The Korean Journal of Applied Statistics (2020) DOI: https://doi.org/10.5351/KJAS.2020.33.3.335
33(3), 335–345

Statistical review and explanation for Lanchester model

Byung Joo Yoo[a,1]

[a]Operations Analysis Branch, Ground Operations Command

(Received March 17, 2020; Revised April 24, 2020; Accepted May 4, 2020)

Abstract

This paper deals with the problem of estimating the log-transformed linear regression model to fit actual battle data from the Ardennes Campaign of World War II into the Lanchester model. The problem of determining a global solution for parameters and multicollinearity problems are identified and modified by examining the results of previous studies on data. The least squares method requires attention because a local solution can be found rather than a global solution if considering a specific constraint or a limited candidate group. The method of exploring this multicollinearity problem can be confirmed by a statistic known as a variance inflation factor. Therefore, the Lanchester model is simplified to avoid these problems, and the combat power attrition rate model was proposed which is statistically significant and easy to explain. When fitting the model, the dependence problem between the data has occurred due to autocorrelation. Matters that might be underestimated or overestimated were resolved by the Cochrane-Orcutt method as well as guaranteeing independence and normality.

Keywords: Lanchester combat model, combat power attrition rate model, autocorrelation, multicollinearity, Ardennes campaign

1. 서론

과거의 역사적 경험을 반영하여 미래를 예측하고자 하는 방법으로 복잡한 현실을 단순화하여 모형화하고 이를 반복적으로 시뮬레이션하여 교훈을 도출하는 경우가 많다. 군사학 분야에서는 다루고자 하는 주제의 특수성과 위험성으로 인하여 실제 실험을 반복적으로 하거나 시행착오를 여러 번 경험할 수 없는 경우가 많다. 이러한 특수한 환경으로 교육 훈련이나 작전계획 분석에 컴퓨터 시뮬레이션을 많이 활용하는데 이러한 도구 개발의 기반 이론이 Lanchester (1916) 모형이다. 이는 두 부대 간의 상대적 전투력의 관계를 미분방정식으로 표현하고 시간 경과에 따라 발생할 수 있는 피해를 예측할 수 있게 해준다.

란체스터 모형을 다양한 상황에 적용하기 위하여 수학적이나 통계적 방법으로 접근한 연구들이 많이 있다. Morse와 Kimball (1951)은 수학적 접근 방법으로 조준사격이나 조우전하에서의 란체스터 선형모형을 발전시켰으며, Weiss (1966)는 미국의 남북전쟁에 관한 연구를, Peterson (1967)은 기갑 전투에서의 란체스터 지수모형 또는 자승모형을 적용하고자 노력하였다. 또한 Dietchman (1962)과 Schaffer (1968)는 비정규전을 위한 란체스터 모형을 제안하였고, Hartley와 Helmbold (1995)은 한국전쟁 당시 인천상륙작전 중 미 해병 1사단과 북한군 자료를 활용하여 란체스터 동질 지수모형을 제시하였다. 이들

[1]Operations Analysis Branch, Ground Operations Command PO Box 505-1-3, Seongsan-ro 57, Cheoin-gu, Yongin-si, Gyeonggi-do 17018, Republic of Korea. E-mail: bjyoo@korea.ac.kr

은 대부분 주어진 자료에 대한 최적화 측면에서 분석을 시도하였으며, 통계적 방법으로 접근한 사례는 Bracken (1995)과 Fricker (1998)가 2차 세계대전 중 Ardennes 전역에 대한 자료를 이용하여 란체스터 모형을 발전시켰으며, Wiper 등 (2000)은 이 자료에 대한 베이지안 추론을 시도하였다.

본 논문에서는 Bracken (1995)과 Fricker (1998)가 적용한 모형을 집중적으로 재조명하고 새로운 모형을 제안하고자 한다. 새로운 모형을 제안하기 위한 기준으로는 첫 번째 통계적으로 유의한 모형이어야 하고, 두 번째는 해석과 설명이 용이한 모형을 우선적으로 고려하여 정확하게 모수를 추정하고자 한다.

2. 란체스터 모형 소개

2.1. 란체스터 일반화모형

Lanchester (1916)는 현대전의 전투손실률(casualty rate)은 전투원의 화기 발사능력 때문에 상대 군의 전투원 수에 비례할 것이라고 생각하였다. 이를 수학적으로 표현하면 아래와 같은 미분방정식이 된다.

$$\begin{aligned} \dot{B}(t) &= \frac{dB(t)}{dt} = \alpha_1 R(t), \\ \dot{R}(t) &= \frac{dR(t)}{dt} = \alpha_2 B(t), \end{aligned} \tag{2.1}$$

여기서 $B(t)$와 $R(t)$은 t시점의 B군과 R군의 전투력이고, $\dot{B}(t)$와 $\dot{R}(t)$은 일정 기간에 발생한 B군과 R군의 손실이며, α_1과 α_2는 각 군의 손실계수이다. $dB(t)/dt$와 $dR(t)/dt$은 단위 시간 t를 기준으로 미분한 값을 의미하고 본 논문에서는 1일 동안의 손실량을 의미한다. 그리고 식 (2.1)을 정리하면 $\alpha_2(B(0)^2 - B(t)^2) = \alpha_1(R(0)^2 - R(t)^2)$이 되어 자승모형이라고도 하는데 이는 쌍방의 전투력 손실 교환율(casualty exchange rate) $dB(t)/dR(t) = \alpha_1 R(t)/\alpha_2 B(t)$이므로 최초부터 종료까지 일정하여 최초 전투력인 $B(0)$ 또는 $R(0)$가 높은 군이 결국 이긴다는 의미이기도 하다.

