

한국 무기체계 개발 실적을 고려한 연구개발 비용추정관계식 개발

어원재 · 이용복[†] · 강성진

국방대학교 운영분석학과

Developing an R&D CER Using Historical Defense Weapon System Data in Korea

Won-Jae Eo · Yong-Bok Lee · Sung-Jin Kang

Department of Operations Research, Korea National Defense University

Currently, cost estimation is very important to the government acquisition programs to support decisions about funding and to evaluate resource requirement as key decision points. Parametric cost estimating models have been used extensively to obtain appropriate cost estimates in early acquisition phase. However, they have many restrictions to ensure the cost estimating result in Korean defense environment because they are developed in the U.S.A. environment. In order to obtain a good R&D cost estimate, developing our own CERs (Cost Estimation Relationships) using historical R&D data is essential. Nevertheless, there has been little research to develop our own CERs.

In this research, we established a CER development process and found some cost drivers in the historical movement weapon system R&D data. The R&D CER is developed using the PCR(Principle Component Regression) method to remove multicollinearity among data and to overcome the restriction of the insufficient number of sample. At least, this research is meaningful as a first attempt in terms of defining the CER development process and obtaining our own R&D CER based on the historical data in Korean weapon system R&D environment.

Keywords : CER, Multicollinearity, PCR, MMRE, PRED(ℓ)

1. 서론

과학기술의 발달은 군의 무기체계에도 첨단화를 요구함에 따라 무기체계의 비용은 갈수록 증가되는 추세에 있다. 그러나 가용한 국방예산은 한정되어 있기 때문에, 제한적인 국방예산을 효율적으로 분배, 활용함으로써 투자효율의 극대화와 국방비의 경제성을 도모하기 위한 노력이 요구되고 있다. 연구개발, 생산, 운영유지로 이어지는 무기체계의 수명주기를 고려할 때 초기 단계인

연구개발 단계에서의 비용추정은 전 주기(Total Life Cycle)의 비용통제를 위한 기준선을 제시한다는 차원에서 매우 중요하다. 또한 무기체계의 형상과 성능이 결정되는 연구개발에 투입되는 비용을 적절히 예측하기 위한 필요성 역시 증가되고 있다.

연구개발 단계에서의 비용추정은 현재 존재하지 않은 체계에 대해 미래에 발생할 비용을 추정함으로써 사업추진의 가부를 판단하는 의사결정을 지원한다. 비용분석 전문가가 판단한 비용추정치가 의사결정에 반영되면

향후에 이를 수정, 보완하기는 쉽지 않기 때문에 연구 개발 단계에서 비용추정은 신중을 기하여야 한다[1]. 연구개발 단계에서는 비용추정 대상 체계와 유사한 경험 자료를 획득하기가 쉽지 않고 투입되는 노력과 재료를 구체적으로 산출할 수 없기 때문에, 하향식 접근 방법(Top-down)인 파라메트릭 방법을 주로 사용하여 비용을 추정한다.

파라메트릭 비용추정법(Parametric Cost Estimating)은 기존의 개발경험을 기반으로 구축된 데이터베이스를 활용하여 통계적 추론과정에 의해 구축된 비용추정관계식(CER : Cost Estimation Relationships)으로 미래의 비용을 추정하는 방법이다. CER을 사용하는 장점은 복잡한 연산과정을 거치지 않으면서 합리적인 비용추정치출을 도출할 수 있는 “간편한 활용성”에 있다. 그러나 CER을 구축하고 이를 실제로 활용하기 위해서는 무기체계 연구개발 실적자료에 대한 충분한 데이터베이스가 확보되어 있어야 한다는 전제조건이 따른다[9].

국내에서 개발되는 무기체계에 대한 비용추정은 국내의 실적자료를 활용함으로써 신뢰성을 보장할 수 있어야 한다. 그러나 국내의 연구개발 자료는 CER을 구축할 수 있을 만큼 충분하지 않다. 따라서 무기체계 분야의 비용분석 역사가 10년이 넘었음에도 불구하고, 독자적인 CER 개발에 대한 시도가 거의 없었던 것이 사실이다.

본 연구에서는 현 수준에서 수집 가능한 자료를 활용하여 CER을 개발하는 전반적인 절차와 사례를 제시하였다. 이를 위해, 정부주관 무기체계 연구개발 사례를 수집하였고, 기동 무기체계 분야를 대상으로 CER을 개발하였다.

