

CER 선형결합을 통한 비용추정 모델 개발

정원일 · 이용복[†] · 김동규 · 강성진

국방대학교 운영분석과

A Cost Estimation Development Methodology via CER's Linear Combination

Won Il Jung · Yong Bok Lee · Dong Kyu Kim · Sung Jin Kang

Department of Operations Research, Korea National Defense University

The acquisition cost of defense weapon system has been continuously increasing because of art-of-technology of it. This phenomenon requires efficiency and transparency in the weapon system acquisition process through cost estimation. Therefore cost estimation is very important to the government acquisition programs to support decisions about funding and to evaluate resource requirement as a key decision point. The Commercial parametric cost estimating models have been using extensively to obtain appropriate cost estimates in early acquisition phase. These models have many restrictions to ensure the cost estimating result in Korean defense environment because they are developed based on foreign R&D data. Also estimation results are different from Korean defense industry accounting system. So, some studies have been tried to develop a CER (Cost Estimation Relationship) based on the Korean historical data. However, there are some restrictions to improve the predictability and ensure the stability of the developed singular CERs which consider the following data characteristics individually. The the abnormal conditions of data that is multicollinearity, outlier and heteroscedasticity under rack of the number of observations.

In this paper, a CER's Linear Combining Model is proposed to overcome those limitations which guarantee more accurate estimation (25.42% higher precision) than other singular CERs. At least, this study is meaningful as a first attempt to improve the predictability of CER with insufficient data. The methodology suggested in this study will be useful to develop a complex Korean version cost estimating model development in future.

Keyword: combining model, cer(cost estimation relationship), multicollinearity, outlier, heteroscedasticity

1. 서론

국방 무기체계 획득에 대한 국·내외적인 환경 변화로 기존의 해외 직구매를 통한 무기체계 획득에서 국내 연구개발에 의한 무기체계 획득으로 패러다임이 변화하고 있다. 또한, 국방 무기체계의 첨단화 및 정밀화에 따라 무기체계 획득비가 증대됨에 따라 획득 초기 단계부터 비용분석을 통한 의사결정지원과

예산 사용의 효율성을 높이는 것이 강조되고 있다. 획득 초기 단계에서의 비용분석은 PRICE, SEER와 같은 해외의 상용 비용 추정 모델에 의존하고 있다. 그러나 해외 모델에 대하여 국내 기관 및 방산업체의 비용분석 전문가를 대상으로 설문을 실시한 결과 다음과 같이 몇 가지 중요한 시사점이 도출되었다 (Kang *et al.*, 2010).

첫째, 상용 모델은 사용자 만족도 측면에서 모델 운용을 위

한 자료 획득 곤란, 다수의 입력값 요구 등 사용자 편의성이 전반적으로 낮은 것으로 나타났다. 둘째, 모델 활용성 측면에서 방산원가 체계와 비용항목의 구성요소가 불일치하여 다른 비용추정 방법과 교차검증이 제한되는 것으로 나타났다. 셋째, 추정결과와 정확성 측면에서 사용자의 모델 이해 수준에 따라 상이한 결과가 발생하고, 운용자의 이해관계에 의해서 비용추정 결과를 임의적으로 맞추는 사례가 빈번한 것으로 나타났다.

결론적으로 해외의 비용추정 모델은 모델의 기본 값(Default)에 객관적인 최소한의 필수 요소만 입력하여 획득 초기단계에서 신속하고 대략적인 예산추정에 활용하는 것이 타당하지만, 근본적으로 한국적 개발 환경이나 생산 여건을 반영하지 못함으로 인해 정확성에 항상 의문이 제기되어 왔다.

이러한 상황에서 국내 실적자료 기반의 비용추정 모델 개발의 핵심인 비용추정관계식(CER : Cost Estimation Relationship)과 관련된 연구가 시도되었으나, 연구 단계가 초보적인 수준이어서 정확하고 안정적인 비용추정이 곤란하다. 이러한 제한이 발생하는 이유는 첫째, 국내 무기체계의 개발 실적 건수는 통계적으로 부족함에도 불구하고 개발된 다수의 CER 중 가장 적합한 하나의 CER만을 선택하는 절차를 일반적으로 사용하고 있었다. 따라서 선정되지 않은 CER 중에 미래 예측에 보다 우수한 CER이 있을 수 있을 가능성이 배제되고 있다. 둘째, 자료수의 부족으로 선정된 2~3개의 비용인자 위주로 비용을 추정함으로써 모델에 포함되지 않은 비용인자 중 비용에 결정적인 요소가 있을 가능성을 내포한 CER 개발 방법을 제시하고 있다.

본 연구에서는 이러한 제한사항을 완화할 수 있는 CER 선형결합 비용추정모델 개발 방법을 제시하였다. 이를 위해 소수의 자료이면서 비정상적인 자료특성에 통계적으로 유의한 회귀모형을 적합하고, 다수의 CER을 정확도에 따라 가중 선형결합 하였다. 본 연구를 통해 개별 CER 개발위주의 비용추정 모델 개발 연구에 정확도와 예측력을 향상시킬 수 있는 새로운 접근법인 결합방법에 대한 연구의 필요성을 제기하고자 하였다.

2. 문헌 연구

2.1 매개변수 비용 추정법

통계적 기법을 이용하여 무기체계 비용을 종속변수, 비용인자를 독립변수로 두어 비용인자와 비용간의 함수관계를 표현한 것을 CER이라고 하며 CER을 이용하여 비용을 추정하는 것을 매개변수 비용 추정법이라 한다. 매개변수에 의한 비용추정은 자료수집, 비용주도요인 식별, CER 개발, 모형구축의 단계로 진행된다(Kang *et al.*, 2010). 이 방법은 입력변수의 조정으로 비용추정을 비교적 쉽게 할 수 있는 반면, 모델의 타당성을 보장하기 위해서는 방대한 양의 자료가 요구되는 한계가 있다.

2.2 결합모델 관련 연구

회귀모형의 결합은 Bates와 Granger가 최초로 제시하였으며,

실험적 분석을 통해 결합 모델이 비결합 모델에 비해 오차를 감소시킨다는 것을 증명했다(Bates *et al.*, 1965). 또한, 1992년 “International Journal of Forecasting”에서 결합 모델관련 검증 실험에 참가한 학자의 83%가 결합 모델이 개별 모델에 비하여 예측 오차가 최소화됨을 입증하였다. Armstrong은 규칙에 기반한 예측 모형(Rule-based forecasting)을 제시함으로써 결합 모델을 효과적으로 적용할 수 있는 방법에 대해 정의하였다(Armstrong, 1989). Armstrong은 30개의 연구결과에 대한 메타 분석결과로서 결합 모델이 개별 모델에 비해 평균적으로 12.5% 정도 오차를 감소시키는 것에 대해 경험적인 방법으로 타당성을 제시함으로써 결합 모델의 우수성을 입증하였다(Armstrong, 2001).