란체스터 모형은 다양하게 변형되어 활용되었는데 Bracken (1995)은 이러한 여러 가지 모형을 전투력 승수를 포함한 모형으로 일반화하였다. 상대군 전투력 승수를 β, 자군 전투력 승수를 γ로 정의하면 란체스터 일반화 모형은 아래와 같다.

$$\begin{aligned} \dot{B}(t) &= \frac{dB(t)}{dt} = \alpha_1 R(t)^\beta B(t)^\gamma f_B(\delta), \\ \dot{R}(t) &= \frac{dR(t)}{dt} = \alpha_2 B(t)^\beta R(t)^\gamma f_R(\delta), \end{aligned} \tag{2.2}$$

여기서 $f(\delta)$는 공격시 전술적으로 얻을 수 있는 이점을 표현한 함수로서 δ를 공격모수라고 하면 식 $f_B(\delta)$의 경우 B군이 공격시 δ이고 방어시 $1/\delta$이며, $f_R(\delta)$은 R군이 공격시 δ이고 방어시 $1/\delta$이다. 통상적으로 공격과 방어를 구분하지 않는 경우의 란체스터 모형에서는 $f(\delta) = 1$로 고려하면 된다.

Morse와 Kimball (1951)이 주장한 제1선형모형은 식 (2.2)에서 $\beta = \gamma = 0$인 경우, 제2선형모형은 $\beta = \gamma = 1$인 경우라고 할 수 있다. 선형모형은 개인 대 개인 또는 무기 대 무기의 조준사격으로 교전하는 경우에 적용할 수 있으며, B군의 전투력 손실은 사격을 가하는 R군의 전투력 수준과 그 지역을 점령하고 있는 B군의 전투력에 비례한다는 의미이다. 그리고 란체스터 자승모형은 $\beta = 1$, $\gamma = 0$인 경우이고 지역사격에 적용할 수 있는 모형이다.

란체스터 지수모형의 경우는 $\beta = 0$, $\gamma = 1$인 경우를 말하는데, Weiss (1966)에 의하면 미국의 남북전쟁에 관한 연구를 하면서 약 1.5만명 이상이 참가하는 전투에는 란체스터 선형모형과 자승모형을 적용하는 것은 적절하지 않고 오히려 전투력 손실은 자군의 전투력에 비례하여 나타나기 때문에 지수모형이 적합하다고 주장하였다. 또한 2차 세계대전의 서유럽 기갑전투에 관한 연구를 한 Peterson (1967)은 기갑

전투에서 첫 번째 파괴는 자군의 전차수에 의존하는 지수모형을 따르며, 두 번째 파괴는 자승모형을 따른다는 주장을 하였다.

2.2. 로그선형모형으로 변환

전쟁은 많은 요소가 우연적이거나 확률적인 사건의 연속으로 이루어지는 경우가 많아서 교전 중인 부대의 전투력 변화를 미분방정식 형태의 란체스터 모형으로 모두 설명하기는 현실적으로 한계가 있다. 그래서 란체스터 모형은 전쟁의 전반적인 부분을 통찰하는 데는 매우 유용하지만 전쟁의 확률적 과정을 반영하지 못하는 결정적 모형(deterministic model)이기 때문에 많은 비판을 받아왔다. 특히 결정적 모형은 지휘관이 선택할 수 있는 많은 전술적 의사결정 내용이 배제되어 있어서 전투력의 증원 및 전환, 화력 할당의 조정 등의 내용을 반영할 수 없는 단점이 있다. 그래서 이러한 단점을 극복하기 위하여 결정적 모형보다는 통계적 모형으로 접근하여 란체스터 모형을 연구하기 시작하였다.

Fricker (1998)는 식 (2.2)를 선형회귀모형으로 변환하여 통계적 접근 방법으로 Bracken (1995)과 Hartley와 Helmbold (1995)의 연구 결과를 검증하고 비판하였는데, 그는 통계적 모형으로 적용하기 위해 자료 행렬을 아래와 같이 로그 변환을 취하였다.

$$Y = \begin{pmatrix} \log(\dot{B}(t)) \\ \log(\dot{R}(t)) \end{pmatrix} \quad X = (X_1, X_2) = \begin{pmatrix} \log(R(t)),\ \log(B(t)) \\ \log(B(t)),\ \log(R(t)) \end{pmatrix}. \tag{2.3}$$

로그 변환한 자료 행렬을 일반적인 선형회귀모형으로 정리하면 아래와 같다.

$$y_i = a + \beta x_{1i} + \gamma x_{2i} + cI_{Ai} + dI_{Ri} + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right),\ i = 1, \ldots, n, \tag{2.4}$$

여기서 I_{Ai}는 공격함수 $f(\delta)$를 로그변환 하였을 때를 표현하는 지시함수로서 y_i에 해당하는 부대가 공격 시 1, 방어 시 -1을 갖는다. I_{Ri}는 B군과 R군의 차이를 표현하기 위한 가변수(dummy variable)로서 y_i가 R군에 해당하면 1이고 아닌 경우는 0이다. 식 (2.4)에 의해 추정한 모수를 식 (2.1)과 식 (2.2)의 모수로 변환하면 $\hat{\alpha}_1 = e^a$, $\hat{\alpha}_2 = e^{(a+d)}$, $\hat{\delta} = e^c$이다.

