对final_vector中的V——M加工变量和Q变量做ARIMA时间序列回归

首先是所有场次的回归，如下：

Q——point：

Model Description

			Model Type
Model ID	Q	Model_1	ARIMA(2,1,4)

Model Summary

Model Fit

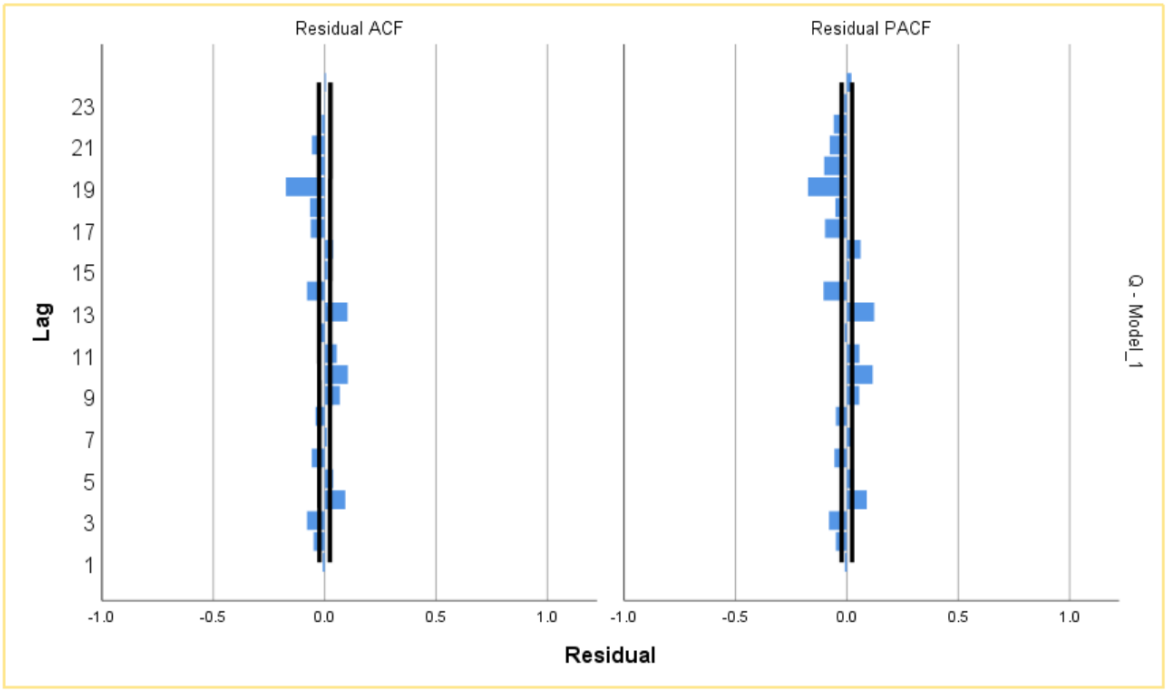
Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.200	.	.200	.200	.200	.200	.200	.200	.200	.200	.200
R-squared	.986	.	.986	.986	.986	.986	.986	.986	.986	.986	.986
RMSE	.011	.	.011	.011	.011	.011	.011	.011	.011	.011	.011
MAPE	.883	.	.883	.883	.883	.883	.883	.883	.883	.883	.883
MaxAPE	185.353	.	185.353	185.353	185.353	185.353	185.353	185.353	185.353	185.353	185.353
MAE	.005	.	.005	.005	.005	.005	.005	.005	.005	.005	.005
MaxAE	.423	.	.423	.423	.423	.423	.423	.423	.423	.423	.423
Normalized BIC	-8.932	.	-8.932	-8.932	-8.932	-8.932	-8.932	-8.932	-8.932	-8.932	-8.932

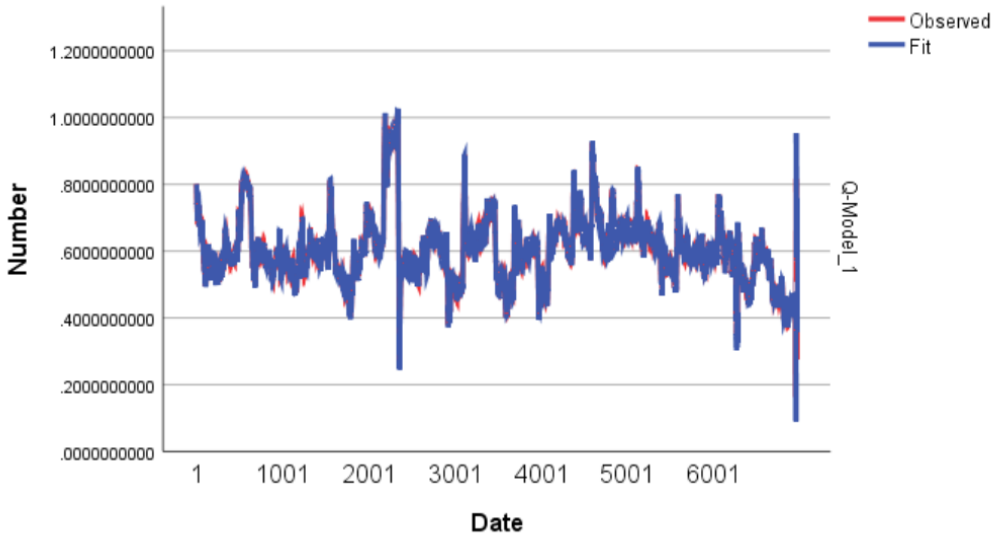
Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
Q-Model_1	0	.200	487.830	12	.000	0

