066

Simulation 4 statistics:

Regression Statistics	
Multiple R	0.912764
R Square	0.833138
Adjusted R Square	0.831526
Standard Error	0.284537
Observations	210

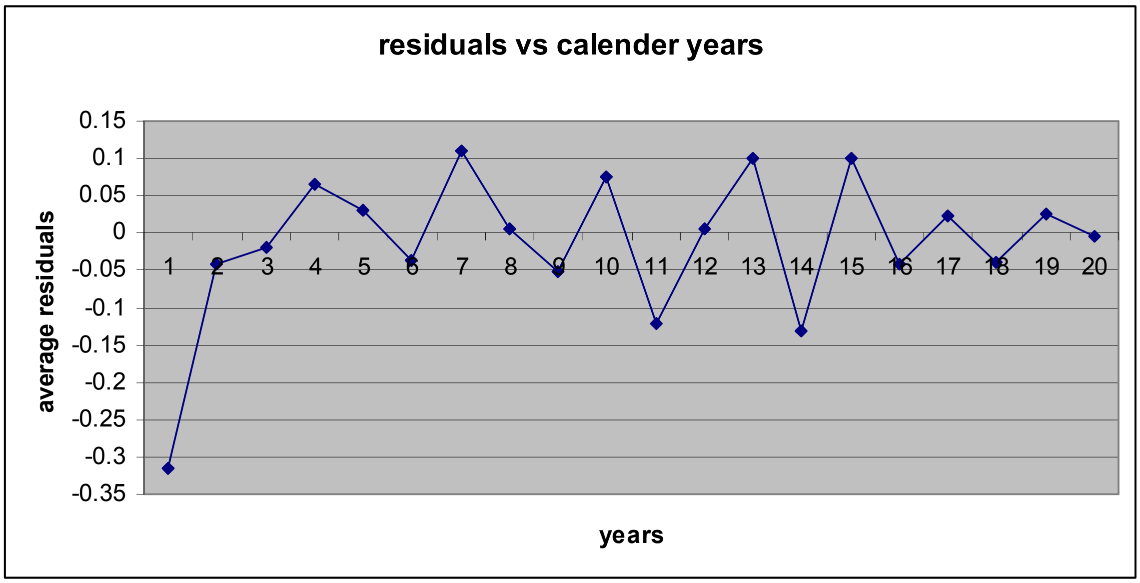
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	83.677	41.8385	516.773	3.27E-81
Residual	207	16.75894	0.080961		
Total	209	100.4359			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Alpha	14.64001	0.055219	265.1243	4.6E-264	14.53114	14.74887
Beta1	-0.14986	0.004705	-31.8518	9.8E-82	-0.15913	-0.14058
Beta2	0.092694	0.004705	19.70172	2.31E-49	0.083418	0.101969

The adjusted R^2 was 0.83 and the standard error was 0.28. The average residuals plotted against calendar years showed a rounded, but noisier V shape, unlike the one seen in run 1 simulation 4. Thus, this showed that the regression was affected by the higher value of sigma.

Simulation 5:



CY residuals:

1	2	3	4	5
-0.3164	-0.04243	-0.01891	0.06601	0.031802
6	7	8	9	10
-0.03533	0.110408	0.006602	-0.05088	0.075308
11	12	13	14	15
-0.12163	0.005486	0.100112	-0.1313	0.100405
16	17	18	19	20
-0.04229	0.024084	-0.03941	0.025639	-0.00408

Simulation 5 statistics:

Regression Statistics	
Multiple R	0.929057
R Square	0.863146
Adjusted R Square	0.861153
Standard Error	0.243702
Observations	210

ANOVA

	df	SS	MS	F	Significance F
Regression	3	77.16388	25.72129	433.0849	1.16E-88
Residual	206	12.23452	0.059391		
Total	209	89.3984			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Alpha	14.92644	0.062267	239.7182	5E-254	14.80367	15.0492
Beta1	-0.14061	0.00403	-34.8937	2.03E-88	-0.14855	-0.13266
Beta2	0.052601	0.006589	7.983192	9.85E-14	0.039611	0.065592
Beta2a	0.017435	0.002448	7.121427	1.75E-11	0.012608	0.022262