통상 선형회귀모형으로 분석하고자 하는 이유는 변수들 간의 함수관계를 쉽게 설명할 수 있고 미래를 예측할 수 있기 때문일 것이다. 선형모형으로 적합시켜 회귀분석을 할 때 모수들에 대한 추정은 일반적으로 최소제곱추정량(least squares estimates)으로 추정하게 되는데 이때 독립성, 정규성, 등분산성 등 모형의 가정에 대한 부분을 진단할 필요가 있다. 예를 들어 설명변수 간의 강한 선형적 연관 관계가 있어 독립성이 훼손되어 다중공선성(multicollinearity) 문제를 내포하고 있는 모형의 경우 변수를 추가하거나 삭제할 때 추정된 회귀계수가 크게 변화하여 모형이 불안정한 현상이 발생하거나 추정된 회귀계수가 이론적으로 기대했던 값과 완전히 다르게 나타나는 현상들이 발생한다.

3. 기존 연구에 대한 고찰

3.1. 분석 자료에 대한 설명

미 육군 개념분석국(U.S. Army Concept Analysis agency; CAA)은 1990년도에 미국, 영국, 독일의 각종 문헌 및 전자 자료로부터 제2차 세계대전 자료를 수집하였다. 그중 1944년 12월 15일부터 1945년 1월 16일까지 Ardennes 전역에서 있었던 연합군과 독일군 전투 자료를 완성하여 Data Memory System Inc. (1990) The Ardennes Campaign Simulation Data Base (ACSDB)를 만들었고, Bracken (1995)은 이 자료를 활용하였다.

Bracken (1995)은 ACSDB를 이용하여 식 (2.2)와 같은 란체스터 모형을 일반화하여 통계적 분석을 시도하였는데 자신의 논문에 이 자료를 7개의 표 형태로 제공하고 있다. 여기에는 연합군(B군)과 독일군(R군)의 전차, 장갑차, 포병, 전투원, 전투력, 총전투원, 총전투력에 대해 일자별로 보유 현황과 손실 현황을 제시하고 있다. 여기서 CAA에서 제시한 무기효과지수에 무기별 수량을 곱하여 합한 가중합을 계산하여 전투력과 총전투력을 도표로 제시하였다. 예를 들어 t일 B군의 전투력 수준 $B(t)$는 무기효과지수에 해당 일자의 무기 수량을 곱하여 $20 \times$ 전차수 $+ 5 \times$ 장갑차수 $+ 40 \times$ 포병수 $+ 1 \times$ 전투원수로 얻은 값이며, 총전투력은 전투원수 대신 전투원과 지원병력을 합한 총전투원수를 반영하여 계산한 값이다. 연합군과 독일군의 전투력 손실 현황인 $\dot{B}(t)$와 $\dot{R}(t)$도 각 무기체계의 손실 현황을 참조하여 동일한 방법으로 계산할 수 있다. 그러나 Bracken이 자신의 논문에서 실질적으로 활용한 데이터는 이 자료 중 초기 12월 16일부터 12월 25일까지 10일간의 자료만을 사용하였는데 이유는 초기 5일은 독일군이 공격한 자료이며, 이후는 연합군이 공격한 결과이므로 자료의 균형을 고려했을 것으로 생각된다(이하 Bracken 자료). Fricker (1998)가 사용한 자료는 Bracken (1995)이 제시한 Ardennes 전역 전체 자료를 포함하여 당시 연합군과 독일군이 투입했던 근접항공지원 소티(Sorties)수를 추가하여 분석하였다. 그는 증원병력과 예비대 포함 여부 등을 확인 검증하여 Bracken이 제시한 자료를 수정하였고, 일부 자료 중 피해가 없는 첫날 12월 15일 자료를 제외한 32일간의 자료를 활용하였다. 결국 분석간 활용한 자료는 32일간의 자료를 식 (2.3)처럼 변환하여 64개의 관찰치를 식 (2.4)의 모형에 적합시킨 것이다. 그의 전투력과 총전투력 계산 방법은 Bracken이 제시한 방법과 동일하며 근접항공지원의 소티수는 무기효과지수를 30으로 계산하여 가중합을 구하였다(이하 Fricker 자료).

3.2. Bracken 모형에 대한 평가

Bracken (1995)은 10일간의 Bracken 자료를 이용하여 4가지 Bracken's model (BM) 모형을 검토하였는데 Table 3.1을 기준으로 설명하면 Case1에 포함된 모형들은 전투력을 기준으로 분석한 모형이고, Case2에 포함된 모형들은 총전투력을 기준으로 분석한 모형이다. 그의 모형 BM1 모형은 전투력 자료를 기준으로 $\beta = \gamma = 1$로 가정하여 공격모수 δ를 추정하는 데 중점을 두었으며, BM2 모형은 총전투력을 기준으로 모수를 추정하였다. BM3 모형은 전투력 자료를 이용하여 공격모수 $\delta = 1$을 가정하여 모형을 적합 시켰고, BM4 모형은 총전투력을 기준으로 $\delta = 1$을 가정하여 모수를 추정하였다. 특히 BM2, BM3, BM4의 경우 란체스터 선형모형인 $\beta = \gamma = 1$에서 확장하여 $\beta + \gamma = 2$라는 제약조건을 기준으로 모수를 추정하였다.