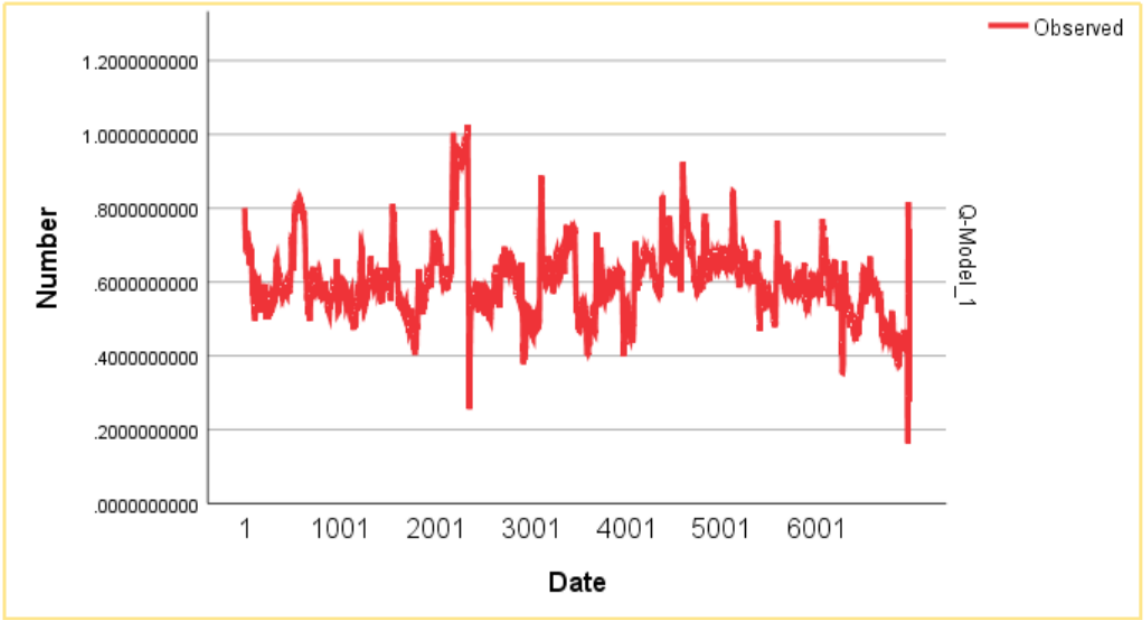
ARIMA Model Parameters

					Estimate	SE	t	Sig.
Q-Model_1	Q	No Transformation	AR	Lag 1	-.406	.007	-58.406	.000
				Lag 2	-.938	.007	-136.167	.000
			Difference		1			
			MA	Lag 1	-.656	.013	-49.274	.000
				Lag 2	-1.126	.015	-75.140	.000
				Lag 3	-.494	.015	-33.644	.000
				Lag 4	-.045	.014	-3.304	.001


Residual ACF
Residual PACF
Lag
23
21
19
17
15
13
11
9
7
5
3
1
-1.0
-0.5
0.0
0.5
1.0
Residual
Q - Model_1


Observed
Fit
Number
1.2000000000
1.0000000000
.8000000000
.6000000000
.4000000000
.2000000000
.0000000000
1
1001
2001
3001
4001
5001
6001
Date
Q-Model_1

（原值如下：）


Observed
Number
1.2000000000
1.0000000000
.8000000000
.6000000000
.4000000000
.2000000000
.0000000000
1
1001
2001
3001
4001
5001
6001
Date
Q-Model_1

最终选择的是ARIMA（2,1,4）模型

可以看到acf和pacf的情况非常之差，也很难想象划分测试集和训练集后的测试情况将多么糟糕

然后是VM的模型：

Model Description

			Model Type
Model ID	VM	Model_1	ARIMA(3,0,15)