Also noted earlier, this was the more accurate regression of continuously changed inflation rate. The adjusted R^2 was 0.86 opposed to 0.83 from previous simulation. F-test would also be used to compare simulation 4 and 5, which would be:

$$F_{q,N-k} = \frac{(R_{UR}^2 - R_R^2)/q}{(1-R_{UR}^2)/(N-k)} \rightarrow F_{1,210-3} = \frac{(0.863146-0.833138)/1}{(1-0.863146)/(210-3)} \approx 45.389$$

The F-test statistic was a lot smaller than the one calculated previously, but it was still greater than the critical values for $F_{1,210-3}$ between 3.92 and 3.84 for 5% significance, and between 6.85 and 6.63 for 1% significance. Either way, this showed that variable *k* introduced in simulation 5 was still statistically significant. Also, just like F-test statistic, the residuals plot against calendar years showed that the seemingly horizontal line was affected more by the higher value of sigma.

Results of 3rd Simulation Run:

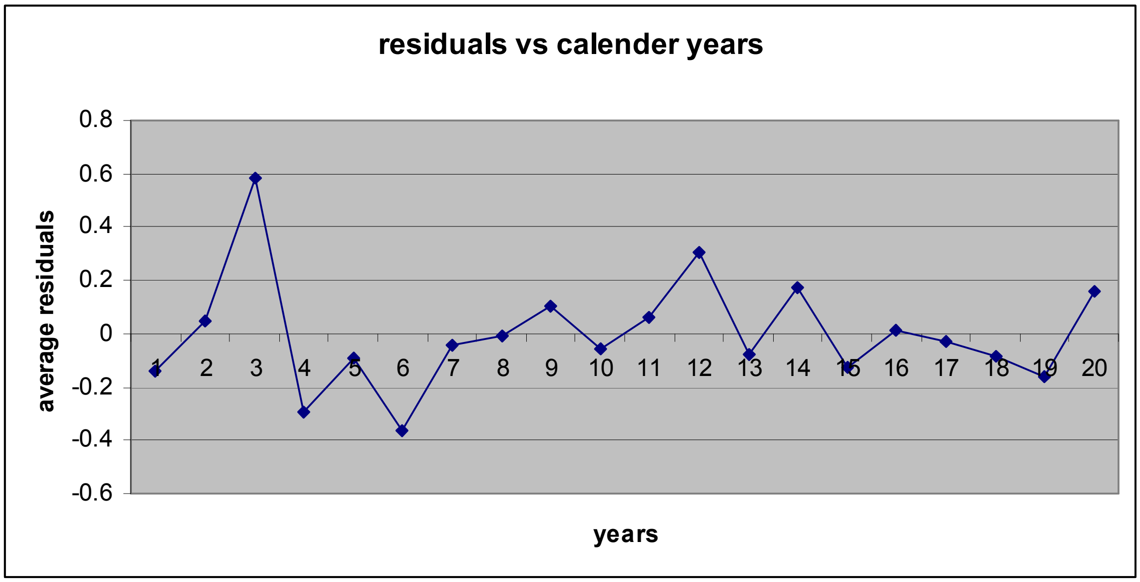
Previously, a moderate value of sigma was used to find out that the simulated random errors were not large enough to mask the significance of new variables introduced in simulation 3 and 5. Now, a larger value of sigma would be used to see if the same patterns would show in each of the simulations. This run was that all simulations used an even higher sigma of 0.6. The other parameters were also the same as the ones in previous each simulation run.