국내에서 결합 모델과 관련된 연구 결과는 거의 알려지지 않았으며, 현재 기상청 예보국 국가태풍센터에서 TAPS-2에 따른 현업 태풍예보 방법으로 결합 모델 방법을 적용하고 있으며, 정확한 예측을 하기에는 비교적 변수가 많은 태풍분석 및 예보시스템에 활용하고 있다(The Meteorological Administration, 2009).

특히 국방 분야에서는 결합 모델과 관련된 연구 실적은 전무하며, 다음과 같이 소수의 개별 CER을 개발한 사례가 있다. 첫째, Eo *et al.*(2010)과 Lee *et al.*(2008, 2006)은 실적 자료 수집의 제한으로 전문가 설문 기반으로 유도무기체계와 어뢰 및 전차 체계 비용추정 모델을 개발하였다. 둘째, Eo *et al.*(2010)은 연구 개발 실적 자료를 분석하여 기동 무기체계에 대한 비용인자를 식별해 내고 다중공선성을 제거하기 위해 주성분 회귀분석을 이용하여 기동 무기체계 연구개발비 CER을 개발하였다.

그러나 상기에서 제시한 국내의 연구들은 설문을 기초로 개발되어 객관성 보장에 한계가 있으며, 자료가 부족한 경우에 흔히 발생하는 비정상적인 자료의 특성을 효과적으로 처리할 수 있는 방법을 제시하지 않은 한계가 있다.

3. CER 선형결합 모델 개발 방법

3.1 결합 모델의 기본 개념

본 연구에서 제시하는 결합 모델은 상기에서 제시한 자료특성을 고려하여 개발된 개별 CER들에 가중치를 부여하여 선형결합하는 방법이다. 결합 모델은 개별 CER보다 많은 정보를 근거로 예측하여 비용추정의 오차를 감소시키며, 정확도에 평균적으로 오차가 적은 관측치에 더 많은 가중치를 부여함으로써 추정값이 개별 모델보다 상대적으로 실제값에 더 가까워지게 하는 경향을 유도한다.

3.2 CER 선형결합 모델 개발 절차 및 방법

(1) CER 개발 절차

CER 선형결합 모델 개발은 <Figure 1>처럼 크게 자료 수집 및 가공, 자료분석, 개별 CER 개발, 선형결합 CER 개발, CER 검증 및 개발 완료 순으로 진행된다.

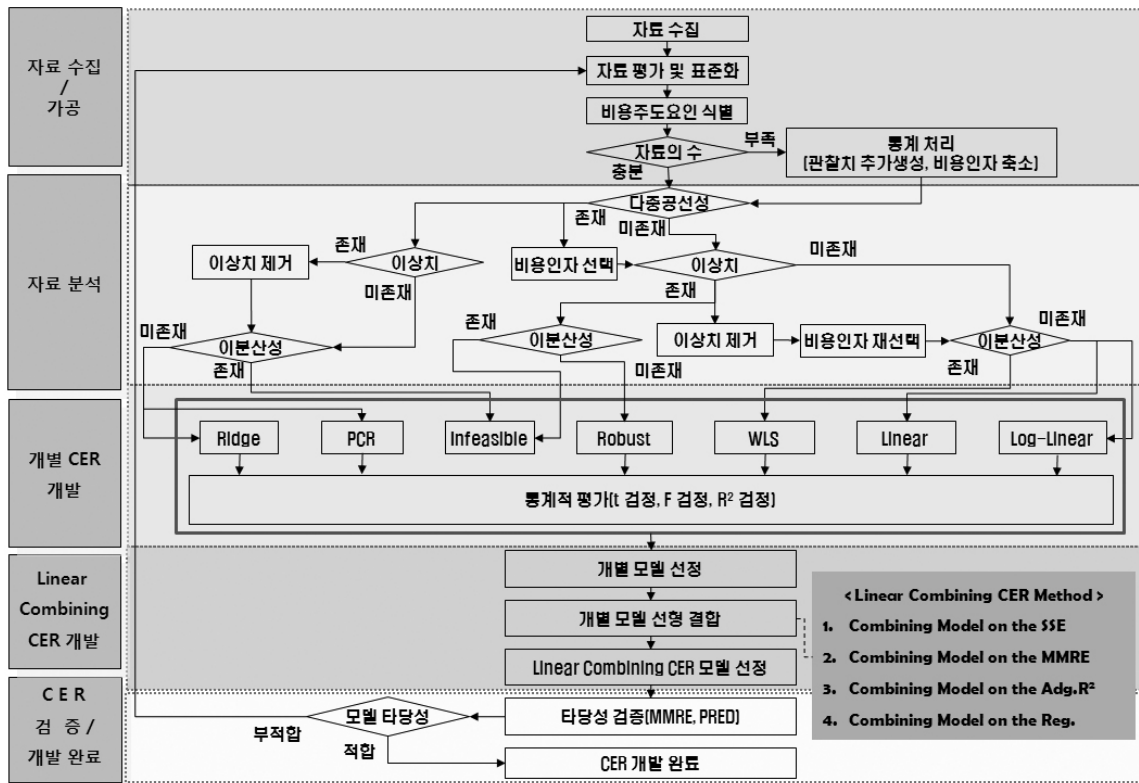


Figure 1. CER 선형결합 모델 개발도

(2) 자료 수집 및 가공

파라메트릭 비용추정 모델의 타당성은 자료의 신뢰성에 의존적이기 때문에 무기체계 획득 관련 권위 있는 기관인 방위사업청, 국방과학연구소 등과 방산업체 등으로부터 비용, 기술, 사업과 관련된 자료를 수집한다.

수집된 자료는 6개 영역으로 분류하여 자료를 평가하고 표준화한다(Kang, 2010). 구체적으로 비용자료는 수량 및 화폐가치를 표준화하며, 규격자료는 중량, 밀도 등의 척도를 표준화한다. 유사한 자료는 유사성의 수준에 따라 가능한 경우에는 동일한 유형으로 보정하고 적용이 불가능한 자료는 제거한다. 또한 무기체계의 운영과정에서 소요되는 반복 및 미반복 비용을 구분하고, 학습곡선 이론을 적용하여 생산 물량을 표준화한다. 마지막으로 무기체계의 운영 환경을 구분하여 표준화한다.

수집된 자료는 비용과의 관계에 따라서 비용주도요인을 식별한다. 이는 무기체계 분야별 전문가들과의 인터뷰 및 비용인자와 비용간의 산포도를 통해 식별한다.

일반적인 회귀 방정식은 $n-k \geq 2(n : \text{자료 수}, k : \text{비용인자 개수})$ 의 조건이 충족될 때 분석이 가능하다. 이때, 자료의 수가 부족하여 직접적인 통계 분석이 불가능한 경우에는 관찰치를 추가 생성(Random Generation)하거나, 비용인자 축소(Variable Reduction)를 통해 문제를 해결한다.