2. 기존연구 고찰 및 이론적 배경

2.1 기존연구 고찰

국내에서 무기체계 CER과 관련된 연구는 거의 이루어지지 않았으나, 국방과학연구소에서 유도 무기, 어뢰 및 전차, 소프트웨어에 대한 비용추정 모델을 개발하기 위한 연구가 최근에 실시되었다.

유도 무기체계 분야는 전문가의 사전지식을 기반으로 비용을 추정하기 위하여 유도무기체계를 총 17개 하부 체계로 나누고 설계요구조건과 성능요구조건에 해당되는 하위체계들을 선별하였다. 이를 기반으로 각 하위체계별 해당 전문가로부터 설문을 통해 비용인자의 종류, 범위, 수준을 결정하여, 확률분포에 의한 비용추정 결과를 제공한다[10].

어뢰 및 전차체계 분야 역시 전문가의 사전지식을 기반으로 비용을 추정하기 위하여 어뢰 및 전차체계를 일반화된 WBS(Work Breakdown Structure)로 구분하고, 해당 WBS상의 세부과제별로 전문가 인터뷰를 통해 설문을 수행하였다. 설문을 기반으로 세부과제별 CER을 구축하였고 CER 구축이 곤란한 부체계들은 비율추정 방법을 활용하여 비용분포 계산방법을 적용하였다[9].

국방 소프트웨어 분야는 전문가를 대상으로 소프트웨어를 개발하기 위한 노력, 개발 환경 등에 대한 설문을 통해 비용인자를 식별하여 CER을 개발하였다[7].

그러나 이러한 연구는 실적자료의 부재로 전문가 설문을 기반으로 추정식을 유도하였으며, 설문조사 과정에서 전문가 교체/부재 등으로 설문결과에 대한 객관성 보장이 제한된다. 또한, 실제 개발과정에 대한 타당성 검증이 이루어지지 않아 현재까지 연구개발비 추정을 위한 CER로 채택되지 않은 상태이다.

2.2 비용추정관계식

비용추정관계식은 종속변수인 비용(Cost)과 독립변수인 비용인자(Cost Drivers)들 간의 관계를 설명하고, 비용인자의 변화에 따라 비용이 어떻게 변화하는가를 표현한 수식이다. CER을 추정하기 위해서는 회귀분석 방법이 가장 적합하다고 알려져 있다[14]. k 개의 독립변수 X_1, X_2, \dots, X_k 와 종속변수 Y 에 대한 일반적인 다중회귀 모형은 식 (1)과 같다.

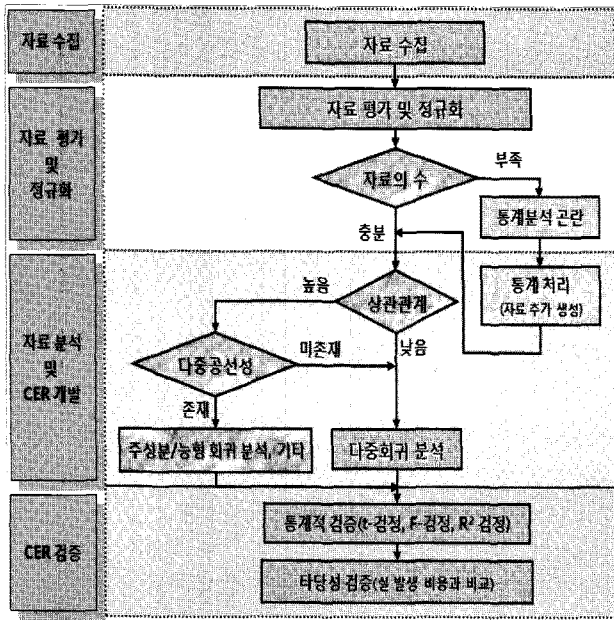
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \quad (1)$$

β 는 모집단의 모수로 회귀계수를 나타내고, ϵ 은 잔차를 나타낸다. Y 를 측정할 때 발생하는 오차(ϵ)에 대한 가정은 다변량 정규분포를 따르며, $E(\epsilon) = 0$, $Var(\epsilon) = \sigma^2$ 로 가정된다. 일반적으로 잔차의 합이 최소가 될 때 실제 값에 가장 근접한 회귀식이라고 볼 수 있는데, 최소제곱법(Least Squares Method)을 사용하여 ‘잔차의 제곱합이 최소’가 되는 회귀선이 실제 값에 가장 적합한 회귀식으로 간주되어진다[6].