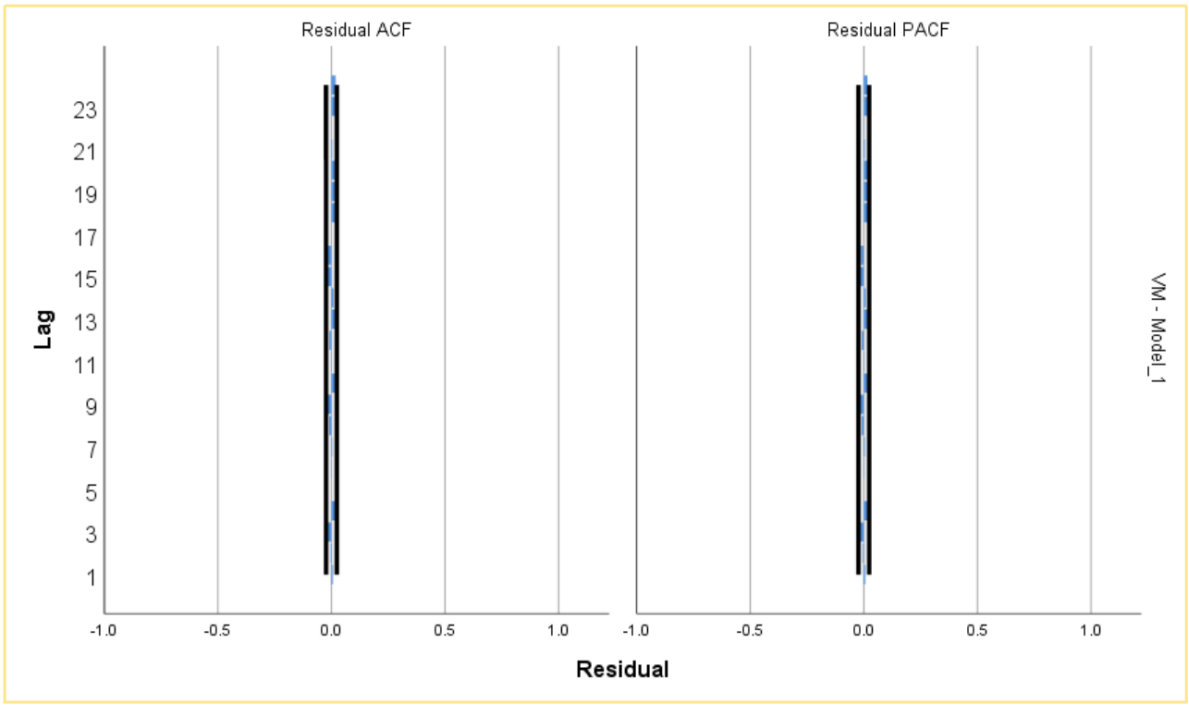
Model Summary

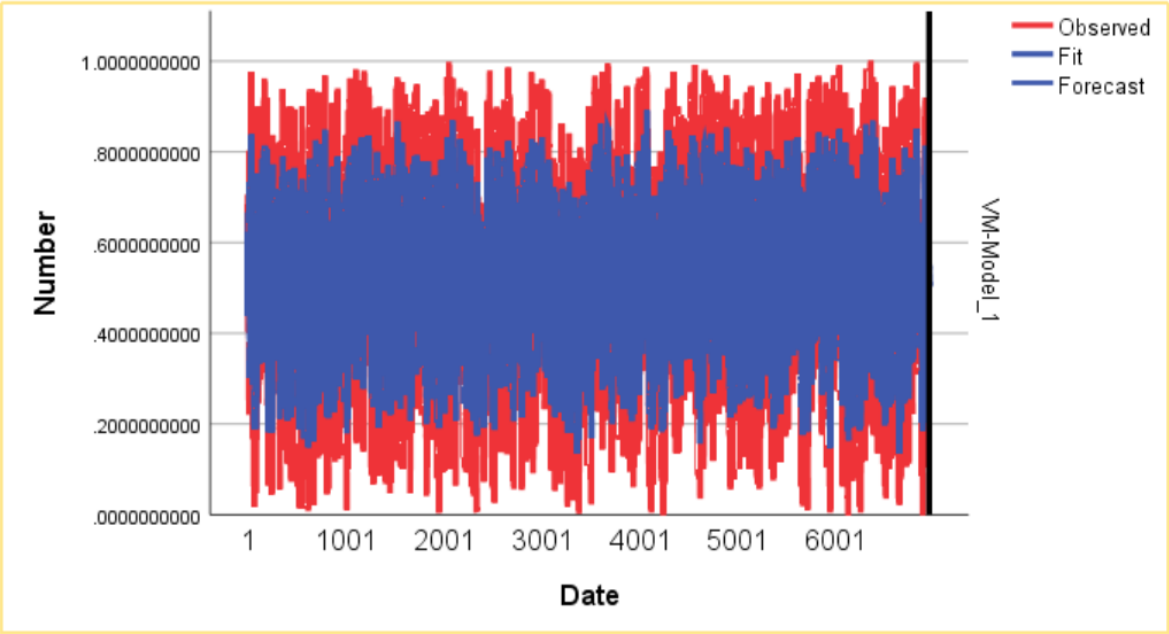
Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.487	.	.487	.487	.487	.487	.487	.487	.487	.487	.487
R-squared	.487	.	.487	.487	.487	.487	.487	.487	.487	.487	.487
RMSE	.148	.	.148	.148	.148	.148	.148	.148	.148	.148	.148
MAPE	53.418	.	53.418	53.418	53.418	53.418	53.418	53.418	53.418	53.418	53.418
MaxAPE	6651.744	.	6651.744	6651.744	6651.744	6651.744	6651.744	6651.744	6651.744	6651.744	6651.744
MAE	.132	.	.132	.132	.132	.132	.132	.132	.132	.132	.132
MaxAE	.406	.	.406	.406	.406	.406	.406	.406	.406	.406	.406
Normalized BIC	-3.813	.	-3.813	-3.813	-3.813	-3.813	-3.813	-3.813	-3.813	-3.813	-3.813

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
VM-Model_1	0	.487	25.506	12	.013	0





可以发现VM的acf和pacf出奇地好，尽管拟合值相对要比真实值紧凑很多（要不然呢，VM本来就是通过近几场的胜负情况推导出来的）

接下来考虑对于单场比赛的分析

仍然选择温网2023决赛的VM与Q，首先分析Q，二者都选取前279个作为训练组，后面的作为测试组

Model Description

			Model Type
Model ID	testQ	Model_1	ARIMA(1,1,5)