Bracken의 모수 추정 방법은 식 (2.4) 모형에서 최소제곱 추정법에 의해 추정하긴 하였지만 각 모수에 대한 5개의 후보 값들을 대입시켜서 그중에 최소분산을 갖는 모형을 추정하였다. 그리고 $\beta + \gamma = 2$라는 제약조건을 기준으로 최소제곱 추정법을 적용했기 때문에 최적해(global solution)가 아닌 지역해(local solution)를 찾아낸 결과가 되었다. Table 3.1에는 당시 Braken이 제시한 모수 추정값은 BM(a) 모형으로 제시하였고, 동일한 자료를 이용하여 위에서 제시한 $\beta + \gamma = 2$라는 제약 조건 없이 추정한 결과를 BM(b) 모형으로 제시하였다. BM2(b), BM3(b), BM4(b)의 경우와 같이 통계적으로 유의하지 않은 변수들이 있다면 모형을 좀더 단순화하여 설명변수와 반응변수 간의 관계 설명이 쉽도록 하는 것이 좋은 방법이라고 생각한다.

Bracken은 여러 가지 종류의 란체스터 모형을 모두 포괄할 수 있는 일반화 모형을 제안하였고 이를 실제 전쟁 자료에 적용하였으며, 전술적 상황을 고려한 공격 모수를 BM1 모형이나 BM2모형에서처럼 추정하여 공격 부대가 0.8 정도로 손실이 적게 발생한다는 점을 입증하는 성과를 얻었다. 반면 당시 컴퓨터의 계산 능력의 한계에 의해서인지 모르지만 모수들이 가질 수 있는 후보 값을 한정하여 $\beta + \gamma = 2$라는 제약 조건을 둠으로써 최적의 해를 찾지 못하였고, 통계적 측면에서 유의하지 않은 변수들까지 모형

Table 3.1. Parameter estimation by model

Case	Model	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$	Non-significant variables (p-value)
Case1	BM1(a)	8.0×10^{-9}	10.0×10^{-9}	1.00	1.00	0.80	$x_1(0.06)$, $I_A(0.91)$
	BM1(b)	7.9×10^{-9}	10.8×10^{-9}	1.00	1.00	0.84	
	BM3(a)	8.0×10^{-9}	10.0×10^{-9}	1.30	0.70	1.00	
	BM3(b)	9.2×10^{3}	9.0×10^{3}	1.64	−1.72	1.00	
	FM1(a)	4.7×10^{-27}	3.1×10^{-26}	0.00	5.00	0.81	
	FM1(b)	1.1×10^{-26}	7.6×10^{-26}	0.00	4.97	0.81	
	YM1	7.0×10^{-14}	7.0×10^{-14}	1.84	1.00	0.90	
Case2	BM2(a)	8.0×10^{-9}	8.0×10^{-9}	0.80	1.20	0.80	
	BM2(b)	1.2×10^{5}	2.8×10^{5}	−1.36	1.11	0.75	
	BM4(a)	8.0×10^{-9}	8.0×10^{-9}	1.20	0.80	1.00	$x_1(0.11)$, $x_2(0.19)$
	BM4(b)	1.9×10^{5}	1.8×10^{5}	0.32	−0.57	1.00	$x_1(0.66)$, $x_2(0.44)$, $I_A(0.77)$
	FM2(a)	1.7×10^{-16}	8.0×10^{-16}	0.00	3.20	0.82	
	FM2(b)	1.0×10^{-15}	5.0×10^{-15}	0.00	3.05	0.85	
	YM2	6.8×10^{-10}	6.8×10^{-10}	1.14	1.00	0.90	
Case3	FM3(a)	2.7×10^{-24}	1.6×10^{-23}	0.00	4.60	0.80	
	FM3(b)	5.6×10^{-24}	3.4×10^{-23}	0.00	4.50	0.81	
	YM3	6.5×10^{-14}	6.5×10^{-14}	1.84	1.00	0.91	
Case4	FM4(a)	1.3×10^{-15}	5.6×10^{-15}	0.00	3.00	0.82	
	FM4(b)	1.8×10^{-14}	8.0×10^{-14}	0.00	2.85	0.85	
	YM4	6.3×10^{-10}	6.3×10^{-10}	1.15	1.00	0.90	

Table 3.2. Comparison of VIF between FM and YM

FM model variables				YM model variables		
Model	x_2	I_A	I_R	Model	x_1	I_A
FM1	2.0	18.9	17.9	YM1	2.00	2.00
FM2	2.0	25.3	24.3	YM2	1.98	1.98
FM3	2.0	17.5	16.4	YM3	2.03	2.03
FM4	2.0	24.6	23.4	YM4	1.99	1.99

에 포함하여 모형을 단순화하지 못하였다.

3.3. Fricker 모형에 대한 평가

Fricker (1998)는 3.1절에서 설명한 Fricker 자료를 이용하여 모형 식 (2.4)를 적용하여 분석하였다. 그는 4가지 Fricker's model (FM) 모형을 검토하였는데 Table 3.1을 기준으로 설명하면 Case1과 Case2는 Bracken 모형처럼 각각 전투력 모형, 총전투력 모형이며, Case3의 경우는 근접항공지원 시 전투력 모형, Case4는 근접항공지원 시 총전투력 모형이다. Fricker의 접근 방법은 Bracken이 시도한 10일간의 전투자료가 아닌 Ardennes 전역 전체 자료 64개 관찰치를 활용했다는 점과 추정하고자 하는 모수에 특별한 제약조건 없이 추정하였다는 점에서 진일보하였다. Fricker가 추정한 모수들은 FM(a) 모형으로 제시하였고 본인이 검증한 모수의 추정치는 FM(b) 모형으로 제시하였다. Table 3.1에서 (a)와 (b)모델 추정치를 비교해보면 전반적으로 유사해 보이지만 어느 한 모수 추정치의 작은 변화에 다른 모수들의 추정치가 많이 변해있음을 확인할 수 있어서 다중공선성 문제를 검증해본 결과를 Table 3.2로 제시하였다.