Parameters for Run #3:

Sigma	0.6
Alpha	15
beta1	-0.15
beta2	0.05
beta1b	-0.15
beta2b	0.2

CY	20
DY	20
x^{th} year inflation changed?	12
beta2a	0.01875

Simulation 1:



CV Residuals:

1	2	3	4	5
-0.13792	0.044396	0.58299	-0.29137	-0.0882
6	7	8	9	10
-0.36587	-0.03937	-0.00637	0.104073	-0.05361
11	12	13	14	15
0.061649	0.306204	-0.07574	0.170693	-0.12479
16	17	18	19	20
0.010614	-0.03149	-0.08407	-0.16291	0.155805

Simulation 1 statistics:

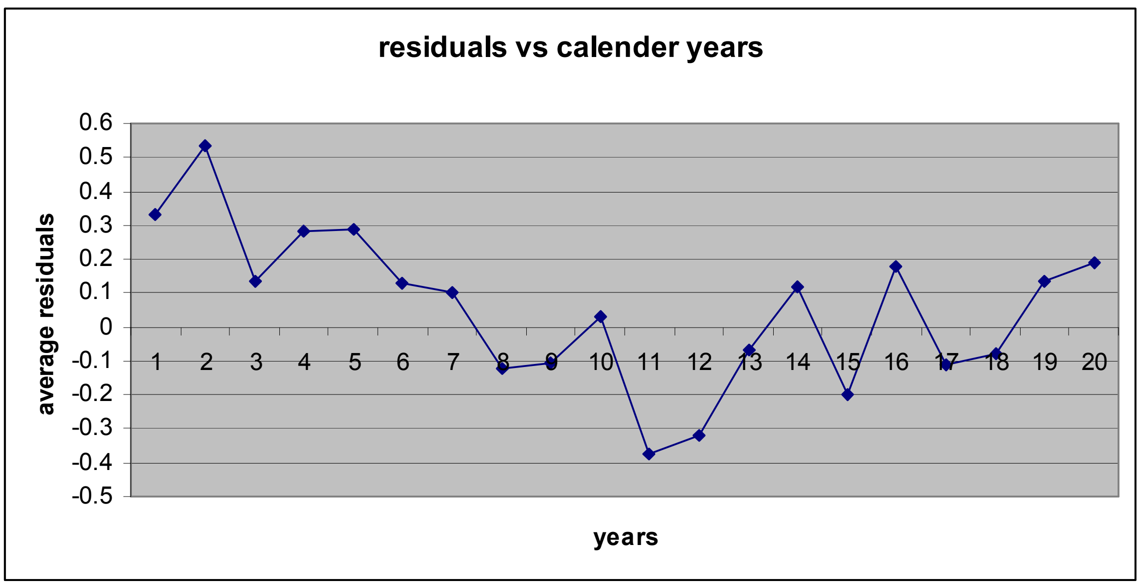
Regression Statistics	
Multiple R	0.767396
R Square	0.588897
Adjusted R Square	0.584925
Standard Error	0.549473
Observations	210

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	89.5265	44.76325	148.2616	1.11E-40
Residual	207	62.49759	0.301921		
Total	209	152.0241			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Alpha	15.01453	0.106635	140.8029	1.7E-207	14.8043	15.22476
Beta1	-0.15273	0.009086	-16.8102	1.47E-40	-0.17064	-0.13482
Beta2	0.046987	0.009086	5.171533	5.46E-07	0.029074	0.064899

Since an even higher sigma was used, the adjusted R^2 is now approximately 0.59 and the standard error is about 0.55. The average residuals plotted against calendar years also showed a higher absolute residual averages, proving that it was affected by the higher value of sigma.