(3) 자료 분석

첫 번째 단계는 다중공선성 판단으로서 다중공선성은 VIF

(Variance Inflation) >10 또는 CI(Condition Index) >30 일 때 심각하다고 판단한다(Kim, 2002; Youm, 2005). 다중공선성의 존재는 CER에서 회귀계수 추정값 상관관계만큼의 변화를 제외하기 때문에 사용할 수 있는 정보량의 감소 즉, 표본의 수를 감소시키는 것과 동일한 효과를 미침으로서 추정계수의 분산과 표준오차를 증가시킨다. 따라서 다중공선성의 심각성 정도는 해당 비용인자에 대한 다른 비용인자들 모형의 결정계수를 이용하여 지수화한 VIF와 CI를 이용하여 판단할 수 있다.

두 번째 단계는 다중공선성 판단에 따른 조치로서 다중공선성이 존재하지 않으면 이상치를 판단하고, 다중공선성이 존재한다면 다음과 같은 절차로 조치한다. 첫 번째는 주성분회귀 및 능형회귀를 적용하기 위해 비용인자를 제거하지 않은 상태에서 이상치를 판단한다. 두 번째는 변수 선택 방법론을 적용하여 비용인자를 제거하여 다중공선성이 없는 비용인자 조합을 선택한다. 변수 선택은 6가지 방법(R-Square, Adjusted R-Square, Forward, Backward, Stepwise, C(p) Selection)으로부터 복수 선정된 비용인자 조합에 대해 다중공선성을 재판단하고 이상이 없을 시 이상치를 판단한다. 이때 복수 선정된 모든 조합에서 다중공선성이 다시 발견되면 R^2 가 0.8 이상인 조합에 대해서 R^2 가 높은 조합부터 차례로 다중공선성을 진단하여 다중공선성이 제거된 첫 번째 조합으로 다음 단계를 수행한다.

세 번째 단계는 이상치 판단으로서 이상치 판단 기준은 표준화 잔차(studentized residual) 및 표준화 제외잔차(studentized deleted residual)를 이용하여 비용에 대한 이상치를 판단하고 지

렛대값(Leverage)를 이용하여 비용인자의 이상치를 판단한다. 이 때 표준화 잔차에 의한 이상치 판단은 Lund가 제시한 임계값을 기준으로 하며, 이를 초과하면 이상치로 판단한다(Lund, 1975). 표준화 제외잔차에 의한 이상치 판단은 국내 연구개발 자료가 부족한 점을 감안하여 유의수준 1%를 기준으로 판단한다. 지렛대값에 의한 이상치 판단은 지렛대값이 $2(p+1)/n$ 보다 클 때 이상치로 간주한다(Kim, 2002; Youm, 2005). 이상치로 판단된 값에 대해서는 쿡의 거리(Cook's Distance)를 적용하여 이상치의 전체 적합값에 대한 영향력을 판단하고 모수값 변화(DFBETAS : Difference in Betas)를 통해 회귀계수의 영향력을 판단한다.

네 번째 단계는 이상치 판단에 따른 조치로서 이상치가 존재하고 영향력이 클 때에는 다음 두 가지 방법으로 문제를 해결한다. 첫째, 로버스트 회귀를 통한 이상치를 해결하기 위해 이상치를 제거하지 않고 이분산성을 판단한다. 둘째, 관측값 제거에 따른 이상치 문제 해결을 위해 이상치를 제거한 후 비용인자 선택을 다시 수행하고 이분산성을 판단한다. 이상치 존재여부에 관계없이 각 요소들의 영향력이 약할 때에는 이분산성을 판단한다.

다섯 번째 단계는 잔차 산포도에 의해 이분산성을 판단하는 것으로서 산포도가 비용인자(X_i)를 기준으로 잔차가 불규칙적으로 분포하면 이분산성이 존재하지 않은 것으로 판단한다. 비용인자에 따라 잔차가 일정한 경향을 보이면 이분산성이 존재하는 것으로 판단한다.

여섯 번째 단계는 이분산성 판단에 따른 조치로서 이분산성이 존재하는 경우에는 가중 회귀를 이용하여 이분산성 문제를 해결한다. 이분산성이 존재하지 않는 경우에는 1~5단계 과정을 통해 선정된 방법에 적합한 능형, 주성분, 로버스트, 선형, 로그선형 회귀를 적용한다.

(4) 개별 CER 개발

개별 CER 개발은 <Table 1>의 적용 기준과 절차에 의해 개발한다. 예를 들어 자료 분석 과정이 “다중공선성 존재 → 이상치 미존재 → 이분산성 미존재”의 경우에는 능형 회귀 또는 주성분 회귀를 적용한다.

CER은 회귀모형이기 때문에 회귀분석간 고려되어야 할 주요한 자료 특성인 다중공선성, 이분산성, 이상치의 문제점을 해결하기 위한 다양한 방법을 적용한다. 본 연구에서는 자료 특성에 따라 다중선형, 로그선형, 능형, 주성분, 로버스트, 가중 회귀를 사용하였으며, 각각의 개념과 장·단점을 간단히 요약하면 <Table 1>과 같다.

1) 다중선형 회귀에 의한 CER 개발

다중선형 회귀는 회귀 방정식을 해결하는 과정에서 오차항(ϵ)의 변동을 최소화하는 회귀계수(β_r)를 추정하기 위하여 식 (1)과 같이 CER의 각 관측치로부터 수직거리의 제곱합을 최소화하는 보통 최소제곱법(Ordinary Least Square Method)을 사용하는 방법이다. 이 방법은 모든 가능한 불편추정량들 중에서 최

Table 1. 회귀모형 적용기준

구 분	자료분석 과정
능형 회귀	◦ 다중공선성 존재 → 이상치 미존재 → 이분산성 미존재
주성분 회귀	◦ 다중공선성 존재 → 이상치 존재 → 이상치 제거 → 이분산성 미존재
로버스트 회귀	◦ 다중공선성 존재 → 비용인자 선택 → 이상치 존재 → 이분산성 미존재 ◦ 다중공선성 미존재 → 이상치 존재 → 이분산성 미존재
가중 회귀	◦ 다중공선성 존재 → 비용인자 선택 → 이상치 미존재 → 이분산성 존재 ◦ 다중공선성 존재 → 비용인자 선택 → 이상치 존재 → 이상치 제거 → 비용인자 재선택 → 이분산성 존재 ◦ 다중공선성 미존재 → 이상치 미존재 → 이분산성 존재 ◦ 다중공선성 미존재 → 이상치 존재 → 이상치 제거 → 비용인자 재선택 → 이분산성 존재
선형 회귀	◦ 다중공선성 존재 → 비용인자 선택 → 이상치 미존재 → 이분산성 미존재 ◦ 다중공선성 존재 → 비용인자 선택 → 이상치 존재 → 이상치 제거 → 비용인자 재선택 → 이분산성 미존재
로그선형 회귀	◦ 다중공선성 미존재 → 이상치 미존재 → 이분산성 미존재 ◦ 다중공선성 미존재 → 이상치 존재 → 이상치 제거 → 비용인자 재선택 → 이분산성 미존재