3. CER 개발 절차 및 고려사항

3.1 CER 개발 절차

무기체계 연구개발 CER 개발 절차는 <그림 1>과 같이 자료 수집 → 자료 평가 및 정규화 → 자료 분석 및 CER 개발 → CER 검증의 순서로 진행된다.



〈그림 1〉 자료 특성을 고려한 CER 개발 방법

3.2 CER 개발시 고려사항

3.2.1 부족한 자료의 수 해결

수집된 자료를 분석하여 비용인자를 식별하고, 통계적 분석 가능성을 판단하는 과정에서 무기체계 개발 건수(n)가 비용인자의 개수(k)와 비교하여 $n-k < 2$ 이면 직접적인 통계적 분석이 불가능하다. 국내에서 유사한 무기체계의 개발 건수는 통계 분석이 가능할 만큼 충분한 경우가 많지 않다. 이때는 통계적으로 유의하게 개발 건수(n)를 추가 생성하여 $n-k \geq 2$ 를 만족하게 되면 분석이 가능하다.

3.2.2 다중공선성 진단

CER 개발 과정에서 독립변수인 비용인자들 사이에 다중공선성이 존재하는 경우가 흔히 발생한다. 이는 비용인자간에 서로 공통적인 정보를 포함하고 있다는 것을 의미하며, 회귀분석에서 회귀계수의 계산을 불가능하게 만들거나, 가능하더라도 회귀계수의 표준오차를 크게 부풀려 정확한 검증을 할 수 없게 한다. 다중공선성의 문제는 회귀모형 설정에 기인하는 것이 아니라 주로 자료의 결합에서 야기된다. 다중공선성은 회귀식의 잔차 분석을 통해서도 감지해낼 수 없기 때문에, 다른 방법을 통해 진단해야 한다[5].

일반적으로 상관관계가 0.7보다 낮으면 다중회귀분석을 시행하고, 0.7보다 크면 다중공선성 존재 여부를 확인한다. 다중공선성을 진단하는 방법은 분산팽창지수

(VIF: Variation Inflation Factor), 공차한계(Tolerance), 상태지수(Condition Index) 등이 있으며, VIF 측정법이 가장 많이 사용된다.

$$VIF_k = 1/(1 - R_k^2) \quad (2)$$

식 (2)에서 R_k^2 는 X_k 를 종속변수로 하고 나머지 독립변수에 대하여 회귀방정식을 적합시킬 때 얻어지는 결정계수를 의미한다. 일반적으로 VIF가 10이상 일 때 두 변수간에 다중공선성이 존재한다고 판단한다[4].

3.2.3 다중공선성의 해결

다중공선성을 해결하기 위해서는 변수를 제거하거나 최소제곱법의 대안으로서 주성분 회귀분석(Principal Components Regression)과 능형회귀(Ridge Regression)를 사용할 수 있다[12, 15].

본 연구의 CER 개발과정에서 사용한 주성분 회귀분석의 개념은 다음과 같다. 주성분 회귀분석은 상관관계가 있는 원래의 변수들을 변환하여 서로 독립적인 새로운 변수들을 유도하는 방법이다. 새로운 변수인 주성분은 원래 변수들의 선형결합으로 표시된다. 이때 각 주성분이 보유하고 있는 정보의 양은 분산으로 측정되기 때문에 분산의 크기에 따라 정보의 양이 가장 큰 주성분이 첫 번째가 된다. 각 주성분 선형결합들의 계수는 다음 세 가지 조건을 만족시키는 값을 선택하게 된다.

- (1) $\text{Var}(C_1) \geq \text{Var}(C_2) \geq \dots \geq \text{Var}(C_p)$
- (2) 두 주성분 사이는 서로 독립이다.
- (3) 각 주성분 계수들의 자승합은 1이다.

식 (3)과 같은 주성분 회귀모형에서 첫 번째 주성분 C_1 은 p 개 변수(X_1, X_2, \dots, X_p)들의 모든 가능한 선형결합 중 계수들의 자승합이 1이라는 조건에서 가장 큰 분산을 갖는 합성변수이다. 또한 두 번째 주성분 C_2 는 첫 번째 주성분과 서로 독립이면서 두 번째 주성분의 계수들의 자승합이 1이라는 조건에서 첫 번째 주성분이 설명하고 남은 분산 중에서 가장 큰 분산을 갖는 합성변수이다. 이러한 과정을 계속하면 i 번째 주성분 C_i 는 이미 구한 주성분들과 서로 독립이면서 나머지 분산 중에서 가장 큰 분산을 갖는 합성변수이다.

