IV. COMPARATIVE AND EXPLORATIVE ANALYSIS OF BATTLE OF KURSK DATA

A. APPLICATION OF PREVIOUS STUDIES

1. Application of Bracken's methodology

This section analyzes the Battle of Kursk data following the same steps used by Bracken in his study, and subsequently applies Bracken's models to the Battle of Kursk data. The Battle of Kursk data is formatted and presented in tables using the same methodology, explained in detail in Chapter 3 and formatting techniques as Bracken did in his study. Tables 15 through 18 present data on combat manpower, APCs, tanks, and artillery consecutively for days 1-15 of the Battle of Kursk, from June 4, 1943 to June 18, 1943, for both the German and the Soviet forces.

Day	Blue manpower	Blue casualties	Red manpower	Red casualties
1	510252	130	307365	800
2	507698	8527	301341	6192
3	498884	9423	297205	4302
4	489175	10431	293960	3414
5	481947	9547	306659	2942
6	470762	11836	303879	2953
7	460808	10770	302014	2040
8	453126	7754	300050	2475
9	433813	19422	298710	2612
10	423351	10522	299369	2051
11	415254	8723	297395	2140
12	419374	4076	296237	1322
13	416666	2940	296426	1350
14	415461	1217	296350	949
15	413298	3260	295750	1054

Table 15. Combat manpower data for both sides. Casualties are killed, wounded, captured/missing in action, and disease and nonbattle injuries. Notice the low casualty rates for day 1, when the offensive had not really started.

Day	Blue APCs	Blue APCs Killed	Red APCs	Red APCs killed
1	511	0	1170	0
2	507	4	1142	29
3	501	6	1128	14
4	490	11	1101	27
5	477	13	1085	16
6	458	19	1073	14
7	463	3	1114	42
8	462	4	1104	16
9	432	30	1099	12
10	424	8	1096	4
11	418	8	1093	6
12	417	1	1089	5
13	417	0	1092	1
14	417	2	1095	1
15	409	8	1098	5

Table 16. APC data for both sides. Killed are destroyed+abandoned and damaged. Notice the high number of German APC losses on day 7 and the high number of Soviet APC losses on day 9.

Day	Blue Tanks	Blue Tanks Killed	Red Tanks	Red Tanks killed
1	2500	0	1178	4
2	2396	105	986	198
3	2367	117	749	248
4	2064	259	673	121
5	1754	315	596	108
6	1495	289	490	139
7	1406	157	548	36
8	1351	135	563	63
9	977	414	500	98
10	978	117	495	57
11	907	118	480	46
12	883	96	426	79
13	985	27	495	23
14	978	42	557	7
15	948	85	588	6

Table 17. Tank data for both sides. Killed are destroyed+abandoned and damaged. Notice the decrease in the number of tank losses after day 9. After day 9, the battle lost its intensity.

Table 19 presents data on total forces, where the data from Tables 16-18 on combat manpower, APCs, tanks, and artillery are weighted by 1, 5, 40, and 20, respectively. Bracken [Ref.8] states in his study that, “The weights given above are

Day	Blue artillery	Blue artillery killed	Red artillery	Red artillery killed
1	718	0	1189	1
2	705	13	1166	24
3	676	30	1161	5
4	661	15	1154	7
5	648	14	1213	13
6	640	9	1210	6
7	629	13	1199	12
8	628	7	1206	15
9	613	16	1194	12
10	606	10	1187	7
11	603	5	1184	5
12	601	5	1183	3
13	600	3	1179	4
14	602	0	1182	2
15	591	4	1182	11

Table 18. Artillery data for both sides. Killed are destroyed+abandoned and damaged. On the initial days of the battle, German artillery losses were higher than the Soviet artillery losses.

consistent with those of studies and models of the U.S. Army Concepts Analysis Agency. Virtually all theater-level dynamic combat simulation models incorporate similar weights, either as inputs or as decision parameters computed as the simulations progress".

Day	Blue forces	Blue losses	Red forces	Red losses
1	591527	130	384335	920
2	586353	11167	373411	11257
3	575769	12993	364265	9532
4	559345	16266	359085	6249
5	545332	16472	372524	5702
6	528552	18071	367444	6043
7	516403	14445	366504	3450
8	507576	10754	365070	4415
9	480033	28492	361965	5112
10	469271	13302	362229	3491
11	459604	11323	359820	3290
12	463159	6201	357522	3047
13	462451	3600	358946	1975
14	461186	2067	360245	1174
15	457943	5160	360280	1639

Table 19. Data on aggregated forces. Forces are combat manpower, APCs, tanks and artillery which are weighted by 1, 5, 20 and 40 respectively. The number of Soviet losses on day 9 is almost three times the amount of Soviet loss on day 8.

a. Estimation of Parameters

The parameters of the model are chosen to minimize the sum of squared residuals between the estimated and actual attrition. Using 15 days of the Battle of Kursk data, where the first 8 days the Germans attack and the last 7 days the Soviets attack, it is desired to minimize:

$$\text{SSR} = \sum_{n=1}^{8}(\dot{B}_n - adR_n^p B_n^q)^2 + \sum_{n=1}^{8}(\dot{R}_n - b(1/d)B_n^p R_n^q)^2$$

$$+ \sum_{n=9}^{15}(\dot{B}_n - a(1/d)R_n^p B_n^q)^2 + \sum_{n=9}^{15}(\dot{R}_n - bdB_n^p R_n^q)^2 \qquad (5)$$

where n denotes the index for the 15 days of the battle. Using the data given in Table 17, the above procedure will give a different SSR value for each set of parameters, i.e. combination of a, b, p, q and d values. It will evaluate $\text{SS}(a_i, b_j, p_k, q_l, d_m)$ for all combinations of i, j, k, l and m where i=1,....,9, j=1,....,9, k=1,....,21, l=1,....,21, and m=1,....,9.

The range of possibilities allowed for the parameters, for the model with the tactical parameter d will be:

$$(a_1, \ldots\ldots\ldots, a_9) = (4\times10^{-9}, \ldots\ldots\ldots, 1.2\times10^{-8}),$$

$$(b_1, \ldots\ldots\ldots, b_9) = (4\times10^{-9}, \ldots\ldots\ldots, 1.2\times10^{-8}),$$

$$(p_1, \ldots\ldots\ldots, p_{21}) = (0.0, \ldots\ldots\ldots, 2.0),$$

$$(q_1, \ldots\ldots\ldots, q_{21}) = (0.0, \ldots\ldots\ldots, 2.0),$$

$$(d_1, \ldots\ldots\ldots, d_9) = (0.6, \ldots\ldots\ldots, 1.4).$$

These ranges were selected by Bracken himself.

There are a total of 9x9x21x21x9 = 321489 combinations of the estimated parameters. The algorithm searches all combinations and determines the parameters that minimize the sum of squared residuals for the data given in Table 17 as:

$$\mathrm{SS}\,(a_9, b_6, p_2, q_{21}, d_4) = 8.65\times10^8$$

with the estimated parameters of:

$$a_9 = 1.2\times10^{-8}, b_6 = 9\times10^{-9}, p_2 = 0.1, q_{21} = 2.0, d_4 = 0.9\,.$$

Notice that the values of the *a* parameter and the *q* parameter are on the boundary.

Now, considering the estimation of parameters for the model without the tactical parameter *d*, the ranges of possibilities allowed will be the same as those in the previous procedure, except for the tactical parameter *d*. There are now a total of 9x9x21x21 = 35721 combinations of parameters. The algorithm searches all combinations and determines the parameters that minimize the sum of squared residuals for the data given in Table 17 as:

$$\mathrm{SS}\,(a_9, b_6, p_4, q_{19}) = 8.88\times10^8$$

with the estimated parameters of:

$$a_9 = 1.2\times10^{-8}, b_6 = 9\times10^{-9}, p_2 = 0.3, q_{21} = 1.8\,.$$

Table 20 gives the sums of squared residuals for different values of *d*, and shows which *a*, *b*, *p*, *q* combinations gives the minimum sums of squared residuals for the various *d* values. Table 20 also shows the sensitivity of the *p* and *q* parameters to the *d* parameter and suggests that the sums of the squared residuals are similar within a wide range of parameter values.

b. *Results*

The best fitting results for the two models for the Battle of Kursk data are:

Bracken's model 1 with tactical parameter *d*,

$$\dot{B} = 1.2\times10^{-8}(\frac{10}{9} \text{ or } \frac{9}{10})R^{0.1}B^{2.0} \quad (6)$$

$$\dot{R} = 9\times10^{-9}(\frac{9}{10} \text{ or } \frac{10}{9})B^{0.1}R^{2.0} \quad (7)$$

d	**SSR**	***a***	***b***	***p***	***q***
0.5	1.38E+9	9.00E-9	6.00E-9	0.1	2.0
0.6	1.15E+9	1.00E-8	7.00E-9	0.1	2.0
0.7	9.84E+8	1.20E-8	8.00E-9	0.1	2.0
0.8	8.87E+8	1.20E-8	9.00E-9	0.1	2.0
0.9	8.65E+8	1.20E-8	9.00E-9	0.1	2.0
1.0	8.88E+8	1.20E-8	9.00E-9	0.3	1.8
1.1	9.34E+8	1.20E-8	8.00E-9	0.5	1.6
1.2	9.90E+8	1.20E-8	7.00E-9	0.7	1.4
1.3	1.05E+9	1.20E-8	7.00E-9	0.8	1.3
1.4	1.10E+9	1.20E-8	6.00E-9	1.0	1.1
1.5	1.16E+9	1.20E-8	5.00E-9	1.2	0.9
1.6	1.21E+9	1.20E-8	5.00E-9	1.3	0.8
1.7	1.25E+9	1.20E-8	4.00E-9	1.5	0.6

Table 20. SSR values for different d values. a and b values are varied between 8×10^{-9} and 1.2×10^{-8} with increments of 1×10^{-9}, p and q values are varied between 0.0 and 2.0 with increments of 0.1. The lowest SSR value is observed to be 8.65E+8 when *d*=0.9.

Bracken's model 3 without the tactical parameter *d*

$$\dot{B} = 1.2\times10^{-8}R^{0.3}B^{1.8} \quad (8)$$

$$\dot{R} = 9\times10^{-9}B^{0.3}R^{1.8} \quad (9)$$

The parameters found in equations IV.A.1.b.(6), IV.A.1.b.(7), IV.A.1.b.(8), IV.A.1.b.(9) suggest that one side's losses are more a function of his own forces rather than a function of the opponent's forces. This result is similar to what Fricker found in his study. There are boundaries set for the search of parameters that give the best fit. There may be other sets of parameters that are out of the range of possibilities allowed by this method, and they may give a better fit for the data. The fact that some of the best fitting parameters are on the boundary supports this hypothesis.

Figures 18 and 19 show the real and fitted values found using the model with the *d* parameter (i.e., using equations IV.A.1.b.(6) and IV.A.1.b.(7), for the Soviet and the German forces respectively). Figures 20 and 21 show the real and fitted values found using the model without the *d* parameter (i.e. using the formulas IV.A.1.b.(8) and IV.A.1.b.(9), for the Soviet and German forces, respectively).

When the plots given in Figures 18 and 20 are examined, there appears to be three phases in the battle. It is also apparent that the battle lost its intensity after July 12. The model underestimates the casualties for the beginning part and the last part of the battle while overestimating the 8 days in a row between these two periods. This pattern suggests that fitting a model with change points may improve the fit to the data.

For the model with the tactical parameter, p-q=-1.9, and for the model without the tactical parameter p-q=-1.5. These two results imply that the Battle of Kursk data does not fit any one of the basic Lanchester linear, square or logarithmic laws.

For both cases, parameters a and b are significantly small and $a > b$. This suggests that individual German effectiveness was greater than individual Russian effectiveness.

For the purpose of comparing a variety of models throughout this thesis, R^2 values are also computed together with the SSR for each model, where R^2 is given as:

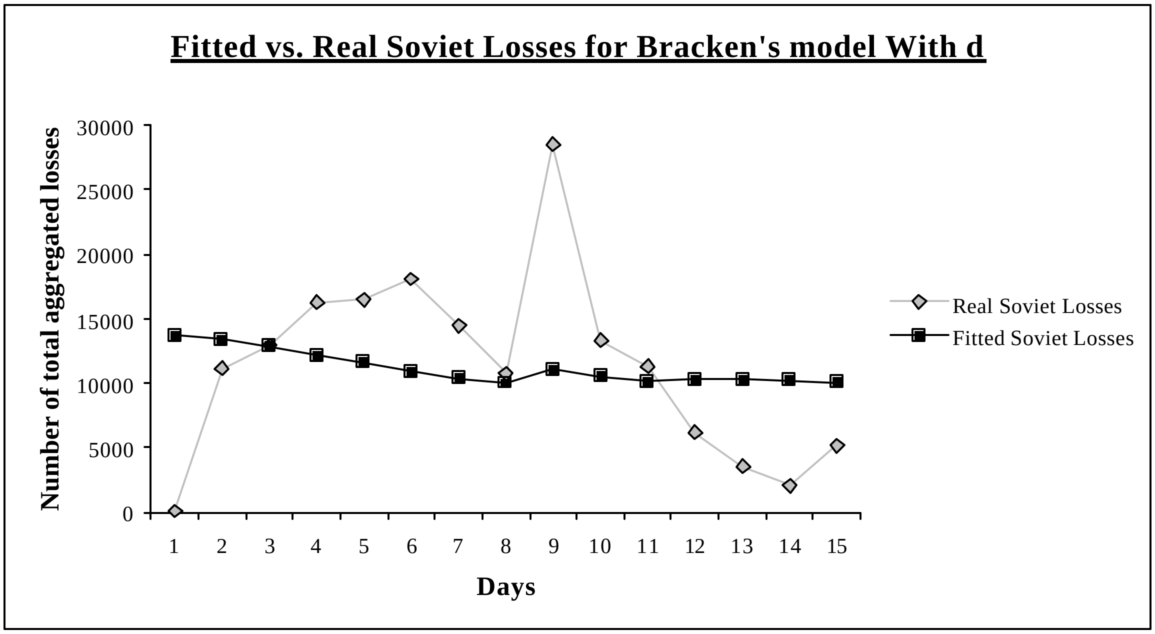

Figure 18. Fitted Soviet losses found by using Bracken's model with parameter *d*, plotted versus real Soviet losses. Notice the three-phase pattern in the model's fit to the battle data where the model overestimates the first two days and the last four days of the battle while underestimating the part between these two phases.

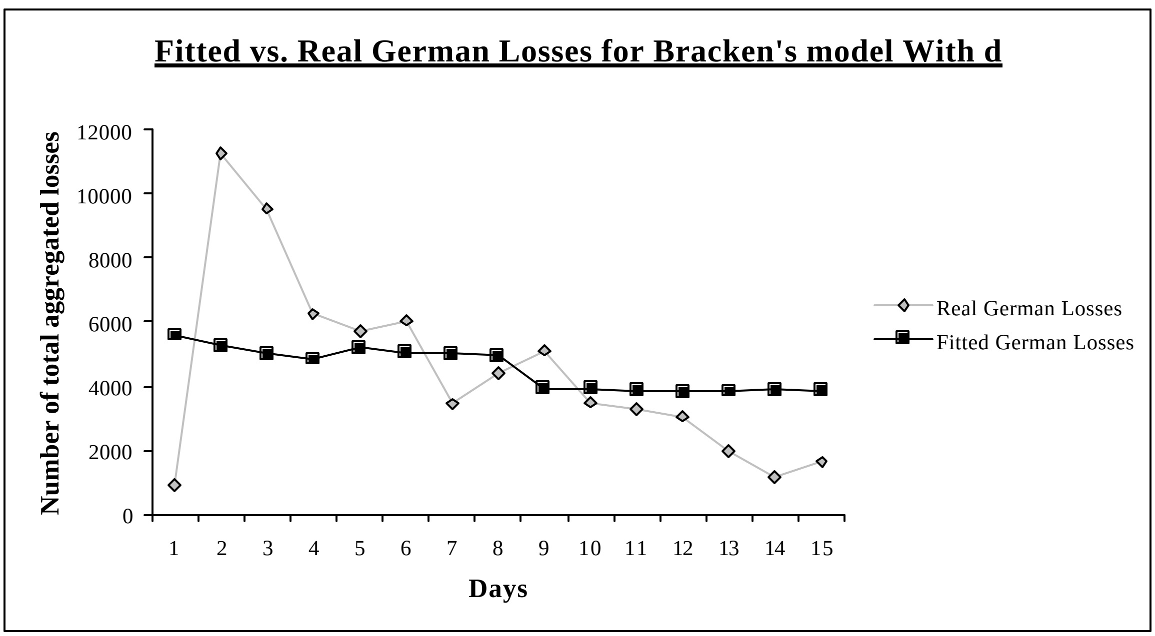

Figure 19. Fitted German losses found by using Bracken's model with parameter *d*, plotted versus real German losses. After the Soviets went into offense, the battle was not as intense.

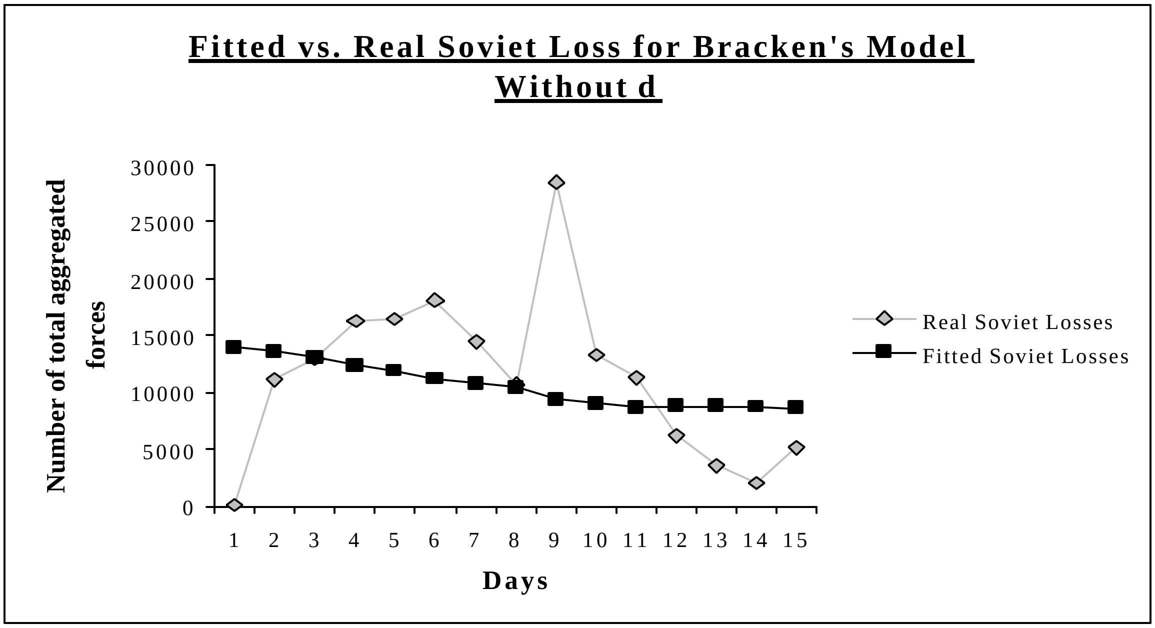

Figure 20. Fitted Soviet losses found by using Bracken's model without parameter d, plotted versus real Soviet losses. Like the plot given in Figure 18, notice the three-phase pattern in the model's fit to the battle where the model overestimates the first two days and the last four days of the battle while underestimating the part between these two phases.

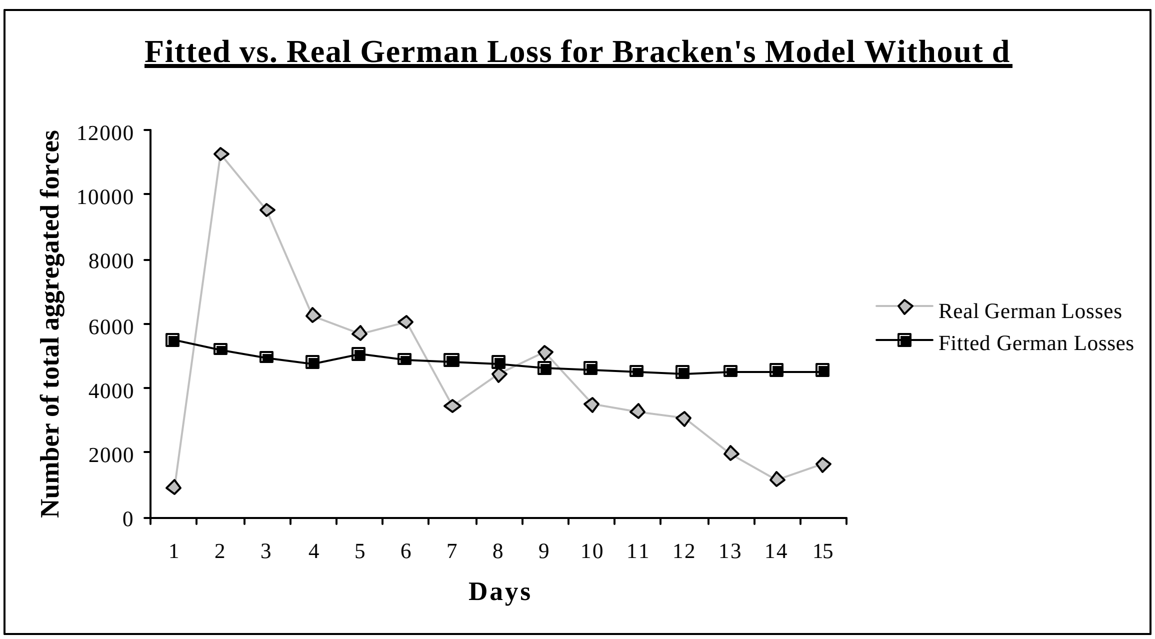

Figure 21. Fitted German losses found by using Bracken's model without parameter *d*, plotted versus real German losses. After the Soviets went into offense, the battle was not as intense.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_i (Y - \hat{Y})^2}{\sum_i (Y - \overline{Y})^2} \tag{10}$$

where $\hat{Y}$, Y and $\overline{Y}$ denote the estimated value, the real value and the mean value of the Y parameter (daily casualties) which are indexed by days. A greater R^2 value indicates a better fit. It is possible to get a negative R^2 value, implying that the fitted model yields worse results than using the average daily losses as an estimate.

Table 21 shows the results for Bracken's models as a whole.

Name of the model	***a***	***b***	***p***	***q***	***d***	***SSR***	R^2
Bracken Model 1 Ardennes	8.0E-9	1.0E-8	1.0	1.0	1.25	1.63E+7	0.2552
Bracken Model 3 Ardennes	8.0E-9	1.0E-8	1.3	0.7	1.0	2.08E+8	0.0493
Bracken Model 1 Kursk	1.2E-8	9.0E-9	0.1	2.0	0.9	8.65E+8	0.0006
Bracken Model 3 Kursk	1.2E-8	9.0E-9	0.3	1.8	1.0	8.88E+8	-0.0266

Table 21. Bracken's results for his models with the tactical parameter *d* for both the Ardennes and Kursk data.

Upon examination of the fits of Bracken's models found in this section, it is clear that the battle did not start until the second day. Thus, the first day of data was dropped in fitting the data to the models in the rest of the thesis. More detailed explanation on this approach is given in Section IV.B.1.

Bracken's Model 1 was refit using only the last 14 days of the data. The new parameter estimates are: a=1.2×10^{-8}, b=1.0×10^{-8}, p=0.1, q=2.0, d=1.0. The SSR value dropped to 6.50×10^{8} and the R^2 value increased to 0.0919.

2. Application of Fricker's methodology

In his study, Fricker presents an alternate way to structure the data that reflects the effects of both previous casualties and incremental reinforcements. His idea is based on the fact that the casualties occur according to the Lanchester equations that use the previous day's force size, and for any given day, the previous day's force size also depends on the transfer of troops in or out of the fighting force.

Because of this phenomenon, Fricker uses the following algorithm in his study to estimate the original total for each resource. The algorithm works sequentially stepping through the whole battle from day 1 to the last day of the battle. By using this algorithm, local reserves (X_{lr}) or the addition of reinforcements (X_r) are accounted for. The algorithm first uses local reserves for any force increase before using reinforcements. As described in Fricker's study [Ref.6], the algorithm is:

For resource *X*:

1. Set $X_r(t) = X_{lr}(t) = 0$

2. Let t=1:

   - If $X(t+1) > X(t) - \dot{X}(t)$ and $X_{lr} = 0$, then

     $X_r(t) = X_r(t) + [X(t+1) - X(t) + \dot{X}(t)]$

   - Else, if $X(t+1) > X(t) - \dot{X}(t)$ and $X_{lr} \geq X(t+1) - X(t) + \dot{X}(t)$,

     then $X_{lr}(t) = X_{lr}(t) - [X(t+1) - X(t) + \dot{X}(t)]$

   - Else, if $X(t+1) > X(t) - \dot{X}(t)$ and

     $0 < X_{lr}(t) < X(t+1) - X(t) + \dot{X}(t)$, then

     $X_r(t) = X_r(t) + [X(t+1) - X(t) + \dot{X}(t)] - X_{lr}(t), X_{lr}(t) = 0$.

- Else, if $X(t+1) < X(t) - \dot{X}(t)$, then

$$X_{lr}(t) = X_{lr}(t) + [X(t) - \dot{X}(t) - X(t+1)]$$

3. If $t < 31$, increment t and go to step #2; else $\tilde{X}(0) = X(0) + X_r(t)$

Then the new daily resources $\tilde{X}(t+1)$ are calculated as:

$$\tilde{X}(t+1) = \tilde{X}(t) - \dot{X}(t) \qquad t = 0,\ldots\ldots\ldots,31 \tag{11}$$

After the data is reformatted using the algorithm given above, Fricker applies linear regression to logarithmically transformed Lanchester equations for estimating the model parameters. After the logarithmic transformation, the basic Lanchester equations, given in I.B.(1) and I.B.(2), will look like:

$$\ln(\dot{B}) = \ln(a) + p\ln(R) + q\ln(B) \tag{12}$$

$$\ln(\dot{R}) = \ln(b) + p\ln(B) + q\ln(R) \tag{13}$$

Below is the Battle of Kursk data reformatted using Fricker's approach. For reformatting the data, the algorithm, which is explained in detail above, is applied to the given Battle of Kursk data.

Tables 22 and 23 present the raw manpower and weapon systems data, respectively. Tables 24 and 25 present the resulting reformatted Kursk data for manpower and weapon systems, respectively. Table 26 presents the aggregated force (except the first day) found by aggregating the data given in Tables 24 and 25.

The air sortie data given in the KOSAVE study [Ref.12] consists of the number of air-air role sorties, ground attack role sorties, reconnaissance role sorties and evacuation role sorties (solely used by Germans). Table 27 presents data on number of ground attack role sorties. Table 28 presents the aggregated force, after the air sortie data is added (except the first day) by using the weight coefficient of 30, as used by Fricker, (i.e., the number of air sorties presented in Table 27 is multiplied by 30 and added to the aggregated force levels given in Table 26 to get the data presented in Table 28).

Day	Blue Available	Blue Killed	Red Available	Red Killed
1	510252	130	307365	800
2	507698	8527	301341	6192
3	498884	9423	297205	4302
4	489175	10431	293960	3414
5	481947	9547	306659	2942
6	470762	11836	303879	2953
7	460808	10770	302014	2040
8	453126	7754	300050	2475
9	433813	19422	298710	2612
10	423351	10522	299369	2051
11	415254	8723	297395	2140
12	419374	4076	296237	1322
13	416666	2940	296426	1350
14	415461	1217	296350	949
15	413298	3260	295750	1054

Table 22. Battle of Kursk manpower data for the Soviet and German forces.

Day	BLUE						RED					
	Available			Killed			Available			Killed		
	Tank	APC	Art.	Tank	APC	Art.	Tank	APC	Art.	Tank	APC	Art.
1	2500	511	718	0	0	0	1178	1170	1189	4	0	1
2	2396	507	705	105	4	13	986	1142	1166	198	29	24
3	2367	501	676	117	6	30	749	1128	1161	248	14	5
4	2064	490	661	259	11	15	673	1101	1154	121	27	7
5	1754	477	648	315	13	14	596	1085	1213	108	16	13
6	1495	458	640	289	19	9	490	1073	1210	139	14	6
7	1406	463	629	157	3	13	548	1114	1199	36	42	12
8	1351	462	628	135	4	7	563	1104	1206	63	16	15
9	977	432	613	414	30	16	500	1099	1194	98	12	12
10	978	424	606	117	8	10	495	1096	1187	57	4	7
11	907	418	603	118	8	5	480	1093	1184	46	6	5
12	883	417	601	96	1	5	426	1089	1183	79	5	3
13	985	417	600	27	0	3	495	1092	1179	23	1	4
14	978	417	602	42	2	0	557	1095	1182	7	1	2
15	948	409	591	85	8	4	588	1098	1182	6	5	11

Table 23. Battle of Kursk data for tanks, APCs, and artillery of the Soviet and German forces.

Day	Blue Available	Blue Killed	Red Available	Red Killed
1	529562	130	331292	800
2	529432	8527	330492	6192
3	520905	9423	324300	4302
4	511482	10431	319998	3414
5	501051	9547	316584	2942
6	491504	11836	313642	2953
7	479668	10770	310689	2040
8	468898	7754	308649	2475
9	461144	19422	306174	2612
10	441722	10522	303562	2051
11	431200	8723	301511	2140
12	422477	4076	299371	1322
13	418401	2940	298049	1350
14	415461	1217	296699	949
15	414244	3260	295750	1054

Table 24. The reformatted Battle of Kursk manpower data for the Soviet and German forces.

Day	BLUE						RED					
	Available			Killed			Available			Killed		
	Tank	APC	Art.	Tank	APC	Art.	Tank	APC	Art.	Tank	APC	Art.
1	3139	524	742	0	0	0	1815	1285	1298	4	0	1
2	3139	524	742	105	4	13	1811	1285	1297	198	29	24
3	3034	520	729	117	6	30	1613	1256	1273	248	14	5
4	2917	514	699	259	11	15	1365	1242	1268	121	27	7
5	2658	503	684	315	13	14	1244	1215	1261	108	16	13
6	2343	490	670	289	19	9	1136	1199	1248	139	14	6
7	2054	471	661	157	3	13	997	1185	1242	36	42	12
8	1897	468	648	135	4	7	961	1143	1230	63	16	15
9	1762	464	641	414	30	16	898	1127	1215	98	12	12
10	1348	434	625	117	8	10	800	1115	1203	57	4	7
11	1231	426	615	118	8	5	743	1111	1196	46	6	5
12	1113	418	610	96	1	5	697	1105	1191	79	5	3
13	1017	417	605	27	0	3	618	1100	1188	23	1	4
14	990	417	602	42	2	0	595	1099	1184	7	1	2
15	948	415	602	85	8	4	588	1098	1182	6	5	11

Table 25. The reformatted Battle of Kursk equipment data for tanks, APCs, and artillery of the Soviet and German forces.