소분산을 갖는 β_r (BLUE : Best Linear Unbiased Estimator)를 추정함으로써 적합도가 높다.

$$\min \sum (y_q - \beta_0 - \beta_1 x_{q1} - \dots - \beta_q x_{qr})^2 \quad (1)$$

그러나 각 β_r 는 \hat{Y}_i 에 대한 X_{ir} 을 제외한 다른 비용인자의 효과를 제거한 효과를 표현하기 때문에 독립변수 간에 다중공선성이 존재하면 보통 최소제곱추정량의 최소 분산 값이 커지는 단점이 있다.

2) 로그선형 회귀에 의한 CER 개발

로그선형 회귀는 분석대상 변수가 평균에 비해서 큰 표준편차를 갖는 경우에 원래의 변수를 그대로 사용하지 않고, 식 (2)와 같이 로그 변환하여 선형성을 부여함으로써 변이를 완화시키면서 비대칭성을 줄일 수 있는 유용한 방법이다. 이 방법은 변수변환을 통해 이분산성의 문제를 효과적으로 해결할 수 있다.

$$\ln y_q = \beta_0 + \beta_1 x_{q1} + \dots + \beta_q x_{qr} + \epsilon_q \quad (2)$$

3) 가중 회귀에 의한 CER 개발

가중 회귀는 다중선형 회귀에서 잔차가 *i.i.d*(각 ϵ_r 의 평균이

0, 분산이 σ^2 인 서로 독립이고 동일한 분포) 가정을 만족하지 못하는 경우에 변수변환을 통해 분산을 안정화한 뒤 OLS를 수행하는 방법이다. 가중 회귀는 식 (3)을 최소화하는 β_r 을 추정함으로써 분산에 반비례하는 가중치를 부여하여 작은 가중치를 갖는 관찰 값이 회귀계수 결정에 미치는 영향을 적게 한다. 이 방법은 일반적으로 이분산성 문제를 효과적으로 해결할 수 있지만, 직접적인 방법으로는 가중치를 미리 알기 어려운 단점이 있다.

$$\sum \omega_q (y_q - \beta_0 - \beta_1 x_{q1} - \dots - \beta_r x_{qr})^2, \quad (3)$$

$$\omega_q = 1/\sigma_q^2$$

4) 주성분 회귀에 의한 CER 개발

주성분 회귀는 식 (4)와 같이 다중공선성의 존재로 인해 OLS 추정 값이 불안정한 경우 원래의 변수(x_{qr})대신 주성분 분석으로 구한 주성분 변수(C_k)를 사용하는 방법이다. 주성분 회귀는 직교 선형 결합된 원 변수들을 통해 변수들이 가지고 있는 전체 분산을 최대한 설명할 수 있는 상호 독립적인 주성분(Principal Components)을 유도하고, 고유 값(Eigenvalue)의 크기를 조사하여 작은 고유 값을 가진 주성분변수를 제거한 후 회귀계수를 추정한다. 따라서 편향은 있지만 더 안정된 회귀계수 추정 값을 얻을 수 있는 장점이 있다. 그러나 데이터의 일부를 제거한 축소된 변수에 의해 회귀식이 표현되는 한계와 주성분이 이상치에 영향을 많이 받는 상관행렬(Correlation Matrix)로부터 계산되는 특성으로 인해 높은 지레점을 갖는 이상치에 대해 크게 영향을 받을 수 있다는 단점이 있다.

$$Y = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \dots + \beta_r C_r + \epsilon \quad (4)$$

$$C_k = a_{1k} X_1 + a_{2k} X_2 + \dots + a_{rk} X_r,$$

$$k = 1, 2, \dots, r$$

5) 능형 회귀에 의한 CER 개발

능형 회귀는 OLS의 최소 분산 값이 매우 커질 때 식 (5)와 같은 편향상수(Ridge Parameter)를 이용하여 편의를 크게 증가시키지 않는 범위에서 고의적으로 편향된 추정(Biased Estimation)을 함으로써 전체 추정량의 분산을 감소시키는 회귀계수를 추정하는 방법이다. 이때 편향상수는 직접 계산 또는 능형 추적법(Ridge Trace Method) 등에 의해 결정할 수 있다.

$$\sum_{r=1}^v \text{Var}(\hat{\theta}_r(c)) = \sigma^2 \sum_{r=1}^v \frac{\lambda_r}{(\lambda_r + c)^2} \quad (5)$$

$\hat{\theta}_r(c)$: X_r 의 능형회귀 계수 추정치
 λ_r : 상관행렬에서 X_r 의 고유값(분산)

능형 회귀는 추정량의 분산을 작게 하여 자료의 작은 변화에 거의 영향을 받지 않게 함으로써 회귀계수의 신뢰성을 높일 수 있지만, 편향상수를 추정함에 있어 주관적이고 어느 정

도의 편향을 인정한다는 한계가 있다.

6) 로버스트 회귀에 의한 CER 개발

로버스트 회귀는 OLS에서 이상치의 잔차에 작은 가중치를 부여하여 영향력을 감소시키는 방법으로서 CER을 강건하게 한다. 이 방법은 M 추정, LTS 추정, S 추정, MM 추정이 있지만, 본 연구에서는 분석대상의 수가 많지 않은 경우에 사용할 수 있는 로버스트 M(Maximum Likelihood Type) 추정법을 사용하였다. 로버스트 M 추정은 간단하고 효과적인 방법으로서 식 (6)과 같이 손실함수를 적용하여 잔차가 큰 값에 대해서는 가중치를 작게 부여하고, 잔차가 작은 값에 대해서는 가중치를 크게 부여하여 회귀계수를 추정한다.

$$\hat{\beta} = \arg \min \sum_{q=1}^v \rho\left(\frac{\epsilon_q}{s}\right) \quad (6)$$

$$= \arg \min \sum_{q=1}^v \rho\left(\frac{y_q - X_q \beta}{s}\right)$$

$\arg \min f(x) : f$ 값을 최소화하는 x
 ρ : 손실 함수(loss function)
 s : ϵ 의 표준편차

로버스트 회귀는 이상치가 존재할 때 이상치를 제거하지 않고, 이상치의 영향력을 줄여서 CER을 개발할 때 사용할 수 있다. 이상치는 가능하면 제거하는 것이 CER의 안정성을 위해 타당하다. 그러나 이상치를 제거할 만큼 표본의 수가 충분하지 않은 경우에는 이상치 제거에 따른 정보의 손실을 유발할 수 있으며, 선택할 수 있는 비용인자의 수가 줄어들기 때문에 고려할 수 있는 비용인자 수의 부족을 초래하여 안정적인 CER 개발을 제한할 수 있다. 따라서 로버스트 회귀는 표본의 수가 작으면서 이상치가 존재할 때 이상치의 영향력을 감소시킬 수 있는 효과적인 방법이다.