$$Y = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \dots + \beta_p C_p + \epsilon \quad (3)$$

$$C_k = a_{1k} X_1 + a_{2k} X_2 + \dots + a_{pk} X_p, \quad k = 1, 2, \dots, p$$

주성분은 변수들의 분산-공분산행렬 또는 상관행렬에 의하여 구할 수 있다. 분산-공분산행렬에서 얻어진 주성

분은 변수들의 사용된 측정단위에 의하여 매우 민감하다. 예를 들어, 어느 특정 변수의 측정단위가 상대적으로 다른 변수들보다 크기 때문에 매우 큰 분산을 가지고 있다면, 첫 번째 주성분은 변수들 사이에 존재하는 상관구조에 관계없이 가장 큰 분산을 갖는 변수에 좌우되는 경향이 있음을 보이게 된다. 이러한 경우에는 원래 변수를 이용하기 보다는 변수를 표준화시킨 다음에 주성분 분석을 하는 것이 바람직하다.

일반적으로 변수들의 측정단위가 서로 다른 경우에는 상관행렬을 이용하여 주성분 회귀분석을 한다[4]. 표준화 변수는 평균으로부터 편차를 표준편차로 나눈 값을 의미하며 표준화변수의 분산-공분산행렬은 변수들의 상관행렬이 된다. 따라서 표준화변수를 사용할 경우 주성분은 분산-공분산행렬 대신에 상관행렬에서 유도된다.

3.2.4 CER의 타당성 확인

일반적으로 회귀모형의 적합성은 회귀계수의 타당성을 평가하기 위한 t-검정, 회귀식의 유의성을 평가하기 위한 F-검정, 회귀식의 설명력을 진단하기 위한 R²검정의 통계적 방법을 사용한다.

그러나 CER은 향후 유사 무기체계에 대한 실질적인 비용을 추정하기 위해서 반드시 실비용 대비 예측비용 비교를 통한 오차 확인, 정확성 검증, 실증분석 등을 이용한 현실적인 타당성을 확인해야 한다.

4. CER 개발

4.1 자료 수집

CER의 가치는 수집된 자료가 얼마나 신뢰할 수 있는가에 달려있다. 무기체계에 대한 자료를 수집하기 위해서는 국가 연구기관, 방산업체 등의 실적자료를 보유하고 있는 출처로부터 공인된 자료를 수집하는 것이 중요하다. 본 연구를 수행하기 위하여 국내에서 무기체계 연구개발을 주도하고 있는 “국방과학연구소”와 협조하여 284건의 개발사례에 대하여 <표 1>과 같은 25종류의 연구개발 자료를 획득하였다.

획득한 자료를 무기체계 대분류(첫 번째 단계)인 8가지

<표 1> 자료 수집 현황

구 분	내 용
비용 자료	4건(R&D 비용, 생산 단가, 도입 단가, 조변액)
규격 자료	17건(중량, 길이, 속도, 탑승인원, 무장, 방호력 등)
사업 자료	5건(R&D 기간, 수량, 제작업체, 국산화율 등)

분야별(지휘통제·통신, 감시·정찰, 기동, 함정, 항공기, 화력, 방호, 기타)로 분류하여 분석한 결과 개체 수가 적당하고 개체간 특성이 유사한 중분류(두 번째 단계) 수준에서 CER을 개발하는 것이 타당하다고 판단되었다. 그러나 27개의 중분류 수준 역시 9개 분야는 무기체계간 유사성이 적고, 16개 분야는 자료의 수가 부족하여 통계적 분석이 제한되었다. 최종 분석결과 기동, 화력 분야에서 통계분석이 가능한 것으로 판단되었다.

4.2 자료 평가 및 정규화

본 연구에서는 개발 건수와 개체 특성의 유사성을 고려하여 기동 무기체계 분야를 <표 2>와 같이 선정하여 CER을 개발하였다.

<표 2> 분석 대상 무기체계

자 료 수	기동 무기체계
4건	K1A1, K-2, K-200, K-21

수집된 자료는 상호 비교하여 활용할 수 있도록 연도와 단위를 정규화하여 일관성을 유지하였다. 예를 들어, 비용 현가화를 위해 연구개발비는 물가상승률을 고려한 2009년도 기준 불변가로 환산하였다.

추정하고자하는 종속변수를 연구개발비로 하고, 비용 인자는 <표 3>과 같이 전투중량, 탑승인원, 엔진출력, 항속거리, 최대속도와 전차와 장갑차를 구분하는 지시변수(Dummy variable)로 선정하였다.