Day	Blue forces	Blue losses	Red forces	Red losses
1	624512	11167	425017	11257
2	613345	12993	413760	9532
3	600352	16266	404228	6249
4	584086	16472	397979	5702
5	567614	18071	392277	6043
6	549543	14445	386234	3450
7	535098	10754	382784	4415
8	524344	28492	378369	5112
9	495852	13302	373257	3491
10	482550	11323	369766	3290
11	471227	6201	366476	3047
12	465026	3600	363429	1975
13	461426	2067	361454	1174
14	459359	5160	360280	1639

Table 26. Data on aggregated forces that are reformatted without air sorties. Forces are combat manpower, APCs, Tanks and artillery weighted by 1, 5, 20 and 40, respectively.

Day	German grnd, att, role sorties	Soviet grnd, att, role sorties
1	160	1
2	1942	600
3	1356	613
4	1499	661
5	1426	669
6	1286	472
7	530	383
8	809	348
9	460	603
10	451	623
11	1147	704
12	541	369
13	278	681
14	122	336
15	18	377

Table 27. Data on number of ground attack role air sorties for German and Soviet forces.

Day	Blue forces	Blue losses	Red forces	Red losses
1	642512	11167	483277	11257
2	631735	12993	454440	9532
3	620182	16266	449198	6249
4	604156	16472	440759	5702
5	581774	18071	430857	6043
6	561033	14445	402134	3450
7	545538	10754	407054	4415
8	542434	28492	392169	5112
9	514542	13302	386787	3491
10	503670	11323	404176	3290
11	482297	6201	382706	3047
12	485456	3600	371769	1975
13	471506	2067	365114	1174
14	470669	5160	360820	1639

Table 28. Data on aggregated forces reformatted with air sorties. Forces are combat manpower, APCs, tanks, artillery and number of ground attack role air sorties which are weighted by 1, 5, 20,40 and 30, respectively.

a. Estimation of Parameters

After reformatting the data, linear regression is applied to logarithmically transformed Lanchester equations to estimate the model parameters which are given in equations IV.A.2.(12) and IV.A.2.(13).

To estimate the parameters of the model, which minimize the sum of squared residuals, 14 days of data given in Table 24, Table 26 and S-PLUS Software are used.

b. Results

Results for the models are:

Fricker's model for the Kursk data without the air sorties, with tactical parameter *d*, with an SSR value of 5.94×10^{8} and an R^2 value of 0.1703 is:

$$\dot{B} = 3.77\times10^{-33}\left(\frac{100}{79} \text{ or } \frac{79}{100}\right)R^{0.0604}B^{6.3066} \tag{14}$$

$$\dot{R} = 1.09\times10^{-32}\left(\frac{79}{100} \text{ or } \frac{100}{79}\right)B^{0.0604}R^{6.3066} \tag{15}$$

Fricker's model for the Kursk data without the air sorties, and without the tactical parameter *d*, with an SSR value of 6.69×10^8 and an R^2 value of 0.0657 is:

$$\dot{B} = 1.19 \times 10^{-32} R^{3.6736} B^{2.6934} \tag{16}$$

$$\dot{R} = 3.44 \times 10^{-33} B^{3.6736} R^{2.6934} \tag{17}$$

It is significant that the resulting parameters are sensitive to the *d* parameter; after adding the *d* parameter, the *p* and *q* parameters change dramatically.

The above parameters are the ones that give the smallest SSR value. It is possible to have smaller SSR values for the model with the tactical parameter *d* if the parameter p or q is allowed to have negative values. In staying consistent with Fricker's approach, negative exponent parameters are not considered in this section. Negative values are looked at in the conclusion section.

Fricker's model for the Kursk data with the air sorties, with tactical parameter *d*, with an SSR value of 6.24×10^8 and an R^2 value of 0.1285 is:

$$\dot{B} = 3.35 \times 10^{-27} \left(\frac{100}{93} \text{ or } \frac{93}{100}\right) R^{0.0955} B^{5.2207} \tag{18}$$

$$\dot{R} = 5.76 \times 10^{-27} \left(\frac{93}{100} \text{ or } \frac{100}{93}\right) B^{0.0955} R^{5.2207} \tag{19}$$

Fricker's model for the Kursk data with the air sorties, and without the tactical parameter *d*, with an SSR value of 7.18×10^8 and an R^2 value of –0.020 is:

$$\dot{B} = 5.01 \times 10^{-27} R^{1.4983} B^{3.8179} \tag{20}$$

$$\dot{R} = 3.85 \times 10^{-27} R^{1.4983} B^{3.8179} \tag{21}$$

Like the models without the air sorties added, the above parameters are the ones that give the smallest SSR value. It is possible to have smaller SSR values for the model with the tactical parameter d if the parameter p or q is allowed to have negative values. That is the algorithm of a force's casualties decreases as one of the force strengths increases, and since this interpretation does not make sense, the negative values are not considered.

Figures 22 and 23 show the fitted losses plotted versus real losses for the Soviet and the German forces, respectively, for Fricker's model without the air sortie data added and using the *d* parameter.

Figures 24 and 25 show fitted losses plotted versus real losses for the Soviet and the German forces, respectively, for Fricker's model without the air sortie data added and without using the *d* parameter.

Figures 26 and 27 show the fitted losses plotted versus real losses for the Soviet and the German forces, respectively, for Fricker's model with the air sortie data added and using the *d* parameter.

Figures 28 and 29 show the fitted losses plotted versus real losses for the Soviet and the German forces, respectively, for Fricker's model with the air sortie data added and without using the *d* parameter.

When the R^2 values above, which are found by using Fricker's methodology, are compared, it is seen that adding the air sortie data improves the fit for the Battle of Kursk data. Using the tactical parameter does not improve the fit to the Battle of Kursk data for the model without the air sorties. On the contrary, using the tactical parameter improves the fit to the Battle of Kursk data for the model with the air sorties.

The *d* parameter is found to be 0.79 and 0.93 for the models without the air sorties and with the air sorties, consecutively. This result implies a defender advantage/attacker

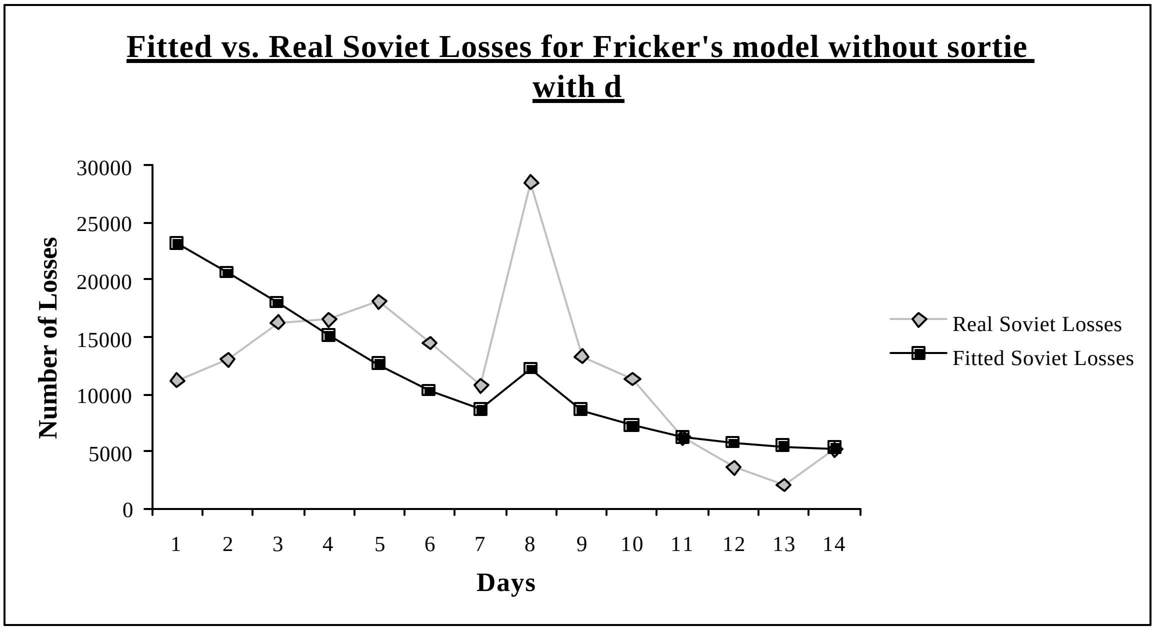

Figure 22. Fitted losses plotted versus real losses for the Soviet forces for Fricker's model without the air sortie data added and using the *d* parameter. Notice the pattern where the model overestimates the initial and the last part of the battle, while underestimating the part in between.

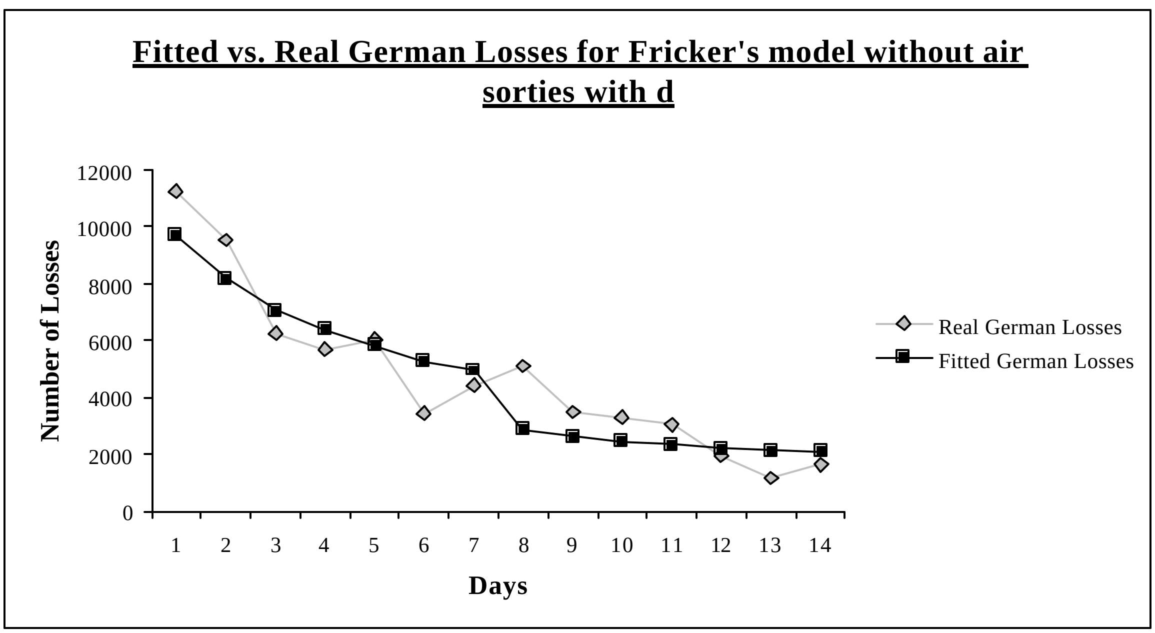

Figure 23. Fitted losses plotted versus real forces for the German forces for Fricker's model without the air sortie data added and using the *d* parameter.

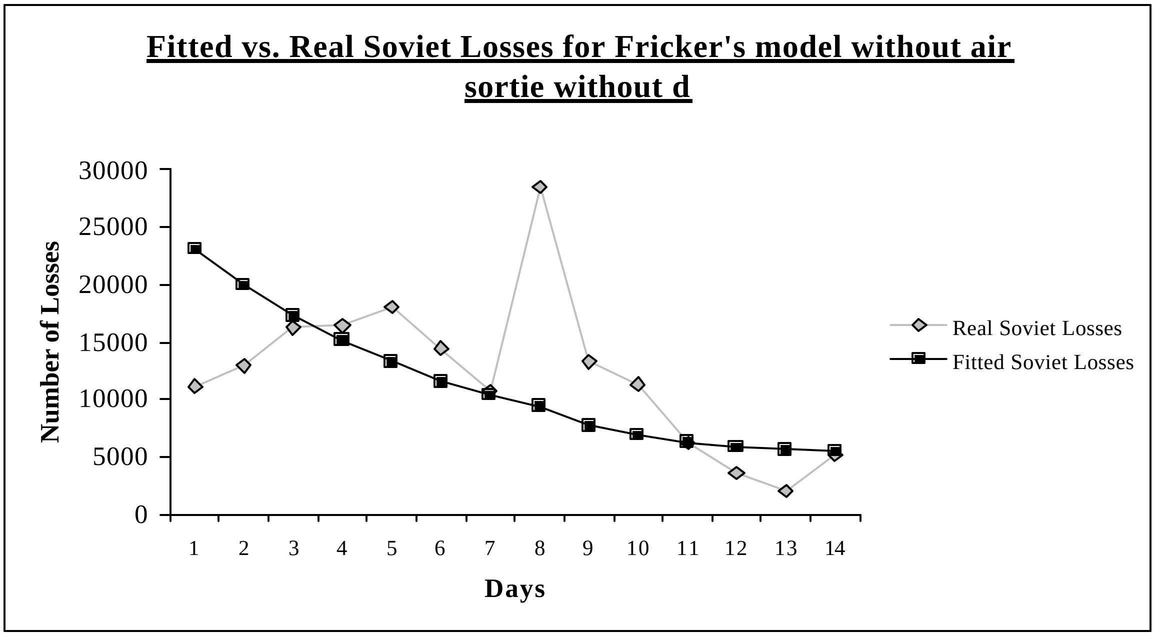

Figure 24. Fitted losses plotted versus real Soviet losses for the Soviet forces for Fricker's model without the air sortie data added and without using the *d* parameter. The same pattern in which the model over/underestimates the battle in three distinctive phases is also observable in this plot.

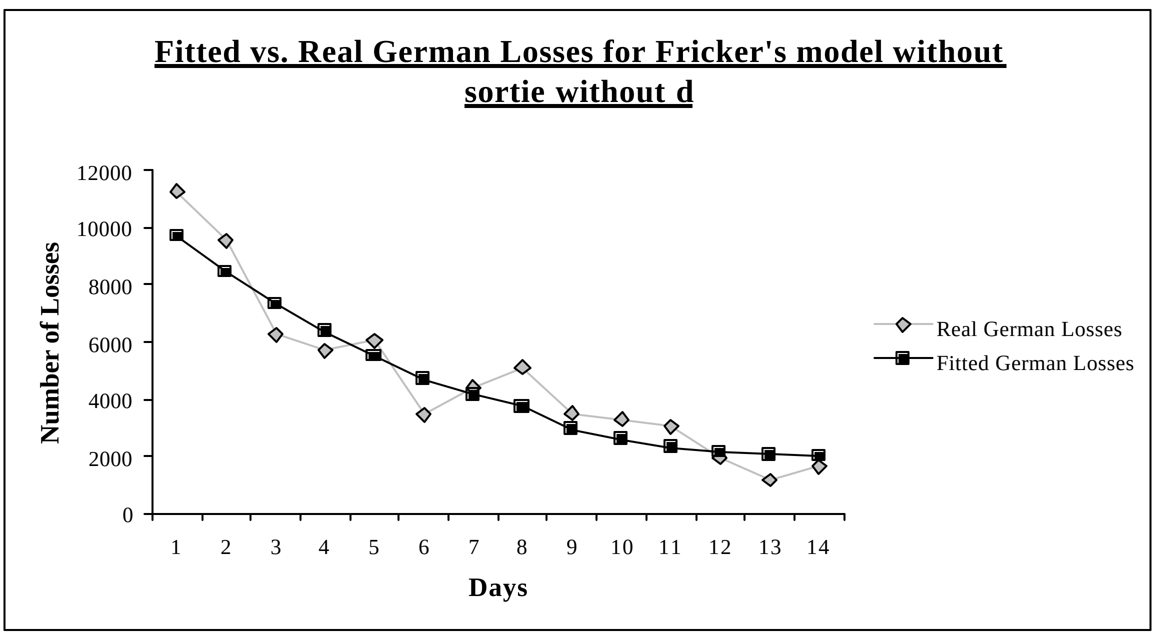

Figure 25. Fitted losses plotted versus real Geman losses for the German forces for Fricker's model without the air sortie data added and without using the *d* parameter.

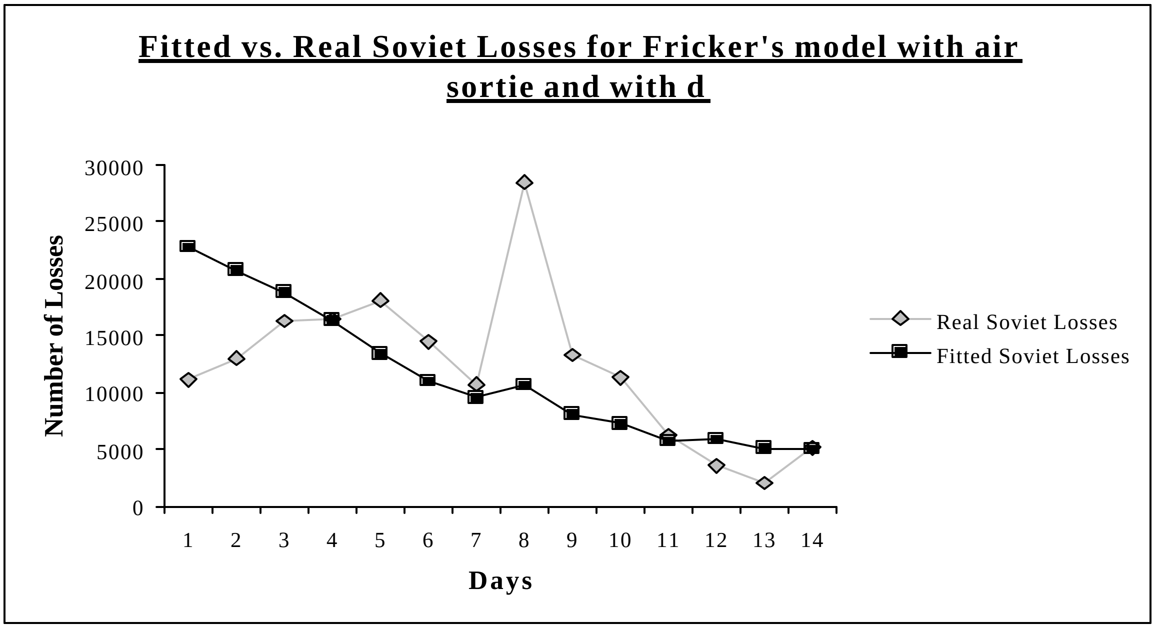

Figure 26. Fitted losses plotted versus real losses for Soviet forces for Fricker's model with the air sortie data added and using the *d* parameter.

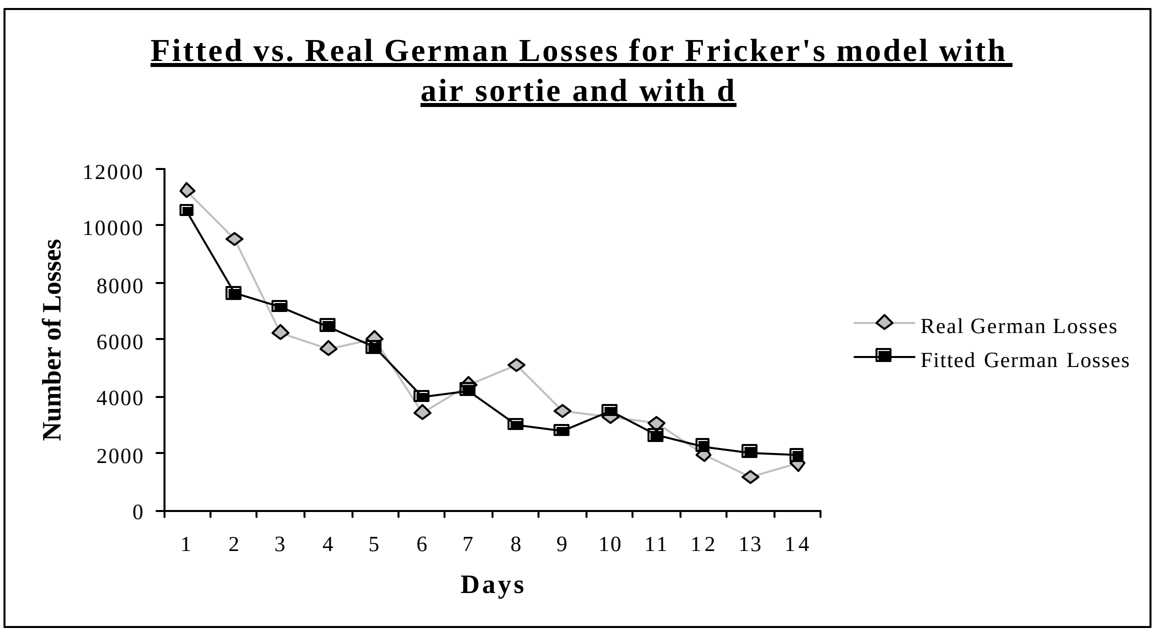

Figure 27. Fitted losses plotted versus real losses for German forces for Fricker's model with the air sortie data added and using the *d* parameter.

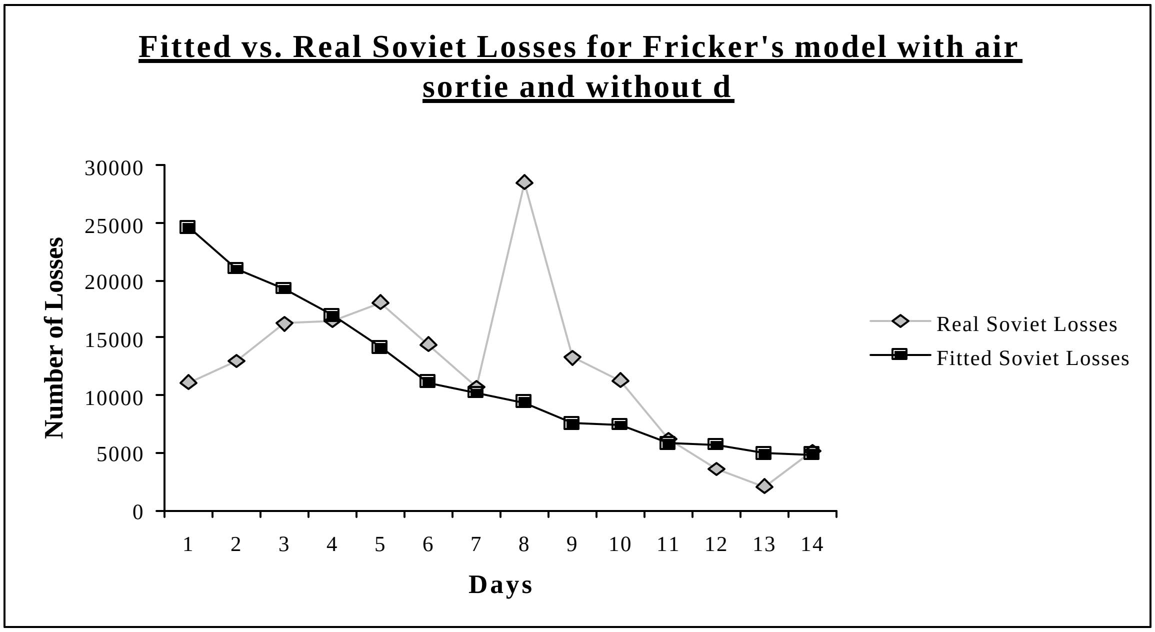

Figure 28. Fitted losses plotted versus real losses for the Soviet forces for Fricker's model with the air sortie data added and without using the *d* parameter.

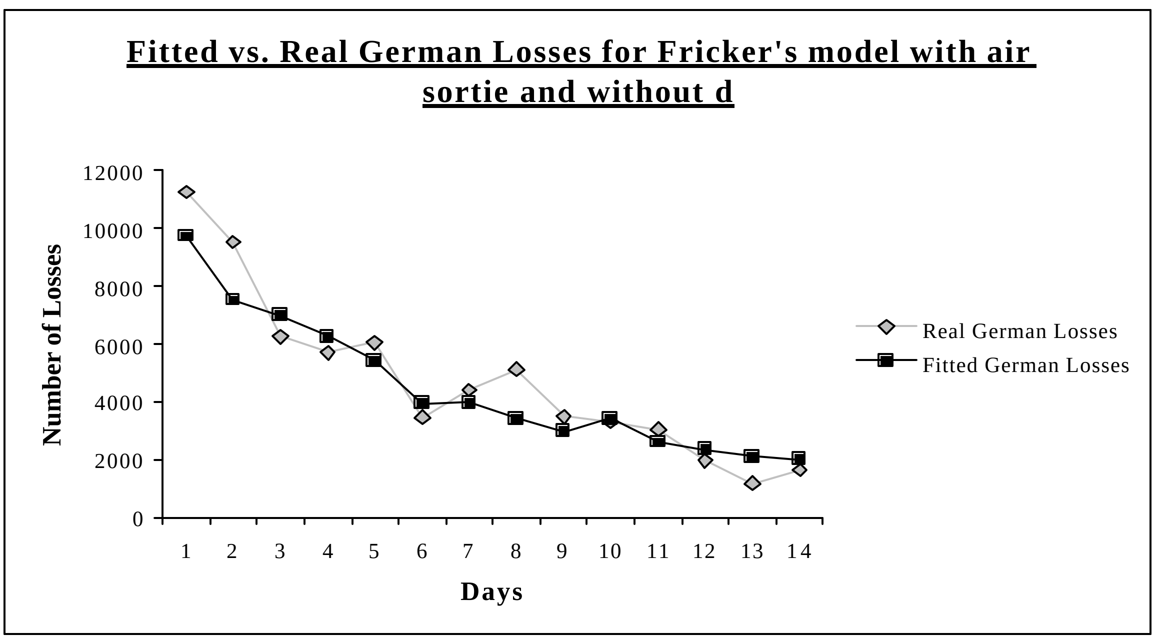

Figure 29. Fitted losses plotted versus real losses for the German forces for Fricker's model with the air sortie data added and without using the *d* parameter.

disadvantage. In both cases, when the tactical parameter is used, $a<b$, and when tactical parameter is not used, $a>b$. Again, in both cases, the a and b parameters are very small.

When the plots given in Figures 22, 24, 26 and 28 are examined, there appears to be three phases in the battle. It is also apparent that the battle lost its intensity after July 12. Notice the pattern where the model overestimates the beginning part and the last part of the battle while underestimating the 8 days in a row between these two parts. This pattern suggests that fitting a model with change points may improve the model's fit to the data. Also, the model provides a much better fit for the German side.

In equations IV.A.2.b.(14), IV.A.2.b.(15), IV.A.2.b.(18), IV.A.2.b.(19) and IV.A.2.b.(20), IV.A.2.b.(21) the q parameter is greater than the p parameter suggesting that one side's loss is more a function of his own forces rather than his opponent's forces. This finding is similar to what Fricker observed in his study.

In equations IV.A.2.b.(16), IV.A.2.b.(17) the p parameter is greater than the q parameter, which suggest that one side's loss is more a function of his opponent's forces rather than his own forces. This finding is different from Fricker's findings.

It is significant that using the tactical parameter d does not improve the fit for the model without the air sortie data when SSR values are compared. This may be interpreted as using the logarithmically transformed equations does not necessarily gives the best fit in the original form. Table 29 shows the results for Fricker's models as a whole for both the Ardennes and the Kursk data. The negative R^2 values found here imply that the fitted model yields worse results than using the average daily losses as an estimate. This finding was communicated with Fricker and it was concluded that the reason for the negative R^2 values are the combination of extreme sensitivity of the results to the precision of parameters and using the rounded off values given in Fricker's study [Ref.6]. For example, for the first model given in Table 29, changing the q parameter from 5.0 to 5.02 increases the R^2 value from –0.7938 to 0.1904, and changing the q parameter from 5.0 to 5.03 increases the R^2 value to 0.4581.

Name of the model	***a***	***b***	***p***	***q***	***d***	*SSR*	R^2
Ardennes w/o sorties with d	4.7E-27	3.1E-26	0	5	0.8093	1.57E+8	-0.7938
Ardennes w sorties with d	2.7E-24	1.6E-23	0	4.6	0.7971	2.64E+7	0.5256
Kursk w/o sorties with d	3.76E-33	1.09E-32	0.0604	6.3066	0.79	5.94E+8	0.1703
Kursk w/o sorties w/o d	1.61E-33	3.44E-33	3.6736	2.6934	-	2.16E+9	0.0657
Kursk with sorties with d	3.35E-27	5.76E-27	0.0955	5.2207	0.93	6.23E+8	0.1294
Kursk with sorties w/o d	5.01E-27	3.85E-27	1.4983	3.8179	-	7.16E+8	-0.0222

Table 29. Fricker's results for his models with/without the tactical parameter *d*, with/without the air sortie added, for both the Ardennes and the Kursk data.

B. EXPLORATORY ANALYSIS OF BATTLE OF KURSK DATA

The fighting on the first day of the battle was sporadic. The extremely low casualty levels represent large outliers; this, including the data of the first day could drastically effect the outcome of the analysis. Thus, the first day of data was dropped in fitting the data to the models. This kind of approach is also supported by the historical account of the Battle of Kursk, because the main offensive did not really begin until July 5, the second day of the battle. Even if there are other days on which large outliers are observed—like July 12—these outliers will not be left out of the analysis as they are a result of the fighting during the Kursk offensive. Therefore, this study will fit only the last 14 days of the aggregated data given in Table 14, excluding the first day. All the results found from the models are summarized as a whole in Table 42 in Section IV.B.10.

1. The scalar aggregation models

Two numerical methods are used to fit parameters to the scalar model of Lanchester equations. One is linear regression and the other is robust LTS regression. Robust LTS regression method performs least-trimmed squares regression [Ref.17]. When the given data in hand contains significant outliers as in our case, robust regression models are useful for fitting linear relationships by discounting outlying data. Both methods minimize the sum of squared residual (SSR) error resulting from the model to the actual data.

a. Linear regression

Linear regression is used for fitting parameters to the logarithmically transformed Lanchester equations. The original form of Lanchester equations are given in equations I.A.(1) and I.A.(2). By taking the logarithm of each side of the equations, we get:

$$\log(\dot{B}) = \log(a) + p\log(R) + q\log(B) \tag{22}$$

$$\log(\dot{R}) = \log(b) + p\log(B) + q\log(R) \tag{23}$$

Only the last 14 days of the data given in Table 19 are used for performing the linear regression analysis.

b. Results of the linear regression model

Results of the linear regression model which gives an SSR value of 6.36×10^{8} and an R^2 value of 0.1126 are:

$$\dot{B} = 1.06\times10^{-47} R^{5.7475} B^{3.3356} \tag{24}$$

$$\dot{R} = 1.90\times10^{-48} B^{5.7475} R^{3.3356} \tag{25}$$

c. Robust LTS regression

We will use Robust LTS regression for fitting parameters to the Lanchester equations. The original form of Lanchester equations are given in equations I.A.(1) and I.A.(2). By taking the log of each side of the equations, we obtain the equations given in IV.A.2.(22) and IV.A.2(23). Only the last 14 days of the data given in Table 19 are used for doing the robust LTS regression analysis.

d. Results of the robust LTS regression

Results for the robust LTS regression model which gives an SSR value of 5.54×10^{8} and an R^2 value of 0.2262 are:

$$\dot{B} = 2.27\times10^{-40} R^{6.0843} B^{1.7312} \tag{26}$$

$$\dot{R} = 1.84\times10^{-41} B^{6.0843} R^{1.7312} \tag{27}$$

Figures 30 and 31 show fitted losses plotted versus real losses for the Soviet and the German forces, respectively, for the linear regression model. Figures 31 and 32 show fitted losses plotted versus real losses for the Soviet and the German forces, respectively, for the robust LTS regression model.