7) 통계적 평가

개별 CER의 통계적 평가는 R^2 검정, t검정, F검정을 이용한다 (Rencher et al., 2008).

R^2 는 회귀분석 모형의 적절성을 평가하는 기준으로 절대평가 기준이 없기 때문에 무기체계 비용분석 전문가들의 의견 수렴과 일반적인 기준을 적용하여 0.8 이상일 때 모형이 적절하다고 판단하였다.

t검정은 개별 회귀계수들의 타당성을 검정하는 기준으로 유의 수준 5% 내의 확률값을 기준으로 타당성을 판단한다.

F검정은 전체 회귀계수의 적합도(goodness of fit)를 검정하는 기준으로 유의 수준 5% 내의 확률값을 기준으로 검정한다.

(5) CER 선형결합 모델 개발

본 연구에서 제시한 CER 선형결합 모델 개발은 개별 CER들의 선형결합과 결합된 모델의 평가를 통해 최종 CER을 선정하는 절차로 진행된다.

첫째, 개별 CER들에 대해서 식 (7)과 같이 정확도에 따라 SSE

기준, 수정 결정계수(R_{adj}^2) 기준, MMRE 기준, 회귀 분석 기준의 4가지 가중치 부여방법을 통해 선형결합한다.

$$C_{Total} = W_1 Model_1 + \dots + W_N Model_N \quad (7)$$

N : 선정된 모델

Table 2. CER 결합 모델 개발을 위한 가중치

구분	모델별 가중치
SSE 기준	$W_k = \frac{1/SSE_k}{\sum_{k=1}^N 1/SSE_k}$ (단, $SSE = \sum (y_i - \hat{y}_i)^2$)
MMRE 기준	$W_k = \frac{1/MMRE_k}{\sum_{k=1}^N 1/MMRE_k}$ (단, $MMRE = \frac{1}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{y_i}$)
R_{adj}^2 기준	$W_k = \frac{R_{adj\ k}^2}{\sum_{k=1}^N R_{adj\ k}^2}$ (단, $R_{adj}^2 = 1 - \frac{SSE/(n - (k+1))}{SST/(n-1)}$)

SSE, MMRE, R_{adj}^2 를 기준으로 한 결합모델은 <Table 2>를 통해 산출된 가중치를 개별 CER에 부여하여 선형결합한다.

회귀 분석을 통한 CER 결합 모델 개발은 개별 모델의 추정비용 \hat{y}_i 를 독립변수로 하고 실제비용 y 를 종속변수로 하여 다중 선형 회귀를 실시하여 모델별 가중치를 추정하고 이를 반영하여 식 (8)과 같이 CER 선형결합 모델을 개발한다.

$$C_{Total} = \beta_0 + \beta_1 Model_1 + \dots + \beta_N Model_N \quad (8)$$

β_N : 선형 회귀에 의한 추정 계수

둘째, 개발된 CER 선형결합 모델 중 평균 자승오차의 제공근(RMSE : Root Mean Squared Error) 및 R^2 를 기준으로 가장 작은 값을 가지는 모델을 최종 CER로 확정한다.

(6) CER 선형결합 모델 타당성 평가

모델의 타당성은 MMRE와 PRED로 평가하고, 평가 결과 타당한 경우에는 CER 개발을 완료하며, 부적합한 경우에는 자료 수집 및 가공 단계부터 다시 개발한다.

MMRE에 의한 평가 기준 25%를 기준으로 하였으며, 식 (9)와 같다(Conte *et al.*, 1986).

$$MMRE \leq 0.25 \quad (9)$$

PRED는 식 (10)과 같으며, $PRED(0.3) \geq 0.3$ 일 때 모델이 타당하다고 판단한다(Boehm *et al.*, 2000).

$$PRED(l) = k/n \quad (10)$$

n : 전체 자료 수

k : $MRE \leq l$ 에 해당되는 자료 수

MRE : $|y_i - \hat{y}_i|/y_i$

4. 포병 무기체계 연구개발비 CER 개발

본 장에서는 앞에서 제안한 개발 방법을 국내 포병 무기체계 연구개발비 CER 개발에 적용한 사례 연구를 제시하였다. 통계 분석 도구는 SAS 9.1, Excel 2007를 사용하였다.

4.1 자료 수집 및 가공

1단계는 자료 수집으로서 국방과학연구소 및 방산업체를 통해 1978년부터 2006년까지 개발된 포병 무기체계 9종에 대한 자료를 획득하였다. 그러나 연구개발과 관련된 자료체계가 미흡하여 연구개발비와 일부의 물리적 형상 자료만 수집할 수 있었다.

2단계는 자료 평가 및 표준화로서 표준화 영역 6개 중 비용 자료 영역에 있어서 연구개발비는 해당 연도에 지출된 경상이기 때문에 이를 한국제조업 생산자 물가지수를 적용하여

Table 3. 포병무기체계 자료 표준화

무기 체계	최대사거리 (km)	구경 (mm)	중량 (kg)	전장 (cm)	최대발사 속도(분)	지속발사 속도(분)	연구개발비 (억 원, 2010년)
1	3.59	60	18	99	30	20	18.2027
2	1.8	60	21	82	30	18	12.7289
3	6.473	81	41	155	30	11	35.2546
4	4.737	81	81	130	12	5	17.8506
5	11.274	105	2,260	231	3	1	37.6372
6	14.7	105	2,650	392	5	2	27.0690
7	18	155	6,890	701	4	2	43.0712
8	18	155	25,000	912	4	1	74.0739
9	41	155	47,000	810	6	2	1,342.847

<Table 3>과 같이 2011년 불변가(억 원)로 환산하였다. 비용을 불변가로 환산하는 것은 동일한 화폐가치를 적용하고, 필요한 경우 경상가로 쉽게 전환할 수 있으며, 가치 전환을 위한 할인 과정을 모델에 표현할 때 수식이 불필요하게 복잡해지는 것을 방지할 수 있다.

3단계는 비용주요인 식별로서 비용분야 전문가와의 인터뷰 및 비용과 비용인자간의 산포도를 통해 최대발사속도 및 지속발사속도는 비용과의 상관관계가 적다고 밝혀짐에 따라 최대사거리, 구경, 중량, 전장을 비용주요인으로 판단하였다.

4단계는 자료 개수 확인으로서 4개의 비용인자 대비 관측치가 9개로서 회귀분석을 수행하기에는 최소조건을 만족하였다.