<표 3> 기동 무기체계 비용인자

무기 체계	전투 중량 (톤)	탑승 인원 (명)	엔진 출력 (hp)	항속 거리 (km)	최대 속도 (km/h)	지시 변수	연구 개발비 (억원)
K1A1	54.5	4	1200	400	60	1	511.08
K-2	55	3	1500	450	70	1	2727.23
K-200	13.2	12	280	480	74	0	169.33
K-21	25	12	750	450	74	0	1054.55

그러나 <표 3>의 자료는 자료 수가 4개인데 비해 변수의 개수가 6개로 회귀분석을 수행하기에는 한계가 있다. 따라서 통계 패키지인 ‘Crystal Ball’의 Correlation matrix를 활용하여 무기체계별 19개씩 자료를 추가 생성하여 전체 자료의 수를 80개로 조정하였다. 변수 생성간 각 변수는 정규분포를 따른다고 가정하였고, 원 변수들과 상관계수가 유사하도록 하였다[3]. 생성된 자료의 기초 통계량은 <표 4>와 같다.

<표 4> 추가 생성한 자료의 기초 통계량

변 수		원변수	생성된 변수	비 율
전투 중량	평 균	36.9	36.8	1.00
	표준편차	21.1	18.7	1.13
탑승 인원	평 균	7.75	7.66	1.01
	표준편차	4.9	4.25	1.16
엔진 출력	평 균	932.5	932.2	1.00
	표준편차	533.1	466.4	1.14
항속 거리	평 균	445.0	444.2	1.00
	표준편차	33.17	38.04	0.87
최대 속도	평 균	69.5	69.5	0.99
	표준편차	6.61	8.28	0.80
지시 변수	평 균	0.5	0.5	1.00
	표준편차	0.58	0.50	1.15

4.3 상관관계 분석

독립변수인 전투중량, 탑승인원, 엔진출력, 항속거리, 최대속도의 5가지 물리적 특성치와 지시변수에 대한 상관관계 분석 결과 <표 5>와 같이 대부분의 변수간 상관관계수가 매우 높아 다중공선성의 존재가 의심되었다.

<표 5> 변수간 상관계수

	전투 중량	탑승 인원	엔진 출력	항속 거리	최대 속도	지시 변수
전투중량	1					
탑승인원	-0.94	1				
엔진출력	0.96	-0.89	1			
항속거리	-0.51	0.38	-0.38	1		
최대속도	-0.51	0.46	-0.34	0.88	1	
지시변수	0.93	-0.98	0.87	-0.31	-0.40	1

정확한 진단을 위해 VIF 값을 계산한 결과 <표 6>에 서처럼 전투중량, 탑승인원, 엔진출력, 지시변수의 VIF 값이 10이상으로 다중공선성이 존재함을 확인하였다.

<표 6> 변수별 VIF 값

변 수	VIF
intercept	0
전투 중량	75.30704
탑승 인원	34.04295
엔진 출력	32.02866
항속 거리	7.72641
최대 속도	6.70667
지시 변수	51.77642

4.4 주성분 회귀분석

본 연구에서는 독립변수의 단위가 다르기 때문에 표준화변수를 이용한 상관행렬로서 주성분 회귀분석을 시행하였다. 상관행렬의 고유치, 주성분별 전체 분산에 대한 설명비율과 누적 설명비율을 계산한 결과는 <표 7>과 같다.

첫 번째 주성분 C₁의 분산 λ_1 는 4.33838이고, 전체 표본분산의 72.31%를 설명한다. 두 번째 주성분 C₂의 분산 λ_2 는 1.34933이고, 전체 표본분산의 22.49%를 설명한다. 주성분 C₁, C₂를 제외한 나머지 네 개의 주성분의 고유값($\lambda_3, \lambda_4, \lambda_5, \lambda_6$)은 0에 가깝고, 설명비율이 매우 작기 때문에 설명력은 거의 없다고 할 수 있다. 따라서 보유 하계될 주성분의 개수는 두 개가 되기 때문에 6차원의 자료를 2차원으로 축소할 수 있으며, 이를 통해 원변수가 포함하고 있는 정보의 손실을 최소화할 수 있다.