When the SSR values found by using linear regression and robust LTS regression techniques are compared, it is observed that using the robust LTS regression technique improves the fit for the Battle of Kursk data. The SSR value, which is found by using the robust LTS regression method, is the smallest for the Kursk data so far.

It should be noted that even if the robust LTS regression technique accounts for the outliers when finding the parameters that minimize the SSR for a given model, the

SSR values computed here include the SSR of the outliers. In other words, when the parameters computed by the robust LTS regression technique are used in the analysis, the

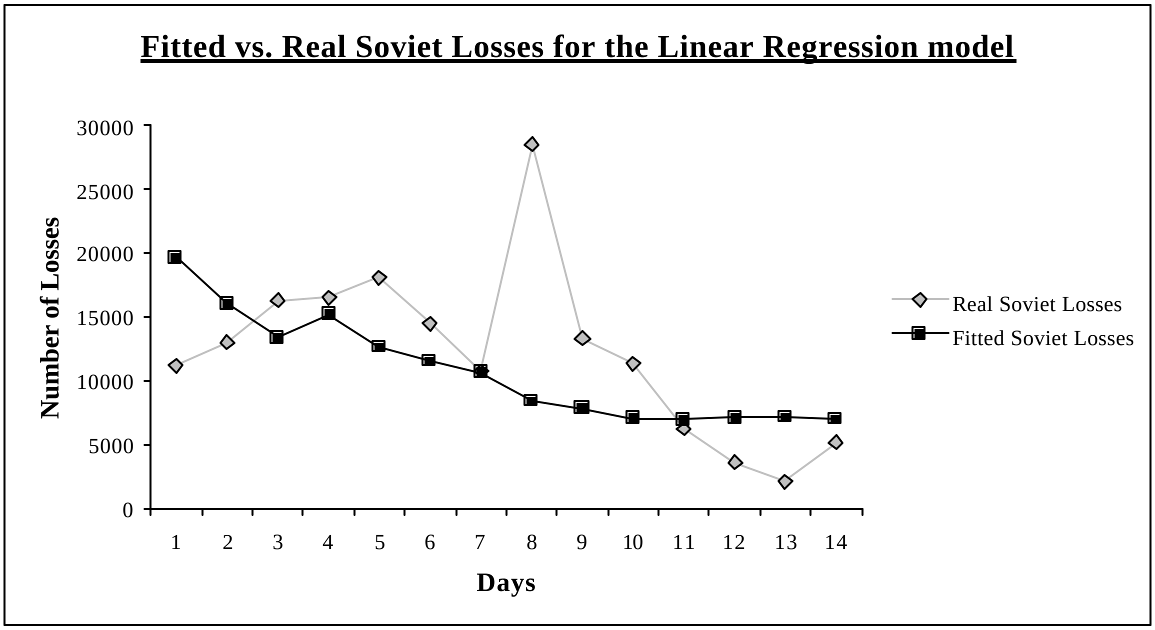

Figure 30. Fitted losses plotted versus real losses for Soviet forces for the linear regression model. The significant outlier on day 8 influences the fit dramatically.

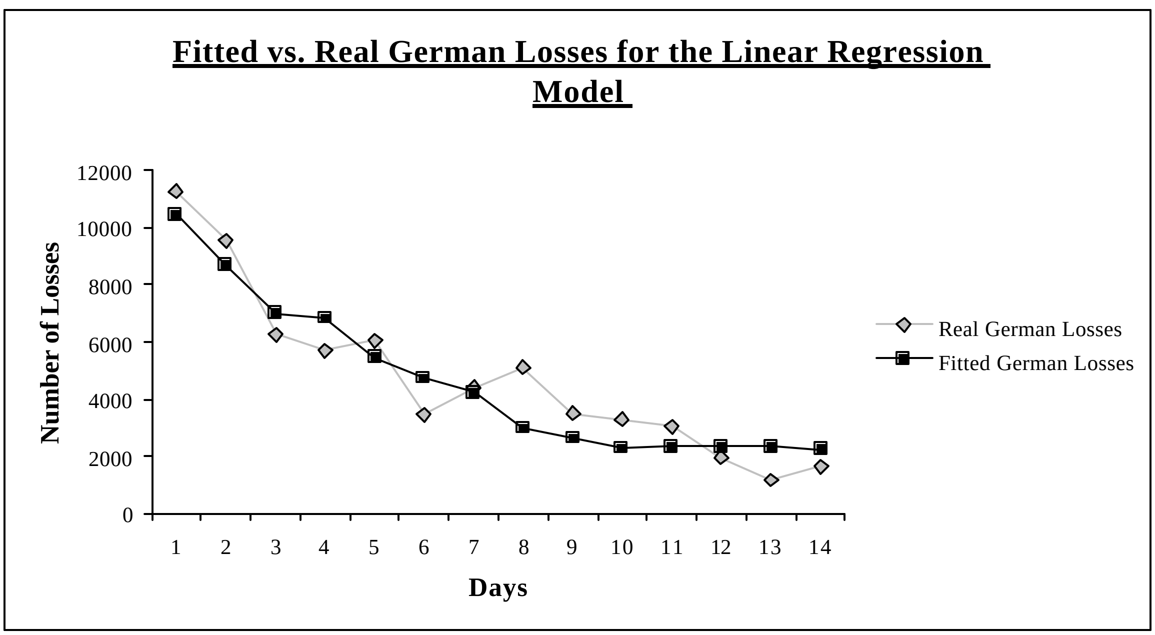

Figure 31. Fitted losses plotted versus real losses for German forces for the robust LTS regression model. The data for the German side, with no significant outliers, gives a better fit for the model.

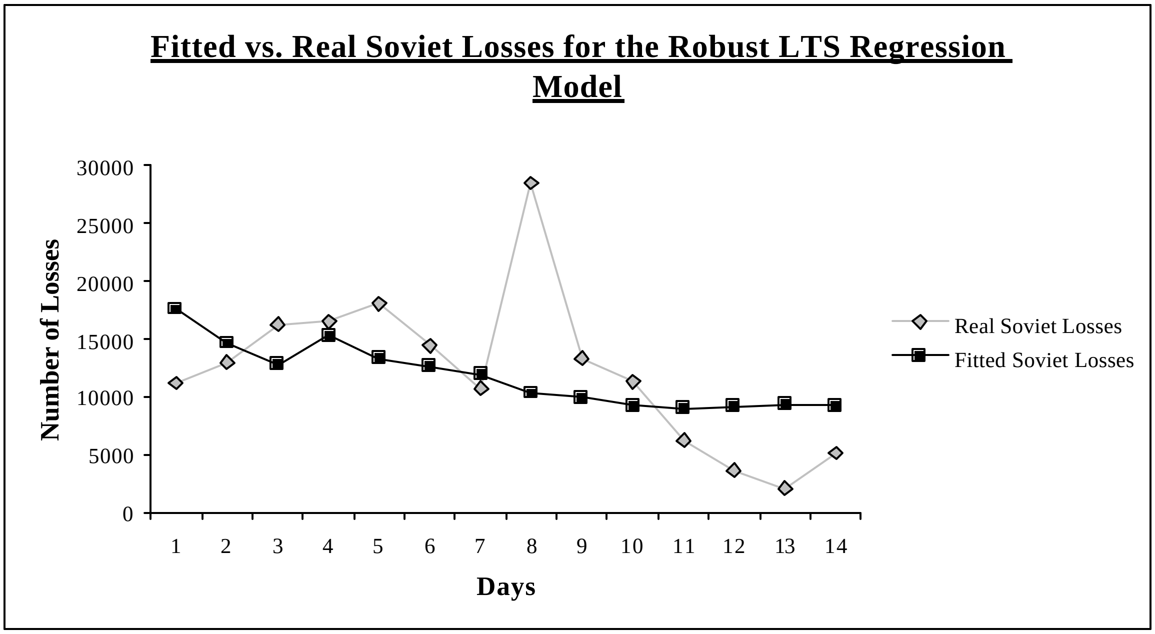

Figure 32. Fitted losses plotted versus real losses for Soviet forces for the robust LTS regression model. The significant outlier on day 8 influences the fit dramatically.

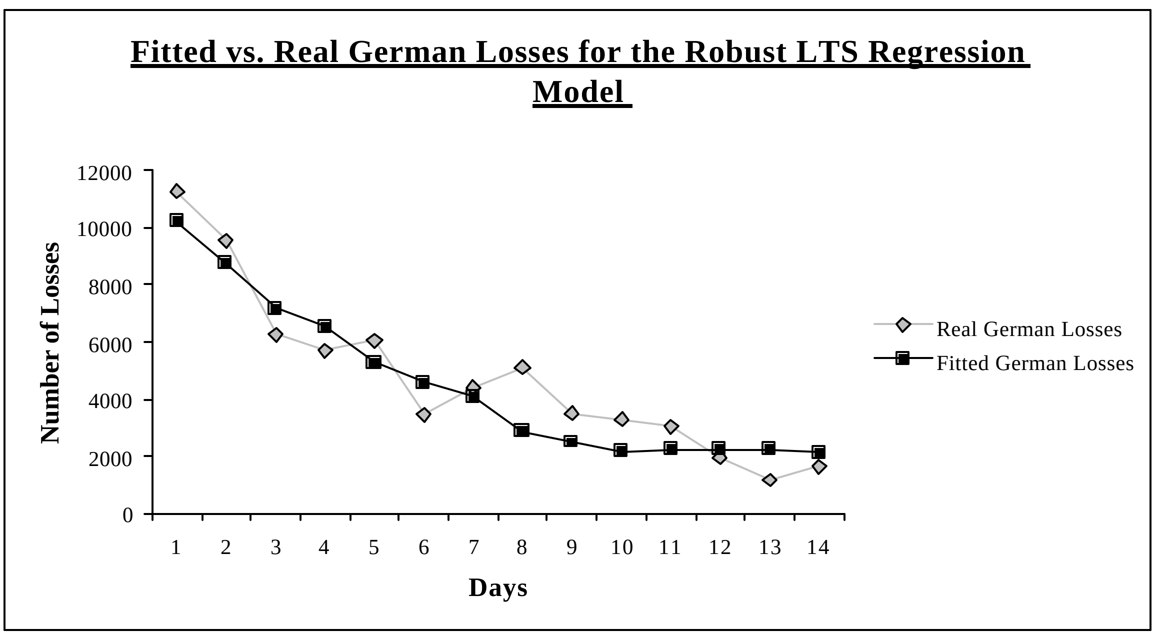

Figure 33. Fitted losses plotted versus real losses for the German forces for the robust LTS regression model. The data for the German side, with no significant outliers, gives a better fit for the model.

outliers are not discounted. They are included for the purpose of computing the SSR value.

When the p and q parameters are compared it is noticed that the p parameter is greater than the q parameter, suggesting that one side's loss is more a function of his opponent's forces rather than his own forces. This interpretation is different from what Fricker found in his study. In both cases, a and b parameters are significantly small, and a > b.

When the plots given in Figures 30 and 32 are examined, there appears to be three distinct phases in the battle. It is also apparent that the battle lost its intensity after July 12. After the Soviets went into offense, the battle was not as intense. There is a clear pattern in Figure 30 where the model overestimates the beginning part and the last part of the battle, while underestimating the attrition for eight days in a row between these two periods.

The pattern seen in Figures 30 and 32 suggests that fitting a model with change points may improve the model's fit to the data. Likewise, leaving out the data given for July 12 when the most intense fighting of the battle took place, it may also be possible to increase the fit to the data, an approach which will be covered in upcoming sections. Also, the model provides a much better fit for the German side.

2. Including air sortie data

As mentioned in IV.A.2, the air sortie data given in the KOSAVE study [Ref.12] consists of the number of air-air role sorties, ground attack role sorties, reconnaissance role sorties and evacuation role sorties (which are solely used by Germans). For aggregating the air sortie data into total aggregated number of forces, we will use the data given in Table 27 that presents data on the number of ground attack role sorties. However, the aggregated data will be different than that given in Table 28, because the data in Table 28 is calculated using the reformatted data by applying Fricker's algorithm.

The data, which we will be using in this section, is given in Table 30 which presents the total number of aggregated forces, including the air data by weighing each sortie by 30. In other words, the number of air sorties presented in Table 27 is multiplied by 30 and added onto the aggregated force levels given in Table 19 in order to compute the data presented in Table 30.

Two regression methods, presented in IV.B.1 are used for fitting the data given in Table 30, namely, linear regression and robust LTS regression.

day	Blue Forces	Blue Losses	Red Forces	Red Losses
1	604353	11167	431671	11257
2	594159	12993	404945	9532
3	579175	16266	404055	6249
4	565402	16472	415304	5702
5	542712	18071	406024	6043
6	527893	14445	382404	3450
7	518016	10754	389340	4415
8	498123	28492	375765	5112
9	487961	13302	375759	3491
10	480724	11323	394230	3290
11	474229	6201	373752	3047
12	482881	3600	367286	1975
13	471266	2067	363905	1174
14	469253	5160	360820	1639

Table 30. Data on aggregated forces. Forces are combat manpower, APCs, tanks, artillery and number of ground-attack role sorties which are weighted by 1, 5, 20, 40 and 30, respectively.

a. Results of linear regression model

Results for the linear regression model, which gives an SSR value of 6.85×10^8 and an R^2 value of 0.0433, are:

$$\dot{B} = 1.40 \times 10^{-35} R^{5.1323} B^{1.7793} \quad (28)$$

$$\dot{R} = 2.09 \times 10^{-36} B^{5.1323} R^{1.7793} \quad (29)$$

b. Results of robust LTS regression model

Results for the robust LTS regression model, which gives an SSR value of 7.58×10^8 and an R^2 value of -0.0579, are:

$$\dot{B} = 1.21 \times 10^{-38} R^{5.3691} B^{2.0883} \quad (30)$$

$$\dot{R} = 1.75 \times 10^{-39} B^{5.3691} R^{2.0883} \quad (31)$$

Figures 34 and 35 show the fitted losses plotted versus real losses for the Soviet and the German forces respectively, for the linear regression model with the air sortie data added.

Figures 36 and 37 show the fitted losses plotted versus real losses for the Soviet and the German forces respectively, for the robust LTS regression model with the air sortie data added.

Following the aggregation of the data using the number of air sorties, it is not appropriate to compare the models using the SSR values because, the increase in the SSR value may be a natural result of adding the air sortie data. For this reason, R^2 values will be used to compare the fit of the model.

Upon the examination of the R^2 values above, which are found by applying linear regression and robust LTS regression techniques to the logarithmically transformed data that includes air sorties, one can determine that considering the air sorties data does not improve the model's fit to the data. The R^2 values, which are found by using the linear regression and the robust LTS regression technique, are both lower than the R^2 values

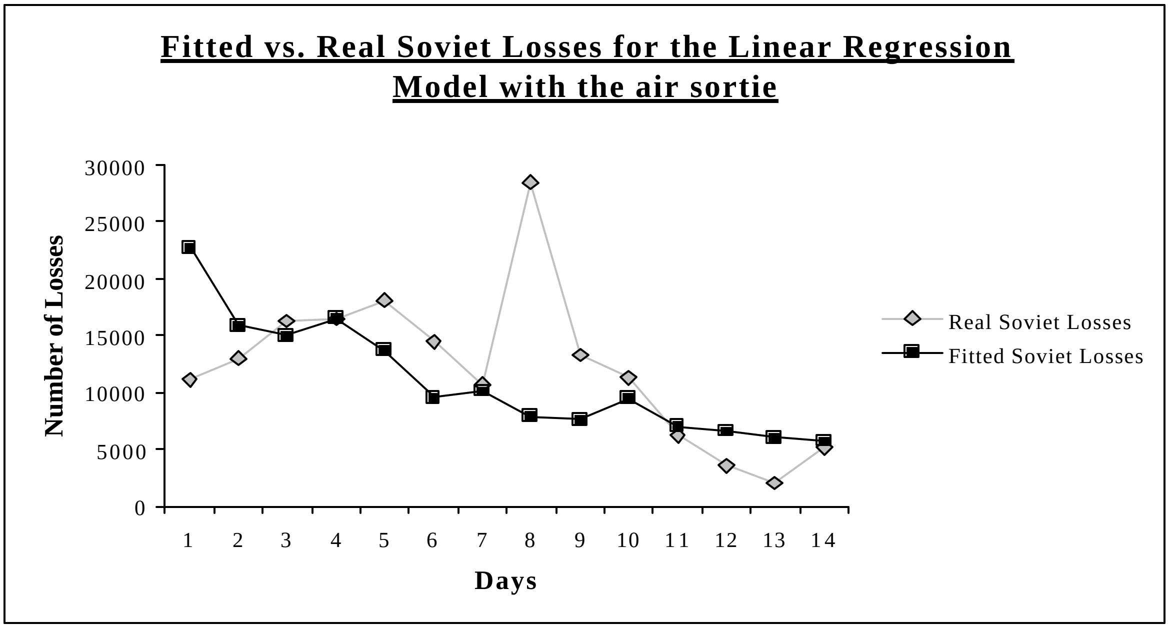

Figure 34. Fitted losses plotted versus real losses for Soviet forces for the linear regression model with the air sortie data added. The significant outlier on day 8 influences the fit dramatically. The same pattern where the model over/underestimates the battle in three distinctive phases is observable in this plot too.

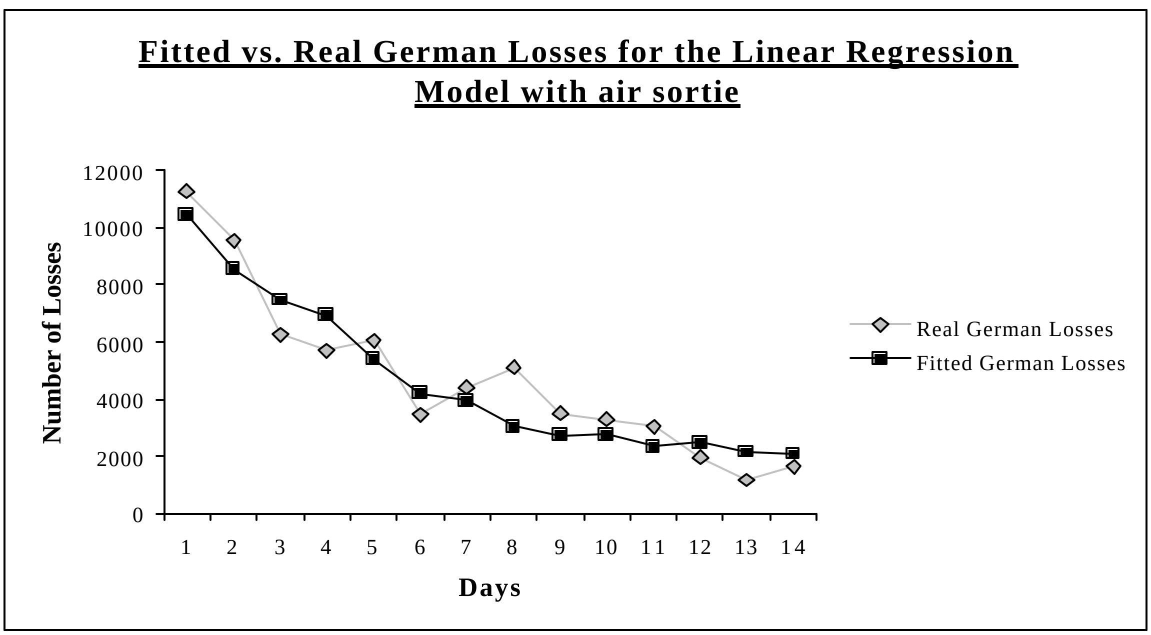

Figure 35. Fitted losses plotted versus real losses for German forces for the linear regression model with the air sortie data added. The data for the German side, with no significant outliers, gives a better fit for the model.

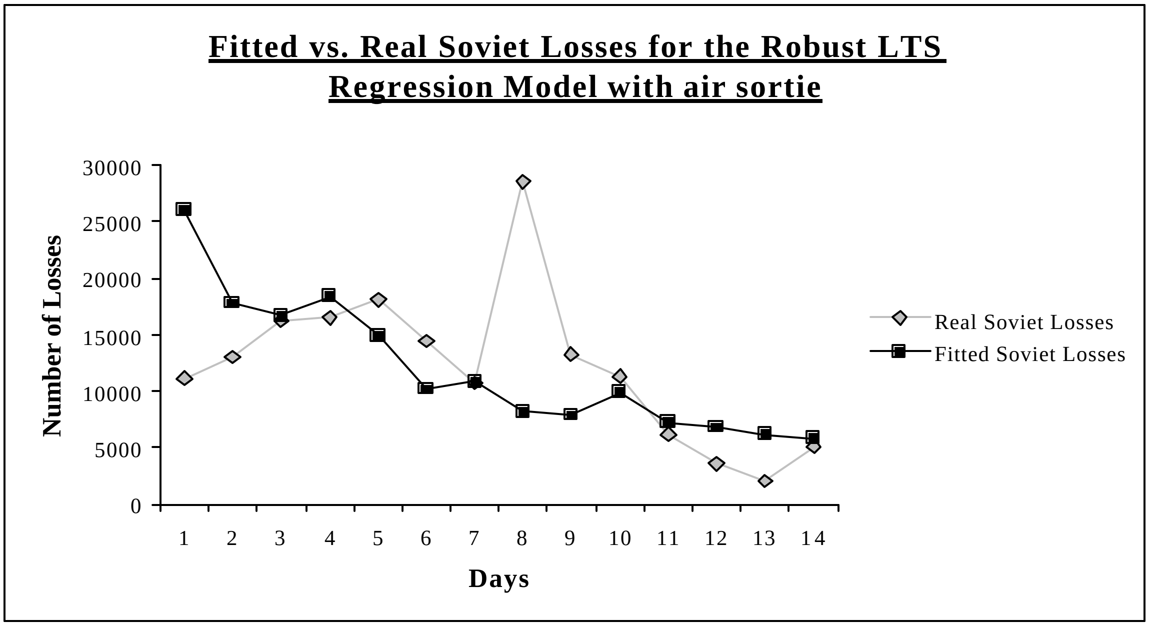

Figure 36. Fitted losses plotted versus real losses for the Soviet forces for the robust LTS regression model with the air sortie data added. The significant outlier on day 8 influences the fit dramatically. The same pattern where the model over/underestimates the battle in three distinctive phases is observable in this plot too.

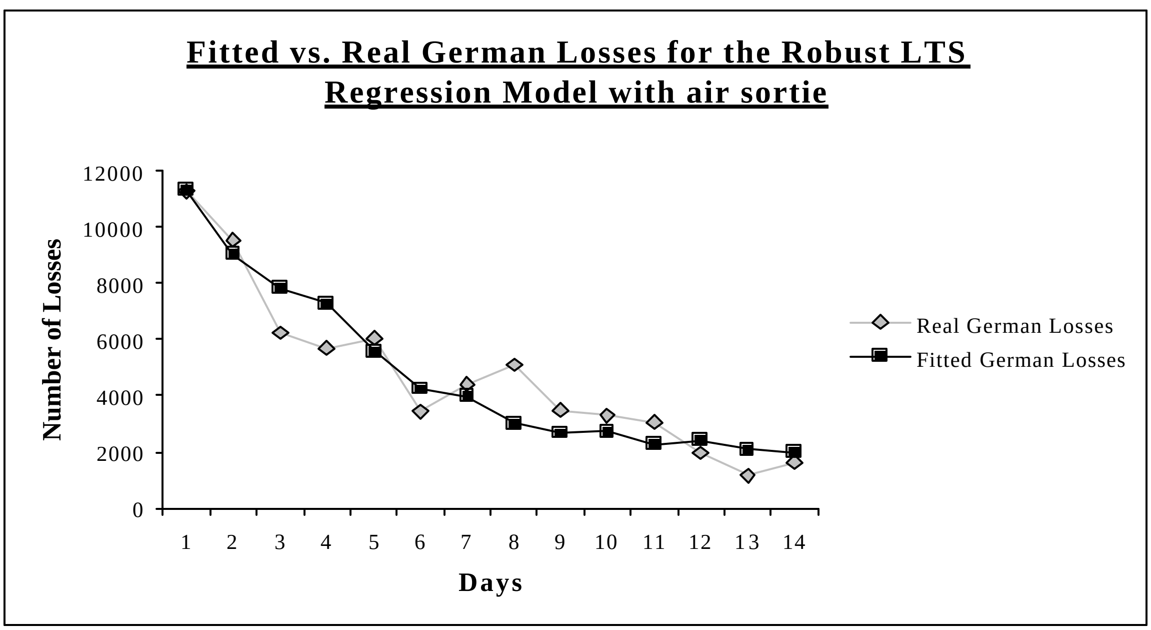

Figure 37. Fitted losses plotted versus real losses for the German forces for the robust LTS regression model with the air sortie data added The data for the German side, with no significant outliers, gives a better fit for the model.

found in Section IV.B.1 which did not include the air sortie data. While the previous best fit found was 0.2262 in section IV.B.1.d, after the air sortie data is added, R^2 is found to be 0.0433 and -0.0579. Adding air sortie data did not improve model's fit to the data.

For both cases, the a and b parameters are significantly small and $a > b$. This suggests that individual German effectiveness was greater than individual Russian effectiveness.

When the p and q parameters are compared it is observed that the p parameter is greater than the q parameter, indicating that one side's losses are more a function of his opponent's forces rather than being a function of his own forces. This result is different from what Fricker found in his study.

When the plots given in Figures 34 and 36 are examined, the resulting pattern is similar to the one seen in the previous section. This pattern again suggests that fitting a model with change points may improve the model's fit to the data. Again, similar to the previous results, it may be possible to increase the fit to the data by leaving out the data given for July 12.

3. Taking into account the change in offensive/defensive roles

By historical account, the German forces generally maintained an offensive posture (this is not valid for all units on the battlefield) through July 12, when the Soviets were able to gain the initiative and launch their counter-offensive. Bracken [Ref.13] introduced an additional parameter d to the standard Lanchester equations (I.B.(1) and I.B.(2)), called a *tactical parameter*, to account for a battle in which defense and offense switch during the course of the campaign.

With d for the defender and $(1/d)$ for the attacker, the Lanchester equations are modified to accept the tactical parameter d and are given as:

$$\dot{B} = (d \text{ or } 1/d)\, aR^{p} B^{q} \qquad (32)$$

$$\dot{R} = (1/d \text{ or } d)\, bB^{p} R^{q} \qquad (33)$$

The logarithmically transformed Lanchester equations which are modified to accept the tactical parameter (for the days that red is the attacker), are given as:

$$\log(\dot{B}/d) = \log(a) + p\log(R) + q\log(B) \qquad (34)$$

$$\log(\dot{R}/(1/d)) = \log(b) + p\log(B) + q\log(R) \qquad (35)$$

Linear regression and robust LTS regression models are used to estimate the model parameters represented above in IV.B.3.(34) and IV.B.3.(35).

a. Linear regression

The last 14 days of the aggregated data given in Table 14 in section IV.A.1 and the S-PLUS software are used to estimate the model's parameters, which minimize the sum of squared residuals of the actual and estimated attrition.

In order to iterate for different d values, linear regression is fit for multiple d values, and then the d value that gives the minimum SSR is selected. The value of tactical parameter d is varied between 0.0 and 9.0 in increments of 0.01.

b. Results of the linear regression model

Results for the linear regression model which gives an SSR value of 6.24×10^{8} and a tactical parameter value of 1.17 and an R^2 value of 0.1295 are:

$$\dot{B} = (\frac{1}{1.17} \text{ or } 1.17)1.88\times10^{-47} R^{7.5038} B^{1.5793} \qquad (36)$$

$$\dot{R} = (1.17 \text{ or } \frac{1}{1.17})1.07\times10^{-48} B^{7.5038} R^{1.5793} \qquad (37)$$

c. Robust LTS regression

For estimating the parameters, which minimize the sum of squared residuals of the actual and estimated attirition, the last 14 days of the aggregated data given in Table 5, in Section IV.A.1 and the S-PLUS software are used.

d. Results of the robust LTS regression

Results for the robust LTS regression model which gives an SSR value of 5.54×10^{8} and a tactical parameter value of 1.00 and an R^2 value of 0.2262 are:

$$\dot{B} = (\frac{1}{1.0} \text{ or } 1.0)2.27\times10^{-40} R^{6.0843} B^{1.7312} \tag{38}$$

$$\dot{R} = (1.0 \text{ or } \frac{1}{1.0})1.84\times10^{-41} R^{6.0843} B^{1.7312} \tag{39}$$

Figures 38 and 39 shows the fitted losses plotted versus real losses of the Soviet and the German forces, respectively, for the linear regression model.

Figures 40 and 41 shows the fitted losses plotted versus real losses of the Soviet and the German forces, respectively, for the robust LTS regression model.

When the SSR values above are examined, it is apparent that taking into consideration the change in offensive/defensive roles improves the fit. The SSR values, which are found by using the linear regression and robust LTS regression technique, are both less than or equal to the SSR values found in section IV.B.1, which did not consider the change in offensive/defensive roles. The best fit found in section IV.B.1 was 6.36×10^{8} for the linear regression model, after the d parameter is included in the model, SSR value is found to be 6.24×10^{8}, suggesting only a 2% improvement in fit. But, this is not the case for robust LTS regression model. While the previous result for robust LTS regression model was found to be 5.54×10^{8} in Section IV.B.1.d, after the change in

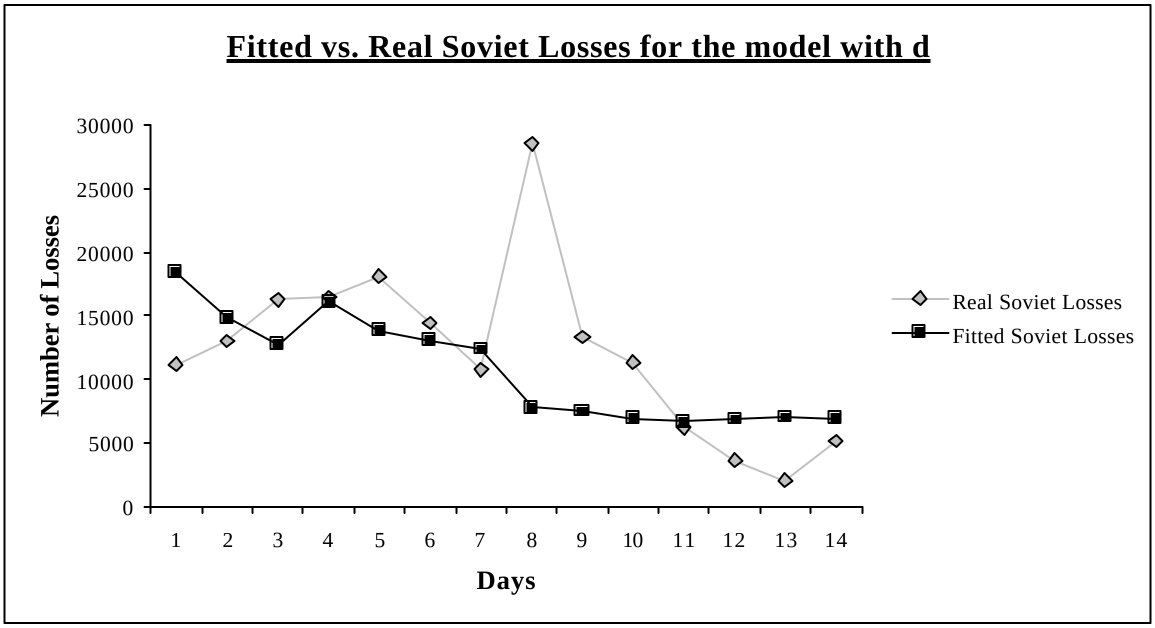

Figure 38. Fitted losses plotted versus real losses for the Soviet forces for the linear regression model with the tactical parameter *d*. The significant outlier on day 8 influences the fit dramatically.