4.2 자료 분석

1단계는 다중공선성 판단으로서 <Table 4>와 같이 VIF 및 CI 산출 결과를 분석한 결과 모든 비용인자들의 VIF 값이 10을 초과하고, 특히 4번째 비용인자인 전장의 CI가 30을 초과하여 다중공선성이 존재한다고 판단하였다.

Table 4. 비용인자별 VIF 및 CI

구 분	최대사거리	구경	중량	전 장
VIF	15.76	30.2	14.54	29.74
CI	2.75	6.7	10.87	42.8

2단계는 다중공선성 판단에 따른 조치로서 <Table 5>와 같이 변수 선택에 의해 다중공선성 제거를 수행하였다. 복수 선정된 비용인자 조합에 대해 판단한 결과 다중공선성의 심각성이 해소되어서 이상치를 판단하였다.

Table 5. 비용인자 선택

R ²	Adj. R ²	Forward	Stepwise	C(p)	Backward
최대사거리, 중량, 전장					구경

※ 비용인자 조합 선택 결과 : 최대사거리, 중량, 전장.

3단계는 이상치 판단으로서 <Table 6>에서와 같이 9번 무기체계가 표준화 제외잔차 검정에서 1% 유의수준을 초과하여 이상치로 의심되며, 영향력 조사에서는 Cook's Distance 및 DFBETAS 값이 기준치 보다 높아 영향력이 크다고 판단하였다.

4단계는 이상치 판단에 따른 조치로서 로버스트 회귀를 적용하기 위해 이상치를 제거하지 않고 이분산성을 판단하였다. 한편, 나머지 회귀분석 방법을 적용하기 위해 <Table 7>과 같이 9번 무기체계를 제거한 후 8개의 무기체계에 대해서 비용인자를 재선택하고 이분산성을 판단하였다.

Table 7. 비용인자 재선택

R ²	Adj. R ²	Forward	Backward	Stepwise	C(p)
구경, 중량		중량	최대사거리, 전장, 구경	중량	구경, 중량

※ 비용인자 조합 선택 결과 : 구경, 중량 & 중량.

5단계는 이분산성 판단으로서 이상치를 제거하지 않은 9개와 이상치를 제거한 8개 무기체계에 대한 잔차 산포도가 불규칙적으로 분포되어 있어서 이분산성이 존재하지 않았다. 따라서 개별 CER 개발 단계를 수행하였다.

4.3 개별 CER 개발

자료분석 결과를 기초로 <Table 1>에서 제시한 절차를 적용하여 개별 CER을 개발한 결과는 <Table 8>과 같다. 변수 조합별 자료 특성에 적합한 7개의 회귀모형을 적용한 결과 주성분 회귀와 선형 회귀(II)를 제외한 나머지 CER은 모델의 적절성 또는 회귀 계수의 신뢰성이 보장되지 않아 개별 CER로서 부적절한 것으로 나타났다. 따라서 선형 결합 대상인 2개 CER 개발 결과를 간략하게 기술하면 다음과 같다.

첫째, 주성분 회귀를 통해 개발된 CER은 4개의 비용인자와 이상치를 제거한 8개의 무기체계를 대상으로 개발하였다. CER은 식 (11)과 같으며, R² 값이 0.82로 모델의 설명력은 적절한 것으로 나타났으며, F 및 t 검정 결과 5% 내에서 회귀계수가 유의하였다.

Table 6. 이상치 및 영향력 검정

무기 체계	표준화 잔차	표준화 제외잔차	Hat Diag	Cook Distance	DFBETAS				
					상수	최대 사거리	구경	중량	전장
1	0.15	0.13	0.41	0.00	0.08	0.03	-0.07	-0.02	0.05
2	0.68	0.62	0.36	0.05	0.24	-0.09	-0.12	0.09	0.07
3	0.15	0.13	0.18	0.00	0.00	-0.01	0.01	0.01	-0.02
4	0.35	0.31	0.34	0.01	-0.09	-0.13	0.14	0.13	-0.15
5	-1.52	-2.04	0.53	0.52	1.54	0.65	-1.68	-0.72	1.74
6	-1.48	-1.91	0.69	0.98	-1.76	-2.44	2.00	2.59	-1.95
7	1.89	5.02	0.62	1.18	-1.86	1.02	1.04	-2.60	0.34
8	-1.94	-7.06	0.89	6.37	-0.44	10.48	-0.53	-6.62	-4.22
9	1.99	19.08	0.98	45.17	2.73	30.47	-3.84	24.81	-16.40

Table 8. 개별 CER 개발 및 통계적 평가

구 분	변수 개수	이상치	$R^2(R_{adj}^2)$	회귀계수 검증값
주성분 회귀	최대사거리(A), 구경(B), 중량(C), 전장(D)	제거	0.82(0.80)	(모형) $Pr>F : 0.002$ (계수) $Pr> t : intercept = 1, Prin1(\text{주성분}) = 0.002$
능형 회귀			0.93(0.83)	(모형) $Pr>F : 0.047$ (계수) $Pr> t : intercept = 0.8337, A = 0.636, B = 0.372, C = 0.099, D = 0.311$
로버스트 회귀	최대사거리(A), 중량(C), 전장(D)	미제거	0.90(0.84)	(계수) $Pr>ChiSq : intercept = 0.0056, A = 0.204, C = 0.032, D = 0.458$
선형 회귀 (I)	구경(B), 중량(C)	제거	0.89(0.85)	(모형) $Pr>F : 0.004$ (계수) $Pr> t : intercept = 0.509, B = 0.187, C = 0.043$
로그선형 회귀 (I)			0.78(0.69)	(모형) $Pr>F : 0.023$ (계수) $Pr> t : intercept = 0.431, B = 0.218, C = 0.832$
선형 회귀 (II)	중량(C)	제거	0.84(0.81)	(모형) $Pr>F : 0.001$ (계수) $Pr> t : intercept = 0.0005, C = 0.0014$
로그선형 회귀 (II)			0.69(0.64)	(모형) $Pr>F : 0.011$ (계수) $Pr> t : intercept = 0.0002, C = 0.01$

$$Y_{PCR} = 5.747 + 0.714range + 0.127caliber + 0.001weight + 0.016length \quad (11)$$

SSE_{LD} : 개별 CER 중 가장 작은 SSE 값
 $SSE_{C.B.}$: 결합 모델의 SSE

선형 회귀(II)는 1개의 비용인자와 이상치를 제거한 8개의 무기체계를 대상으로 개발하였다. CER은 식 (12)와 같으며, R^2 값이 0.84로 적절한 모델로 평가되었으며, F 및 t검정을 결과 회귀 계수 신뢰성이 5% 내에서 유의하였다.

$$Y_{Lin II} = 23.52163 + 0.0021weight \quad (12)$$

MMRE 기준의 결합 모델 역시 <Table 2>를 이용하여 산출한 0.42964와 0.57036의 가중치를 이용하여 식 (15)와 같이 개발하였다.

$$C_{Total} = 0.42964Model_1 + 0.57036Model_2 \quad (15)$$

4.4 CER 선형결합 모델 개발

(1) 모델 개발

선정된 두 개의 CER을 이용하여 4가지 방법에 따라 선형결합 모델을 개발하였다. 이 때, 주성분 회귀에 의한 CER을 $Model_1$ 로, 선형 회귀(II)에 의한 CER을 $Model_2$ 로 하였다.