<표 7> 상관행렬의 고유치

	Eigenvalue(λ_i)	Proportion	Cumulative
C ₁	4.33838	0.7231	0.7231
C ₂	1.34933	0.2249	0.9480
C ₃	0.21312	0.0355	0.9835
C ₄	0.06318	0.0105	0.9940
C ₅	0.02941	0.0049	0.9989
C ₆	0.00657	0.0011	1.0000

주성분 C₁, C₂에 대한 값을 구하기 위하여 λ_i 에 대응되는 고유벡터행렬을 계산한 결과는 <표 8>과 같다. 첫 번째 고유값에 대응되는 고유벡터의 원소는 첫 번째 주성분에서 표준화변수의 선형결합계수로 사용된다. 같은 방법으로 두 번째 고유값에 대응되는 고유벡터의 원소는 두 번째 주성분에서 표준화 변수의 선형결합계수로 사용된다.

<표 8> 표준화 변수에 대한 주성분

	Eigenvectors					
	C1	C2	C3	C4	C5	C6
전투중량	0.4700	0.1259	0.1984	0.1272	0.5274	-0.6555
탑승인원	-0.4554	-0.2123	0.3077	0.2838	0.6812	0.3290
엔진출력	0.4372	0.2516	0.5767	0.4223	-0.2624	0.4072
항속거리	-0.2959	0.6472	-0.4026	0.5668	-0.0016	-0.1010
최대속도	-0.3109	0.6207	0.4350	-0.5698	0.0478	-0.0441
지시변수	0.4432	0.2676	-0.4267	-0.2811	0.4321	0.5331

따라서 추정된 두 개의 주성분은 식 (4)와 같다.

$$C_1 = 0.4700Z_1 - 0.4554Z_2 + 0.4372Z_3 - 0.2959Z_4 - 0.3109Z_5 + 0.4432Z_6$$

$$C_2 = 0.1259Z_1 - 0.2123Z_2 + 0.2516Z_3 + 0.6472Z_4 + 0.6207Z_5 + 0.2676Z_6$$

이때, $Z_1 \sim Z_6$ 은 원 변수인 전투중량, 탑승인원, 엔진 출력, 항속거리, 최대속도, 지시변수의 표준화 변수를 의미하며, $Z_1 \sim Z_6$ 값을 대입하여 주성분 변수 C_1, C_2 의 점수를 구하면 <표 9>와 같다.

<표 9> C_1, C_2 변수에 대한 주성분 점수

무기체계	연구개발비	C_1	C_2
KIA1	511.08	2.0717	-0.4180
:	:	:	:
K-2	2727.23	1.8828	1.0233
:	:	:	:
K-200	169.33	-2.4412	-0.2950
:	:	:	:
K-21	1054.55	-1.5516	-0.2966
:	:	:	:

두 개의 주성분 점수 C_1, C_2 에 대한 다중회귀분석 결과는 식 (5)와 같다.

$$\text{연구개발비} = 1127.94 + 210.1C_1 + 567.15C_2 \quad (5)$$

4.5 CER 검증

4.5.1 통계적 검증

첫째, CER에 대한 분산분석 결과 <표 10>과 같이 F 값이 61.13으로 큰 값을 가지며, p-value 값이 0.0001로 나타났다. 따라서 모형의 유의성과 설명력이 좋기 때문에, 연구개발 비용추정 관계식이 실적 자료 값들을 잘 설명한다고 할 수 있다.

둘째, R^2 값은 0.6139이고, Adj R^2 값이 0.6035로서 비

<표 10> 분산분석 Table

Source	DF	SS	MS	F	p
Model	2	49416851	24708425	61.13	0.0001
Error	77	31121572	404176		
Total	79	80538422			

$R^2 = 0.6136, \text{Adj } R^2 = 0.6035$

용추정 함수의 변동 중에서 60%가 모형의 독립변수 값으로 설명된다고 할 수 있다. 따라서 추정된 연구개발 비용추정 관계식에 의한 설명력은 실적자료임을 고려시 적절하다고 할 수 있다.

셋째, 회귀계수에 대한 t-검정 값은 <표 11>과 같이 p-value 값이 0.0001이하로 유의수준 1%에서 유의하기 때문에, 3개 변수 모두 채택 가능하다고 할 수 있다.