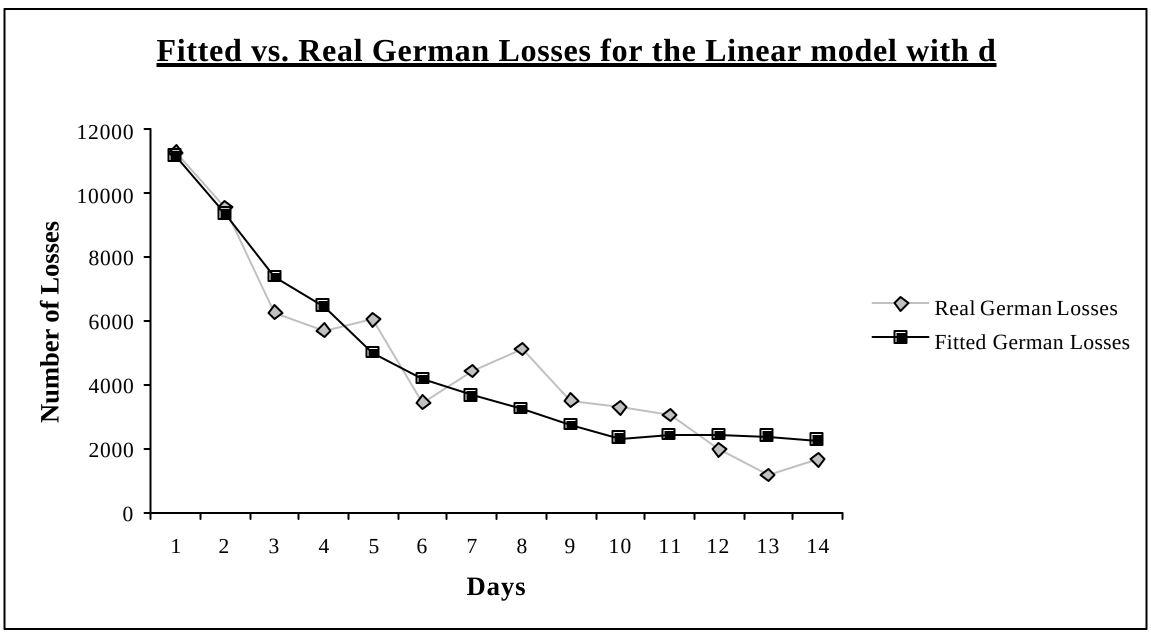

Figure 39. Fitted losses plotted versus real losses for the German forces for the linear regression model with the tactical parameter *d*. The data for the German side, with no significant outliers gives a better fit for the model.

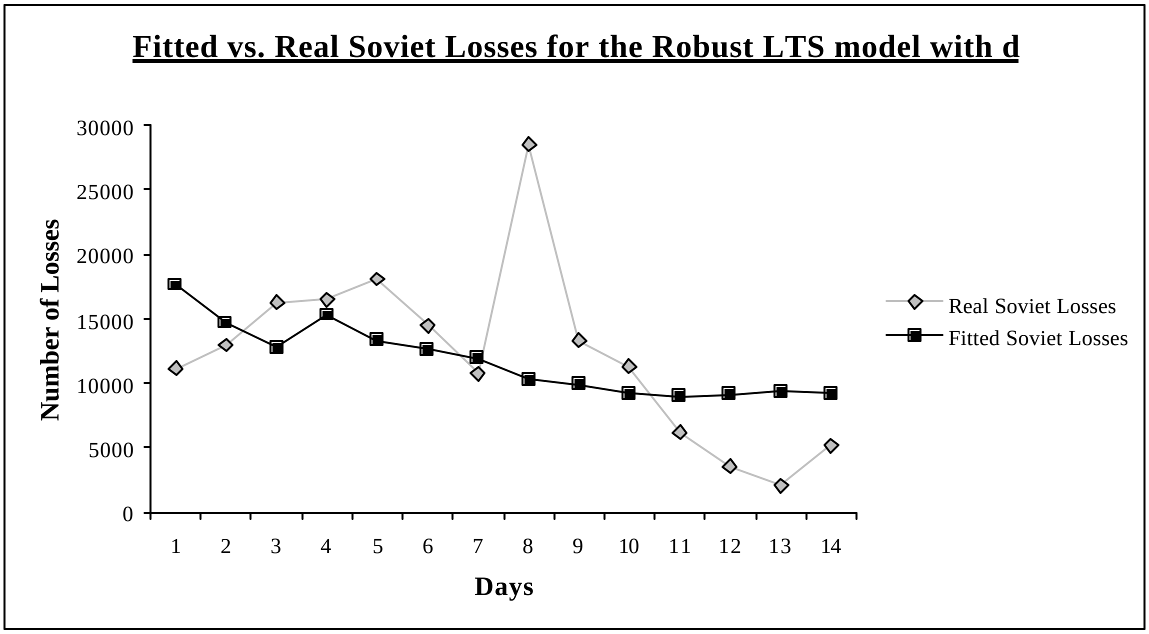

Figure 40. Fitted losses plotted versus Real losses for the Soviet forces for the robust LTS model with the tactical parameter *d*.

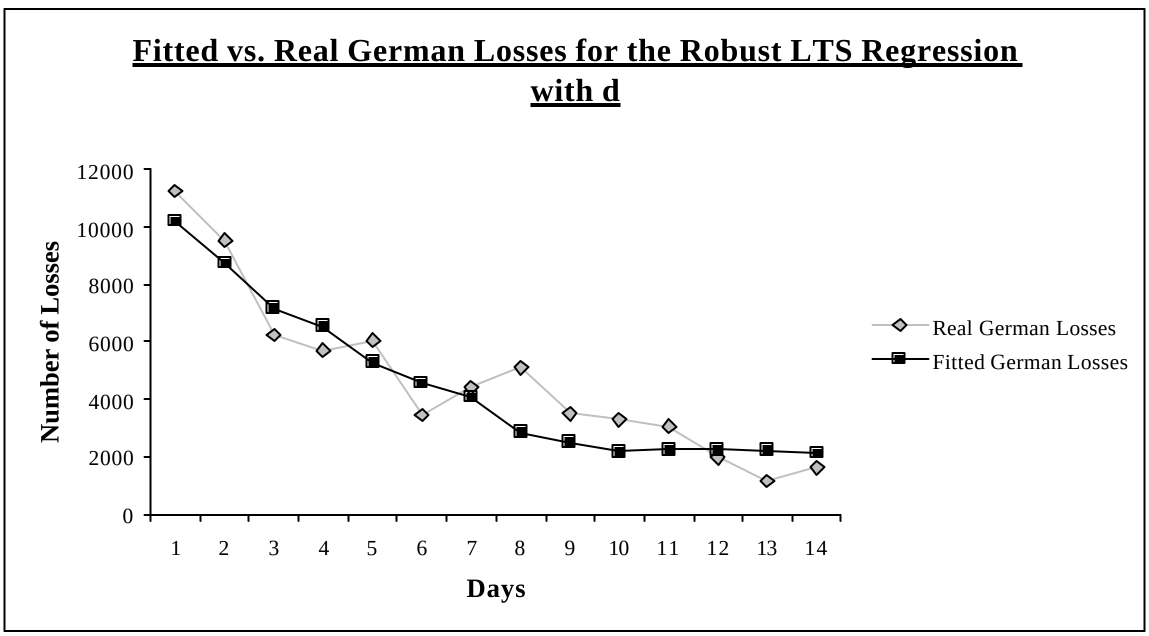

Figure 41. Fitted losses plotted versus Real losses for the German forces for the robust LTS model with the tactical parameter *d*.

offensive/defensive roles is taken into account, it is again found to be 5.54×10^8. In other words, taking into account the change in offensive/defensive roles does not change the fit for the robust LTS regression model.

Following a search of the tactical parameter *d* value, performed in increments of 0.01, 1.0 is found to be the optimal *d* value that gives the smallest SSR value for the robust LTS regression model. This result indicates that in the context of the Battle of Kursk, one side's status as the defender or attacker does not affect the number of losses which either of the sides is going to suffer. This reasoning may not intuitively make sense, but further analysis made in the following sections will provide additional rationale.

For both cases, the *a* and *b* parameters are significantly small, and $a>b$. This suggests that individual German effectiveness is greater than individual Russian effectiveness.

The *d* parameter with a value of 1.17 signifies that the attacker has an advantage. This result is somewhat unexpected and implies that it is the attacker who will suffer fewer casualties. (The *d* parameter is investigated more closely in upcoming sections).

When *p* and *q* parameters are compared, it is observed that the *p* parameter is greater than the *q* parameter, suggesting that one side's losses are more a function of the opponent's forces rather than a function of its own forces. This finding is different from what Fricker found in his study.

When the plots given in Figures 38 through 41 are examined, the pattern seen in these plots are similar to the results observed in the previous section. This pattern, again, suggests that fitting a model with change points may improve the fit and again the models fit better for the Germans.

4. Considering the tactical parameter *d* of the campaign

The findings in the previous sections suggest that fitting models with different d values for separate phases of the battle might improve the fit to the data and this section focuses on that aspect of our findings and will analyze the battle in separate time periods.

The tactical parameter found in the previous section, $d=1.17$, is similar to Bracken's [Ref.8] findings which also implied an attacker advantage. Since $d>1$, implying that if Blue is defending, then blue has a defender disadvantage, and if red is attacking when $d>1$, then red has an attacker advantage. This intuitively does not make much sense because the defender is usually dug in, and the attacker is out in the open and easily detected by the enemy. It should be the defender who has the advantage rather than the attacker when attrition rates are considered. In this situation, it may not make sense to have only one d for the whole campaign.

A closer look at the battle data may find a better fit for the model. The very first day of the battle, the Germans run into the heavily fortified Soviet positions and minefields and have a very rough day. This first day, the Germans obviously have an attacker disadvantage, while the Russians have a defender advantage. July 6, 1943 is the day when things begin to run smoothly for the Germans, as they are not up against a fortified defense, dense barriers and minefields. This scenario continues until July 12, when the Soviets launch their counter-attack. Even on that day, the Germans were not aware of the Soviets' intention to make such a move [Ref.16]. July 12, 1943 can be viewed as the day, when neither side was a defender. Both sides attacked each other resulting in the bloodiest day of the campaign. The Soviets especially suffered heavy casualties. From July 13 on, the Soviets continued their counter-attacks until they recaptured the ground they had lost. During this time Germans use a hasty defense.

This type of approach is also justified by the historical account of the battle, which is explained in detail in [Ref.15] and [Ref.16]. As a result of the clearly defined phases of the battle, the data will be handled in four different time periods. A different d value will be used for each part of the campaign (i.e. there will be four different d parameters for the campaign). A weakness of this approach is the fact that it requires fitting 8 parameters with 14 days of data.

- *First period* July 5: Germans attack heavily fortified Soviet positions.
- *Second period* July 6-July 11: Germans continue a more organized attack.
- *Third period* July 12: Soviets counterattack when Germans were continuing their attack.
- *Fourth period* July 13-July 18: Soviets attack and Germans make a hasty defense).

A different d parameter is fit to each of the four parts of the campaign using the same a, b, p, q parameters shown in equations IV.B.3.b.(34) and IV.B.3.b.(35) for the data in Table 19. This will be referred to as Model 1 for this section. The results are as follows.

The first period had the smallest SSR value when d=0.91. The second period had the smallest SSR value when d=1.24. The third period is considered to have the tactical parameter d=1 because there was no defender during the third period. The fourth period had the smallest SSR value when d=1.17.

The interpretation of the *d* values found is that the *d* value of 0.91 for the first period (i.e. the defender having the advantage), definitely makes sense because the Germans were attacking against the heavily fortified Soviet positions, and as a result, the Soviets inflicted heavier casualties on the Germans than the Germans did on the Soviets.

By intuition, it is likely that Soviets will continue to have the defender advantage through the second period as well. But this is not the case, since d=1.24, meaning that even if it were the Germans attacking they were more advantageous than the Soviets who were in their defensive postures. That is, it was the Soviets who were losing more.

The third period is considered to be the day that neither side is defending, so no interpretation is needed.

The fourth period has a *d* value of 1.17, which again indicates an attacker advantage. The value 1.17 indicates a slightly smaller attacker advantage than the Germans had during the second period. The Soviets had an attacker advantage during the fourth period, but not one so great as the Germans had during the second period.

The SSR values of the first, second, third and fourth periods mentioned above are 1.93×10^7, 3.70×10^7, 3.83×10^8, 9.53×10^7, respectively. The overall sum of the SSR values is 5.34×10^8 for the whole campaign, which gives a 4% better fit than the previous results. Figures 42 and 43 show the fitted versus real losses for the Soviet and German forces, respectively, for Model 1.

Overall, these results interpreted above indicate that for the Battle of Kursk, other than on the first day, it was always advantageous to be the attacker.

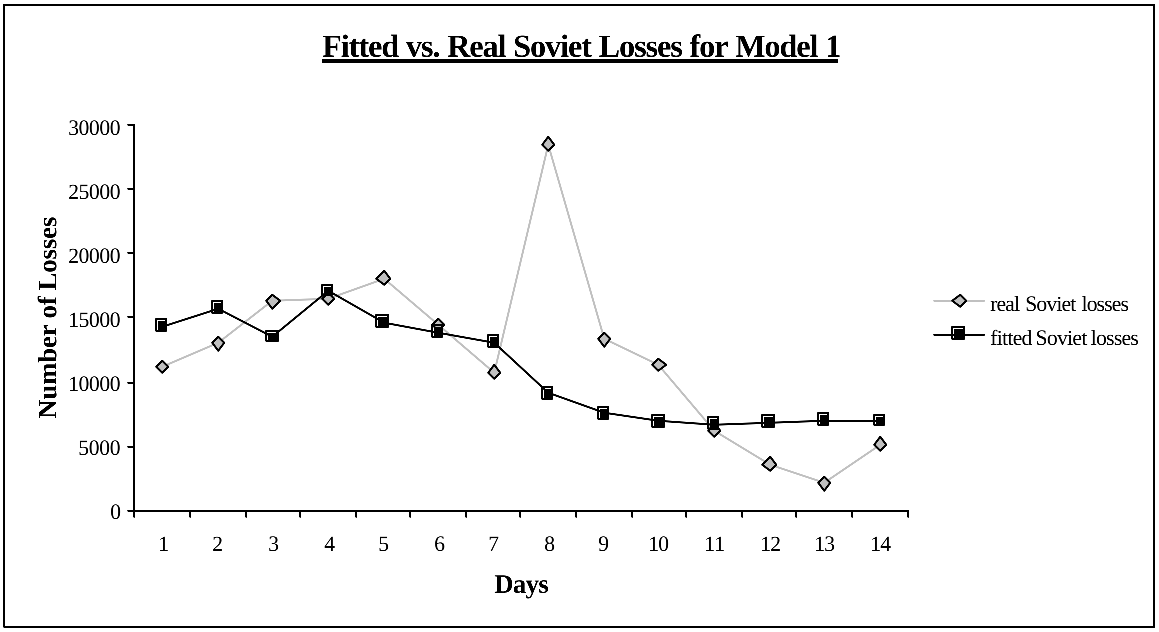

Figure 42. Fitted losses plotted versus real losses for the Soviet forces for model 1, which has four periods, and d=1 for the 8th day of the battle.

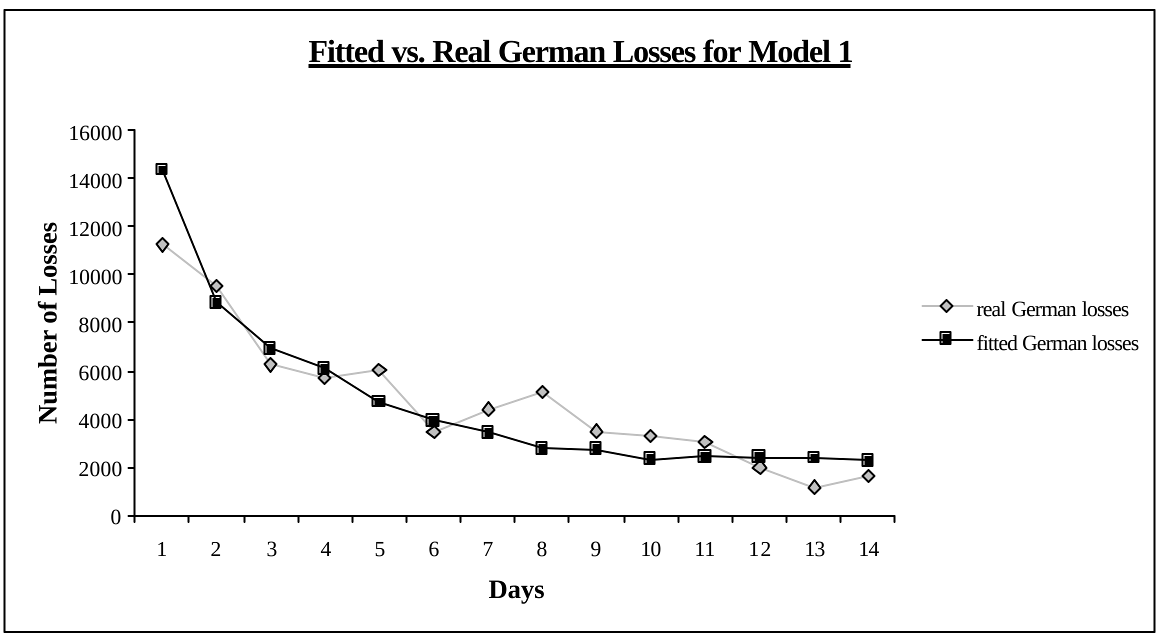

Figure 43. Fitted losses plotted versus real losses for the German forces for model 1, which has four periods, and d=1 for the 8th day of the battle.

One could argue that the third period, having no tactical parameter, does not make sense. If the third period is considered to have a tactical parameter of its own that is independent from the others, assuming that it was the day on which Soviets attacked, it is found to be *d*=0.32. This result obviously indicates an absolute defender advantage for the Germans and attacker disadvantage for the Russians. This will be referred as Model 2 for this section. In such an approach, the SSR value for the third period will be 1.78×10^{7} giving an overall SSR value of 1.69×10^{8} — almost a 70% better fit than the result found for Model 1 above. This is a much better fit because the biggest outlier now has its own unique *d* parameter, and is essentialy removed. This is also a clear indication of the tremendous effect of one outlier on the fit of the models. Figures 44 and 45 show the fitted versus real losses for the Soviet and German forces, respectively for Model 2.

Based on the results above, it can be concluded that considering the campaign in four different parts definitely helps to find a better fit. So, for combat modeling purposes, the tactical parameter values should depend on the situation of the battle.

Another approach is to leave out only the data for July 12, and not to divide the campaign into four periods, (i.e. considering it as a whole, using the same *a*, *b*, *p*, *q* parameters and fitting a new *d* parameter under these given circumstances). This model is referred as Model 3 for this section. By following this methodology, *d* is found to be 1.14 with an SSR value of 1.89×10^{8} which is a 12% worse fit than Model 2, but still a 65% better fit than Model 1. In Model 2, a different *d* parameter for period 3 essentially removed the outlier.

Figures 46 and 47 show the fitted versus real losses for the Soviet and German forces, respectively, for Model 3.

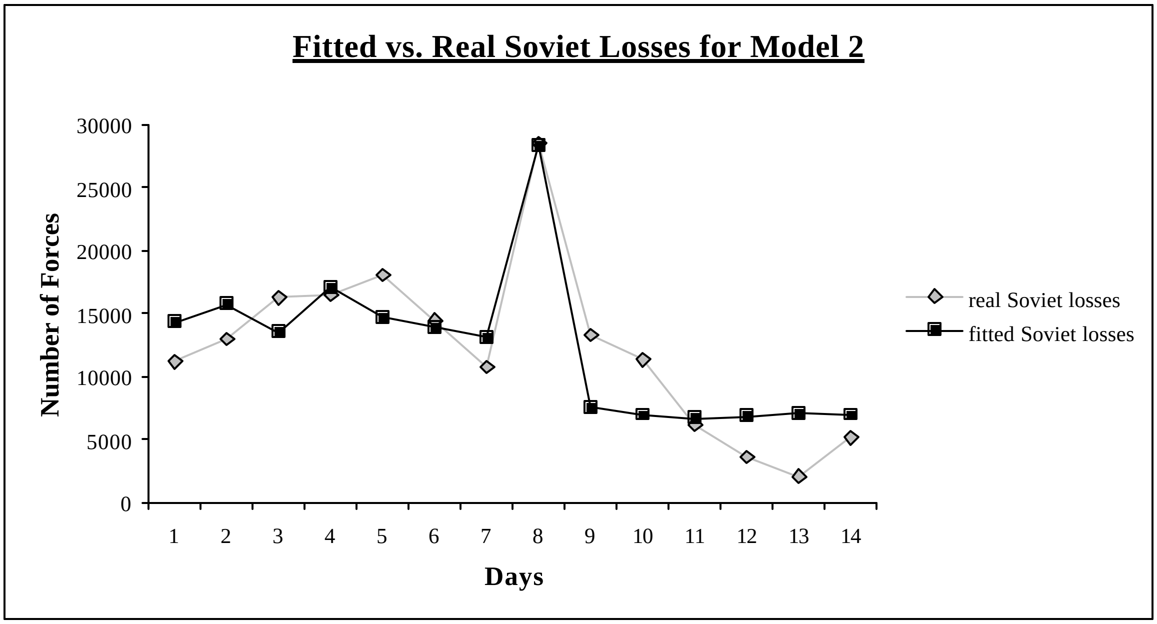

Figure 44. Fitted losses plotted versus real losses for the Soviet forces for model 2, which has four periods, and the Soviets as the attacker for the 8th day of the battle.

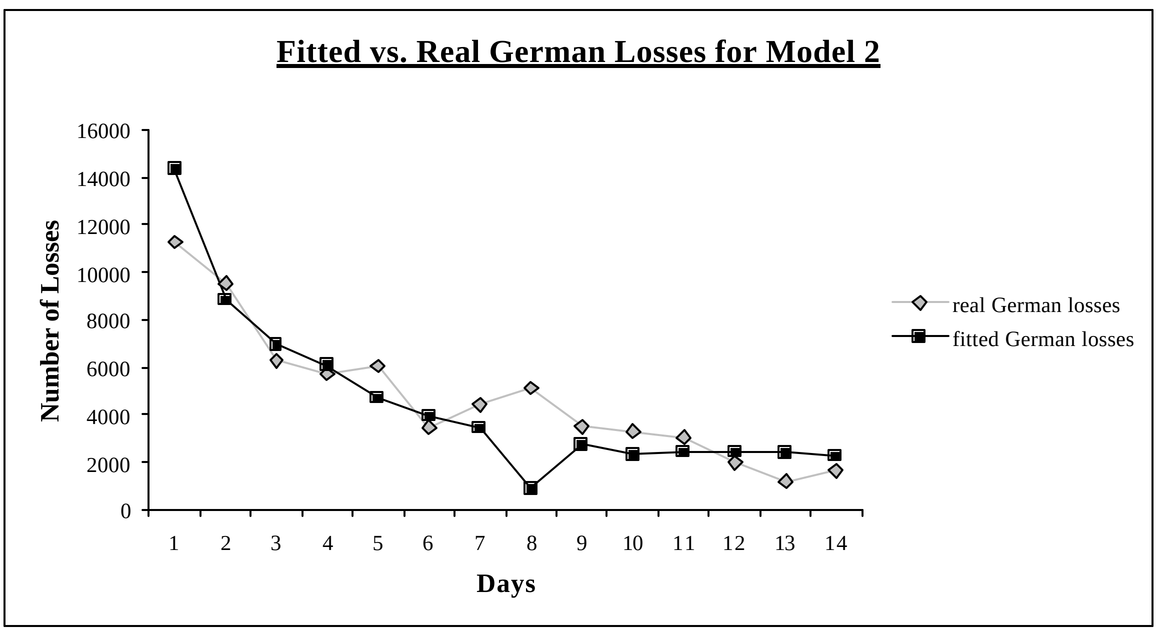

Figure 45. Fitted losses plotted versus real losses for the German forces for model 2, which has four periods, and the Soviets as the attacker for the 8th day of the battle.

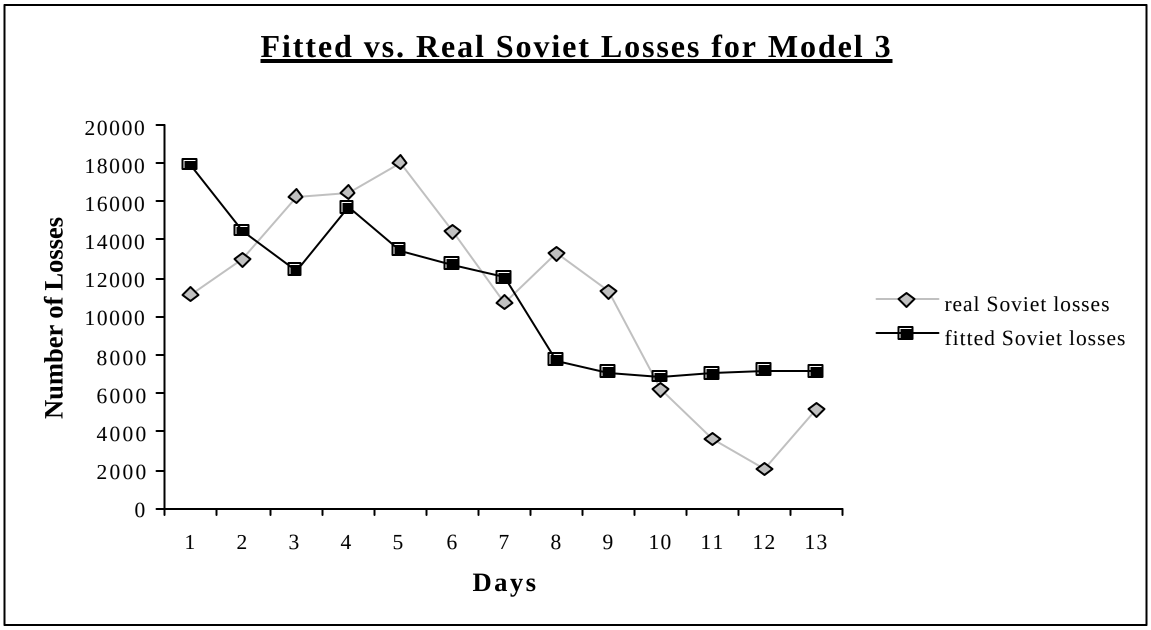

Figure 46. Fitted losses plotted versus real losses for the Soviet forces for Model 3 which leaves out the 8th day of the battle, does not divide the campaign into 4 periods, uses the same parameters as Model 1 and Model 2 and fits a new *d* parameter.

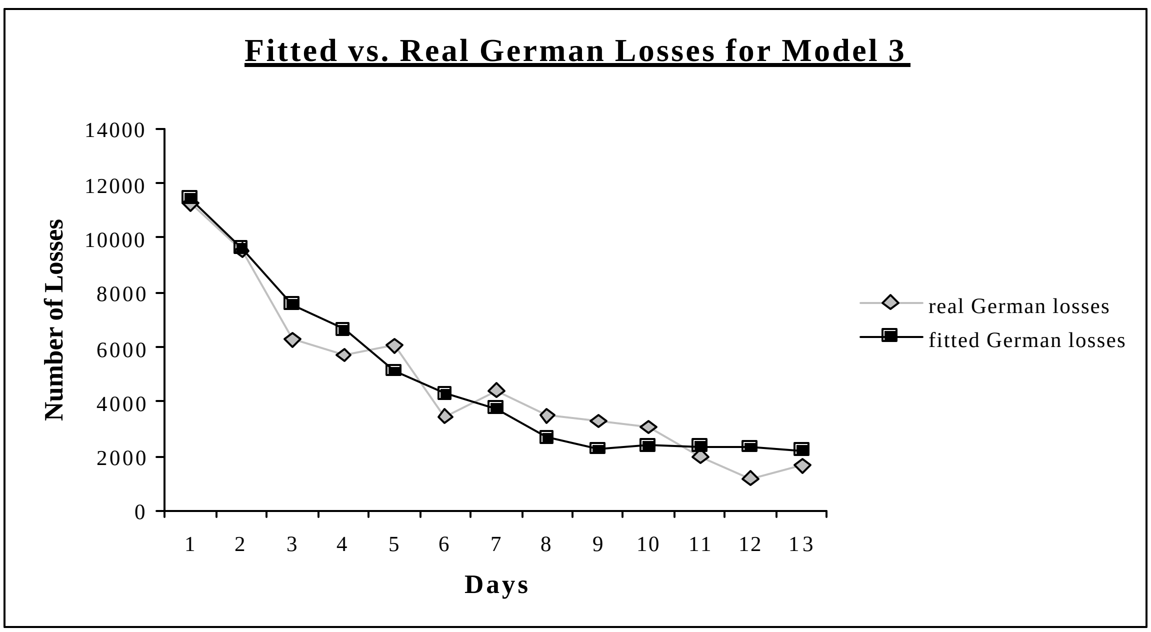

Figure 47. Fitted losses plotted versus real losses for the German forces for Model 3 which leaves out the 8th day of the battle, does not divide the campaign into 4 periods, uses the same parameters as Model 1 and Model 2 and fits a new *d* parameter.

This results in the question; What if a whole new regression analysis is done to the data, leaving out the eighth day? This model is referred to as Model 4 and by doing so, the resulting model with an SSR value of 1.90×10^{8} is found to be:

$$\dot{B} = 1.85\times10^{-51} R^{9.6853} B^{0.1458} \tag{40}$$

$$\dot{R} = 3.56\times10^{-53} B^{9.6853} R^{0.1458} \tag{41}$$

These results are far better than those found in previous sections that contained the outlier. But, they do not however, provide a better fit than the ones found in this section which are adjusted for the outlier. Also, it is significant that there is a big difference in the size of the p and q parameters. Figures 48 and 49 show the fitted versus real losses for the Soviet and German forces, respectively, for model 4.

Handling the data in parts and fitting different tactical parameters definitely improves the fits of all models given in this section. This result is consistent with what Hartley and Helmbold found in their studies [Ref.10].

Model 2 with an SSR value of 1.69×10^{8} has the smallest SSR value thus far. This result largely depends on considering July 12, which is the largest outlier apart from the rest of the data, causing a considerable decrease from the previous lowest SSR value of 5.54×10^{8} to a much lower SSR value of 1.69×10^{8}.

Model 3 finds d to be 1.14, which means an attacker advantage/defender disadvantage. But, this circumstance again largely depends on still using the same parameters that we had when the tactical parameter d is 1.17. Once more, this d value indicates an attacker advantage/defender disadvantage situation.

In Model 4, leaving even only one day out (the largest outlier), improves the model's fit tremendously when compared to the previous SSR values.