Model Summary

Model Fit

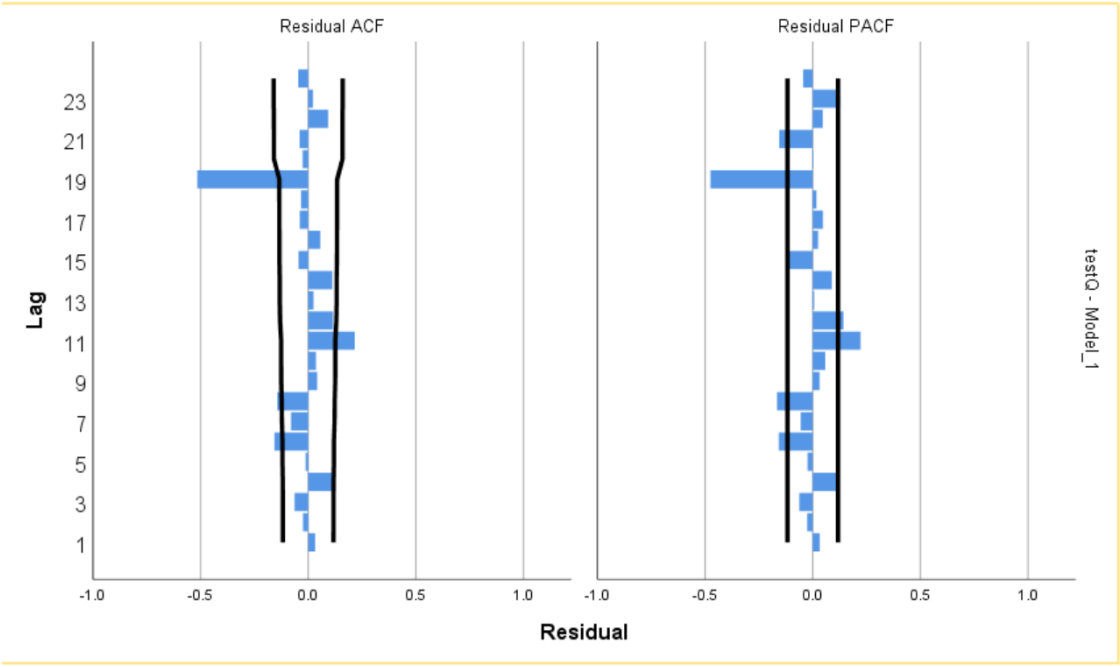
Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.478	.	.478	.478	.478	.478	.478	.478	.478	.478	.478
R-squared	.994	.	.994	.994	.994	.994	.994	.994	.994	.994	.994
RMSE	.004	.	.004	.004	.004	.004	.004	.004	.004	.004	.004
MAPE	.657	.	.657	.657	.657	.657	.657	.657	.657	.657	.657
MaxAPE	3.156	.	3.156	3.156	3.156	3.156	3.156	3.156	3.156	3.156	3.156
MAE	.003	.	.003	.003	.003	.003	.003	.003	.003	.003	.003
MaxAE	.017	.	.017	.017	.017	.017	.017	.017	.017	.017	.017
Normalized BIC	-10.973	.	-10.973	-10.973	-10.973	-10.973	-10.973	-10.973	-10.973	-10.973	-10.973

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
testQ-Model_1	0	.478	44.593	16	.000	0

ARIMA Model Parameters

					Estimate	SE	t	Sig.
testQ-Model_1	testQ	No Transformation	AR	Lag 1	.693	.045	15.540	.000
			Difference		1			
			MA	Lag 5	.207	.061	3.389	.001



一样，acf和pacf效果很差，拟合结果也罕见地出现了立即收束的情况；不过欣慰的是（？），预测值相比于测试组并没有太大的残差，测试组的数据也在预测组的拟合值（约为0.4314）上下浮动

Q	VM	point	testQ	testVM	Predicted _testQ_Model_1
.45583289	.26224066	304	.	.	.43138422
.44827643	.54224066	305	.	.	.43138437
.43932958	.44890733	306	.	.	.43138448
.43565920	.26178221	307	.	.	.43138455
.43283965	.49511555	308	.	.	.43138460
.43176827	.37844888	309	.	.	.43138464
.42941962	.56511555	310	.	.	.43138466
.43384928	.65844888	311	.	.	.43138468
.43922392	.71178221	312	.	.	.43138469
.43769585	.61277174	313	.	.	.43138470
.43630247	.72943841	314	.	.	.43138470
.43428079	.49610507	315	.	.	.43138471
.43285385	.68277174	316	.	.	.43138471
.42305414	.54277174	317	.	.	.43138471
.41407693	.46277174	318	.	.	.43138471
.41378244	.16294192	319	.	.	.43138471
.41424549	.51294192	320	.	.	.43138471
.41857273	.74627525	321	.	.	.43138471
.42634650	.79294192	322	.	.	.43138471
.43378809	.81627525	323	.	.	.43138471
.43300979	.63953544	324	.	.	.43138471
.43461854	.40620211	325	.	.	.43138471
.42985431	.28953544	326	.	.	.43138471
.42459633	.56953544	327	.	.	.43138471
.41943910	.47620211	328	.	.	.43138471
.42279721	.27587617	329	.	.	.43138471
.42582583	.50920950	330	.	.	.43138471
.42981888	.74254284	331	.	.	.43138471
.44051911	.50920950	332	.	.	.43138471
.45081288	.62587617	333	.	.	.43138471
.46094103	.69254284	334	.	.	.43138471

接下来是对VM的预测：

Model Description

			Model Type
Model ID	testVM	Model_1	ARIMA(1,0,0)