<표 11> 회귀계수에 대한 t-검정

Variable	DF	Parameter Estimate	Standard Error	T	p
Intercept	1	1127.94	71.08	15.87	<0.0001
C_1	1	210.10	34.34	6.12	<0.0001
C_2	1	567.15	61.58	9.21	<0.0001

4.5.2 최종 CER 모형

주성분 회귀분석을 통해 얻은 CER은 각 독립변수의 표준화 변수와 주성분 C_1, C_2 를 통해 회귀계수를 추정 하였으므로 이를 식 (6)의 과정을 통해 원 변수로 변환 하여야 한다.

$$C_1 = 0.4700 \times \frac{X_1 - 36.81}{18.73} - 0.4554 \times \frac{X_2 - 7.66}{4.25} + 0.4372 \times \frac{X_3 - 932.18}{466.40} - 0.2959 \times \frac{X_4 - 444.18}{58.05} - 0.3109 \times \frac{X_5 - 69.53}{9.28} + 0.4432 \times \frac{X_6 - 0.5}{0.5}$$

$$C_2 = 0.1259 \times \frac{X_1 - 36.81}{18.73} - 0.4554 \times \frac{X_2 - 7.66}{4.25} + 0.4372 \times \frac{X_3 - 932.18}{466.40} + 0.6472 \times \frac{X_4 - 444.18}{58.05} + 0.6207 \times \frac{X_5 - 69.53}{9.28} + 0.2676 \times \frac{X_6 - 0.5}{0.5}$$

$X_1 \sim X_6$ 은 전투중량, 탑승인원, 엔진출력, 항속거리, 최대 속도, 지시변수를 의미하며, 식 (6)을 식 (5)에 대입하여 식 (7)과 같이 최종 CER을 산출하였다.

$$\begin{aligned} \text{연구개발비} &= -4,010.67 + 9.08 \times \text{전투중량} \\ &\quad - 50.80 \times \text{탑승인원} + 0.50 \times \text{엔진출력} \\ &\quad - 5.25 \times \text{항속거리} + 30.89 \times \text{최대속도} \\ &\quad - 486.60 \times \text{지시변수} \end{aligned} \quad (7)$$

4.5.3 타당성 검증

산출된 최종 CER은 오차 확인, 정확성 검증, 실증분석 등의 타당성 검증이 요구된다. 본 연구에서는 무기체계 개발 건수의 제한으로 척도에 의한 정확성 검증을 수행

하였다. 척도에 의한 검증 결과가 명시된 기준보다 우수한 경우에는 검증의 성공을 의미하고, 실제적인 프로젝트에 CER을 사용할 수 있다. 정확성이 기준보다 못 미치는 경우에는 검증의 실패를 의미하고 이전 단계로 돌아가서 CER을 다시 개발해야 한다.

CER 모형의 정확성을 나타내는 대표적인 척도로 MRE(Magnitude of Relative Error)와 MMRE(Mean Magnitude of Relative Error), PRED(ℓ) (Prediction at level ℓ)가 주로 사용된다[2].

MRE는 가장 기본적으로 사용되는 척도로서 식 (8)로 표현되며, 상대오차의 절대값인 MRE가 작을수록 모델의 정확도는 높아진다.

$$MRE = |y - \hat{y}|/y, \quad y: \text{실제값}, \hat{y}: \text{추정값} \quad (8)$$

MMRE는 비용추정치 정확도의 평균으로 식 (9)와 같으며, 선정된 기준에 따라 정확성 정도를 정량적으로 명시한다.

$$MMRE = (1/N) \sum_{i=1}^N |y_i - \hat{y}_i|/y_i \quad (9)$$

PRED(ℓ)은 식 (10)과 같으며, ℓ 은 오차 범위를 의미하고, k 는 $MRE \leq \ell$ 에 포함된 개수를 의미한다[11, 13].

$$PRED(\ell) = k/n \quad \text{식 (10)}$$

MMRE, PRED(0.25)의 계산 결과는 <표 12>와 같다.

<표 12> MMRE, PRED(0.25) 결과

MMRE	PRED(0.25)
1.12	0.29

CER에 대한 MMRE를 평가하기 위하여 현재 사용하고 있는 비용추정 모델들을 조사한 결과 소프트웨어 비용추정 모델로 널리 활용되고 있는 COCOMO II 모델의 MMRE 값이 1.555로 확인되었다[11]. 본 연구에서 제시한 CER의 MMRE 값은 1.12로 상용화된 모델보다 더 낮게 나타났다. 또한, PRED(0.25) 값 역시 29%로서 Conte et al.(1986)이 제시하고 있는 27%[13]와 유사한 수준을 보이고 있기 때문에 척도에 의한 CER의 정확성은 적절하다고 할 수 있다.

5. 결론 및 향후연구 방향

제안한 기동 무기체계 연구개발비 CER은 국내 연구

개발 자료를 활용하여 최초로 CER을 구축한 사례로서 다음과 같은 의의를 갖는다.