Simulation 2:



CV Residuals:

1	2	3	4	5
0.330059	0.5336	0.132923	0.281276	0.289645
6	7	8	9	10
0.128156	0.104203	-0.12024	-0.10606	0.030295
11	12	13	14	15
-0.37248	-0.32101	-0.06566	0.117487	-0.19634
16	17	18	19	20
0.178872	-0.11374	-0.08068	0.134882	0.190512

Simulation 2 statistics:

Regression Statistics	
Multiple R	0.783955
R Square	0.614586
Adjusted R Square	0.610862
Standard Error	0.552527
Observations	210

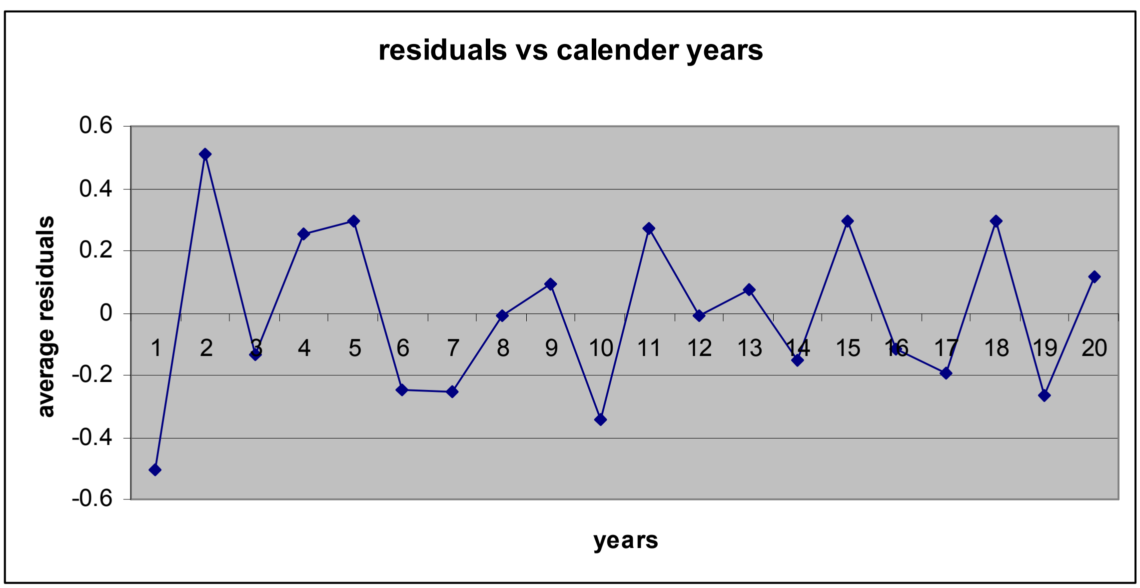
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	100.7704	50.38521	165.0424	1.39E-43
Residual	207	63.19431	0.305287		
Total	209	163.9647			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Alpha	14.42334	0.107228	134.5112	1.9E-203	14.21194	14.63474
Beta1	-0.15283	0.009136	-16.7281	2.64E-40	-0.17084	-0.13482
Beta2	0.132504	0.009136	14.50333	2.35E-33	0.114492	0.150516

Just as the sigma increased, the adjusted R^2 dropped even lower to 0.61 and the standard error increased to 0.55. Furthermore, it was harder to tell whether the average residuals plotted against calendar years showed a V shape, or a noisy horizontal trend as in simulation 1. Therefore, this also showed that the regression was affected by the higher value of sigma.

Simulation 3:



CV Residuals:

1	2	3	4	5
-0.5057	0.50791	-0.13461	0.253055	0.294982
6	7	8	9	10
-0.24481	-0.25221	-0.01013	0.08991	-0.34051
11	12	13	14	15
0.269486	-0.00996	0.072392	-0.15393	0.296632
16	17	18	19	20
-0.11868	-0.19671	0.294367	-0.2679	0.113597

Simulation 3 statistics:

Regression Statistics	
Multiple R	0.744036
R Square	0.553589
Adjusted R Square	0.547088
Standard Error	0.596402
Observations	210

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	90.86542	30.28847	85.15279	7.18E-36
Residual	206	73.2733	0.355696		
Total	209	164.1387			