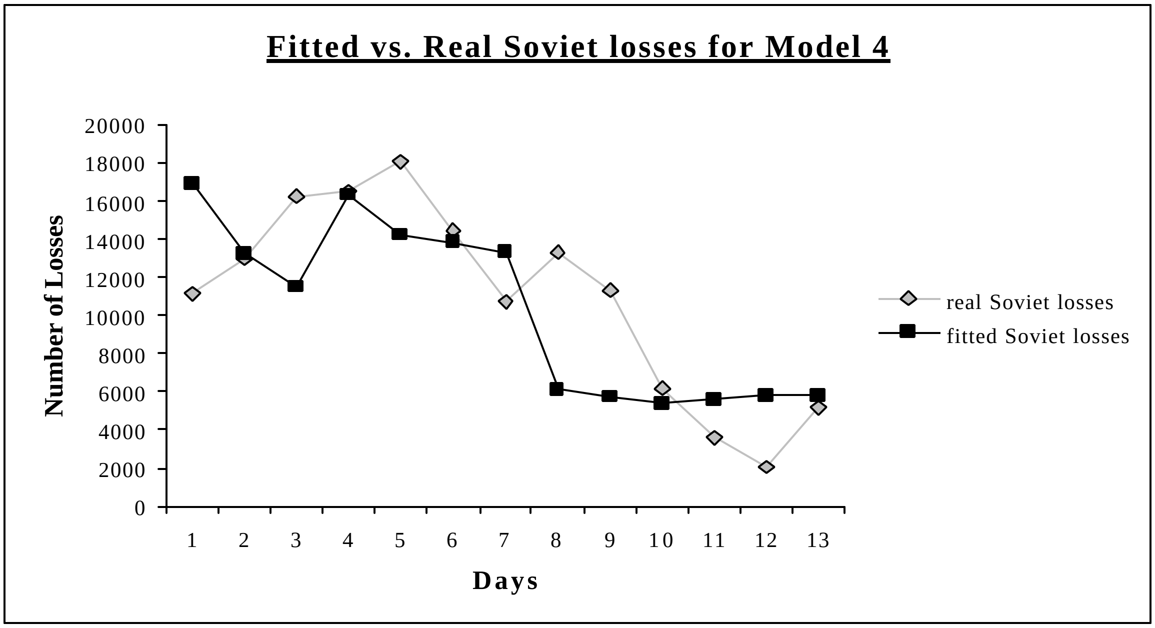

Figure 48. Fitted losses plotted versus real losses for the Soviet forces for model 4, which leaves out the 8^{th} day of the battle, does not divide the campaign into 4 periods, fits a whole new regression model.

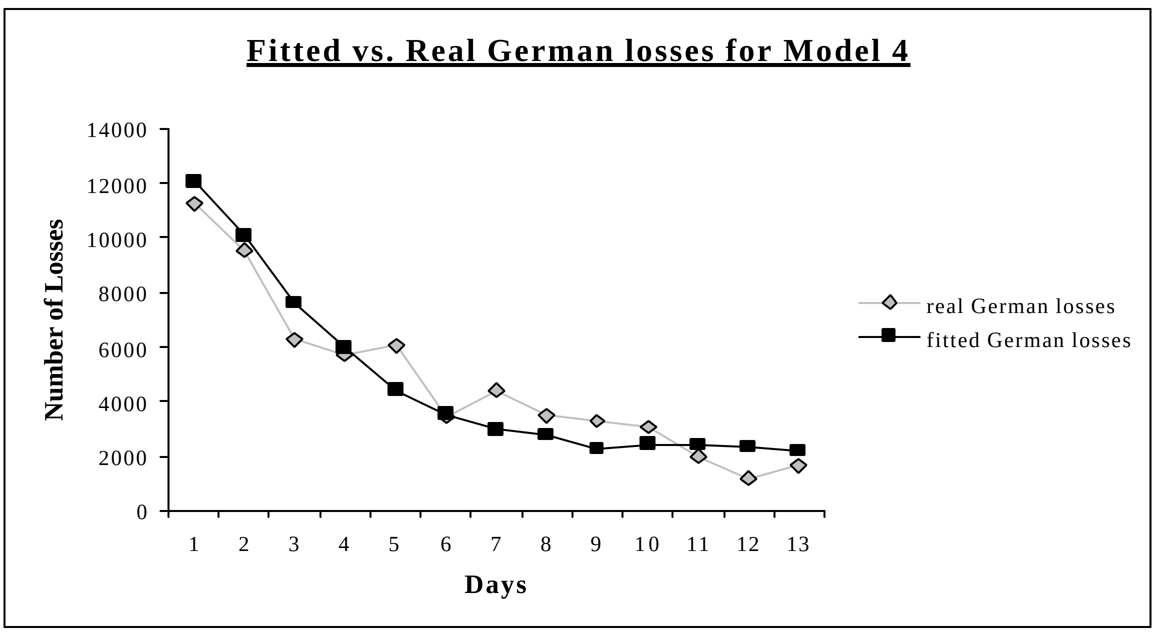

Figure 49. Fitted losses plotted versus real losses for the German forces for model 4, which leaves out the 8^{th} day of the battle, does not divide the campaign into 4 periods, fits a whole new regression model.

For all models, a and b parameters are significantly small and $a>b$. This result suggests that individual German effectiveness was greater than individual Russian effectiveness.

When the p and q parameters are compared, it is observed that the p parameter is greater than the q parameter suggesting that one side's losses are more a function of his own forces rather than being a function of the opponent's forces. This observation is different from what Fricker found in his study.

The results for the four different models are given in Table 30.

Name of the model	***a***	***b***	***p***	***q***	***d***	***SSR***	R^2
Campaign in four Parts	1.88E-47	1.07E-48	7.5038	1.5793	4 periods *d*=0.91,1.24, 1.0,1.17	5.34E+8	-2.3410
Campaign in four Parts	1.88E-47	1.07E-48	7.5068	1.5793	4 periods *d*=0.91,1.24, 0.32,1.17	1.69E+8	-0.0607
Campaign in four Parts	1.88E-47	1.07E-48	7.5038	1.5793	1.14	1.89E+8	0.5689
Campaign in four Parts	1.85E-51	3.56E-53	9.6853	0.1458	-	1.90E+8	0.5658

Table 30. The results for the model which considers the battle in separate parts.

The negative R^2 values are mainly a result of considering certain days in the campaign solely on their own. This results in SST value for that day being zero. This result (i.e, the SST value being zero for a certain day) is the main reason for negative R^2 values in this section.

5. Considering change points in the model

The findings in the previous sections suggest that fitting models for separate phases of the battle might improve the fit to the data. This section considers one or more attrition change points for each side. At each chosen point in the phase of battle all the parameters pertaining to that particular side will change.

When the historical account of the battle is taken into account, it is apparent that the Germans generally attacked between July 5, and July11, for the first seven days, and the Soviets attacked between the days July 12, and July 18, for the last seven days. This is the first change point to be considered and will be referred as change point 7/7.

Another approach is considering that the Germans attacked between July 5, and July 12, for the first eight days, and the Soviets attacked between July 13, and July 18, for the last six days. This is the second change point to be considered, and will be referred to as change point 8/6. This type of approach (considering change points for fitting the model to the data) is similar to what Hartley and Helmbold did in their study [Ref.10].

No tactical parameter will be considered, and only linear regression will be used in fitting the data to the model with change points. For estimating the parameters of the model that minimize the sum of squared residuals of the actual and estimated attrition, S-PLUS software and the last 14 days of the aggregated data given in Table 14 in Section IV.A.1 are used.

Results for the first half of the Linear Regression model for change point 7/7 with an SSR value of 6.53×10^{7} are:

$$\dot{B} = 8.91 \times 10^{-30} R^{6.4117} B^{-0.4323} \tag{42}$$

$$\dot{R} = 2.62 \times 10^{-31} B^{6.4117} R^{-.0.4323} \tag{43}$$

Results for the second half of the Linear Regression model for change point 7/7 with an SSR value of 8.78×10^{7} are:

$$\dot{B} = 1.90 \times 10^{-292} R^{18.0587} B^{34.4502} \qquad (44)$$

$$\dot{R} = 4.37 \times 10^{-291} B^{18.0587} R^{34.4502} \qquad (45)$$

where both halves add up to a total SSR value of 1.53×10^8, and result in an R^2 value of 0.7448.

Results for the first half of the Linear Regression model for change point 8/6 with an SSR value of 1.65×10^8 are:

$$\dot{B} = 7.75 \times 10^{-5} R^{4.4212} B^{-2.8454} \qquad (46)$$

$$\dot{R} = 1.91 \times 10^{-6} B^{4.4212} R^{-2.8454} \qquad (47)$$

Results for the second half of the Linear Regression model for change point 8/6 with an SSR value of 7.78×10^7 are:

$$\dot{B} = 1.94 \times 10^{-246} R^{25.7652} B^{18.7674} \qquad (48)$$

$$\dot{R} = 1.32 \times 10^{-247} B^{25.7652} R^{18.7674} \qquad (49)$$

where both halves add up to a total SSR value of 2.43×10^8 and result in an R^2 value of 0.3488.

The SSR value for the change point 7/7 is the smallest SSR value we have seen. It gives a 9% better fit than Model 2 of Section IV.B.4 which is 1.69×10^8. It is also almost a 56% better fit than the one found in section IV.B.1, where only one set of parameters is fit to the whole data. This model has the highest R^2 value we have seen thus far, and easily the best fit we have obtained. We can conclude that fitting the model using the change points definitely improves the fit, and this is consistent with the result Hartley and Helmbold [Ref. 10] found in their study.

However the only concern is that the q parameter for both the change point 7/7 and change point 8/6 are negative, meaning that the number of a force's casualties decreases as one of the force strengths increases. The p and q parameters found in the models are extremely high. Doubling the force size results in a dramatic change in the outcome and this does not intuitively make sense. Since this analogy is both illogical and unlikely, we resolve that even if the change point approach gives the lowest SSR value of 1.53×10^{8}, with the change point 7/7 model, we cannot accept this fit as the best one. This result also suggests a wide range of parameters gives similar fits to the data

In all the models explored in this section, the a and b parameters are significantly small, and except the equations given in IV.B.5.(44), IV.B.5.(45), $a>b$. This suggests that individual German effectiveness was greater than individual Russian effectiveness.

When the p and q parameters are compared, it is observed that except the model given in equations IV.B.5.(44), IV.B.5.(45), the p parameter is greater than the q parameter. This comparison suggests that one side's losses are more a function of his own forces rather than being a function of the opponent's forces, and is different from what Fricker found in his study.

6. Using different weights

This section considers different weights for aggregating the battle data. Bracken [Ref.8] states in his study that, "The given weights are consistent with those of studies and models of the U.S.Army Concepts Analysis Agency. Virtually all theater-level dynamic combat simulation models incorporate similar weights, either as inputs or as decision parameters computed as the simulations progress." Although Bracken's points are well taken, this study will try to fit models by using different weights for exploratory purposes. The different weights are selected on a wholly intuitive basis and are a result of many different trial and error calculations.

The first weight combination will use the weights 1, 5, 20 and 40; the second weight combination will use the weights 1, 5, 15 and 20; the third weight combination will use the weights 1, 5, 30 and 40; the fourth weight combination will use the weights 1, 5, 20 and 30 for manpower, APC, artillery and tanks, respectively.

Note that tanks are weighted more because the Battle of Kursk was a major tank battle. Both linear and robust LTS regression models are used to fit the data, which is aggregated using the different weight combinations given above.

Table 31 presents the aggregated data obtained using the first weight combination. Table 32 presents the aggregated data obtained using the second weight combination. Table 33 presents the aggregated data obtained using the third weight combination. Table 34 presents the aggregated data obtained using the fourth weight combination.

4. First weight combination

The result for the linear regression model that gives an SSR value of 1.15×10^{9} and an R^2 value of 0.0870, is:

$$\dot{B} = 1.25\times10^{-38} R^{5.2298} B^{2.2746} \tag{50}$$

$$\dot{R} = 1.60\times10^{-39} B^{5.2298} R^{2.2746} \tag{51}$$

The result for the robust LTS regression model that gives an SSR value of 1.07×10^{9} and an R^2 value of 0.1514, is:

$$\dot{B} = 7.26\times10^{-35} R^{5.5312} B^{1.3268} \tag{52}$$

$$\dot{R} = 5.53\times10^{-36} B^{5.5312} R^{1.3268} \tag{53}$$

b. Second weight combination

The result for the linear regression model that gives an SSR value of 6.24×10^{8} and an R^2 value of 0.0975, is:

Day	Blue Forces	Blue Losses	Red Forces	Red Losses
1	620173	13007	369811	14737
2	609589	14733	356025	14392
3	587405	21146	349465	8529
4	567452	22492	360184	7602
5	545652	23671	353044	8703
6	531943	17325	353484	3930
7	522036	13314	352210	5375
8	487313	36452	348085	6832
9	476711	15442	348389	4491
10	465684	13583	345740	4110
11	468799	8021	342382	4567
12	470151	4080	345266	2355
13	468706	2907	347745	1274
14	465083	6780	348400	1539

Table 31. Data on forces which are aggregated by using weight combination 1. Forces are combat manpower, APCs, artillery and tanks, which are weighted by 1, 5, 20 and 40 respectively. Here, a tank is considered to be twice as valuable as an artillery piece.

Day	Blue Forces	Blue Losses	Red Forces	Red Losses
1	568728	10842	344261	10657
2	558869	12243	335240	9407
3	542820	15891	330235	6074
4	529132	16122	342199	5377
5	512552	17846	337194	5893
6	500678	14120	336529	3150
7	491876	10579	334920	4040
8	464708	28092	332115	4812
9	454121	13052	332554	3316
10	444529	11198	330220	3165
11	448134	6076	327947	2972
12	447451	3525	329471	1875
13	446136	2067	330695	1124
14	443168	5060	330730	1364

Table 32. Data on forces which are aggregated by using weight combination 2. Forces are combat manpower, APCs, artillery and tanks, which are weighted by 1, 5, 15 and 20 respectively. Here a tank is considered to be 33% more valuable than artillery.

Day	Blue Forces	Blue Losses	Red Forces	Red Losses
1	627223	13137	381471	14977
2	616349	15033	367635	14442
3	594015	21296	361005	8599
4	573932	22632	372314	7732
5	552052	23761	365144	8763
6	538233	17455	365474	4050
7	528316	13384	364270	5525
8	493443	36612	360025	6952
9	482771	15542	360259	4561
10	471714	13633	357580	4160
11	474809	8071	354212	4597
12	476151	4110	357056	2395
13	474726	2907	359565	1294
14	470993	6820	360220	1649

Table 33. Data on forces which are aggregated by using weight combination 3. Forces are combat manpower, APCs, artillery and tanks, which are weighted by 1, 5, 30 and 40 respectively. Here an artillery piece is considered to be six times more effective than an APC and a tank is considered to be eight times more valuable than an APC.

Day	Blue Forces	Blue Losses	Red Forces	Red Losses
1	596213	11957	359951	12757
2	585919	13563	348535	11912
3	566765	18556	342735	7319
4	549912	19342	354224	6522
5	530702	20781	348144	7313
6	517883	15755	348004	3570
7	508526	11964	346580	4745
8	477543	32312	343085	5852
9	466931	14272	343439	3921
10	456614	12403	340940	3650
11	459969	7061	338122	3777
12	460301	3810	340316	2125
13	458926	2487	342175	1204
14	455603	5930	342520	1479

Table 34. Data on forces which are aggregated by using weight combination 4. Forces are combat manpower, APCs, artillery and tanks, which are weighted by 1, 5, 20 and 30 respectively. Here an artillery piece is considered to be four times more effective than an APC and a tank is considered to be six times more valuable than an APC.

$$\dot{B} = 2.50 \times 10^{-46} R^{5.7638} B^{3.1222} \tag{54}$$

$$\dot{R} = 3.49 \times 10^{-47} B^{5.7638} R^{3.1222} \tag{55}$$

The result for the robust LTS regression model that gives an SSR value of 5.48×10^{8} and an R^2 value of 0.2072, is:

$$\dot{B} = 7.85 \times 10^{-36} R^{5.8613} B^{1.1899} \tag{56}$$

$$\dot{R} = 4.75 \times 10^{-37} B^{5.8613} R^{1.1899} \tag{57}$$

c. Third weight combination

The result for the linear regression model that gives an SSR value of 1.15×10^{9} and an R^2 value of 0.0926, is:

$$\dot{B} = 3.78 \times 10^{-39} R^{5.2293} B^{2.3513} \tag{58}$$

$$\dot{R} = 5.34 \times 10^{-40} B^{5.2293} R^{2.3513} \tag{59}$$

The result for the robust LTS regression model that gives an SSR value of 1.06×10^{9} and an R^2 value of 0.1637, is:

$$\dot{B} = 1.46 \times 10^{-35} R^{5.9619} B^{1.0159} \tag{60}$$

$$\dot{R} = 9.33 \times 10^{-37} B^{5.9619} R^{1.0159} \tag{61}$$

d. Fourth weight combination

The result for the linear regression model that gives an SSR value of 8.63×10^{9} and an R^2 value of 0.0943, is:

$$\dot{B} = 2.89 \times 10^{-42} R^{5.4863} B^{2.666} \tag{62}$$

$$\dot{R} = 3.91 \times 10^{-43} B^{5.4863} R^{2.666} \tag{63}$$

The result for the robust LTS regression model that gives an SSR value of 7.74×10^{8} and an R^2 value of 0.1873, is:

$$\dot{B} = 5.05\times10^{-35} R^{5.6294} B^{1.2631} \tag{64}$$

$$\dot{R} = 3.51\times10^{-36} B^{5.6294} R^{1.2631} \tag{65}$$

Using different weights to aggregate the data can improve the fit to the data. The SSR value observed for the second weight combination when the data is fitted using the Robust LTS Regression model is the lowest SSR value found for models without the tactical parameter d. But, this result may be due to the small size of the weights used for aggregating the data. Comparing SSR values makes sense as long as the weights used for aggregating the data are constant for all models compared, but this is not the case in our discussion. In such circumstances, the R^2 value is a better parameter to use for comparison purposes rather than the SSR value because the R^2 value adjusts to scale. How the R^2 value is computed is given in equation IV.A.1.b.(10).

The parameters and the R^2 values for each weight combination are given in Table 34 for both linear regression and robust LTS regression models. When the R^2 values are compared for the models presented in this section, it is observed that weight combination 2 gives the best fit when the robust LTS regression technique is used, with the greatest R^2 value of 0.2072. The second best fit is found when weight combination 4 is used with the robust LTS regression technique, and the third best fit is found when weight combination 3 is used, again with robust LTS regression technique. These models with different weight combinations do not give a better fit as a whole when compared to the two models given in IV.B.1.d.(26), IV.B.1.d.(27) and IV.B.3.d.(38), IV.B.3.d.(39) where both models have an R^2 value of 0.2262 and use the weight combination of 1, 5, 20 and 40, for combat manpower, APCs, tanks and artillery, respectively.

When the p and q parameters are compared, it is evident that for all the models discussed in this section, the p parameter is greater than the q parameter. This result suggests that one side's losses are more a function of the opponent's forces rather than being a function of his own forces, resembling earlier findings.

Except for the model given in IV.B.6.d.(64) and IV.B.6.d.(65), the a and b parameters are significantly small and $a>b$ for all the models discussed in this section. This result suggests that individual German effectiveness was greater than individual Russian effectiveness.

One can easily argue that tanks are more effective during an offensive then they are during a defense. Likewise, artillery can be considered to have different effects on the outcome of the battle depending on the type of a campaign. The weights used in the second weight combination may give a better fit than the models which use the other three weight combinations. However the relevance of the weights used is another topic of discussion in itself. In short, it is clear according to our examples that changing the weights can help find a better fit, but one must be careful in doing so that the issue of relevancy to the real world is not ignored. Further investigation is recommended for determining weight combinations.

Figures 49 and 50 show fitted losses plotted versus real losses for the Soviet and the German forces respectively, for the robust LTS regression model using the second weight combination which gives the best fit.

For ease of comparison, the results for all the models using different weight combinations and the previous two results are given in Table 35.

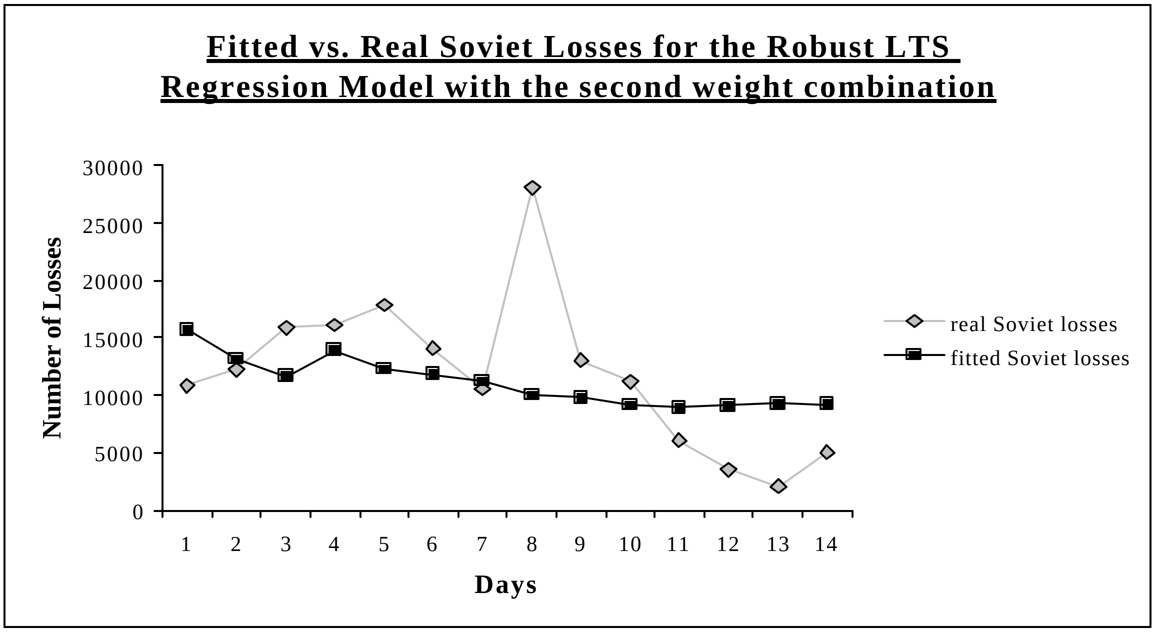

Figure 50. Fitted losses plotted versus real losses for the Soviet Forces for the robust LTS regression model using the weight combination 2. The same pattern where the model over/underestimates the battle in three distinctive phases is observable in this plot.

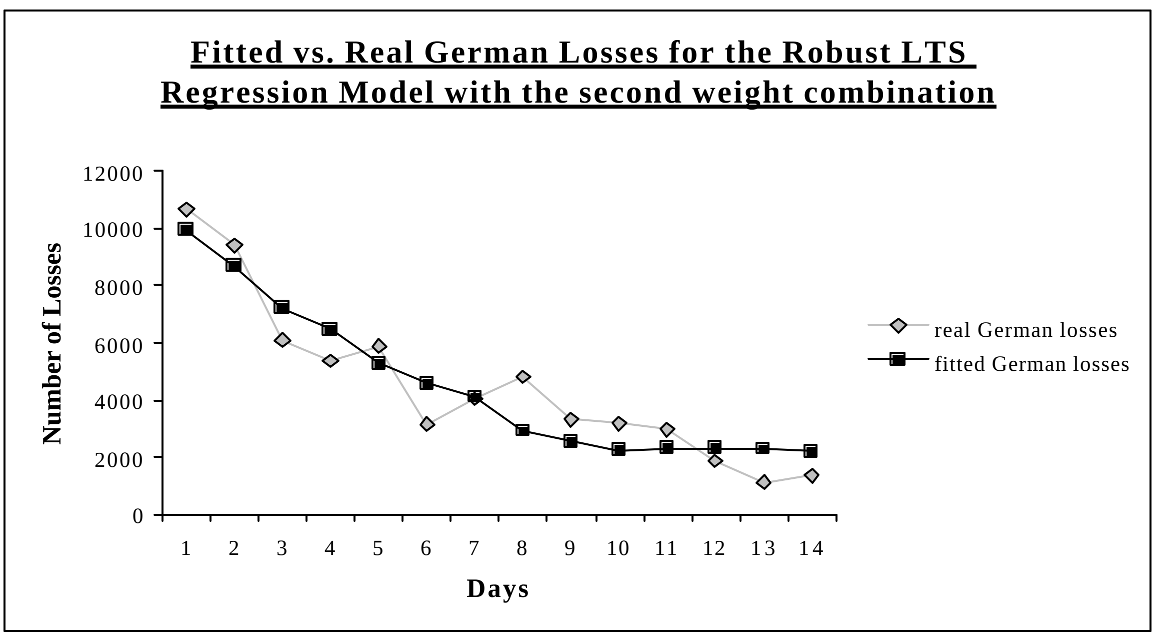

Figure 51. Fitted losses plotted versus real losses for the German Forces for the robust LTS regression model with the weight combination 2.

Type Of the model	*a*	*b*	*p*	*q*	*d*	*SSR*	R^2
Previous Best Result	2.27E-40	1.84E-41	6.0843	1.7312	-	5.54E+8	0.2262
Weight Comb.1 Lin.Reg.	1.25E-38	1.60E-39	5.2298	2.2746	-	1.15E+9	0.0870
Weight Comb.1 Rob.LTS	7.26E-35	5.53E-36	5.5312	1.3268	-	1.07E+9	0.1514
Weight Comb.2 Lin.Reg.	2.50E-46	3.49E-47	5.7638	3.1222	-	6.24E+8	0.0975
Weight Comb.2 Rob.LTS	7.85E-36	4.75E-37	5.8613	1.1899	-	5.48E+8	0.2072
Weight Comb.3 Lin.Reg.	3.78E-39	5.34E-40	5.2293	2.3513	-	1.15E+9	0.0926
Weight Comb.3 Rob.LTS	1.46E-35	9.33E-37	5.9619	1.0159	-	1.06E+9	0.1637
Weight comb.4 Lin.Reg.	2.89E-42	3.91E-43	5.4863	2.6660	-	8.63E+9	0.0943
Weight Comb.4 Rob.LTS	5.05E-35	3.51E-36	5.6294	1.2631	-	7.74E+8	0.1873

Table 35. The results for the models using different weight combinations. Weight combination 2 gives the best fit.

7. Force ratio and fractional exchange ratio models

In this section, Force Ratio (FR) and Fractional Exchange Ratio (FER) models are explored and analyzed. The reason for including this approach in our discussion is that both analysts and military staff use force ratios in models for combat outcomes and decisions. For this purpose, five different models are investigated. The first model uses the FR of aggregated forces as a predictor to predict the percent of casualties for each side. The FR of blue forces is equal to the total number of aggregated blue forces divided by the total number of aggregated red forces, and likewise for the FR of the red forces.

The percent of casualties of the blue forces is equal to the total number of aggregated blue losses divided by the total number of aggregated blue forces.

Figures 52 and 53 show loss ratio plotted against the FR for Soviet and German forces, respectively. The representation of Model 1 looks like:

$$(\dot{B}/B) = I_1 + I_2 + (B/R) \tag{66}$$

$$(\dot{R}/R) = I_1 + I_2 + (R/B) \tag{67}$$

where I_1 is an indicator of the blue force or red force, and I_2 indicates the difference between the attacker and defender, and are given as:

$$\begin{aligned} I_1 &= 1 \text{ if Blue} \\ I_1 &= 0 \text{ if Red} \end{aligned} \tag{68}$$

$$\begin{aligned} I_2 &= 1 \text{ if attacker} \\ I_2 &= 0 \text{ if defender} \end{aligned} \tag{69}$$

The resulting model for Model 1 with the intercept that gives an SSR value of 3.09×10^{-3} and an R-squared value of 0.2296 (given by the S-PLUS software) is:

$$PC = -0.0103 - 0.0074 I_1 + 0.0068 I_2 + 0.0275(OFR) \tag{70}$$

where PC denotes the percent of casualties as given in IV.B.7.(64), IV.B.7.(65), and OFR denotes the opponent's FR for a given side.

The R-squared value given above is not calculated using the formula given in equation IV.A.1.(10) but given by the S-PLUS software and will be used for all the models throughout this section.

Here, indicator variables are mainly used for the purpose of adjusting the intercept. When the intercept term is used in the model, the correlation matrix of the estimated coefficients shows a high correlation between the estimates that, due to high

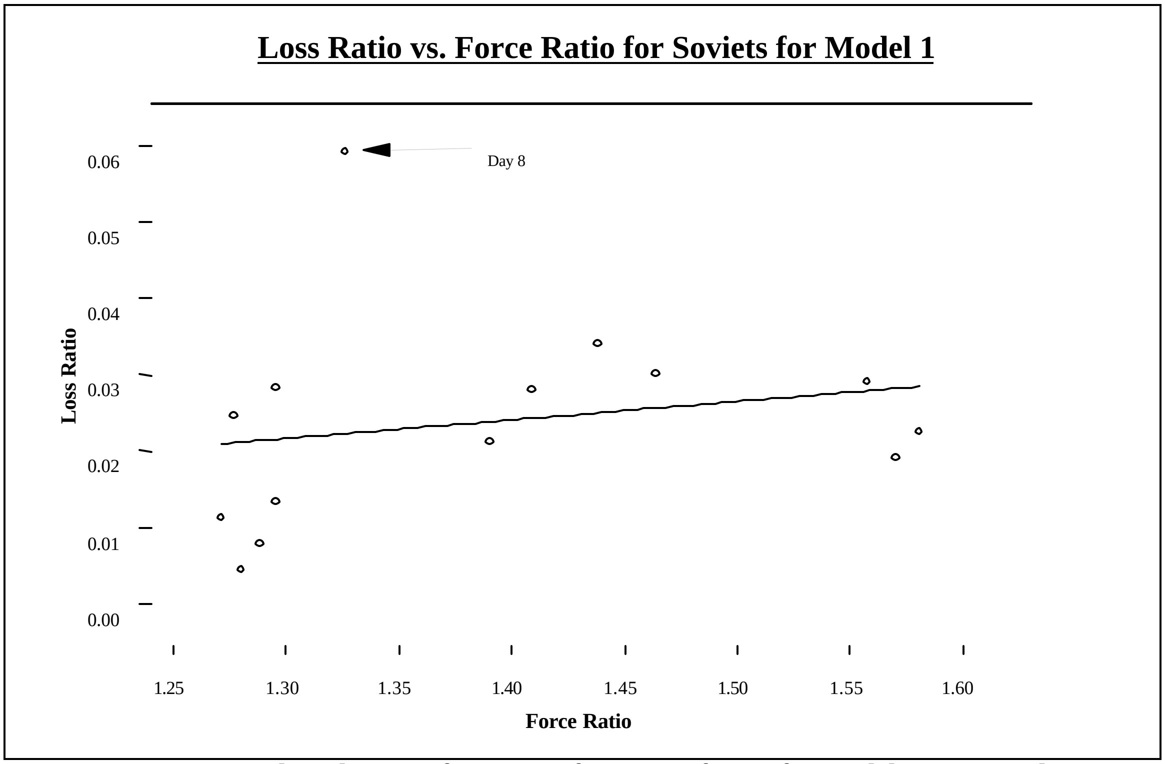

Figure 52. Loss ratio plotted versus force ratio for Soviet forces for model 1. Soviets lost a higher percentage of their forces as their force ratio increased.