Silvey (1969)와 Chatterjee와 Hadi (2000)에 의하면 다중공선성을 탐색하는 방법은 설명변수 간의 선

형적 연관 관계를 측정할 수 있는 분산확대인자(variance inflation gactor; VIF)로 알려진 통계량에 의해 확인이 가능하며 VIF값이 10 이상인 경우 심각한 다중공선성 문제가 있다는 신호가 된다고 한다. Table 3.2를 확인해보면 Fricker 모형은 모두 다중공선성 문제를 가지고 있음을 알 수 있다.

Fricker의 분석은 Ardennes 전역 전체 자료로 확대하여 모형을 구축하고 근접항공지원 분야까지도 모형에 포함해 분석을 시도한 성과가 있었다. 그리고 그는 Ardennes 자료의 경우 란체스터 선형모형이나 자승모형으로 적합 시키는 것은 적절하지 않다고 주장하였다. 상대군의 전투력 승수인 $\beta = 0$로 고려하여 모형을 적합 시켰으며 자군의 손실은 상대군의 전투력보다는 자군의 전투력에 의해 영향을 많이 받는다고 설명하고 있다. 일례로 Gulf전에서 이라크군의 손실은 다국적군의 인원수보다 이라크군의 인원수의 함수에 더 영향을 많이 받았을 것이라고 주장하였다. 그러나 이는 그가 제시한 모형 자체가 다중공선성의 문제를 가지고 있어서 그의 주장을 뒷받침할 만큼 충분한 증거를 가지고 있지 못하다.

4. 전투력 손실률 모형의 제안과 해석

4.1. 전투력 손실률 모형의 제안

새로운 모형을 제안하고자 하는 기준은 두 가지 이다. 첫 번째는 통계적으로 유의한 모형이어야 하고, 두 번째는 해석과 설명이 용이한 모형이어야 한다. 통계적으로 유의한 모형 측면에서 살펴보면 Fricker (1998)의 FM 모형들의 분석 사례에서 확인할 수 있듯이 다중공선성 문제로 독립변수 간 상관관계에 의해 모수 추정값들이 불안정하고 왜곡될 수 있음을 알 수 있다. 사실 모형 식 (2.4)에는 유사하고 의미 없는 모수들이 포함되어 있다. 예를 들어 Table 3.1에서 보면 모형 자체에 문제가 있는 BM 모형들을 제외하면 α_1과 α_2의 차이는 최대 2.9×10^{-9}이므로 이는 B군과 R군의 손실이 10억 명 중에 2.9명의 차이가 있다는 의미이다. Ardennes 전역에 참여한 연합군이나 독일군의 최대 인원수가 107만 명 정도이고 최대 총전투력이 142만 점이라는 것을 고려하면 두 모수의 구분은 의미 없는 것이다. 그래서 모형 식 (2.4)에서 $\alpha_1 = \alpha_2 = \alpha$로 고려하기로 한다.

두 번째 해석과 설명이 용이한 모형 측면에서 살펴보면 전술적 측면에서 Fricker의 주장처럼 자군의 전투력 손실이 자군의 전투력에만 의존한다는 것은 상대방과 상호 교전을 해야 하는 전술적인 상황을 고려한다면 납득하기 어렵다. 사실 그의 논문 토의 부분에는 손실계수인 α_1과 α_2에 이미 상대군 전투력이 반영된 것이 아니냐는 의견을 제시하였다. 그러나 위에서도 설명한 것처럼 유의하지 않은 전투력 손실계수의 차이는 이러한 주장을 뒷받침하기 어렵다. 그래서 $\gamma = 1$로 가정하고 이를 좌변으로 옮기면 자연스럽게 전투력 손실 부분이 전투력 손실률인 $\dot{B}(t)/B(t)$과 $\dot{R}(t)/R(t)$으로 변한다. 그래서 제안하고자 하는 모형은 위의 두 가지 조건, 즉 유사한 모수들은 단순화하고 전술적 상황을 고려한 자군과 상대군 전투력이 모두 고려된 Yoo's model (YM) 모형을 아래와 같이 제안하고자 한다.

$$y_{\mathrm{new}.i} = y_i - x_{2i} = a + \beta x_{1i} + cI_{Ai} + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \ i = 1, \ldots, n. \tag{4.1}$$

이 모형을 기초로 식 (2.2)를 다시 정리하면 아래와 같다.

$$\begin{aligned} \frac{\dot{B}(t)}{B(t)} &= \alpha R(t)^{\beta} f_B(\delta), \\ \frac{\dot{R}(t)}{R(t)} &= \alpha B(t)^{\beta} f_R(\delta). \end{aligned} \tag{4.2}$$

결국 모형의 좌변은 시간 t 시점에 있는 자군의 전투력 손실률이 되고 우변은 공통적으로 영향을 받는 손실계수, 상대군의 전투력 승수, 그리고 전술적 상황인 공격모수에 의존하는 모형이라고 할 수 있다.