첫째, 연구에서 제시한 자료의 가용수준에 따라 CER을 개발하고 검증하는 전반적인 절차는 향후 CER 개발의 표준절차로 활용할 수 있을 것이다.

둘째, 실적 자료가 부족한 현실을 극복하여 비용추정 관계식을 개발할 수 있는 방법론을 보였다. 특히 자료의 수가 적은 경우에 자료를 추가 생성하여 분석하는 방법과 주성분 회귀분석을 통해 변수를 제거하지 않고 다중공성선을 해결하는 절차는 자료의 수가 적은 국내 현실을 고려시 유용한 방법론으로 활용될 수 있을 것이다.

셋째, 개발된 CER을 활용하여 기동 분야의 무기체계 개발시 시스템 수준에서의 기본적인 물리적 특성을 활용하여 적정 연구개발비를 추정할 수 있을 것이다.

본 연구에서는 부족한 자료 현황을 극복할 수 있는 CER 개발 방법론을 제시하였다. 그러나 보다 효율적이고 타당성을 확보할 있는 CER 개발을 위해서는 다음과 같은 내용에 대한 추가적인 연구와 지속적인 자료축적이 병행되어야 할 것이다.

첫째, 동일 계열에 대한 국내 무기체계 개발 사례의 부족으로 기동 무기체계 분야만을 분석하였다. 따라서 다른 분야에 대한 CER 개발의 효율성을 제고하기 위해서는 본 연구에서 시도한 자료의 수를 추가 생성하는 방법 외의 자료부족 상황을 해결할 수 있는 다양한 접근을 시도할 필요가 있다.

둘째, 유사 사례의 제한으로 본 연구에서 제시한 CER의 실증분석을 통한 타당성 평가가 제한되었다. 따라서 이를 극복할 수 있는 해외 유사 사례의 발굴 등의 추가적인 노력이 필요할 것이다.

마지막으로, 회귀모형의 단점인 경험자료의 사용은 복잡소재의 사용, 신개념에 따른 SW와 HW의 통합 난이도 증가 등 첨단 기술을 적시적으로 반영하는 것이 제한된다. 따라서 무기체계 첨단화 등을 반영할 수 있는 CER 개발을 통해 미래의 비용을 정확히 추정할 수 있는 실질적인 연구가 계속되어야 할 것이다.

참고문헌

- [1] 강성진; *효율적인 비용관리 방법론*, 국방대학교, 2008.
- [2] 고민주; “제안서 작성 단계에서의 효과적인 소프트웨어 규모추정 방안”, 석사학위논문, 강릉대학교, 2007.
- [3] 김기호, 김영일; *크리스탈볼을 이용한 재무시뮬레이션*, 이레테크, 2007.
- [4] 김종섭; *SAS를 이용한 통계분석의 이해*, 한울출판사, 2003.
- [5] 김진욱; “다중공선성 상태의 주성분회귀와 능형회귀”,

- 한국체육학회지, 45(4) : 547-556, 2006.
- [6] 박성현, *회귀 분석*, 민영사, 2007.
- [7] 백종문 외 3명; “국방 소프트웨어 비용추정모델 개발 방안 연구”, 국방과학연구소, 2009.
- [8] 이재용 외 3명; “전문가 사전지식기반 국방연구개발 인력추정모델 개발”, 국방과학연구소, 2008.
- [9] 이재용 외 3명; “전문가 사전지식기반 국방연구개발 비용추정 모델(2단계) 개발”, 국방과학연구소, 2008.
- [10] 이재용 외 7명; “전문가 사전지식 기반 국방연구개발 비용추정모델 개발”, 국방과학연구소, 2006.
- [11] Boehm, B. W. et al.; *Software Cost Estimation With COCOMO II*, Prentice Hall, 2000.
- [12] Chatterjee, S., A. S. Hadi, B. Price; *Regression analysis by example*, John Wiley and Sons, 2000.
- [13] Conte S. D., H. E. Dunsmore, V. Y. Shen; *Software Engineering Metrics and Models*, Benjamin-Cummings, 1986.
- [14] ISPA(International Society of Parametric Analysts), *Parametric Estimating Handbook*, ISPA Asian Chapter, 2007.
- [15] Montgomery, D. C., E. A. Peck., G. G. Vining; *Introduction to linear regression analysis*, John Wiley and Sons, 2001.
- [16] Tomas, G. B., R. L. Finney; *Calculus and analytic geometry*, Addison-Wesley, 1996.