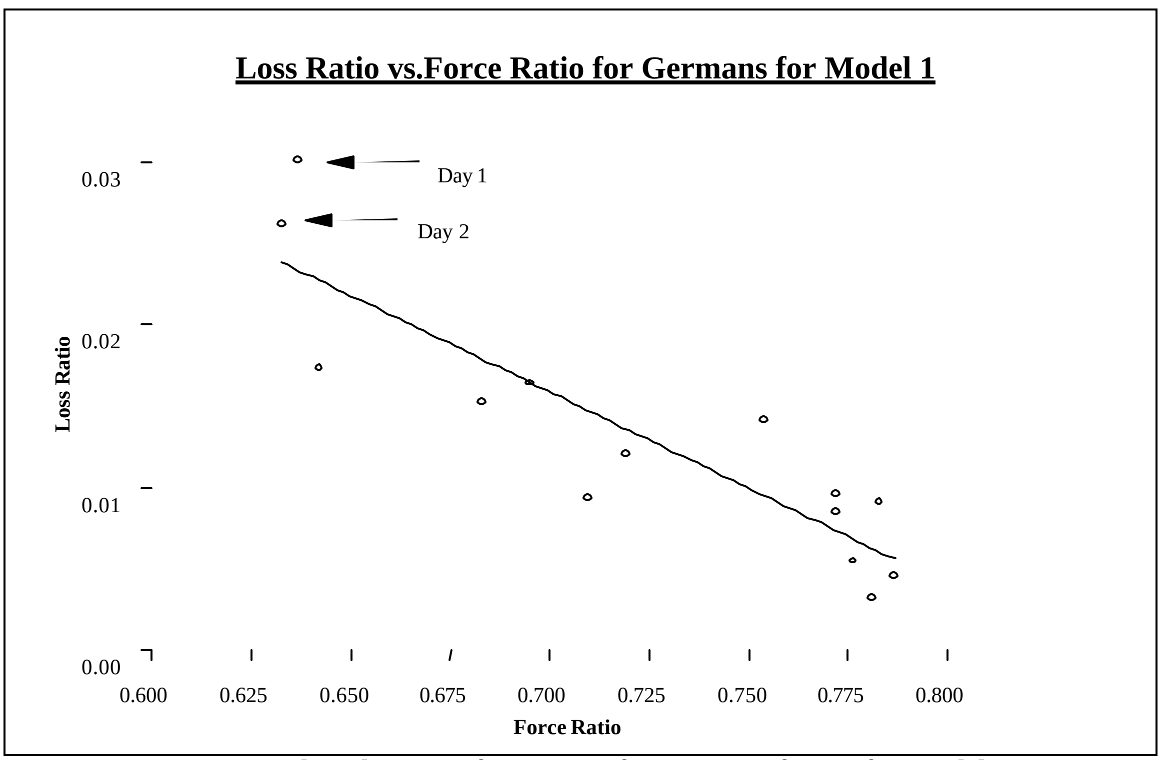

Figure 53. Loss ratio plotted versus force ratio for German forces for model 1. Germans lost a lower percentage of their forces as their force ratio increased.

	Intercept	I_1	I_2
I_1	0.9754		
I_2	-0.8662	-0.8291	
	-0.9945	-0.9895	0.8379

Table 36. Correlation matrix of the estimated coefficients of the first model. Notice the high correlation between the model's coefficients, and especially the correlation between the intercept and the force ratio, which can result in a very bad fit.

collinearity can result in very inaccurate estimates [Ref.18]. Because of this result, an intercept term is not used in the following models. The correlation matrix of the estimated coefficients is given in Table 36.

Concern over whether or not leaving the intercept term out is correct or not can be addressed by doing a hypothesis test. The null hypothesis will be, $H_0: intercept = 0$, and the alternative hypothesis will be, $H_a: intercept \neq 0$. With a significance level of $a = 0.1$ and 24 degrees of freedom, the null hypothesis will be rejected if $t \geq t_{0.05,24} = 1.711$ or if $t \leq -t_{0.05,24} = -1.711$. The t-statistics of the intercept of Model 1 is $t = -0.2899$ which is not in the rejection region. So, the null hypothesis is not rejected, and the intercept will be assumed to be zero throughout the models.

The resulting model for Model 1 without the intercept gives an SSR value of 3.105×10^{-3} and a multiple R-squared value of 0.7699 and looks like:

$$PC = 0.001 I_1 + 0.0048 I_2 + 0.0147(OFR) \qquad (71)$$

Table 37 shows the coefficients, standard errors, and t values for Model 1.

	Value	Std. Error	t value	Pr(>|t|)
I1	0.0010	0.0064	0.1534	0.8793
I2	0.0048	0.0039	1.2385	0.2270
OFR	0.0147	0.0045	3.2419	0.0034

Table 37. Important statistical values of the estimated coefficients for Model 1.

The positive coefficient of the indicator variable I_1 indicates a German advantage (though insignificant), where the positive coefficient of the indicator variable I_2 indicates a defender advantage, and again is insignificant. The positive coefficient of the force ratio variable indicates that as the force ratio increases, so do the losses. Even though statistically significant, this result does not intuitively make much sense.

The second model uses the total aggregated force ratios as a predictor to predict the fractional exchange ratios for each side. FER for the blue forces is equal to the percent of blue casualties divided by the percent of red casualties, and likewise for the FER of the red forces. Figures 53 and 54 show the FER plotted against force ratio for Soviet and German forces, respectively. The representation of Model 2 looks like:

$$(\dot{B}/B)/(\dot{R}/R) = I_1 + I_2 + (B/R) \tag{72}$$

$$(\dot{R}/R)/(\dot{B}/B) = I_1 + I_2 + (R/B) \tag{73}$$

where I_1 indicates the difference between the blue force and red force, and I_2 indicates the difference between attacker and defender (time of battle) and have the values given in IV.B.7.(68) and IV.B.7.(69).

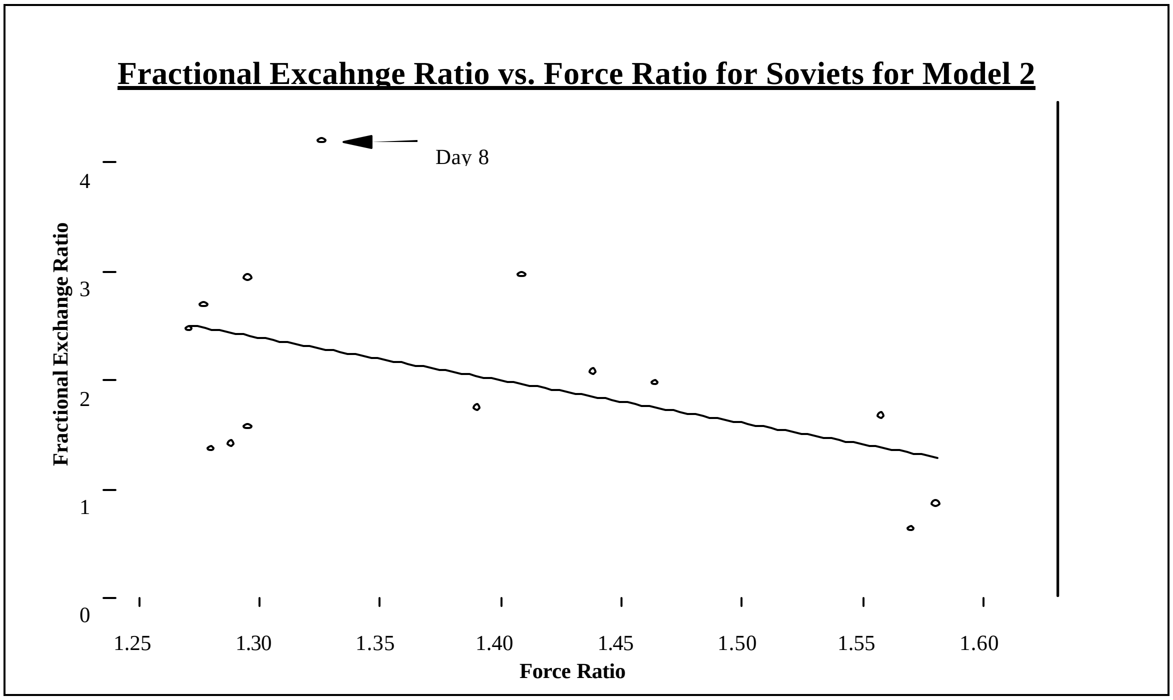

Figure 54. Fractional exchange ratio plotted versus force ratio for Soviet forces for model 2.

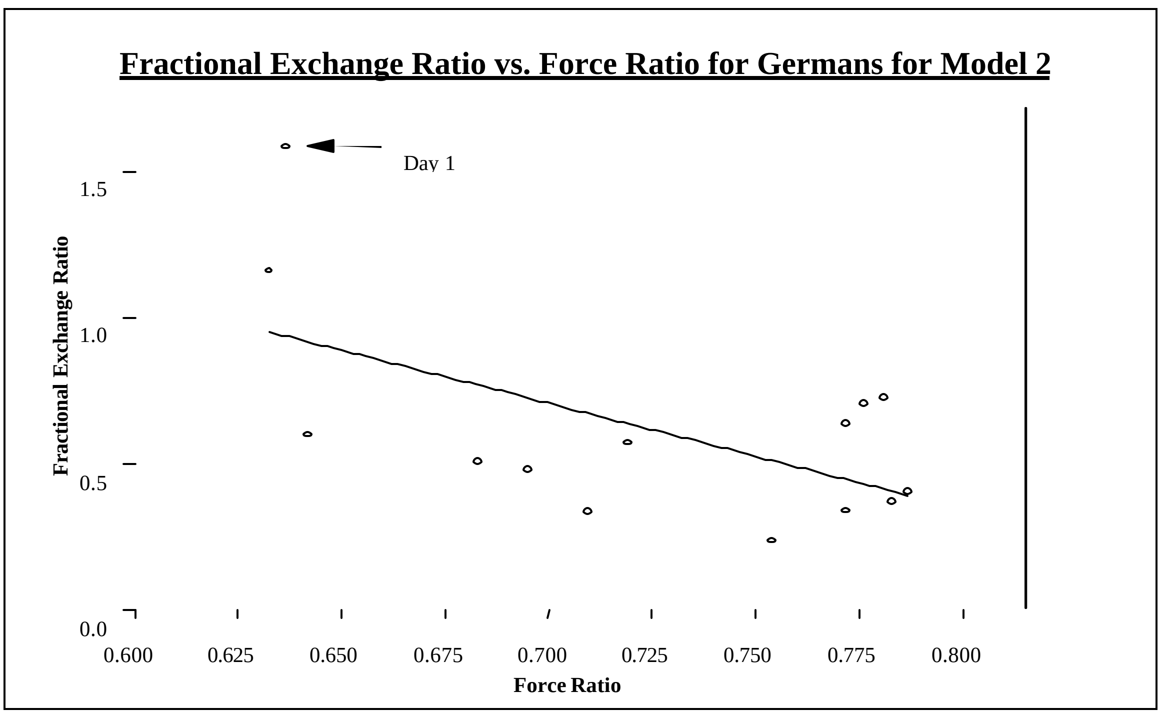

Figure 55. Fractional exchange ratio plotted versus force ratio for German forces for model 2.

Table 38 shows the coefficients, standard errors, and t values for Model 2.

	Value	Std. Error	t value	Pr(>|t|)
I1	1.1849	0.4019	2.9483	0.0068
I2	0.5647	0.2409	2.3441	0.0273
OFR	0.4153	0.2837	1.4638	0.1557

Table 38. Important statistical values of the estimated coefficients for model 2.

The resulting model for Model 2 that gives an SSR value of 12.120 and a multiple R-squared of 0.6963, is:

$$FER = 1.1849 I_1 + 0.5647 I_2 + 0.4153(OFR) \tag{74}$$

where FER denotes the fractional exchange ratio as given in IV.B.7.(72), IV.B.7.(73), and OFR denotes the opponent's FR for a given side.

Similar to the results found for Model 1, the positive coefficient of indicator variable I_1 indicates a German advantage and is significant, where the positive coefficient of indicator variable I_2 indicates a defender advantage and is significant too. The positive coefficient of the force ratio variable indicates that as the force ratio increases so do the losses. Again, the coefficient is not significant and does not intuitively make much sense.

Model 3 uses the force ratio of tanks as a predictor to predict the percent of tank losses for each side. Figures 56 and 57 show the tank loss ratio plotted against the tank force ratio for Soviet and German forces, respectively. The representation of Model 3 looks like:

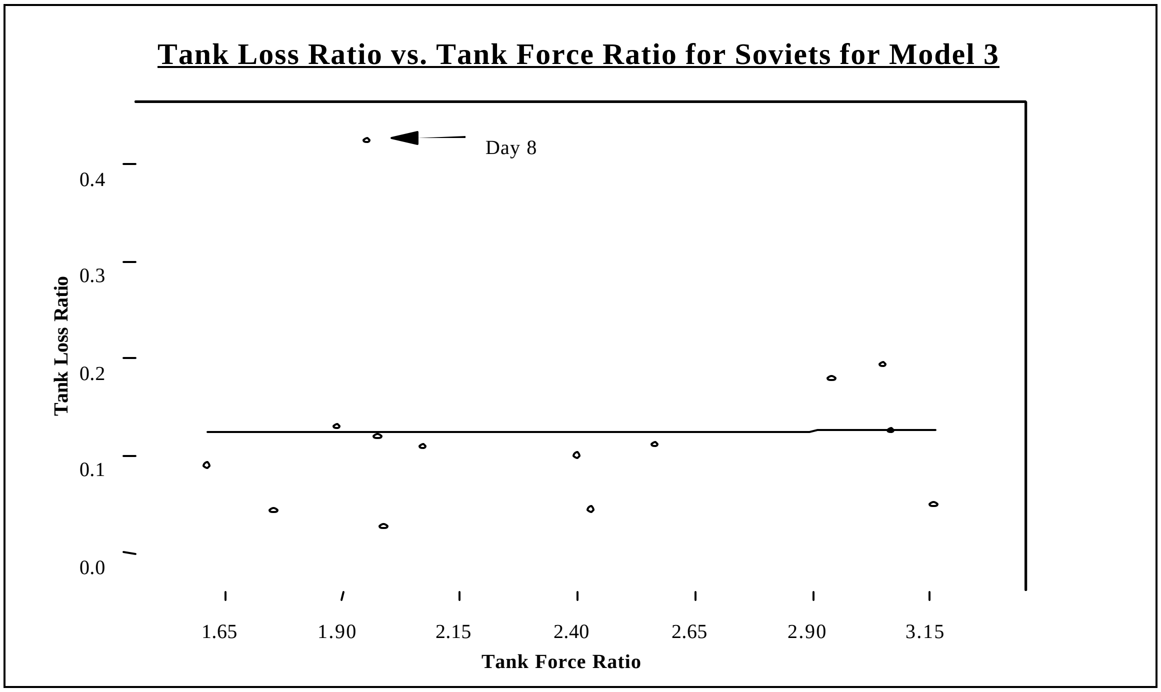

Figure 56. Tank loss ratio plotted versus tank force ratio for Soviet forces for model 3.

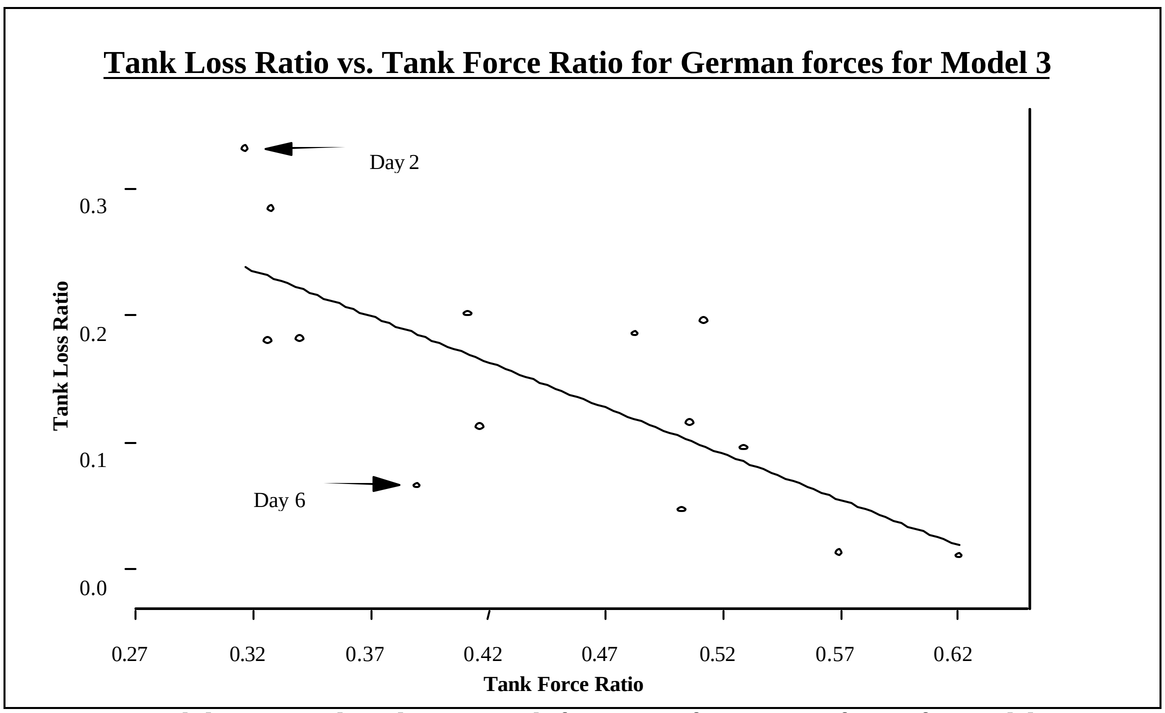

Figure 57. Tank loss ratio plotted versus tank force ratio for German forces for model 3.

$$(BTL / BT) = I_1 + I_2 + (BT / RT) \tag{75}$$

$$(RTL / RT) = I_1 + I_2 + (RT / BT) \tag{76}$$

where I_1 indicates the difference between blue force and red force, and I_2 indicates the difference between attacker and defender (time of battle) and have the values given in IV.B.7.(68) and IV.B.7.(69). BTL, RTL, BT and RT denote blue tank loss, red tank loss, number of blue tanks and number of red tanks, respectively.

Table 39 shows the coefficients, standard errors and t values for Model 3.

	Value	**Std. Error**	**t value**	**Pr(>|t|)**
I1	-0.2703	0.092	-2.9377	0.007
I2	0.1442	0.0291	4.9549	0
OTFR	0.1375	0.0367	3.747	0.0009

Table 39. Critical statistical values of the estimated coefficients of model 3.

The resulting model for model 3, which gives an SSR value of 0.220 and a Multiple R-Squared value of 0.7077 is:

$$PTL = -0.2703 I_1 + 0.1442 I_2 + 0.1375(OTFR) \tag{77}$$

where PTL and OTFR denote the percent of tank losses and opponent's tank force ratio, respectively for a given side.

In contrast to the results we found for Model 1 and Model 2, the negative coefficient of indicator variable I_1 indicates a Soviet advantage, and is significant. The positive coefficient of indicator variable I_2 indicates a defender advantage, and is also significant. The positive coefficient of the force ratio variable indicates that as the force ratio increases, so do the force's losses. Again this is statistically significant but intuitively does not make much sense.

The fourth model uses the total aggregated tank force ratios as a predictor to predict the FER of tanks for each side. The FER of tanks for the blue forces is equal to the percent of blue tank losses divided by percent of red tank losses, and likewise for the FER of tanks for the red forces. Figures 58 and 59 show the FER of tanks plotted against Force ratio of tanks for Soviet and German forces, respectively. The representation of Model 4 looks like:

$$(BTL/BT)/(RTL/RT) = I_1 + I_2 + (BT/RT) \tag{78}$$

$$(RTL/RT)/(BTL/BT) = I_1 + I_2 + (RT/BT) \tag{79}$$

where I_1 indicates the difference between the blue force and red force, and I_2 indicates the difference between attacker and defender and have the values given in IV.B.7.(68) and IV.B.7.(69). BTL, RTL, BT and RT denote blue tank loss, red tank loss, the number of blue tanks and the number of red tanks, respectively.

Table 40 shows the coefficients, standard errors, and t values for Model 4.

	Value	**Std. Error**	**t value**	**Pr(>|t|)**
I1	-0.1242	1.8843	-0.0659	0.948
I2	2.2276	0.5959	3.7382	0.001
OTFR	0.2865	0.7517	0.3811	0.7064

Table 40. Important statistical values of the estimated coefficients of model 4.

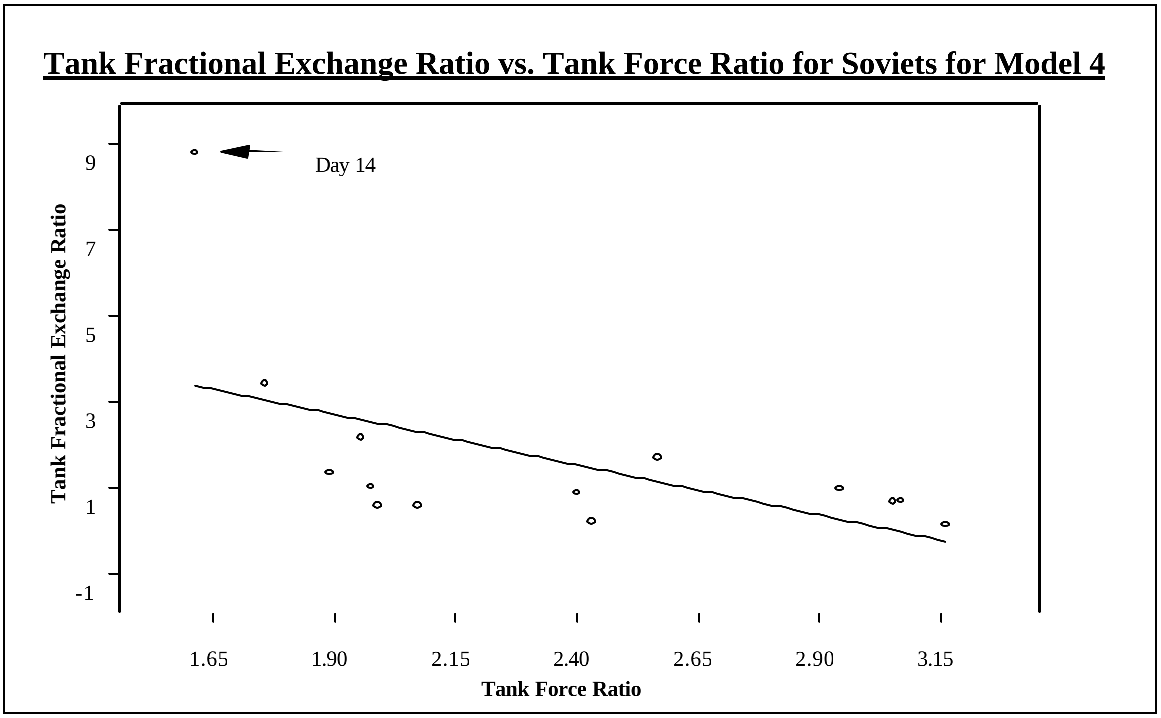

Figure 58. Tank fractional exchange ratio plotted versus tank force ratio for Soviet forces for model 4.

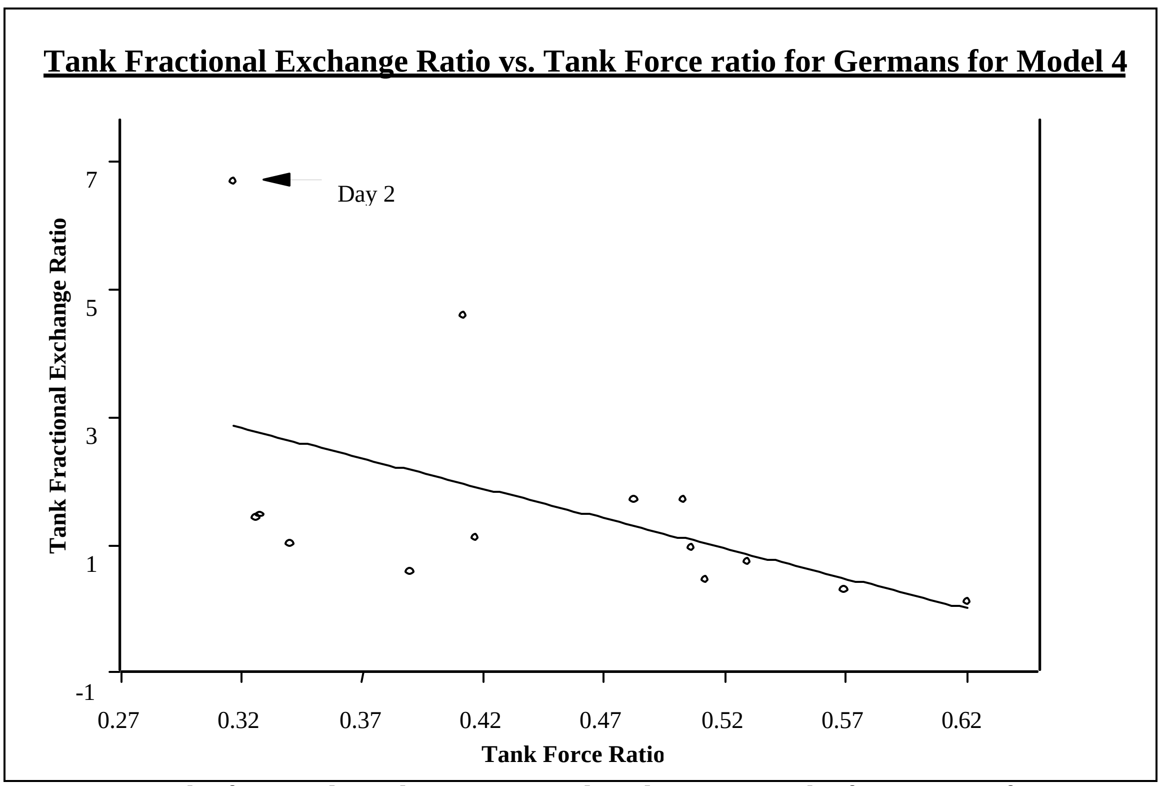

Figure 59.Tank fractional exchange ratio plotted versus tank force ratio for German forces for model 4.

The resulting model for Model 4, which gives an SSR value of 92.637 and a multiple R-squared value of 0.4941 is:

$$TFER = -0.1242 I_1 + 2.2276 I_2 + 0.2865(OTFR) \qquad (80)$$

where TFER and OTFR denote the FER of tanks and the opponent's tank FR for a given side.

Similar to the results we found for Model 3, the negative coefficient of indicator variable I_1, indicates a Soviet advantage and is not significant. The positive coefficient of indicator variable I_2 indicates a defender advantage and is significant. The positive coefficient of the force ratio variable indicates that as the force ratio increases so does your losses. Again the coefficient is not significant and intuitively does not make much sense.

The fifth model uses the same setup as Model 1, but it will do so by using the different weights first introduced in section IV.B.6 as the second weight combination, namely 1, 5, 15 and 20 for manpower, APC, artillery and tanks, respectively. Figures 60 and 61 show the loss ratio plotted against the force ratio for Soviet and German forces, respectively, using these weights. The presentation of model 5 looks like:

$$(\dot{B}/B) = I_1 + I_2 + (B/R) \qquad (81)$$

$$(\dot{R}/R) = I_1 + I_2 + (R/B) \qquad (82)$$

where I_1 indicates the difference between Blue force and Red force, and I_2 indicates the difference between attacker and defender and have the values given in IV.B.7.(68) and IV.B.7.(69).

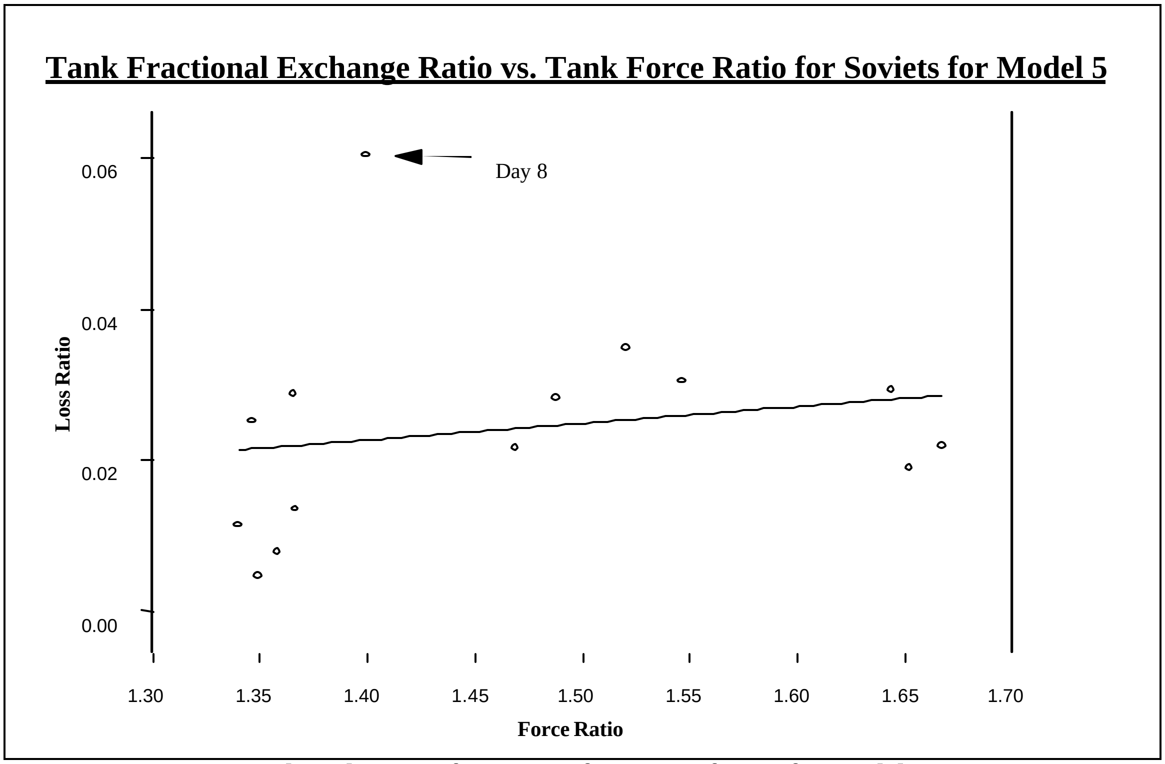

Figure 60. Loss ratio plotted versus force ratio for Soviet forces for model 5.