Table 4.1. Regression diagnostics

Model	Independent-errors assumption		Normality assumption		Constant variance assumption	
	DW statistic	p-value	SW statistic	p-value	NCV statistic	p-value
YM1	1.11	< 0.01	0.99	0.93	0.43	0.51
YM2	1.09	< 0.01	0.95	0.03	0.14	0.71
YM3	1.17	< 0.01	0.99	0.96	0.43	0.51
YM4	1.11	< 0.01	0.96	0.02	0.43	0.51

Table 4.2. Parameter estimation after removal of autocorrelation

Model	DW statistic	p-value	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$
YM1(a)	1.91	0.30	0.48	1.5×10^{-11}	1.44	0.85
YM2(a)	1.84	0.21	0.42	7.1×10^{-9}	0.97	0.86
YM3(a)	1.92	0.31	0.44	3.0×10^{-12}	1.56	0.86
YM4(a)	1.87	0.24	0.42	5.3×10^{-9}	0.99	0.86

이 전투력 손실률 모형은 식 (2.2)보다 2개의 모수를 제거하여 단순화시킨 모형이 되었다. 식 (4.1) 모형을 적용하여 기존 Fricker 자료를 활용하여 모수를 추정한 것이 Table 3.1에 YM 모형으로 반영되어 있고 이 모형들은 각각 전투력, 총전투력, 근접항공지원 하 전투력, 근접항공지원 하 총전투력 모형이다. 모든 모형이 안정적이며 Table 3.2에서 제시한 바와 같이 모든 변수의 VIF는 2.0 이하로 안정적이어서 다중공선성 문제는 없는 것으로 확인하였다.

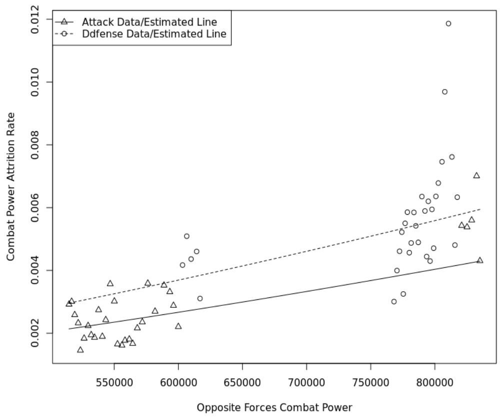
4.2. 모형에 대한 가정위반 탐색과 조치

선형회귀 모형에 적합시키기 위해서는 식(4.1)의 우변에서 확인할 수 있듯이 해당 자료가 독립성, 정규성, 등분산성이 보장된다는 가정을 전제로 한다. 이러한 가정을 위반하였을 때는 적절한 조치를 통하여 부정확한 추정을 방지하여야 한다. 제시된 전투력 손실률 모형인 YM 모형들에 대하여 이러한 가정 위반 여부를 검증한 결과 Table 4.1로 제시하였다. Chatterjee와 Hadi (2000)와 Faraway (2015)를 참조하여 독립성 검증은 Durbin Watson 통계량을 이용하고, 정규성 검증은 Shapiro Wilk 통계량을 이용하며, 등분산성 검증은 Non-constant Variance Score Test를 이용하여 검증한 결과이다.

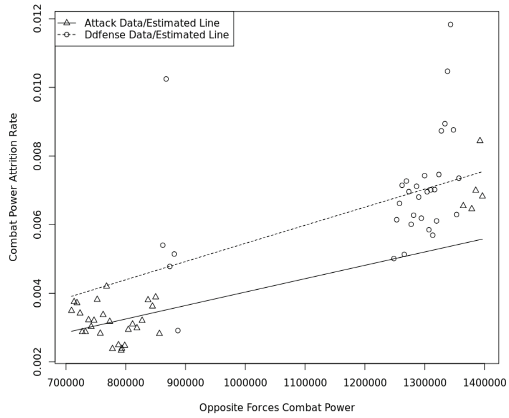
Table 4.1을 참조하면 모든 모형이 독립성을 위반하고 있고 YM2와 YM4 모형의 경우는 정규성도 위반하고 있다. 이는 Ardennes 자료 자체가 일자별 전투 결과를 반영하기 때문에 하루 전 전투력과 손실된 부분은 다음날 전투력에 반영되어 일자별 자료가 당연히 독립적일 수 없다. 이렇게 자기상관(autocorrelation)이 강한 자료에 대한 조치는 Cochrane과 Orcutt (1949)가 제시한 방법으로 서로 상관된 부분에 대한 상관계수(ρ)를 추정하고, 식 (4.1)에서 새로운 $y^*_{\text{new}.i} = y_{\text{new}.i} - \rho y_{\text{new}.i-1}$, $x^*_{1i} = x_{1i} - \rho x_{1i-1}$로 변수 변환하여 $y^*_{\text{new}.i}$와 x^*_{1i}로 모수를 추정하고 수정해주는 과정을 반복하여 Durbin Watson 통계량이 귀무가설을 기각하지 못할 때까지 반복하는 과정을 거친다. 자기상관된 모형을 수정하여 추정한 결과 Table 4.2에 YM(a) 모형으로 그 결과를 제시하였으며 모든 모형의 독립성이 개선된 것을 알 수 있다. Table 4.1에서 정규성에도 문제가 있던 YM2와 YM4 모형의 경우 YM2(a)와 YM4(a) 모형에서는 정규성을 재검증한 결과 Shapiro Wilk 통계량이 모두 0.97 p-value가 0.21로 정규성에 문제가 없음을 확인할 수 있었다.

