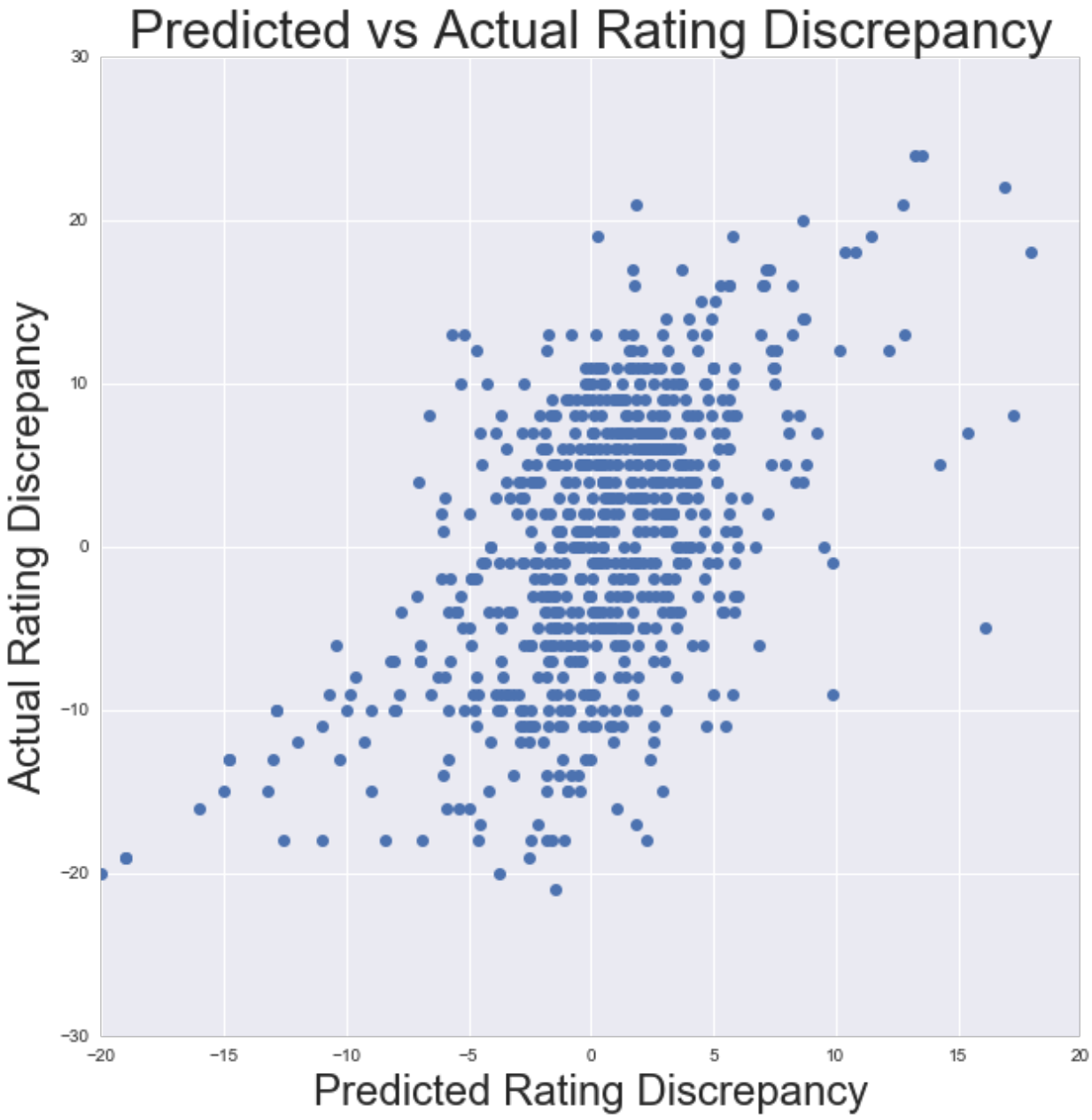
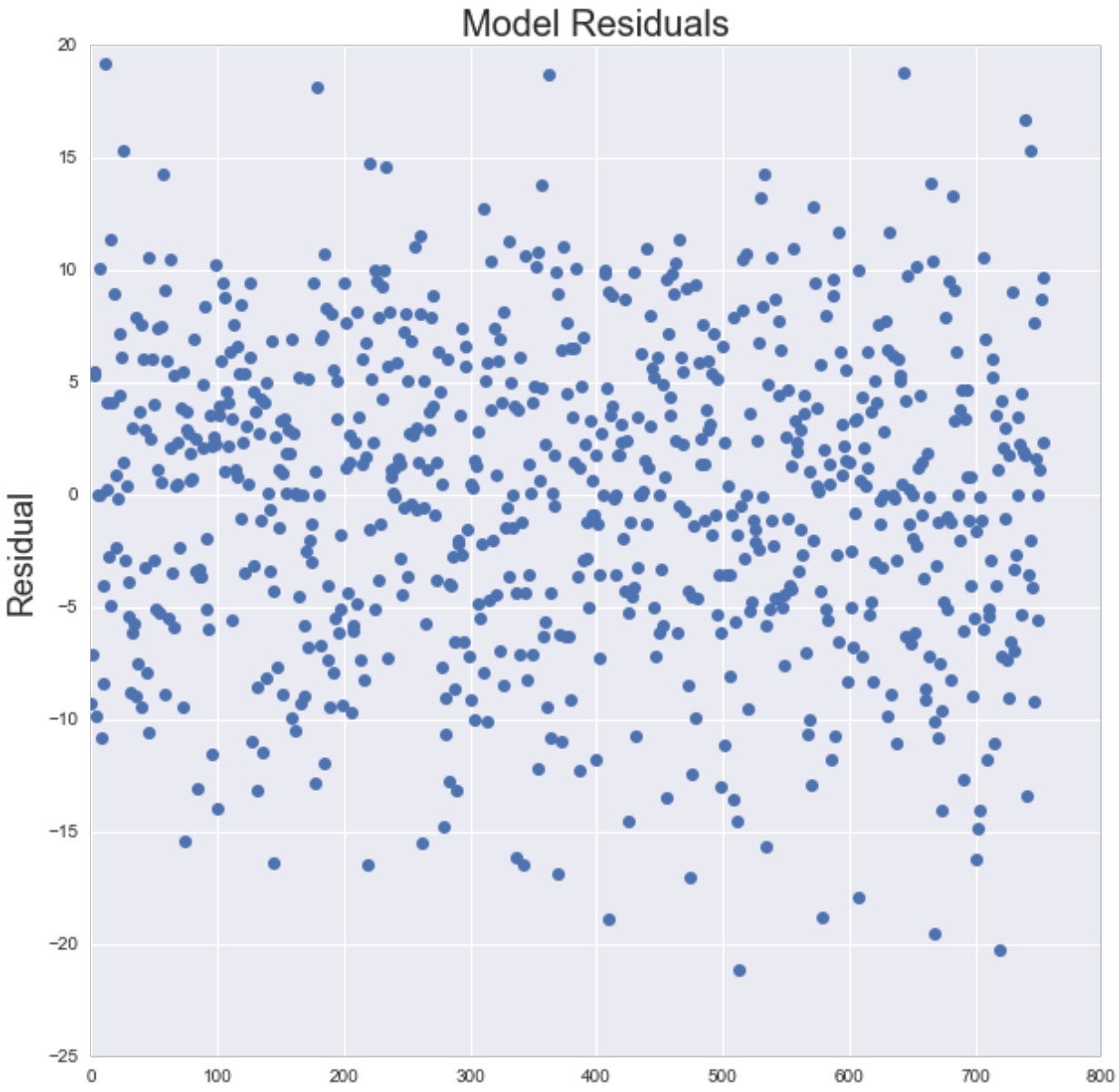
For this project I am attempting to predict the discrepancy between critic opinion and the general public's opinion for movies. I am using Metacritic scores as a proxy for critic opinion, and IMDB user scores as a proxy for general public opinion as the majority of IMDB users are not professional movie critics. The features of my model include gross revenue, budget, run time, the amount of Oscars the film won, the amount of IMDB reviews that the film has on the IMDB website, and fixed effects for MPAA rating, genre, and year it was released. After looking at the pairwise plot for all of the variables, I decided to add a log term for gross revenue and a quadratic term for Oscar wins. The following tables and graphs show my model performance:

Dep. Variable:	crit_discrep	**R-squared:**	0.294
Model:	OLS	**Adj. R-squared:**	0.225
Method:	Least Squares	**F-statistic:**	4.211
Date:	Tue, 04 Oct 2016	**Prob (F-statistic):**	1.11e-22
Time:	17:31:03	**Log-Likelihood:**	-2530.5
No. Observations:	755	**AIC:**	5199.
Df Residuals:	686	**BIC:**	5518.
Df Model:	68		
Covariance Type:	nonrobust		

log_gross	-0.5318	0.223	-2.380	0.018	-0.971 -0.093
square_oscars	0.1845	0.081	2.292	0.022	0.026 0.343
Run_Time	0.0411	0.016	2.563	0.011	0.010 0.073
Oscar_Wins	-3.5281	0.526	-6.703	0.000	-4.562 -2.495
IMDB_Reviews	1.361e-05	1.73e-06	7.849	0.000	1.02e-05 1.7e-05
Gross_Revenue	-8.332e-10	3.69e-09	-0.226	0.821	-8.07e-09 6.41e-09

Omnibus:	7.574	**Durbin-Watson:**	2.020
Prob(Omnibus):	0.023	**Jarque-Bera (JB):**	7.734
Skew:	-0.241	**Prob(JB):**	0.0209
Kurtosis:	2.886	**Cond. No.**	2.22e+19


Predicted vs Actual Rating Discrepancy
Actual Rating Discrepancy
Predicted Rating Discrepancy
30
20
10
0
−10
−20
−30
−20
−15
−10
−5
0
5
10
15
20



Clearly my model will have to be refined further, as the residuals do not hover around zero and predicted vs actual rating discrepancy line is loosely fit.