Model Summary

Model Fit

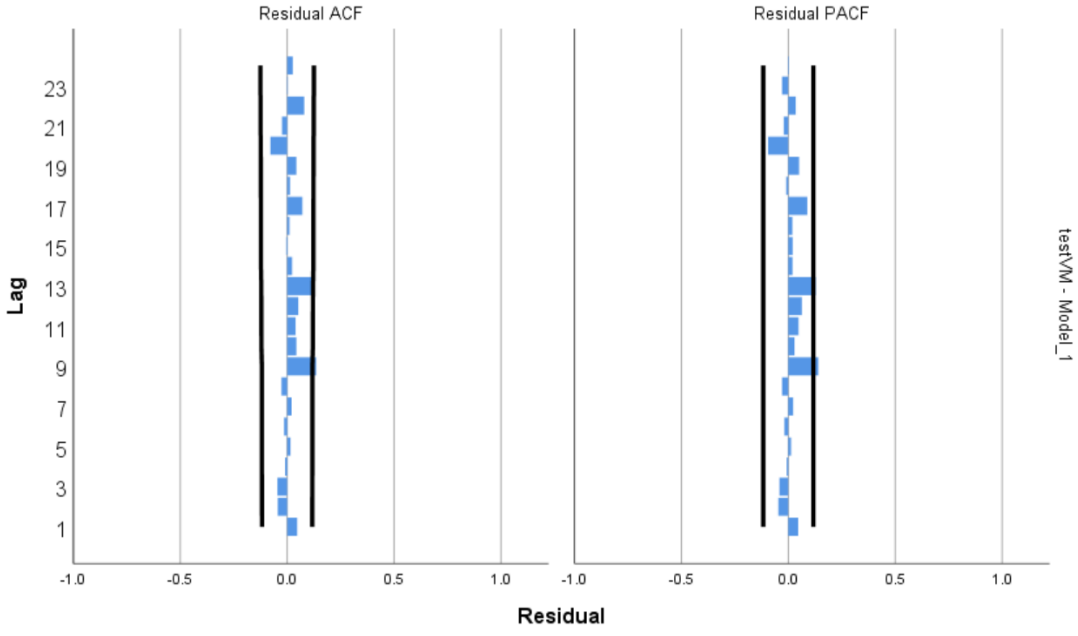
Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.462	.	.462	.462	.462	.462	.462	.462	.462	.462	.462
R-squared	.462	.	.462	.462	.462	.462	.462	.462	.462	.462	.462
RMSE	.149	.	.149	.149	.149	.149	.149	.149	.149	.149	.149
MAPE	93.267	.	93.267	93.267	93.267	93.267	93.267	93.267	93.267	93.267	93.267
MaxAPE	6169.371	.	6169.371	6169.371	6169.371	6169.371	6169.371	6169.371	6169.371	6169.371	6169.371
MAE	.130	.	.130	.130	.130	.130	.130	.130	.130	.130	.130
MaxAE	.366	.	.366	.366	.366	.366	.366	.366	.366	.366	.366
Normalized BIC	-3.761	.	-3.761	-3.761	-3.761	-3.761	-3.761	-3.761	-3.761	-3.761	-3.761

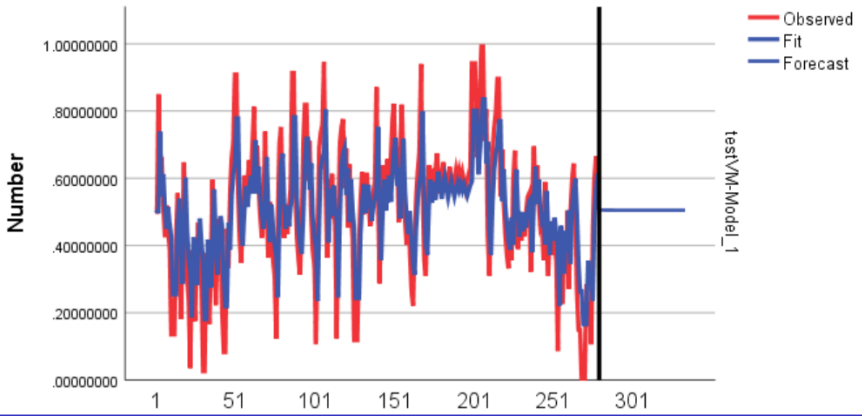
Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
testVM-Model_1	0	.462	15.145	17	.585	0

ARIMA Model Parameters

				Estimate	SE	t	Sig.
testVM-Model_1	testVM	No Transformation	Constant	.505	.028	18.340	.000
			AR Lag 1	.678	.044	15.378	.000





事实证明ARIMA压根不适合拟合20个以后的预测数值，在这里VM在测试组发生了巨大的震荡，而拟合值却迅速收束

于是打算拿出前320来预测后14

首先是Q

Time Series Modeler

Model Description

			Model Type
Model ID	testQ	Model_1	ARIMA(1,1,6)

Model Summary

Model Fit

					Percentile						
Fit Statistic	Mean	SE	Minimum	Maximum	5	10	25	50	75	90	95
Stationary R-squared	.496	.	.496	.496	.496	.496	.496	.496	.496	.496	.496
R-squared	.994	.	.994	.994	.994	.994	.994	.994	.994	.994	.994
RMSE	.004	.	.004	.004	.004	.004	.004	.004	.004	.004	.004
MAPE	.652	.	.652	.652	.652	.652	.652	.652	.652	.652	.652
MaxAPE	3.083	.	3.083	3.083	3.083	3.083	3.083	3.083	3.083	3.083	3.083
MAE	.003	.	.003	.003	.003	.003	.003	.003	.003	.003	.003
MaxAE	.016	.	.016	.016	.016	.016	.016	.016	.016	.016	.016
Normalized BIC	-11.005	.	-11.005	-11.005	-11.005	-11.005	-11.005	-11.005	-11.005	-11.005	-11.005