4.3. 추정 결과를 이용한 군사적 해석

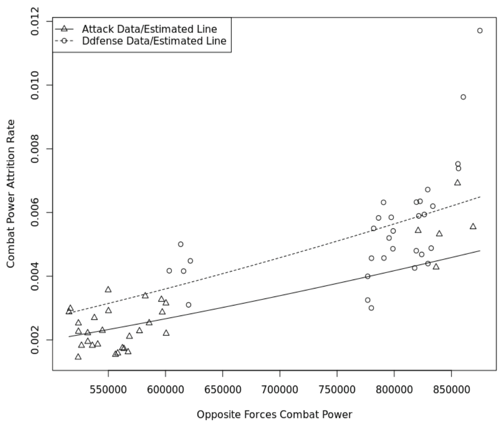
Ardennes 자료에 대하여 불필요한 변수들을 포함하여 야기할 수 있는 다중공선성 문제를 해결하면서



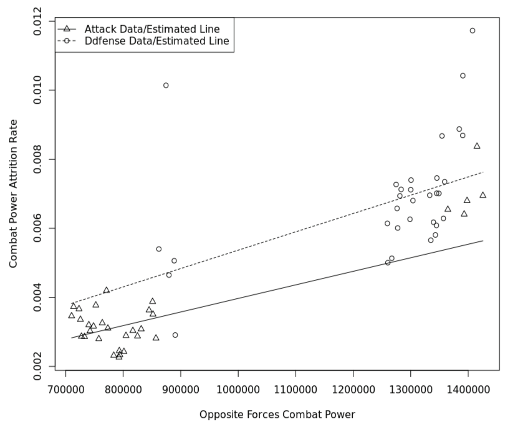
(a) YM1(a) combat power model



(b) YM2(a) total combat power model



(c) YM3(a) combat power model



(d) YM4(a) total combat power model

Figure 4.1. Fitting a regression line to Ardennes data.

모형을 단순화하고, 자기상관으로 인하여 야기될 수 있는 부정확한 추정 문제를 해결하여 통계적으로 유의한 모형으로 전투력 손실률 모형인 YM(a) 모형들로 정리하였다. 그리고 어떤 특정 모형을 수치적으로 적합시켜 추정하는 것도 중요하지만 설명과 해석이 용이한 부분도 매우 중요하다. 그래서 CAR(Combat Power Attrition Rate)은 전투력 손실률, TAR(Total Combat Power Attrition Rate)은 총전투력 손실률, OFCP(Opposite Forces Combat Power)는 상대군 전투력으로 정의하고 설명과 해석을 위해 최종적으로 정리된 YM(a) 모형들에 대하여 추정식 형태로 정리하면 식 (4.3)과 같고 Ardennes 자료와 함께 제시한 최종 추정식을 도식하면 Figure 4.1과 같다.

$$\begin{aligned}
&\mathrm{YM1(a)}: \ \mathrm{CAR} = 1.5 \times 10^{-11} \times \mathrm{OFCP}^{1.44} \times (\mathrm{Attack}: 0.85, \mathrm{Defense}: 1.18)\\
&\mathrm{YM2(a)}: \ \mathrm{TAR} = 7.1 \times 10^{-9} \times \mathrm{OFCP}^{0.97} \times (\mathrm{Attack}: 0.86, \mathrm{Defense}: 1.16)\\
&\mathrm{YM3(a)}: \ \mathrm{CAR} = 3.0 \times 10^{-12} \times \mathrm{OFCP}^{1.56} \times (\mathrm{Attack}: 0.86, \mathrm{Defense}: 1.16)\\
&\mathrm{YM4(a)}: \ \mathrm{TAR} = 5.3 \times 10^{-9} \times \mathrm{OFCP}^{0.99} \times (\mathrm{Attack}: 0.86, \mathrm{Defense}: 1.16) \qquad (4.3)
\end{aligned}$$

그리고 지금까지 분석한 결과들을 기초로 군사적 측면에서 모형을 해석하면 아래와 같다.

1. 미영 연합군과 독일군의 전투력 손실률의 차이는 유의하지 않았다.
2. 특별한 태세가 없는 부대에 비하여 공격부대는 86% 수준으로 피해가 감소하였고 방어부대는 1.16배의 피해가 더 발생하였다.
3. 전투원에 대한 전투력 손실률은 상대군 전투력의 약 1.5승에 비례하며, 지원병력을 포함한 총전투력 손실률은 상대군 전투력에 비례한다.
4. 총전투력 모형은 Fricker (1998)의 주장과 다르게 란체스터 제2선형모형과 거의 유사하다.

군사적 측면에서 논의하면 일반적으로 공격부대가 노출이 많아서 방어부대보다 손실이 클 것으로 생각한다. 그러나 여기서 분석한 결과는 특이하게 공격부대가 적은 피해를 받았다. 그 이유는 Ardennes 전역의 전투, 즉 Bulge 전투라고도 알려진 전사의 특징에서 찾아볼 수 있다. 이 전투는 제2차 세계대전 중 미군과 영국군을 분리할 목적으로 서부전선에서 독일군이 시행한 최후의 대반격이었으며 Dupuy 등 (1994)에 의하면 히틀러의 마지막 도박으로 알려져 있다. 초기 공격부대인 독일군은 기습효과로 연합군에게 큰 피해를 주었고, 12월 23일부터 기상이 호전되어 연합군은 근접항공지원의 출격 수를 압도적으로 늘려 독일군에게 피해를 많이 줄 수 있었다. 그래서 공격부대가 방어부대보다 피해가 적은 이유는 독일군의 기습 효과와 연합군의 근접항공지원에 의한 것으로 평가된다.

5. 결론

란체스터 모형에 대하여 통계적 방법으로 접근하였던 기존 연구를 고찰하면서 몇 가지 문제점을 식별하여 해결하였다. 최소제곱 추정법에서 최적해를 찾지 못하고 몇 가지 후보군을 대입하여 결론을 내리면서 지역해를 찾아낸 사례가 있었으며, 모형을 너무 복잡하게 적용함으로써 다중공선성 문제를 간과한 문제들을 가지고 있었다. 그리고 선형회귀 모형을 적합 시키기 위한 전제 조건인 독립성, 정규성, 등분산성을 진단하고 조치했는데, Ardennes 자료는 자기상관 문제를 내포하고 있어서 Cochrane-Orcutt 방법으로 과소 과대 추정될 수 있는 문제를 해결하였다.