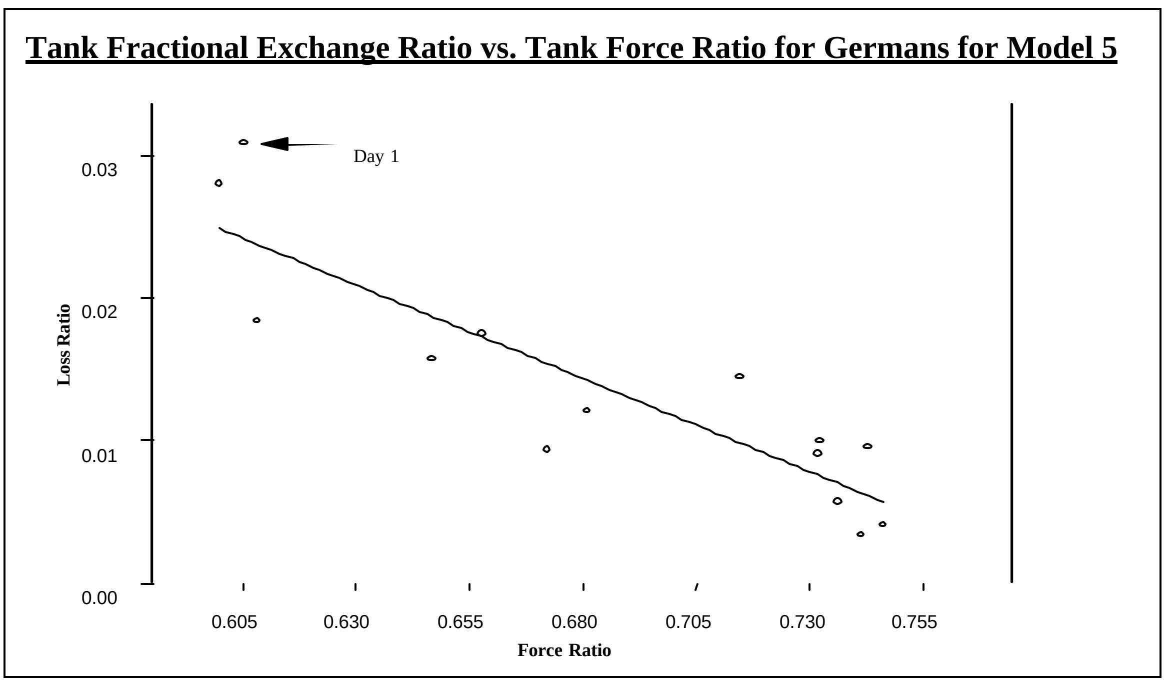

Figure 61. Loss ratio plotted versus force ratio for German forces for model 5.

where I_1 indicates the difference between Blue force and Red force, and I_2 indicates the difference between attacker and defender (time of battle) and have the values given in IV.B.7.(68) and IV.B.7.(69).

Table 41 shows the coefficients, standard errors, and t values for Model 5.

	Value	**Std. Error**	**t value**	**Pr(>|t|)**
I1	-0.0018	0.0072	-0.2544	0.8013
I2	0.0053	0.0039	1.3592	0.1862
OFR	0.0159	0.0049	3.2633	0.0032

Table 41. Important statistical values of the estimated coefficients of Model 5.

The resulting model for Model 5, which gives an SSR value of 0.0032546 and a multiple R-squared value of 0.7679 is:

$$PC = -0.0018I_1 + 0.0053I_2 + 0.0159(OFR) \tag{83}$$

where the notation has the same meaning as in Model 1.

Similar to the results we found for Model 3 and Model 4, the negative coefficient of indicator variable I_1 indicates a Soviet advantage and is not significant. The positive coefficient of indicator variable I_2 indicates a defender advantage and is not significant. The positive coefficient of the force ratio variable indicates that as the force ratio increases so do the losses. The coefficient is statistically significant and again, this interpretation intuitively does not make much sense.

In general, in the models we investigated in this section, the indicator variable I_2 is always positive, different from the result we found in the sections, which investigated the tactical parameter d. This observation suggests that it is advantageous to be the defender, not the attacker. Another interesting, yet ironic, result is the positive force ratio coefficient found in the models throughout the section, suggesting that the more powerful you are, the more you lose, which intuitively does not make much sense.

When the plots are investigated it is seen that, the higher the force ratio or FER is, the less the loss is, except for the Soviets in Model 1, Model 3 and Model 5. So, the results are telling somewhat different than what the plots are telling. This may be due to the interpretation that fitting the logarithmically transformed equations does not necessarily gives the best fit in the original form.

Table 42 summarizes the results found in this section.

	I1	**I2**	**Predictor**	**Multiple R-squared**
Model 1	0.001	0.0048	0.0147	0.7699
Model 2	1.1849	0.5647	0.4153	0.6963
Model 3	-0.2703	0.1442	0.1375	0.7077
Model 4	-0.1242	2.2276	0.2865	0.4941
Model 5	-0.0018	0.0053	0.0159	0.7679

Table 42. Results for the section investigating the force ratio and the fractional exchange ratio models.

When the overall results given in Table 42 are examined it is seen that Model 1 and Model 2 have positive I_1 coefficients, which indicates a German advantage while the rest of the models have negative I_1 coefficients, which indicates a Soviet advantage. All models have positive I_2 coefficients, which indicates a defender advantage. The first model with the highest multiple R-squared value gives the best fit.

8. Fitting the standard Lanchester equations

This section fits the basic Lanchester Equations, (i.e., Lanchester Linear, Lanchester Square and Lanchester Logarithmic models), to the Battle of Kursk data. The basic Lanchester equations are given in I.B.(1) and I.B.(2).

For the Lanchester linear model where p=q=1, the loss for one side will be equal to the product of the existing number of forces of both sides, and a coefficient. The Lanchester linear model will look like;

$$\dot{B} = aRB \tag{84}$$

$$\dot{R} = bBR \tag{85}$$

This model is solved like a typical regression through the origin equation and the resulting model for the Lanchester linear model, which gives an SSR value of 6.24×10^{8} is:

$$\dot{B} = 6.6834\times10^{-8}RB \tag{86}$$

$$\dot{R} = 2.6893\times10^{-8}BR \tag{87}$$

Figures 62 and 63 show fitted losses plotted versus real losses for Soviet and German forces, respectively, for the Lanchester linear model.

For the Lanchester Square Model, where p=1 and q=0, the loss for one side will be equal to the product of the existing number of forces of the opponent and a coefficient. The Lanchester square model will look like;

$$\dot{B} = aR \tag{88}$$

$$\dot{R} = bB \tag{89}$$

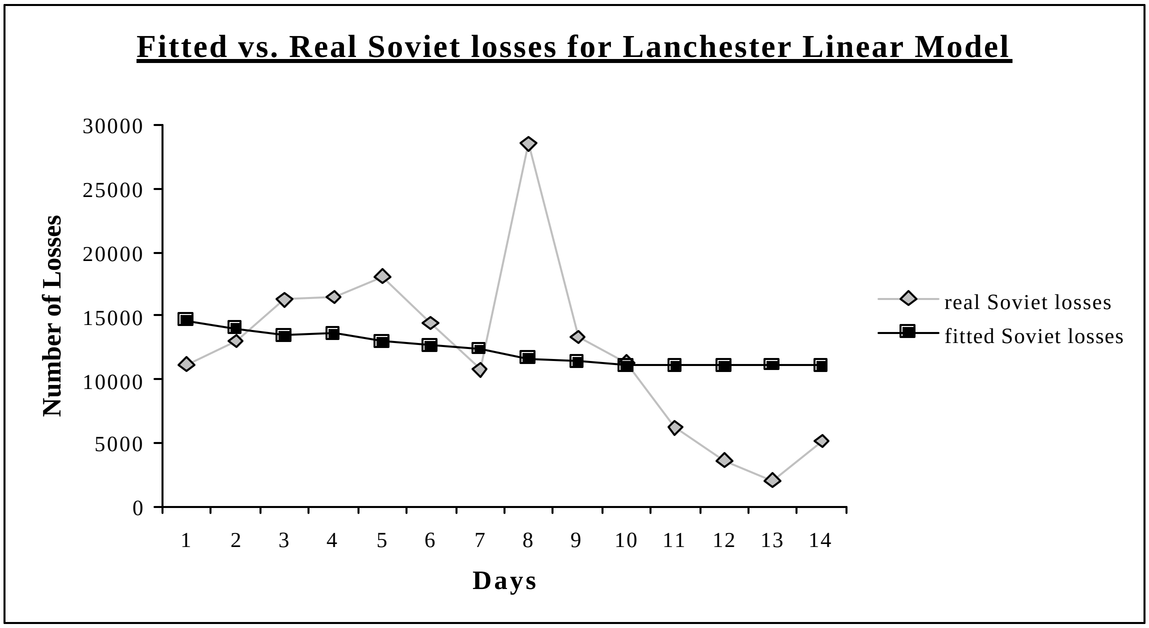

Figure 62. Fitted losses plotted versus real losses for Soviet forces for the Lanchester linear model. The same three-phase pattern where the model over/underestimates the battle in three distinctive phases is observable in this plot for the model, which uses the Lanhester linear model, too.

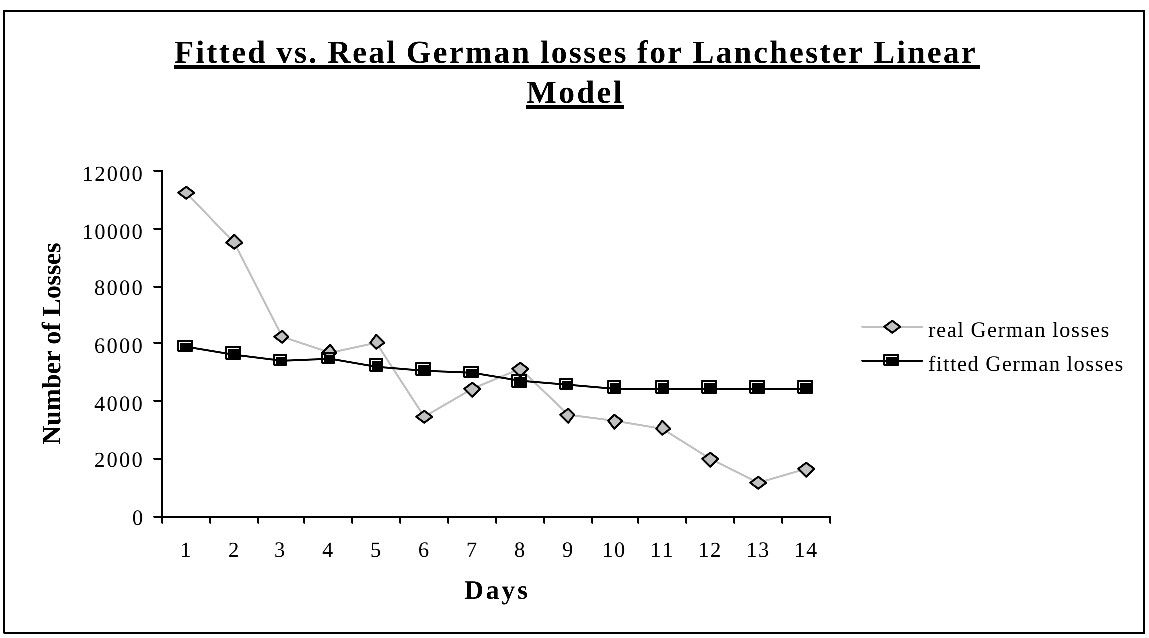

Figure 63. Fitted losses plotted versus real losses for German forces for the Lanchester linear model. Eight days are underestimated while six days are overestimated.

The resulting model for Lanchester square model that gives an SSR value of 6.79×10^8 is:

$$\dot{B} = 0.0335R \tag{90}$$

$$\dot{R} = 0.0098B \tag{91}$$

The high value of the a parameter in the above equation indicates that the Germans fought three times better than the Soviets.

Figures 64 and 65 show fitted losses plotted versus real losses for Soviet and German forces, respectively, for the Lanchester square model.

For the Lanchester logarithmic model where $p=0$ and $q=1$, the loss for one side will be equal to the product of the existing number of forces of its own and a coefficient. Lanchester logarithmic model will look like:

$$\dot{B} = aB \tag{92}$$

$$\dot{R} = bR \tag{93}$$

The resulting model for Lanchester logarithmic model, which gives an SSR value of 6.57×10^8 is:

$$\dot{B} = 0.0243B \tag{94}$$

$$\dot{R} = 0.0131R \tag{95}$$

Figures 66 and 67 show fitted losses plotted versus real losses for Soviet and German forces, respectively, for the Lanchester logarithmic model.

The basic Lanchester Equations do not give the best fit for the Battle of Kursk data. Out of the three Lanchester Models analyzed, the Lanchester linear model gives the best fit (i.e., smallest SSR value).

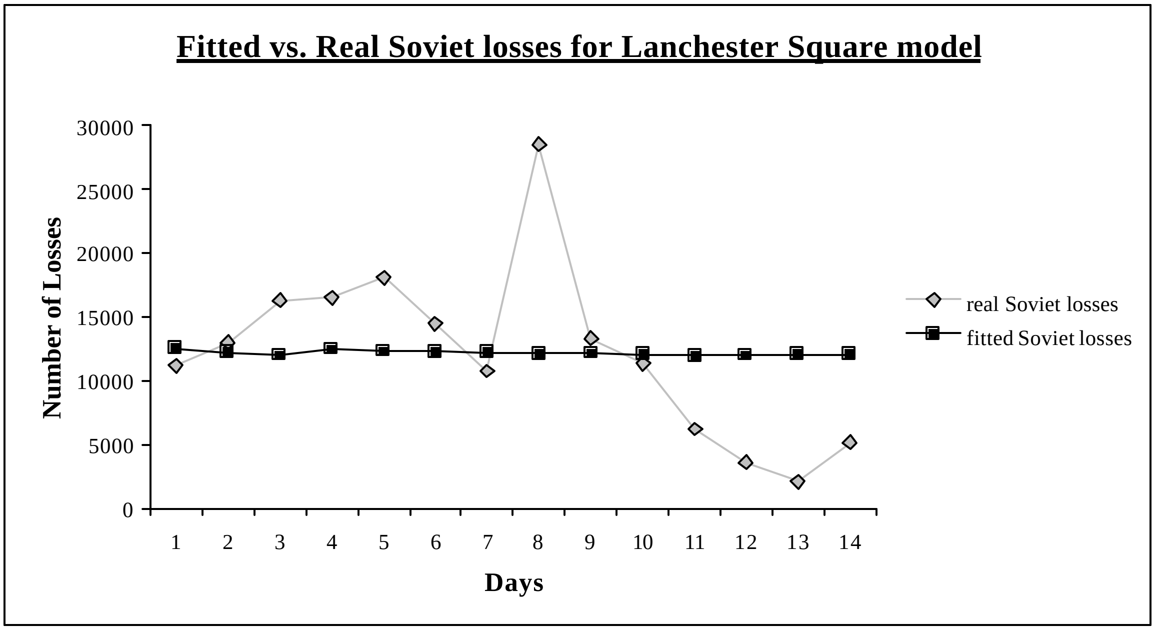

Figure 64. Fitted losses plotted versus real losses for Soviet forces for the Lanchester square model. The same three-phase pattern where the model over/underestimates the battle in three distinctive phases is observable in this plot for the model, which uses the Lanhester square model, too.

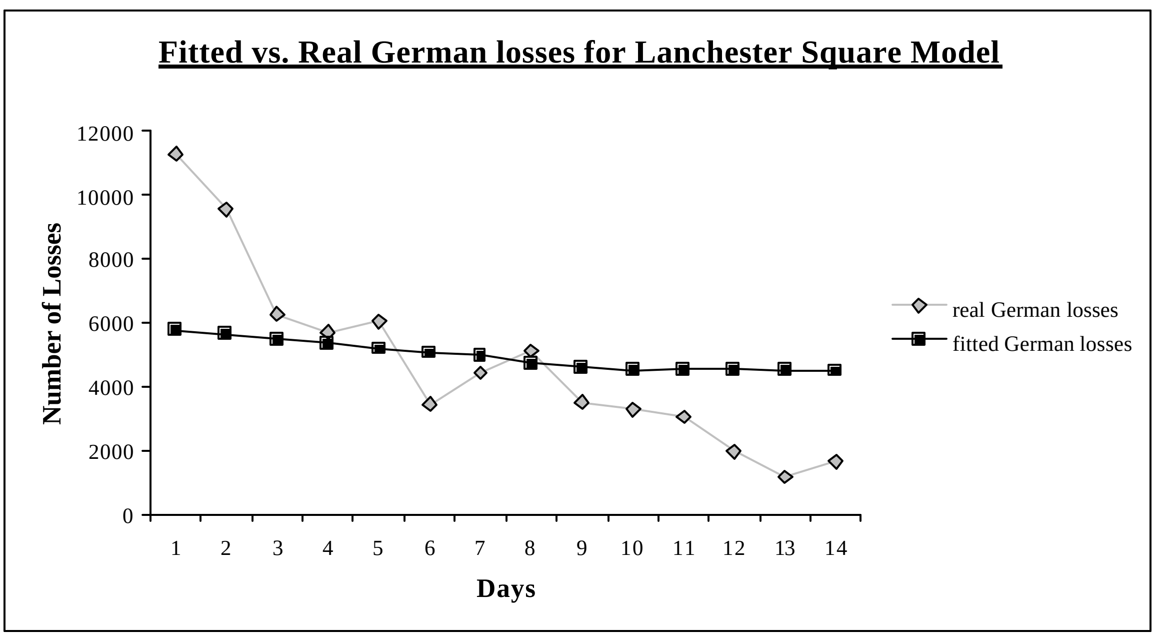

Figure 65. Fitted losses plotted versus real losses for German forces for the Lanchester square model. Eight days are underestimated while six days are overestimated.

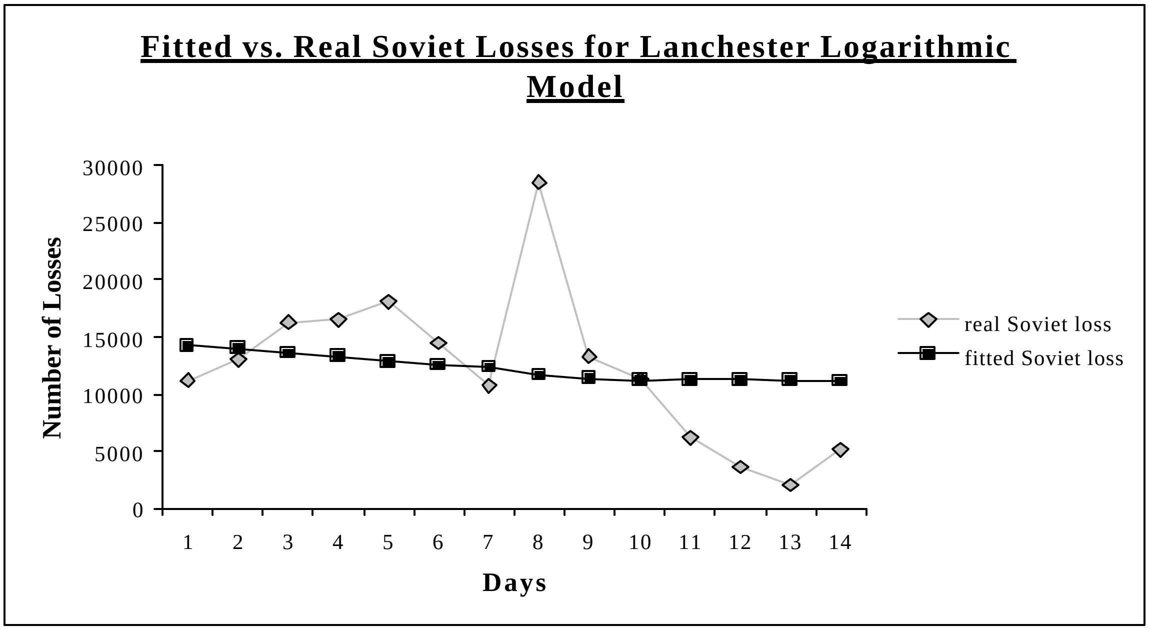

Figure 66. Fitted losses plotted versus real losses for Soviet forces for the Lanchester logarithmic model.

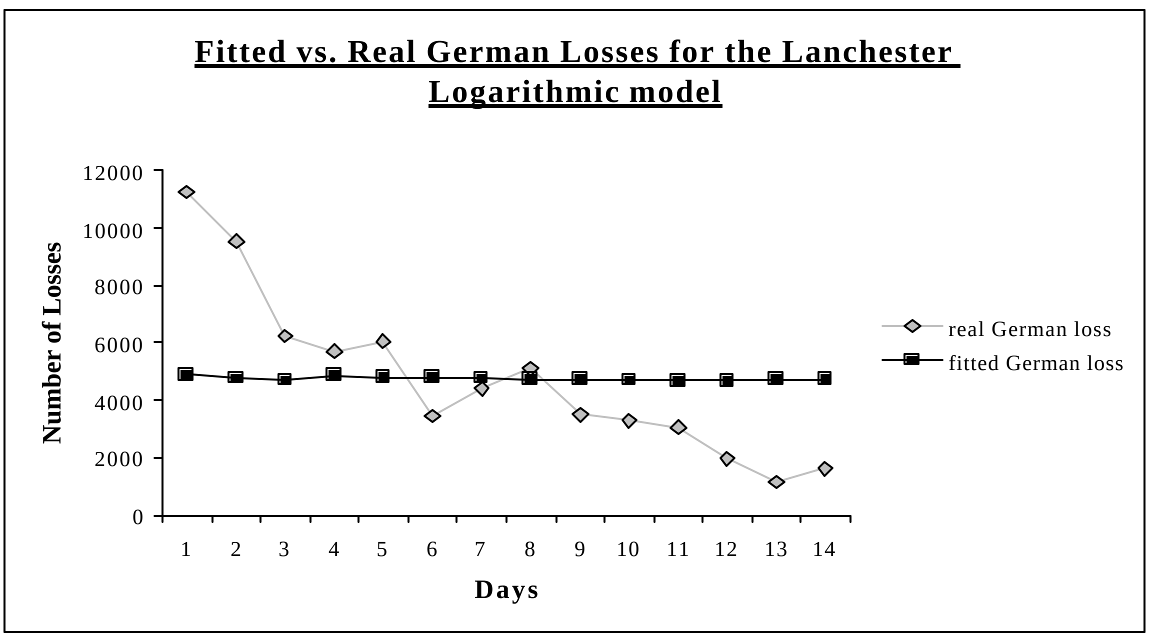

Figure 67. Fitted losses plotted versus real losses for German forces for the Lanchester logarithmic model.

Again in all Lanchester Models, the a and b parameters are significantly small and $a>b$.

Fricker's findings were closest to Lanchester's logarithmic model, while Bracken's findings were closest to Lanchester's linear model. Out of the three basic Lanchester models, it is the Lanchester linear model that best fits the Battle of Kursk data. The Lanchester logarithmic model gives the second best fit for the Battle of Kursk data, while the Lanchester square model gives the third best (i.e., the worst) fit for the Battle of Kursk data.

9. Fitting Morse-Kimball equations

This section will fit the Morse-Kimball Equations to the Battle of Kursk data. Morse and Kimball suggest that one side's losses do not depend solely on the opponent's forces, losses also depend on one's own failures and other mechanical breakdowns too, like the case in the logarithmic law. The Morse-Kimball Equations are:

$$\dot{B} = aR + \alpha_1 B \tag{96}$$

$$\dot{R} = bB + \alpha_2 R \tag{97}$$

These equations are fit separately for the Germans and the Soviets, and the resulting model for the Morse-Kimball Equations, which gives an SSR value of 5.51×10^8 and an R^2 value of 0.2297 is:

$$\dot{B} = -0.0412R + 0.0537B \tag{98}$$

$$\dot{R} = 0.0603B - 0.0707R \tag{99}$$

Figures 68 and 69 show fitted losses plotted versus real losses for Soviet and German forces, respectively, for the Morse-Kimball Equations model.

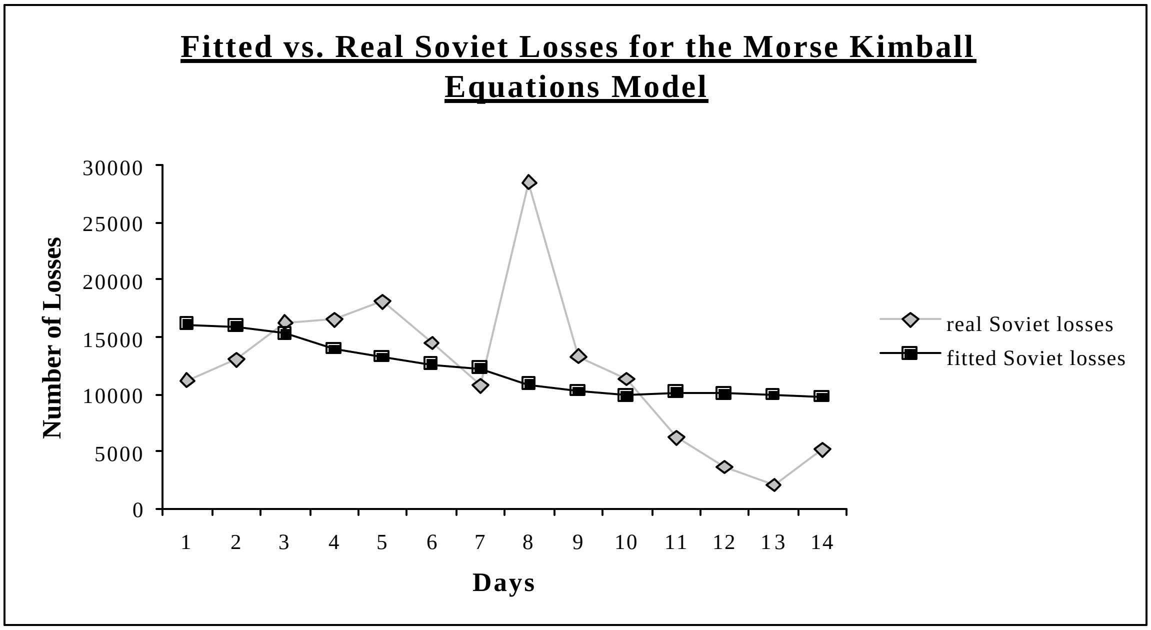

Figure 68. Fitted losses plotted versus real losses for the Soviet forces for the model using Morse Kimball Equations. The same three-phase pattern where the model over/underestimates the battle in three distinctive phases is observable in this plot for the model, which uses Morse Kimball equations, too.

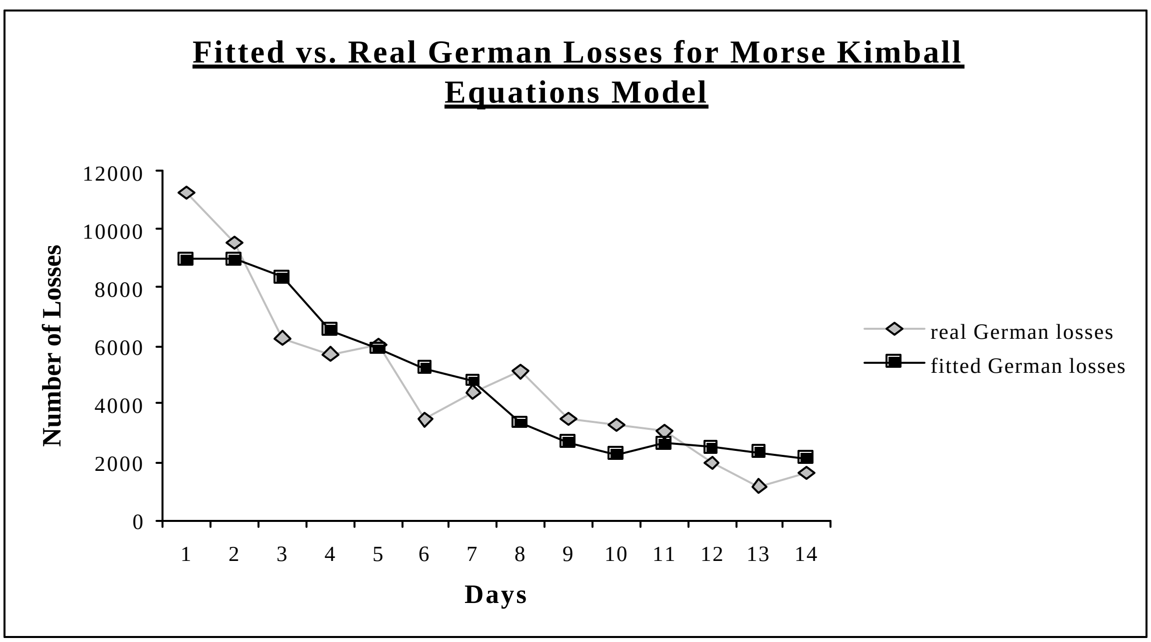

Figure 69. Fitted losses plotted versus real losses for the German forces for the Morse Kimball Equations Model.

Fitting Morse-Kimball Equations to the Battle of Kursk data improves the fit. The SSR value of 5.51×10^8 is one of the lowest SSR values we have so far. But, just as in the models used for the change points approach for each side in section IV.B.5, the parameters physically do not make sense.

For the blue force, the negative a parameter indicates that the more the red forces there are, the less the number of blue casualties. For the red force, the negative a_2 parameter indicates that the greater the number of the red forces is, fewer red casualties are going to be. This physically does not make much sense; so, even if fitting Morse-Kimball equations give a low SSR value of 5.51×10^8, we cannot accept this fit.

10. Fitting the parameters found by Bracken and Fricker

In this section, the parameters for the Ardennes data found in Bracken and Fricker's studies will be used to fit the Battle of Kursk data.

a. Bracken's parameters

In his study, Bracken's conclusion for the Lanchester Model with the tactical parameter is given as:

$$\dot{B} = 8 \times 10^{-9} \left(\frac{8}{10} \text{ or } \frac{10}{8}\right) R^1 B^1 \tag{100}$$

$$\dot{R} = 1 \times 10^{-8} \left(\frac{10}{8} \text{ or } \frac{8}{10}\right) B^1 R^1 \tag{101}$$

Figures 70 and 71 show fitted losses plotted versus real losses for Soviet and German forces, respectively, for the parameters of Bracken's model (with the tactical parameter) given above, which yields an SSR value of 2.39×10^9 for the Battle of Kursk data.

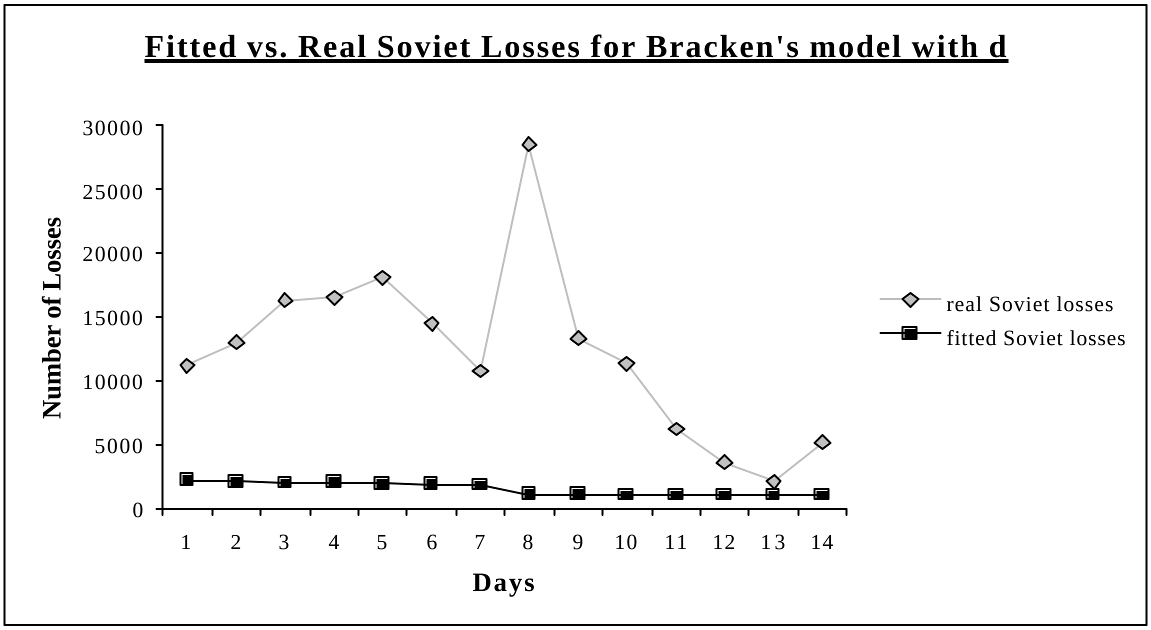

Figure 70. Fitted losses plotted versus real losses for Soviet forces for Bracken's model with the tactical parameter *d*. Bracken's Ardennes parameters always underestimated the Soviet losses for the Battle of Kursk.