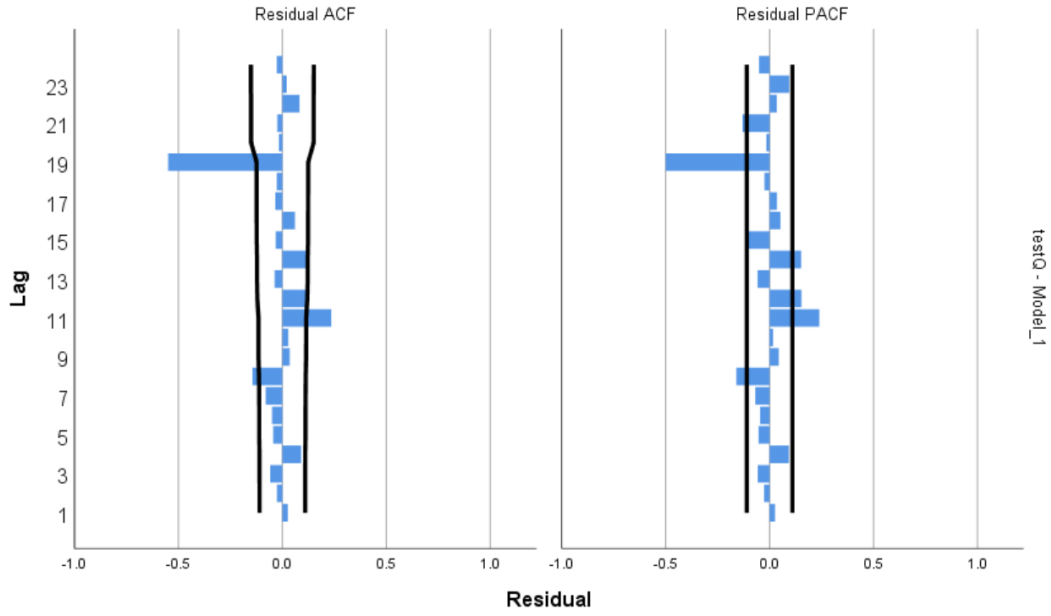
Model Statistics

		Model Fit statistics	Ljung-Box Q(18)			
Model	Number of Predictors	Stationary R-squared	Statistics	DF	Sig.	Number of Outliers
testQ-Model_1	0	.496	46.461	15	.000	0

ARIMA Model Parameters

					Estimate	SE	t	Sig.
testQ-Model_1	testQ	No Transformation	AR	Lag 1	.697	.042	16.743	.000
			Difference		1			
			MA	Lag 5	.162	.057	2.870	.004
				Lag 6	.136	.056	2.429	.016

Lag 6	.136	.056	2.429	.016



.41857273
.42634650
.43378809
.43300979
.43461854
.42985431
.42459633
.41943910
.42279721
.42582583
.42981888
.44051911
.45081288
.46094103

.41467298
.41614902
.41866414
.41994570
.42007900
.42012301
.42015367
.42017502
.42018990

.42020026
.42020748
.42021250
.42021600
.42021844 这两行分别为测试组和拟合值在321~334的值，可以发现较大的误差

下面是VM：

Time Series Modeler

Model Description

			Model Type
Model ID	testVM	Model_1	ARIMA(1,0,0)

Model Summary

Model Fit

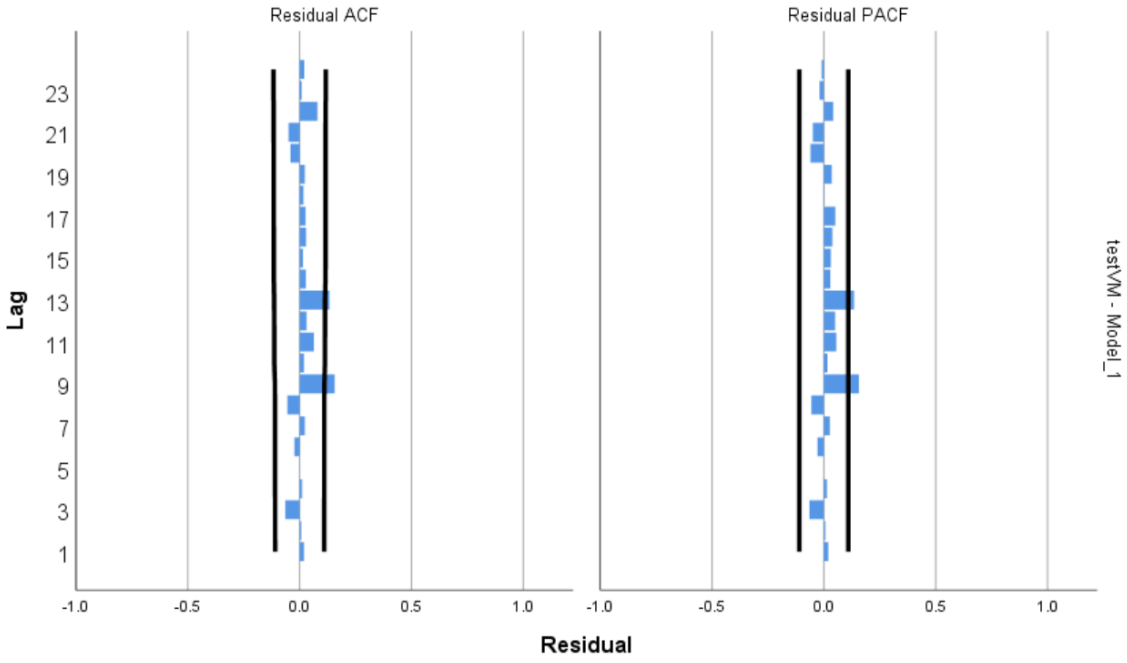
Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.444	.	.444	.444	.444	.444	.444	.444	.444	.444	.444
R-squared	.444	.	.444	.444	.444	.444	.444	.444	.444	.444	.444
RMSE	.151	.	.151	.151	.151	.151	.151	.151	.151	.151	.151
MAPE	88.130	.	88.130	88.130	88.130	88.130	88.130	88.130	88.130	88.130	88.130
MaxAPE	6318.505	.	6318.505	6318.505	6318.505	6318.505	6318.505	6318.505	6318.505	6318.505	6318.505
MAE	.132	.	.132	.132	.132	.132	.132	.132	.132	.132	.132
MaxAE	.365	.	.365	.365	.365	.365	.365	.365	.365	.365	.365
Normalized BIC	-3.742	.	-3.742	-3.742	-3.742	-3.742	-3.742	-3.742	-3.742	-3.742	-3.742