SSE 기준의 결합 모델은 <Table 2>를 이용하여 $Model_1$ 과 $Model_2$ 에 대하여 0.499906과 0.5000937의 가중치를 얻었다. 이를 결합 모델로 표현하면 식 (13)과 같다.

$$C_{Total} = 0.499906Model_1 + 0.5000937Model_2 \quad (13)$$

식 (13)을 적용하여 $Model_1$, $Model_2$, 결합 모델에 대한 SSE 값을 비교한 결과 $Model_1$, $Model_2$, $Model_3$ 에서 각각 475.2, 437.4, 328.4로 계산되었다. 이는 식 (14)에 의해 SSE 기준의 결합 모델이 개별 CER 중 SSE 값이 가장 작은 $Model_2$ 에 비해 정확도를 24.91% 향상시켰음을 알 수 있다.

$$The\ Elevation\ Rate = \frac{SSE_{LD}}{SSE_{C.B.}} \times 100 \quad (14)$$

식 (15)를 적용하여 결합 모델에 대한 SSE 값을 산출한 결과 328.4로서 $Model_2$ 에 비해 24.94%의 정확도가 향상되었다.

R_{adj}^2 기준의 결합 모델 또한 <Table 2>를 이용하여 각각 0.495와 0.505의 가중치를 이용하여 식 (16)과 같이 개발했다.

$$C_{Total} = 0.495Model_1 + 0.505Model_2 \quad (16)$$

결합 모델에 대한 SSE가 328.3로서 $Model_2$ 에 비해 24.95%의 정확도가 향상되었다.

회귀분석을 이용한 결합 모델은 $Model_1$ 에 의해 추정된 값과 $Model_2$ 에 의해 추정된 값을 독립변수로 두고 실제 연구개발비를 종속변수로 두어 선형 회귀 분석을 실시함으로써 $Model_1$ 과 $Model_2$ 의 계수를 추정하였으며, 그 결과값인 0.469와 0.543을 $Model_1$ 과 $Model_2$ 의 가중치로 하여 식 (17)과 같이 개발하였다.

$$C_{Total} = 0.469Model_1 + 0.543Model_2 \quad (17)$$

결합 모델의 SSE가 326.2로서 $Model_2$ 에 비해 25.42%의 정확도가 향상되었다.

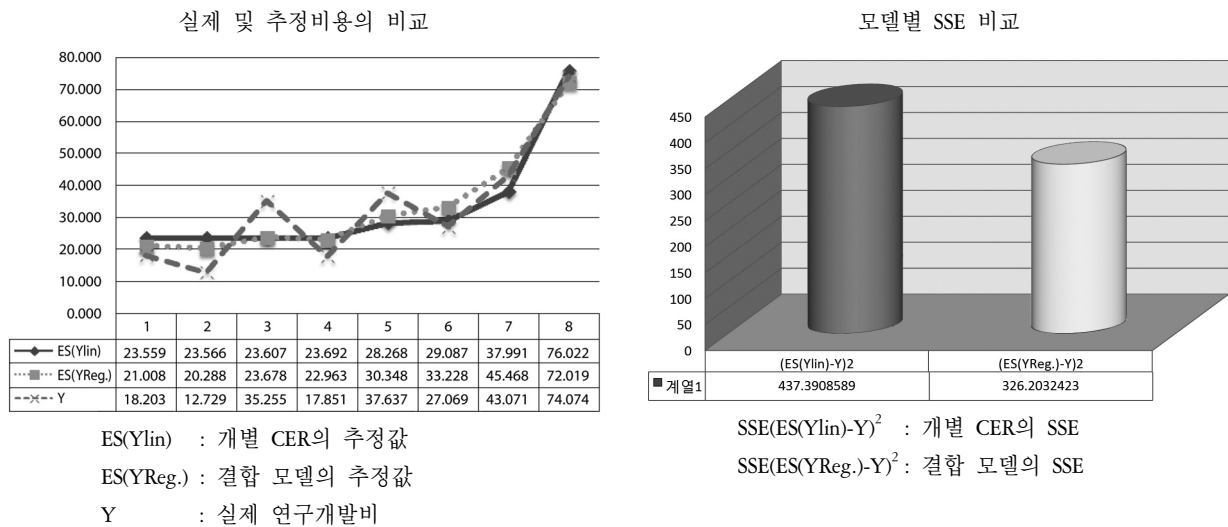


Figure 2. 선형결합 모델의 추정 오차 비교

(2) 모델 선정

모델선정을 위해 4가지의 CER 선형결합 모델과 2가지의 개별 CER에 대해 RMSE 및 R²를 비교한 결과는 <Table 9>와 같다.

Table 9. 모델간 RMSE 및 R² 비교

구분	CER 선형결합 모델				개별 CER	
	SSE	MMRE	R ² _{adj}	Regression	PCR	Lin
RMSE	7.399	7.397	7.397	7.373	12.582	8.538
R ²	0.879			0.880	0.825	0.839

비교 결과 RMSE가 최소이면서 R²가 최대인 “회귀 분석을 통한 결합 모델”을 최종 CER 선형결합 모델로 선정하였으며, CER은 식 (18)과 같다.

$$C_{Total} = 15.46 + 0.335range + 0.0598scaliber + 0.0014weight + 0.0074length \quad (18)$$

(3) 모델 타당성 검증

최종 선정된 모델에 대한 타당성 검증은 MMRE와 PRED(0.3)를 이용하였다.

첫째, MMRE를 계산하기 위해 MRE를 계산한 결과는 <Table 10>과 같으며, MMRE는 <Table 2>에 의해 0.23로서 MMRE ≤ 0.25를 만족하여 신뢰성 있는 모델이라고 판단하였다.

Table 10. 무기체계별 MRE

무기체계	1	2	3	4
MRE	0.154	0.594	0.328	0.286
무기체계	5	6	7	8
MRE	0.194	0.228	0.056	0.028

둘째, <Table 10>에 의해 k = 6, n = 8이므로 식 (10)을 이용

하여 PRED(0.3)을 계산하면 PRED(0.3) = 0.75이다. 따라서 일반 상용 전산모델의 PRED(0.3)가 약 0.3임을 고려할 때 개발된 CER 선형결합 모델의 신뢰성이 높다고 판단할 수 있다.