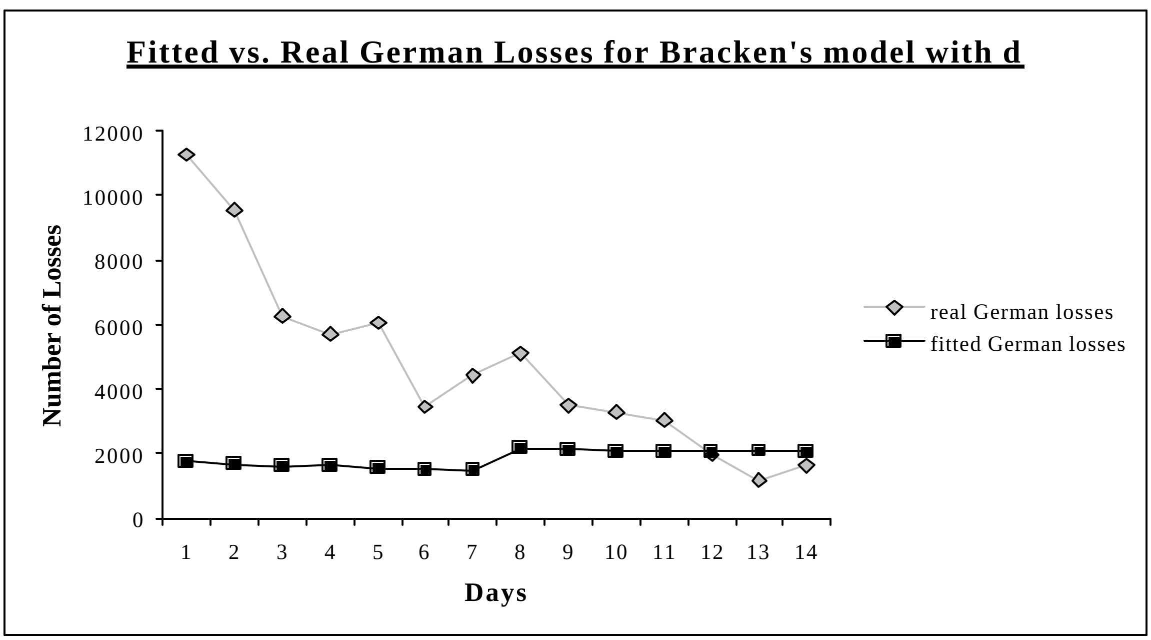

Figure 71. Fitted losses plotted versus real losses for German forces for Bracken's model with the tactical parameter *d*. Except the last three days of the battle, Bracken's Ardennes parameters always underestimated the German losses for the whole Battle of Kursk data.

Bracken's conclusion for the Lanchester model without the tactical parameter is given as:

$$\dot{B} = 8\times10^{-9} R^{1.3} B^{0.7} \tag{102}$$

$$\dot{R} = 1\times10^{-8} B^{1.3} R^{0.7} \tag{103}$$

Figures 72 and 73 show fitted losses plotted versus real losses for Soviet and German forces, respectively, for the parameters of Bracken's model (without the tactical parameter) given above, which yields an SSR value of 2.46×10^{9} for the Battle of Kursk data.

Fitting Brackens's parameters to the Battle of Kursk data does not improve the model's fit and gives the highest SSR value thus far. It is significant that Bracken's parameters always underestimates the real casualties for the Battle of Kursk data.

b. Fricker's parameters

In Fricker's study, the conclusion for the Lanchester model with the tactical parameter is given as:

$$\dot{B} = 4.7\times10^{-27}\left(\frac{1}{0.8093} \text{ or } 0.8093\right)B^{5} \tag{104}$$

$$\dot{R} = 3.1\times10^{-26}\left(0.8093 \text{ or } \frac{1}{0.8093}\right)R^{5} \tag{104}$$

Figures 74 and 75 show fitted losses plotted versus real losses for Soviet and German forces, respectively, for the parameters of Fricker's model (with the tactical parameter) given above that yields an SSR value of 3.02×10^{9} for the Battle of Kursk data.

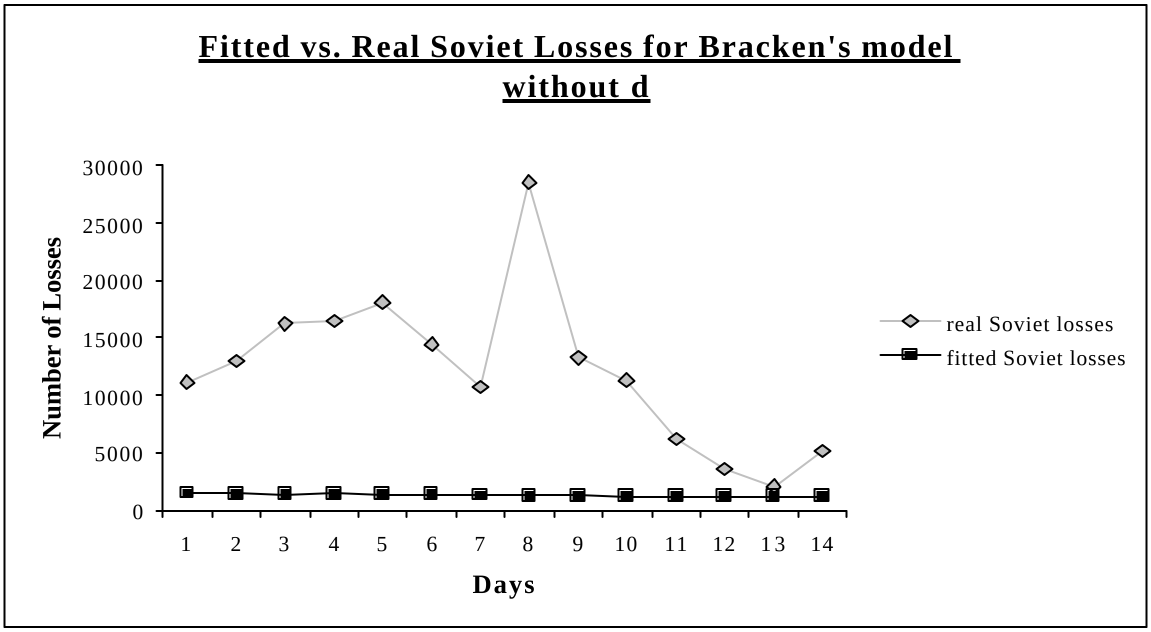

Figure 72. Fitted losses plotted versus real losses for the Soviet forces for Bracken's model without the tactical parameter *d*. Bracken's Ardennes parameters always underestimated the Soviet losses for the Battle of Kursk data.

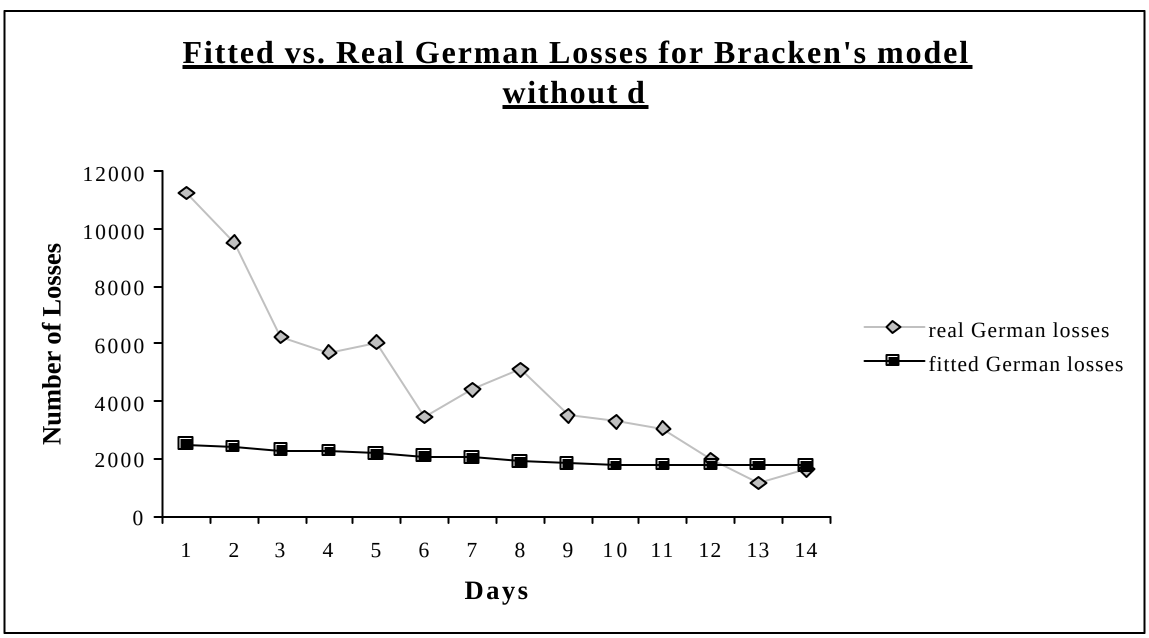

Figure 73. Fitted losses plotted versus real losses for the German forces for Bracken's model without the tactical parameter *d*. With the exception of the last three days of the battle, Bracken's Ardennes parameters always underestimated the German losses for the whole Battle of Kursk data.

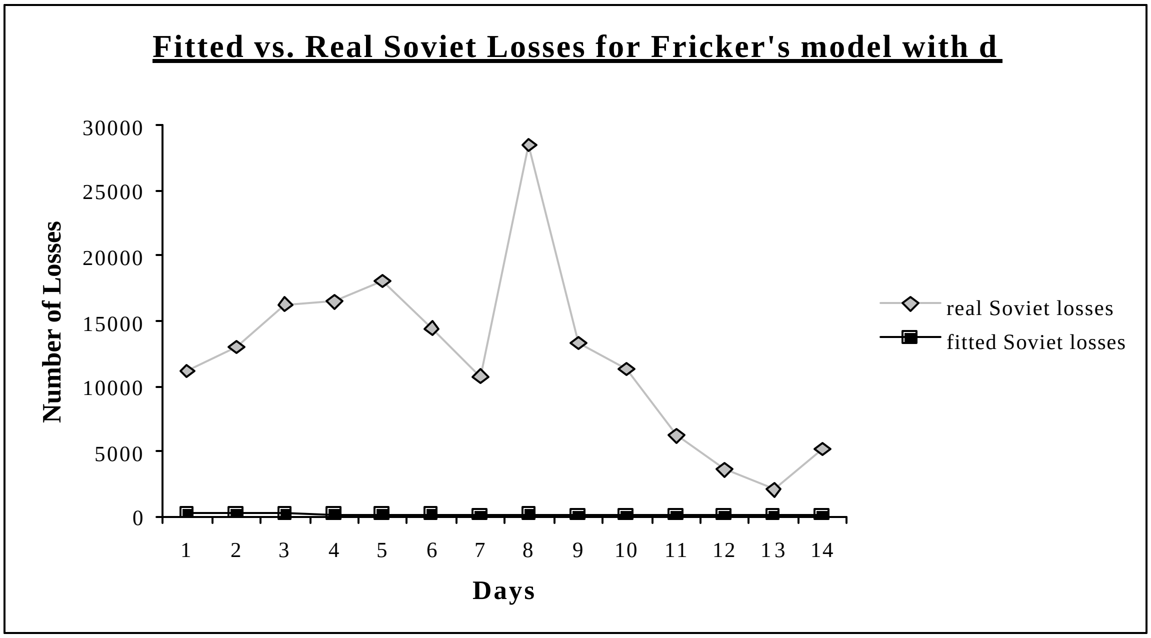

Figure 74. Fitted losses plotted versus real losses for Soviet forces for Fricker's model with the tactical parameter *d*. Fricker's Ardennes parameters always underestimated the Soviet losses for the Battle of Kursk.

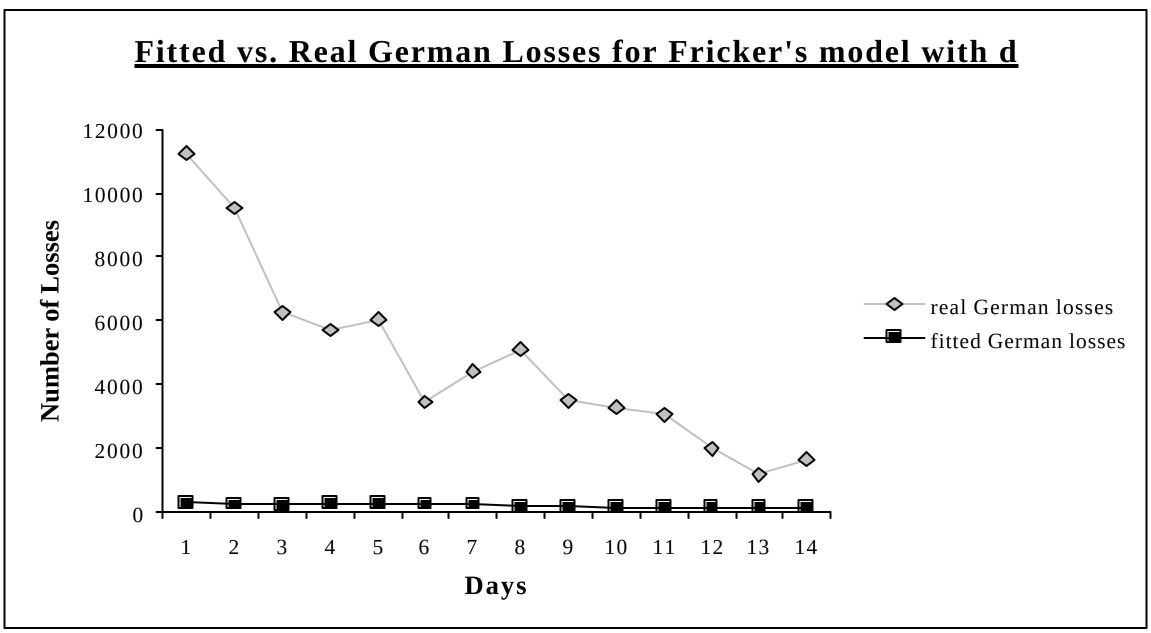

Figure 75. Fitted losses plotted versus real losses for Soviet forces for Fricker's model with the tactical parameter *d*. Fricker's Ardennes parameters always underestimated the German losses for the Battle of Kursk.

Fricker's conclusion for the Lanchester model with the air sortie data added is given as:

$$\dot{B} = 2.7 \times 10^{-24} \left(\frac{1}{0.7971} \text{ or } 0.7971 \right) B^{4.6} \quad (106)$$

$$\dot{R} = 1.6 \times 10^{-23} \left(0.8093 \text{ or } \frac{1}{0.8093} \right) R^{4.6} \quad (107)$$

Figures 76 and 77 show fitted losses plotted versus real losses for Soviet and German forces, respectively, for the parameters of Fricker's model (with air sortie data added) given above, which yields an SSR value of 2.77×10^{9} for the Battle of Kursk data.

Like Bracken's models, fitting Fricker's parameters to the Battle of Kursk data does not improve the model's fit, it gives the highest SSR value in this study so far. Fricker's parameters always underestimate the real casualties for the Battle of Kursk data. This finding is similar to the one for Bracken's parameters.

In general, fitting Bracken's or Fricker's Ardennes parameters to the Battle of Kursk data does not improve the fit; they both give the highest SSR value we have in this study so far. This result suggests that the parameters of one battle data cannot be used to predict another. Each battle has its own unique parameters which cannot be applied to another one battle.

Another interesting finding is that when Bracken's or Fricker's Ardennes parameters are applied to Kursk data, they always underestimate the daily attrition rates. This finding suggests that Battle of Kursk was a much more intense battle than the Ardennes campaign.

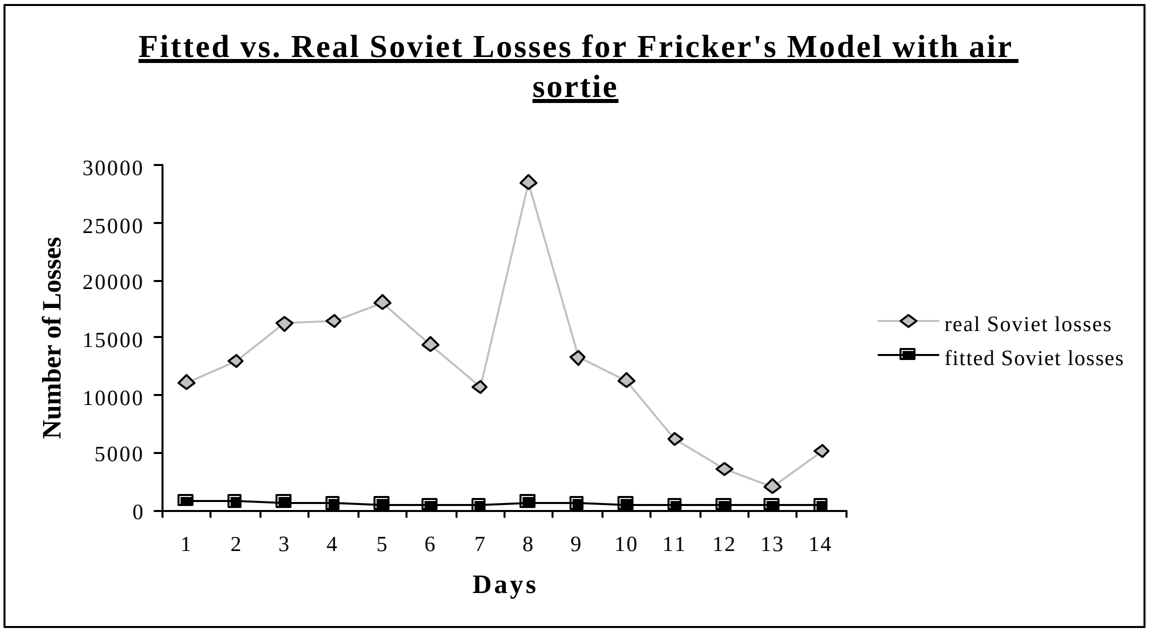

Figure 76. Fitted losses plotted versus real losses for Soviet forces for Fricker's model with the air sortie data added. Notice that Fricker's Ardennes parameters always underestimated the Soviet losses for the Battle of Kursk.

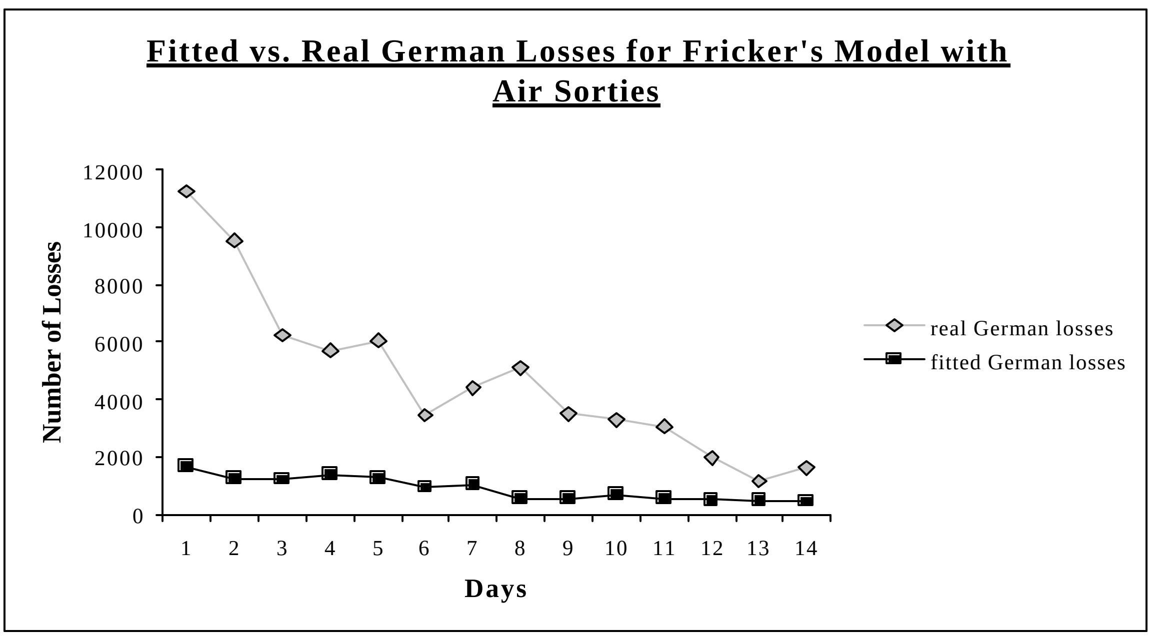

Figure 77. Fitted losses plotted versus real losses for German forces for Fricker's model with air sortie data added. Notice that Fricker's Ardennes parameters always underestimated the German losses for the Battle of Kursk.

11. Summary of results

This section summarizes the results of all the models explored in previous chapters.

Name of the model	*a*	*b*	*p*	*q*	*d*	*SSR*	R^2
Bracken Model 1 Ardennes	8.0E-9	1.0E-8	1.0	1.0	1.25	1.63E+9	0.2552
Bracken Model 2 Ardennes	8.0E-9	1.0E-8	1.3	0.7	1.0	2.08E+8	0.0493
Bracken Model 1 Kursk	1.2E-8	9.0E-9	0.1	2.0	0.9	8.65E+8	0.0006
Bracken Model 3 Kursk	1.2E-8	9.0E-9	0.3	1.8	1.0	8.88E+8	-0.0266
Frick.Ard. w/o sorties with d	4.7E-27	3.1E-26	0	5	0.8093	1.57E+8	-0.7938
Frick.Ard. w sorties with d	2.7E-24	1.6E-23	0	4.6	0.7971	2.64E+7	0.5256
Frick.Kursk w/o sorties with d	3.76E-33	1.09E-32	0.0604	6.3066	0.79	5.94E+8	0.1703
Frick.Kursk w/o sorties w/o d	1.61E-33	3.44E-33	3.6736	2.6934	-	2.16E+9	0.0657
Frick.Kursk with sorties with d	3.35E-27	5.76E-27	0.0955	5.2207	0.93	6.23E+8	0.1294
Frick.Kursk with sorties w/o d	5.01E-27	3.85E-27	1.4983	3.8179	-	7.16E+8	-0.0222
Clemens Linear Regression	6.92E-49	6.94E-48	5.3157	3.6339	-	1.13E+9	0.9975
Clemens Newton-Raphson	3.73E-6	5.91E-6	0.0	1.6178	-	1.04E+9	-0.6242
Linear Regression Model	1.06E-47	1.90E-48	5.7475	3.3356	-	6.36E+8	0.1126
Robust LTS Regression	2.27E-40	1.84E-41	6.0843	1.7312	-	5.54E+8	0.2262

Name of the model	***a***	***b***	***p***	***q***	***d***	***SSR***	R^2
Lin.Reg. With Air sorties	1.40E-30	2.09E-36	5.1323	1.7793	-	6.85E+8	0.0433
Robust LTS with Air sorties	1.21E-38	1.75E-39	5.3691	2.0883	-	7.58E+8	-0.0579
Linear Regression With *d*	1.88E-47	1.07E-48	7.5038	1.5793	1.17	6.24E+8	0.1295
Robust LTS With *d*	2.27E-40	1.84E-41	6.0843	1.7312	1.0	5.54E+8	0.2262
Campaign in four Parts	1.88E-47	1.07E-48	7.5038	1.5793	4 periods *d*=0.91,1.24, 1.0,1.17	5.34E+8	-2.3410
Campaign in four Parts	1.88E-47	1.07E-48	7.5068	1.5793	4 periods *d*=0.91,1.24, 0.32,1.17	1.69E+8	-0.0607
Campaign in four Parts	1.88E-47	1.07E-48	7.5038	1.5793	1.14	1.89E+8	0.5689
Campaign in four Parts	1.85E-51	3.56E-53	9.6853	0.1458	-	1.90E+8	0.5658
Change Point 7/7	8.91E-30	2.62E-31	6.4117	-0.4323	-	1.53E+8	0.7448
Change Point 7/7	1.90E-232	4.37E-291	18.0587	34.4502	-	1.53E+8	0.7448
Change Point 8/6	7.75E-5	1.91E-6	4.4212	-2.8454	-	2.43E+8	0.3488
Change Point 8/6	1.94E-246	1.32E-247	25.7652	18.7674	-	2.43+E8	0.3488
Weight comb.1 Lin.Reg.	1.25E-38	1.60E-39	5.2298	2.2746	-	1.15E+9	0.0870
Weight Comb.1 Rob.LTS	7.26E-35	5.53E-36	5.5312	1.3268	-	1.07E+9	0.1514
Weight comb.2 Lin.Reg.	2.50E-46	3.49E-47	5.7638	3.1222	-	6.24E+8	0.0975
Weight Comb.2 Rob.LTS	7.85E-36	4.75E-37	5.8613	1.1899	-	5.48E+8	0.2072
Weight comb.3 Lin.Reg.	3.78E-39	5.34E-40	5.2293	2.3513	-	1.15E+9	0.0926

Name of the model	*a*	*b*	*p*	*q*	*d*	*SSR*	R^2
Weight Comb.3 Rob.LTS	1.46E-35	9.33E-37	5.9619	1.0159	-	1.06E+9	0.1637
Weight comb.4 Lin.Reg.	2.89E-42	3.91E-43	5.4863	2.6660	-	8.63E+9	0.0943
Weight Comb.4 Rob.LTS	5.05E-35	3.51E-36	5.6294	1.2631	-	7.74E+8	0.1873
Lanchester Linear model	6.68E-8	2.68E-8	1.0	1.0	-	6.24E+8	0.1290
Lanchester Square model	0.0335	0.0098	1.0	0	-	6.79E+8	0.0521
Lanchester Logarithmic model	0.0243	0.0131	0	1.0	-	6.57E+8	0.0831
Morse Kimball Equations	a=-0.041	a_1=0.053	b=0.060	a_2 =-0.07	-	5.51E+8	0.2297
Bracken's Parameters with *d*	8.0E-9	1.0E-8	1.0	1.0	1.25	2.39E+9	-2.4235
Bracken's Parameters w/o *d*	8.0E-9	1.0E-8	1.3	0.7	-	2.46E+9	-2.4430
Fricker's Parameters with *d*	4.7E-27	3.1E-26	0	5.0	0.8093	3.02E+9	-3.2123
Fricker's Par.s with air sortie	2.7E-24	1.6E-23	0	4.6	0.7971	2.79E+9	-2.9021

Table 42. Results of all the models explored and investigated in Chapter IV.

The R^2 value (0.9975) given for Clemens' linear regression model is the self reported value by Clemens and must have been calculated differently than the R^2 values calculated throughout the thesis. When recomputed, a negative R^2 value is found.

Clemens provided four digits of precision in his estimates of *p* and *q*, while Bracken and Fricker gave two. The R^2 values that are found for the models, which do not use the parameters from other studies, are calculated using parameters with four digits of precision. The slightly negative R^2 values found for some of these models are not a result of using low precision.

When the above results are examined, it is seen that the best fitting model for the Battle of Kursk data is the robust LTS regression model used in section IV.B.1, with an SSR value of 5.54×10^8 and an R^2 value of 0.2262. This finding is true for the models that handle the battle in one phase.

When the models which consider the change points are examined, it is seen that the model with the change point 7/7 is the one with the best fit, with an SSR value of 1.53×10^8 and an R^2 value of 0.7448.

Figure 78 shows the p and q values plotted for every model whose parameters are given in Table 42, except for the models with the change points since they have very large p and q parameters. The p and q values are also excluded for the model using the Morse-Kimball equations since these equations do not use p and q parameters.

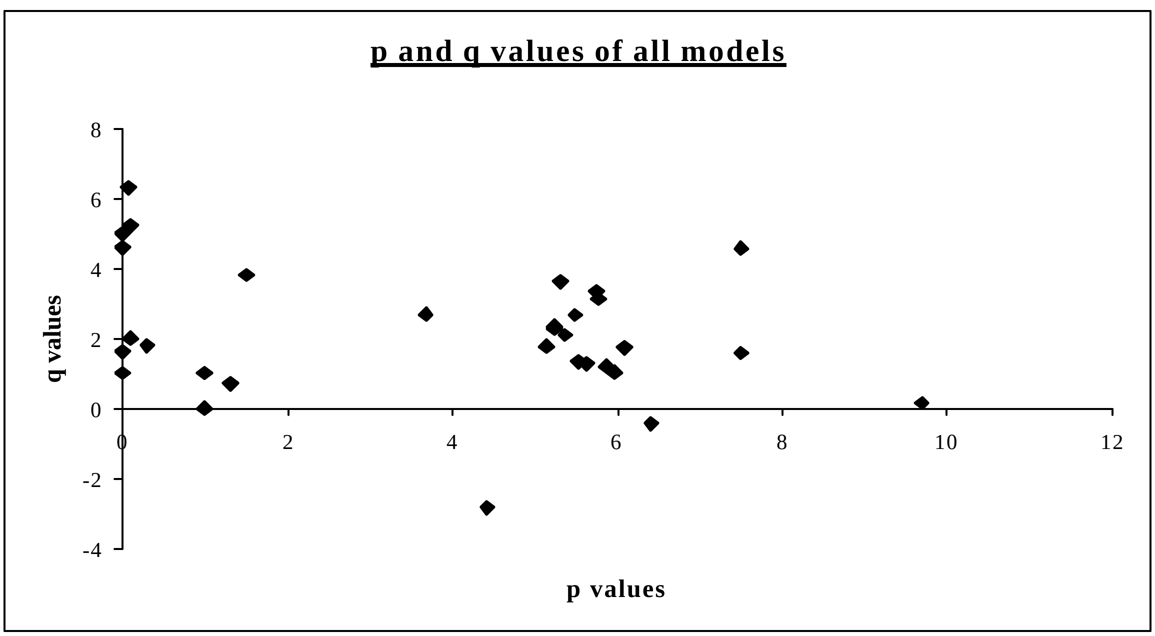

Figure 78. p and q parameters plotted for all the models given in Table 42.

When the pattern seen in Figure 78 is examined it is apparent that p and q parameters are clustered in two regions—one around the $p=5\text{-}6$, $q=1\text{-}4$ region, and the other around the $q=1\text{-}6$, $p=0$ region.