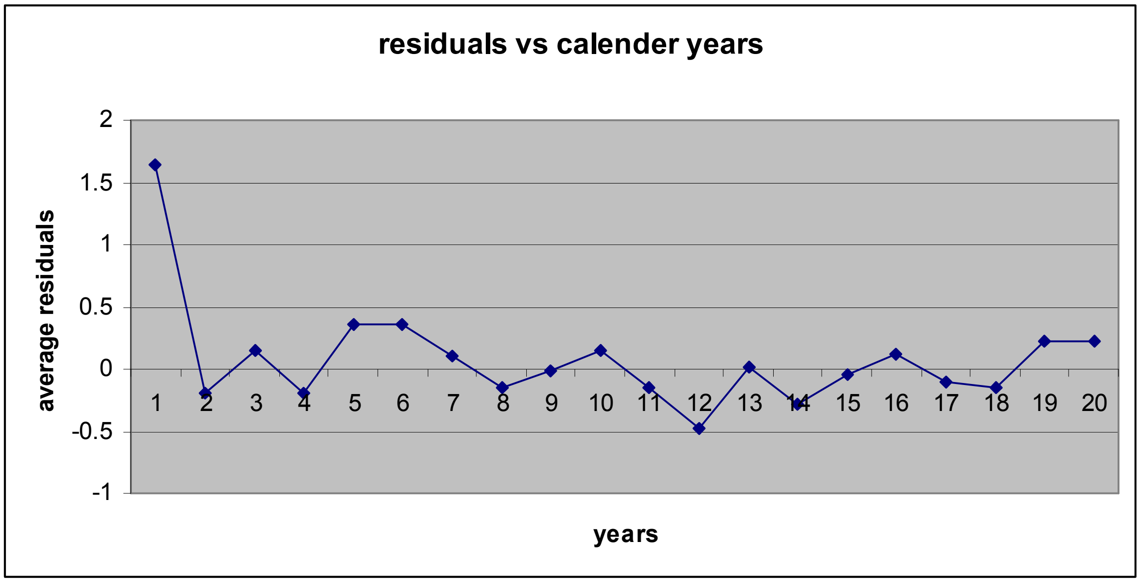
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Alpha	14.93305	0.172135	86.75188	3.8E-164	14.59368	15.27243
Beta1	-0.14734	0.009862	-14.9412	1.11E-34	-0.16679	-0.1279
Beta2	0.063215	0.020336	3.108469	0.002146	0.023121	0.10331
Beta2b	0.157707	0.01756	8.980932	1.72E-16	0.123086	0.192328

The adjusted R^2 was 0.55 opposed to 0.61 from simulation 2 and the standard error was 0.60. Just like the previous runs, F-test would be used to compare simulation 2 and 3, which is:

$$F_{q,N-k} = \frac{(R_{UR}^2 - R_R^2)/q}{(1-R_{UR}^2)/(N-k)} \rightarrow F_{1,210-3} = \frac{(0.553589 - 0.614586)/1}{(1-0.553589)/(210-3)} \approx -28.2842$$

The F-test statistic was less than the critical values for $F_{1,210-3}$ between 3.92 and 3.84 for 5% significance, and between 6.85 and 6.63 for 1% significance. The introduced variable *k* in simulation 3 was no longer statistically significant. Also, it was much more difficult to tell if the residuals plot against calendar years showed even a seemingly horizontal line. Both these pointed out that stochasticity was large enough to mask the significance of the newly introduced variable.

Simulation 4:



CV Residuals:

1	2	3	4	5
1.636707	-0.19047	0.150081	-0.19462	0.359813
6	7	8	9	10
0.357098	0.098883	-0.14522	-0.01069	0.154124
11	12	13	14	15
-0.15645	-0.47019	0.013869	-0.28754	-0.04835
16	17	18	19	20
0.124117	-0.10039	-0.15633	0.216635	0.225385

Simulation 4 statistics:

Regression Statistics	
Multiple R	0.70827
R Square	0.501646
Adjusted R Square	0.496831
Standard Error	0.616375
Observations	210