란체스터 방정식은 1세기 전에 개발한 미분방정식 형태의 모형이지만 아직도 다양한 형태로 변형되어 응용되고 있는 것은 간단하지만 강력하고 시사하는 바가 크기 때문일 것이다. 그래서 실질적인 적용 측면에서는 가장 간단한 형태로 모수를 단순화할 필요가 있고 통계적으로 유의한지 충분히 검토하여 적용해야 할 것이다. 이런 측면에서 본 논문에서는 전투력 손실률 모형을 제안하였으며 Ardennes 전역 자료에 대해서 군사적 측면의 의미 있는 해석이 가능하였고 단순화한 만큼 설명이 용이한 부분을 확인하였다.

References

Bracken, J.(1995). Lanchester models of Ardennes Campaign, *Naval Research Logistics*, **42**, 559–577.

Chatterjee, S. and Hadi, A. S. (2000). *Regression Analysis by Example*, John Wiley & Sons, New York.

Cochrane, D. and Orcutt, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms, *Journal of American Statistical Association*, **44**, 32–61.

Dietchman, S. J. (1962). A Lanchester model of guerrilla warfare, *Operations Research*, **10**, 818–827.

Dupuy, T. N., Bongard, D. L., and Anderson, R. C. (1995). *Hitler's Last Gamble: The Battle of the Bulge, December 1944 - January 1945*, Harper Perennial, New York.

Faraway, J. J. (2015). *Linear Models with R* (2nd ed), Chapman and Hall, London.

Fricker, R. D. (1998). Attrition models of the Ardennes Campaign, *Naval Research Logistics*, **45**, 1–22.

Hartley, D. S. (1995). A mathematical model of attrition data, *Naval Research Logistics*, **42**, 585–607.

Hartley, D. S. and Helmbold, R. L. (1995). Validating Lanchester's square law and other attrition models, *Naval Research Logistics*, **42**, 609–633.

Lanchester, F. W. (1916). *Aircraft in Warfare: The Dawn of the Fourth Arm*, Constable and Company, London.

Morse, P. M. and Kimball, G. E. (1951). *Methods of Operation Research*, Wiley, New York.

Peterson, R. H. (1967). On the logarithmic law of attrition and its application to tank combat, *Operations Research*, **15**, 557–558.

Schaffer, M. B. (1968). Lanchester models of guerrilla engagements, *Operations Research*, **16**, 457–488.

Silvey, S. D. (1969). Multicollinearity and Imprecise Estimation, *Journal of the Royal Statistical Society, Series B*, **31**, 539–552.

Weiss, H. K. (1966). Combat models and historical data: the civil war, *Operations Research*, **14**, 759–790.

Wiper, M. P., Pettit, L. I., and Young, K. D. S. (2000). Bayesian inference for a Lanchester type combat model, *Naval Research Logistics*, **47**, 541–558.

란체스터 모형에 대한 통계적 고찰과 해석

유병주[a,1]

[a]지상작전사령부 작전분석과

(2020년 3월 17일 접수, 2020년 4월 24일 수정, 2020년 5월 4일 채택)

요 약

본 논문에서는 과거 2차 세계대전 자료 중 Ardennes 전역에서 있었던 실제 전투 자료를 란체스터 모형에 적합 시키기 위하여 로그변환된 선형회귀모형을 추정하는 문제를 다루었다. 먼저 동일한 자료에 대하여 기존 연구 결과를 고찰하여 모수에 대한 최적해(Global Solution) 결정 문제와 다중공선성 문제들을 확인하였다. 최소제곱 추정법에 의한 모수 추정은 특정 제약조건이나 제한된 후보군을 고려할 경우 최적해를 찾지 못하고 지역해(Local Solution)를 찾을 수 있음으로 주의가 필요하고, 모형에 포함된 변수들은 통계적으로 충분히 유의성을 검토하여 포함해야지 그렇지 않았을 때 모수 추정값들이 왜곡될 수 있다. 모형에 과도하게 많은 설명 변수를 포함하는 경우 변수 간의 상관관계로 인하여 추정값이 왜곡되고 변수의 추가나 제거 시 불안정한 현상들이 발생한다. 이런 다중공선성 문제를 탐색하는 방법은 설명 변수 간의 선형적 연관 관계를 측정할 수 있는 분산확대인자(VIF)로 알려진 통계량에 의해 확인이 가능하며 이를 조치하기 위해서는 상호 연관된 설명 변수들을 제거하여 모형을 단순화해야 한다. 그래서 이러한 문제가 발생하지 않도록 모형을 단순화하고 이해와 설명이 용이한 전투력 손실률 모형을 제안하였고 Ardennes 자료에 대하여 적합한 결과 모수 추정이 안정적이고 자료에 대한 설명과 해석이 용이하다는 점을 입증하였다. 특히, 모수 추정간 선형회귀 모형의 기본적인 가정사항인 독립성, 정규성, 등분산성을 검증하여 자기상관(Autocorrelation) 문제로 독립성이 훼손되어 과대 과소 추정될 우려가 있는 사항을 Cochrane-Orcutt 방법에 의해 변환하여 독립성과 정규성을 보장하였다.

주요용어: 란체스터 전투 모형, 전투력 손실률 모형, 자기상관, 다중공선성, 아르덴느 전역

[1]경기도 용인시 처인구 성산로 57 사서함 505-1-3, 지상작전사령부 작전분석과. E-mail: bjyoo@korea.ac.kr