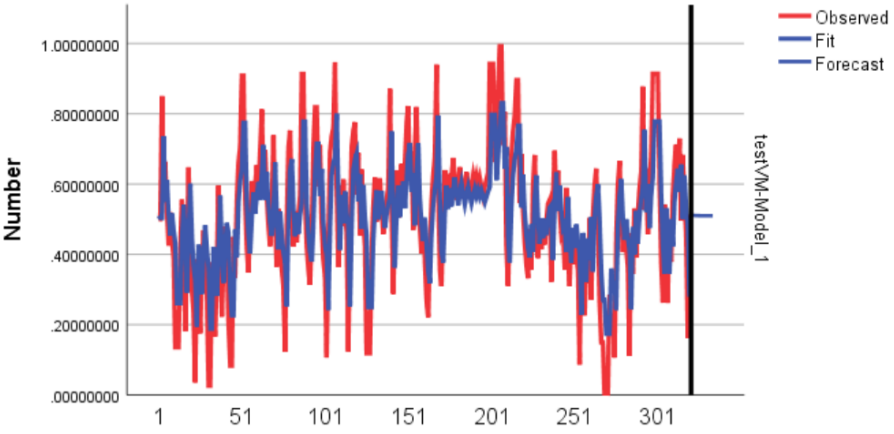
Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
testVM-Model_1	0	.444	19.749	17	.287	0

ARIMA Model Parameters

					Estimate	SE	t	Sig.
testVM-Model_1	testVM	No Transformation	Constant		.510	.025	20.373	.000
			AR	Lag 1	.664	.042	15.885	.000





可以发现拟合值同样迅速收束，压根没法预测

当然只要说明有时间序列的特性即可，因此选用决赛的数据（上面有全部比赛的数据，但全部比赛的无论是技术指标涉及人员还是场次都跨多场，有时间序列特性才怪

首先是Q：

Model Description

			Model Type
Model ID	Q	Model_1	ARIMA(1,1,6)

Model Summary

Model Fit

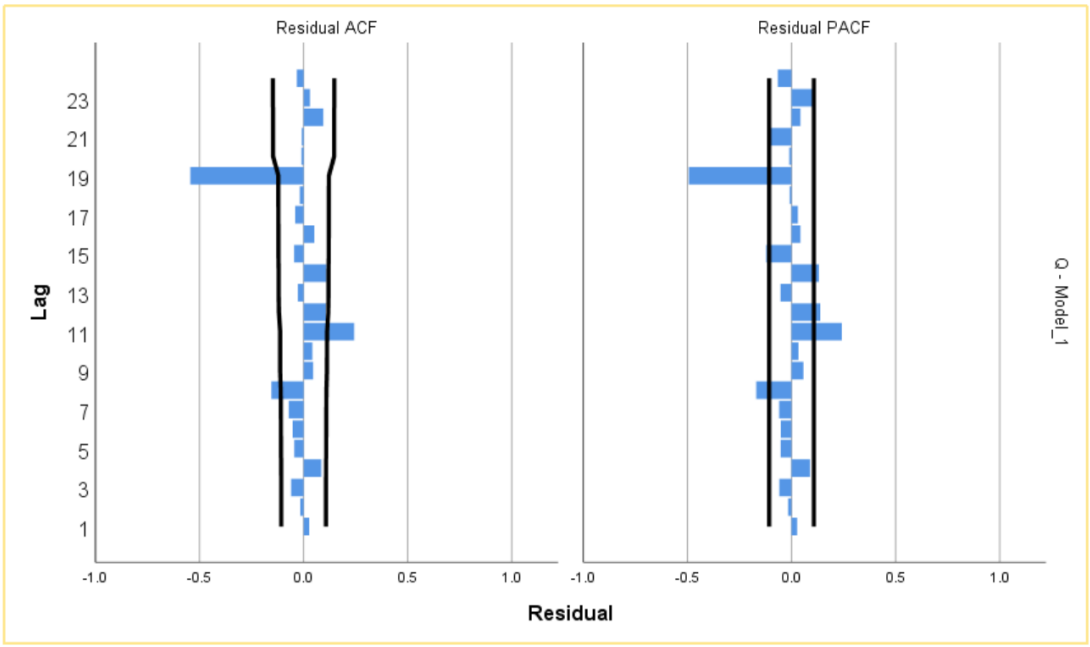
Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.506	.	.506	.506	.506	.506	.506	.506	.506	.506	.506
R-squared	.994	.	.994	.994	.994	.994	.994	.994	.994	.994	.994
RMSE	.004	.	.004	.004	.004	.004	.004	.004	.004	.004	.004
MAPE	.655	.	.655	.655	.655	.655	.655	.655	.655	.655	.655
MaxAPE	3.093	.	3.093	3.093	3.093	3.093	3.093	3.093	3.093	3.093	3.093
MAE	.003	.	.003	.003	.003	.003	.003	.003	.003	.003	.003
MaxAE	.016	.	.016	.016	.016	.016	.016	.016	.016	.016	.016
Normalized BIC	-11.010	.	-11.010	-11.010	-11.010	-11.010	-11.010	-11.010	-11.010	-11.010	-11.010

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
Q-Model_1	0	.506	48.182	15	.000	0

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Q-Model_1	Q	No Transformation	AR	Lag 1	.703	.041	17.309	.000
			Difference		1			
			MA	Lag 5	.181	.055	3.290	.001
				Lag 6	.130	.055	2.388	.017