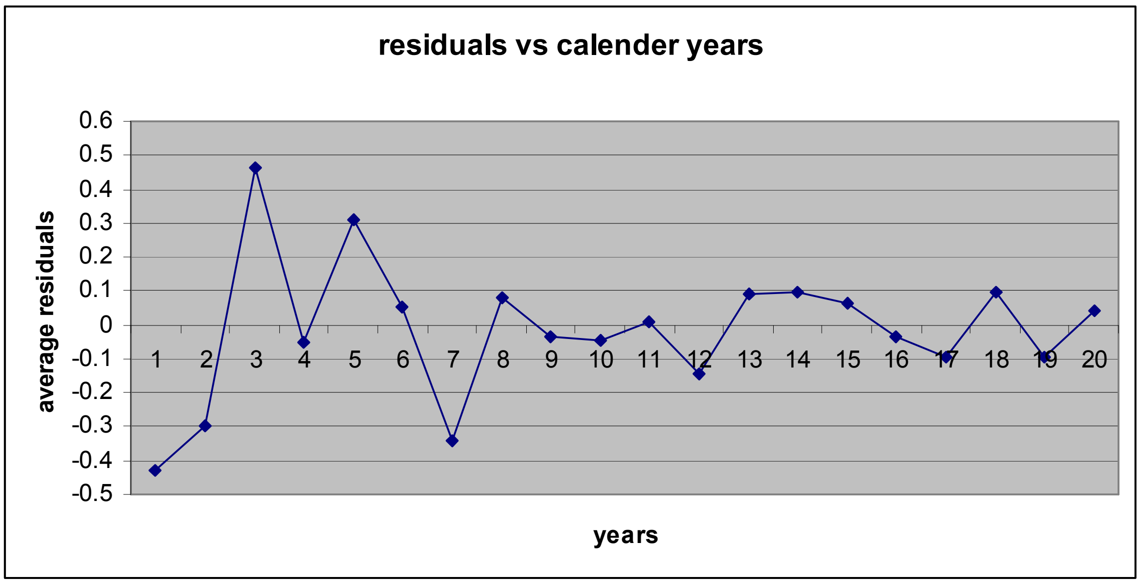
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	79.16252	39.58126	104.1838	4.96E-32
Residual	207	78.64296	0.379918		
Total	209	157.8055			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Alpha	14.68205	0.119619	122.7407	2.6E-195	14.44623	14.91788
Beta1	-0.1468	0.010192	-14.4034	4.85E-33	-0.16689	-0.1267
Beta2	0.081814	0.010192	8.027417	7.36E-14	0.061721	0.101907

The adjusted R^2 was 0.50 and the standard error was 0.62. The difference here was that the average residuals plotted against calendar years did not show a V shape. This showed that the V shape of the regression had less impact than the stochasticity introduced in this simulation.

Simulation 5:



CV Residuals:

1	2	3	4	5
-0.43061	-0.29875	0.460812	-0.05165	0.308879
6	7	8	9	10
0.054682	-0.34103	0.077769	-0.03484	-0.0483
11	12	13	14	15
0.009954	-0.14466	0.088647	0.096968	0.06386
16	17	18	19	20
-0.03404	-0.09481	0.097962	-0.09387	0.043833

Simulation 5 statistics:

Regression Statistics	
Multiple R	0.716379
R Square	0.513199
Adjusted R Square	0.50611
Standard Error	0.602512
Observations	210

ANOVA

	df	SS	MS	F	Significance F
Regression	3	78.83771	26.27924	72.39043	5.18E-32
Residual	206	74.7823	0.363021		
Total	209	153.62			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Alpha	14.87595	0.153943	96.63259	1.5E-173	14.57244	15.17946
Beta1	-0.14434	0.009963	-14.4879	2.91E-33	-0.16398	-0.1247
Beta2	0.065125	0.01629	3.997854	8.9E-05	0.033009	0.097242
Beta2a	0.011235	0.006053	1.85613	0.064863	-0.0007	0.023168