(4) 사례 적용 결과 분석

1) 실제비용과 추정비용 비교

실제 및 추정비용에 대하여 무기체계별로 비교한 결과 <Figure 2>와 같이 6, 8번 무기체계를 제외하고 CER 선형결합 모델에 의해 추정된 비용이 개별 CER에 의해 추정된 비용보다 실제비용에 더 근접함을 확인할 수 있다.

2) 모델별 SSE 비교

모델별 SSE를 <Figure 2>와 같이 분석한 결과 CER 선형결합 모델에 의해 계산된 SSE가 개별 CER에 의해 계산된 SSE 보다 더 적음을 알 수 있다. 이는 본 연구에서 제시한 결합 모델이 개별 모델 보다 정확도가 더 높다는 것을 증명해 준다.

5. 결론 및 향후 연구방향

본 연구는 무기체계 비용추정 분야에서 CER 선형결합 모델 개발 절차를 최초로 제시한 연구로서 다음과 같은 의의를 가진다.

첫째, 연구에서 제시한 개별 CER 개발 절차는 다양한 회귀분석 방법을 소수의 실적자료 특성에 따라 누락 없이 고려할 수 있는 방법을 체계적으로 종합한 것으로서 향후 국방 무기체계의 CER 개발을 위한 일반적인 방법론으로 활용할 수 있다.

둘째, CER 선형결합 모델 개발을 통해 무기체계 CER 개발에 있어서 실적 자료의 수가 부족한 국내의 환경적인 한계를 극복하면서 안정성과 예측력을 향상시켰다. 즉, 다수의 개별 CER을 선형결합 함으로서 개별 CER에서 발생할 수 있는 비용에 중요하게 영향을 미칠 수 있는 비용인자의 누락 가능성을 해소

하고, 개별 모델보다 오차를 향상시킴으로서 CER의 정확도를 높였다. 본 연구에서 제시한 25.42% 향상된 정확도는 무기체계의 획득 비용이 수백~수천억 원임을 고려했을 때 비용적으로 상당한 추정 편차 줄일 수 있다.

본 연구는 국내 실적자료 기반의 파라메트릭 기반 비용추정 모델 개발이 초기 단계임을 고려할 때 이론적, 실무적으로 완성도가 높은 방법론을 제시했다는데 의의가 있지만, 다음과 같은 연구들이 추가되어야 할 것이다.

첫째, 자료의 수가 부족할 때 사용할 수 있는 변수 생성 또는 Bootstrapping 방법 등을 적용하여 회귀모형의 안정성을 보장할 수 있는 연구가 필요하다.

둘째, 선형결합 모델 개발 방법론에 대한 적용이 아직 초보 단계임을 볼 때 선형결합 모델 개발 방법론의 핵심 부분인 가중치 부여 방법에 대해 추가 연구가 필요하다.

참고문헌

Armstrong, J. S. (2001), *Combining Forecasts : Principles of Forecasting, A Handbook for researchers and Practitioners*, Kluwer Academic Publishers, 417-439.
 Armstrong, J. S. (1989), *Combining Forecasts : The End of the Beginning or the Beginning of the End*, *International Journal of Forecasting*.
 Bates, J. M. and Granger, C. W. J. (1965), *The Combination of Forecasts*, *Operational Research Quarterly*, 20, 451-68.
 Boehm, B., Abts, C., Winsor, B. A., Chulani, S., Clark, B., Horowitz, E., Madachy, R., Reifer, D., and Steece, B. (2000), *Software Cost Estimation With COCOMO II*, Prentice Hall.

Chatterjee, S., Hadi, A. S. and Price B. (2000), *Regression analysis by example*, John Wiley and Sons.
 Conte, S. D., Dunsmore, H. E., and Shen, V. Y. (1986), *Software Engineering Metrics and Models*, Benjamin/Cummings, Menlo Park CA.
 Eo, W. J., Lee, Y. B., and Kang, S. J. (2010), *Developing an R&D Using historical Defense Weapon System Data in Korea*, *Journal of Society of Korea industrial and Systems Engineering*, 33(3), 55-62.
 Hoerl, A. E. and Kennard, R. W. (1970), *Ridge Regression: Biased Estimation for Nonorthogonal Problems*, *Technometrics*, 12, 69-82.
 ISPA(International Society of Parametric Analysts) (2007), *Parametric Estimating Handbook*, ISPA Asian Chapter.
 Kang, S. J. (2010), *A conceptual study of the development of Korean version cost estimation model*, Korea National Defense university.
 Kang, S. J. (2010), *Cost Estimation*, Dunam.
 Kim, J. D. (2002), *Linear regression analysis using SAS*, Jayou academy.
 Lee, J. Y., Shin, C., Kim, Y. K., Han, T. I., Park, S. K., and Moon, J. L. (2008), *Estimation of the RDT&E Cost of a Weapon System Using a Prior Elicitation Method(2nd step)*, Agency for Defense Development.
 Lee, J. Y., Lee, S. Y., Kim, Y. K., Kim, D. I., Park, S. K. Jung, B. R., Kim, H. R., and Lee, T. H. (2006), *Estimation of the RDT&E Cost of a Weapon System Using a Prior Elicitation Method*, Agency for Defense Development.
 Lund, R. E. (1975), *Tables for an approximate test for outliers in linear regression*, *Technometrics*, 17.
 Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001), *Introduction to linear regression analysis*, John Wiley and Sons.
 Rencher, A. C. and Schaalje, G. B. (2008), *Linear Models in Statistics*, Wiley-Interscience.
 Sung, N. K. (2009), *SAS/STAT linear regression analysis*, Jayou academy, 375.
 The National Typoon Center (2009), *An application of the system for typhoon analysis and prediction*, Technical note of the national typhoon center.
 Youm, J. K. (2005), *Linear regression analysis*, Jayou academy.



정원일
 육군사관학교 공학사
 한국교육개발원 군수관리학 학사
 국방대학교 운영분석과 석사
 현재 : 육군 대위
 관심분야 : 자원관리, 군수전사분석, 비용분석, 체계분석, M&S



이용복
 육군사관학교 공학사
 한양대학교 산업공학 석사
 국방대학교 운영분석과 박사
 현재 : 육군 소령
 관심분야 : 비용분석, 비용대 효과분석, 체계분석, 국방 M&S, 위계임



김동규
 육군사관학교 공학사
 모스크바 국립대학교 화학공학 석사
 현재 : 국방대학교 운영분석과 박사과정
 관심분야 : 비용분석, 비용대 효과분석, 사업관리



강성진
 육군사관학교 공학사
 미 해대원 운영분석 석사
 텍사스 A&M 산업공학 박사
 현재 : 국방대학교 운영분석과 교수
 관심분야 : 체계분석, 비용대 효과분석, 자원할당, EVMS, CAIV 등