然后是VM：

Model Description

			Model Type
Model ID	VM	Model_1	ARIMA(1,0,0)

Model Summary

Model Fit

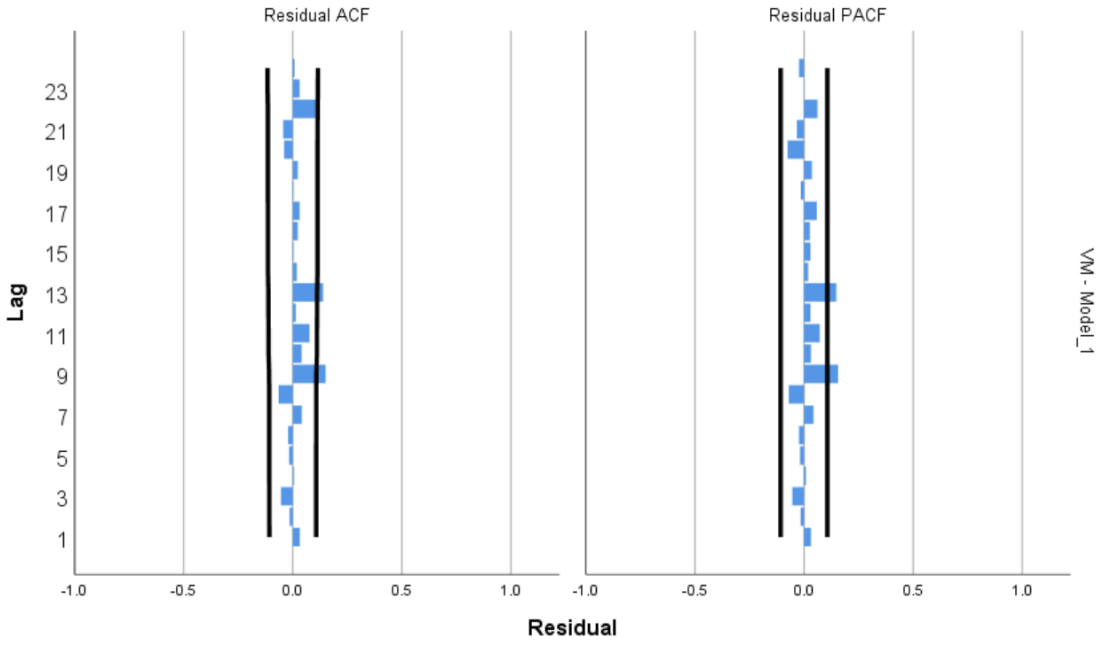
Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.436	.	.436	.436	.436	.436	.436	.436	.436	.436	.436
R-squared	.436	.	.436	.436	.436	.436	.436	.436	.436	.436	.436
RMSE	.152	.	.152	.152	.152	.152	.152	.152	.152	.152	.152
MAPE	86.719	.	86.719	86.719	86.719	86.719	86.719	86.719	86.719	86.719	86.719
MaxAPE	6394.565	.	6394.565	6394.565	6394.565	6394.565	6394.565	6394.565	6394.565	6394.565	6394.565
MAE	.133	.	.133	.133	.133	.133	.133	.133	.133	.133	.133
MaxAE	.364	.	.364	.364	.364	.364	.364	.364	.364	.364	.364
Normalized BIC	-3.737	.	-3.737	-3.737	-3.737	-3.737	-3.737	-3.737	-3.737	-3.737	-3.737

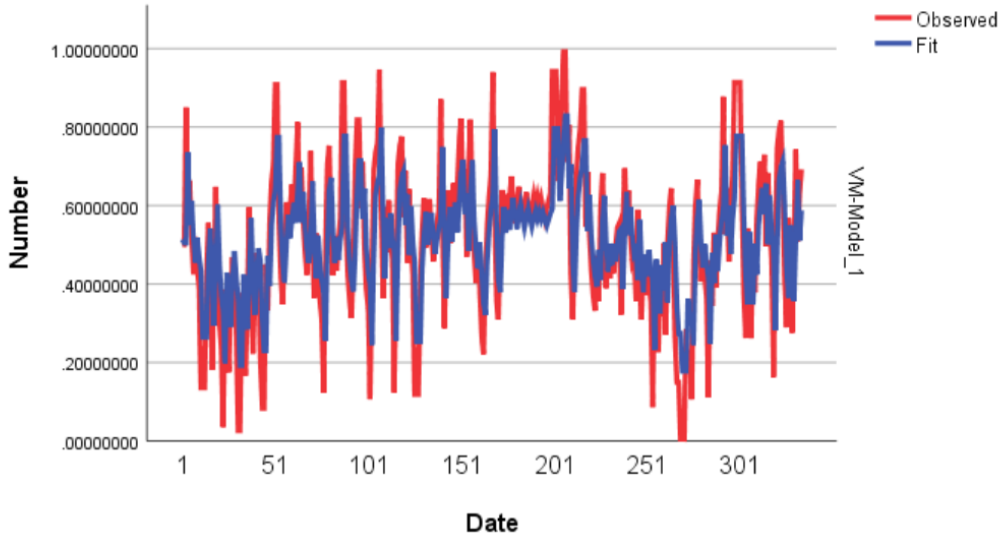
Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
VM-Model_1	0	.436	21.708	17	.196	0

ARIMA Model Parameters

				Estimate	SE	t	Sig.
VM-Model_1	VM	No Transformation	Constant	.514	.024	21.223	.000
			AR Lag 1	.659	.041	15.967	.000





在这一范围内，ARIMA对于Q和VM的建模效果都很好；可以说明momentum并不是随机的，而是能够解释为随着point的深入而相关