The adjusted R^2 was 0.51 opposed to 0.50 from simulation 4 and the standard error was 0.60. Just like the previous runs, F-test would be used to compare simulation 4 and 5, which is:

$$F_{q,N-k} = \frac{(R_{UR}^2 - R_R^2)/q}{(1-R_{UR}^2)/(N-k)} \rightarrow F_{1,210-3} = \frac{(0.513199-0.501646)/1}{(1-0.513199)/(210-3)} \approx 4.913$$

The F-test statistic was greater than the critical values for $F_{1,210-3}$ between 3.92 and 3.84 for 5% significance, but less than the range between 6.85 and 6.63 for 1% significance. The introduced variable *k* in simulation 5 was statistically significant enough at 95% level of confidence. Also, it was much more difficult to tell if the residuals plot against calendar years showed even a seemingly horizontal line. These also pointed out that stochasticity was large enough to mask the significance of the newly introduced variable.

Conclusion:

These simulations have shown that when a change in the inflation rate was introduced in the paid loss triangle simulation, two factors would affect the regression analysis: the inclusion of the extra variable being the year when the change occurred, and the severity of stochasticity affecting the data points. As long as the effect of stochasticity was small enough, the regression equation including the extra variable would be the better fit.

Appendix:

Discrete change

year 0: $\alpha + B_1(DY)$ * cumulative rate

year 1: $\alpha + B_1(DY) + B_2$

year 2: $\alpha + B_1(DY) + 2B_2$

$\vdots$

year k: $\alpha + B_1(DY) + kB_2$ * inflation rate change to B_{2b} starting y

year k+1: $\alpha + B_1(DY) + kB_2 + 1B_{2b}$

year k+2: $\alpha + B_1(DY) + kB_2 + 2B_{2b}$

$\vdots$

year k+l: $\alpha + B_1(DY) + kB_2 + lB_{2b}$

convert to eqn with only dummy var, DY, CY.

For CY from 0 to k, D = 0

$\alpha + B_1(DY) + B_2(CY) + \sout{0(\alpha_2 + B_{2b}\,CY)}$ same as regression w/o coef inf.

For CY from k+1 to k+l, D = 1

$\alpha + B_1(DY) + B_2(CY) + 1(\alpha_2 + B_{2b}(CY))$

$\sout{\alpha + B_1(DY)} + B_2(CY) + B_{2b}(CY) + \alpha_2$

$= \sout{\alpha + B_1(DY)} + kB_2 + lB_{2b}$ where $k + l = CY$ $\nearrow CY - k = l$, $\searrow CY - l = k$

$\Rightarrow B_2(CY) - B_2(l) + B_{2b}(CY) - B_{2b}(k) = B_2(CY) + B_{2b}(CY) + \alpha_2$

$\alpha_2 = -B_2(l) - B_{2b}(k)$

must introduce new variable k where year k+1 has inflation rate cha

if D=1: $\alpha + B_1(DY) + B_2(k) + B_{2b}(CY - k)$

Continuous change:

let B_{2s} = inflation rate where β_2 changed towards

so: CY

yr 0: $\alpha + B_1(DY)$ * culmulative rate

yr 1: $\alpha + B_1(DY) + B_2$

yr 2: $\alpha + B_1(DY) + 2B_2$

⋮

yr k: $\alpha + B_1(DY) + kB_2$ * inflation rate change towards B_{2b} starti[ng] year k+1

let $r = \dfrac{B_{2s} - B_2}{CY_{max} - k}$

yr k+1: $\alpha + B_1(DY) + kB_2 + (B_2 + r)$

yr k+2: $\alpha + B_1(DY) + kB_2 + (B_2 + r) + (B_2 + 2r)$

yr k+3: $\alpha + B_1(DY) + kB_2 + (B_2 + r) + (B_2 + 2r) + (B_2 + 3r)$

⋮

yr k+l: $\alpha + B_1(DY) + kB_2 + lB_2 + r\sum_{i=1}^{l} i$

$= \alpha + B_1(DY) + B_2(CY) + r\left(\dfrac{l(l+1)}{2}\right)$