

A complementary study on analysis of simulation results using statistical models

Ji-Hyun Kim^{1,a}, Bongseong Kim^a

^aDepartment of Statistics and Actuarial Science, Soongsil University

Abstract

Simulation studies are often conducted when it is difficult to compare the performance of nonparametric estimators theoretically. Kim and Kim (2021) showed that more systematic and accurate comparisons can be made if you analyze the simulation results using a regression model. This study is a complementary study on Kim and Kim (2021). In the variance-covariance matrix for the error term of the regression model, only heteroscedasticity was considered and covariance was ignored in the previous study. When covariance is considered together with the heteroscedasticity, the variance-covariance matrix becomes a block diagonal matrix. In this study, a method of estimating and using the block diagonal variance-covariance matrix for the analysis was presented. This allows you to find more pairs of estimators with significant performance differences while ensuring the nominal confidence level.

Keywords: variance-covariance matrix, block-diagonal matrix, heteroscedasticity, simultaneous confidence intervals

1. 서론

통계연구를 위해 모의실험을 실시해야 하는 경우가 많다. 비모수적 추정량의 성능을 이론적으로 알아보기 힘들 때가 그런 경우에 속한다. 다양한 실험조건에서 비교 대상이 되는 여러 추정량의 성능을 알아보기 위해서 모의실험을 실시했다고 하자. 추정량의 성능 우열이 실험조건에 따라 다르고 비교해야 할 추정량의 수가 많아지면 모의실험에서 얻은 결과로부터 객관적인 결론을 내리기가 어려울 때가 있다. Kim과 Kim (2021)은 다양한 실험조건에서 얻어진 추정량의 성능측도를 반응변수로 두고 실험조건과 추정량의 종류를 설명변수로 둔 회귀모형을 이용하면 보다 체계적이고 효과적인 분석을 할 수 있음을 보였다. 실험조건과 추정량의 종류에 따라 성능측도의 분산이 달라질 수 있으므로 회귀모형에서 이분산성(heteroscedasticity)을 고려해야 한다는 점을 강조하였다.

본 연구는 Kim과 Kim (2021)에 대한 후속연구이자 보완연구이다. 선행연구 논문의 심사과정에서 한 심사위원이 동일한 실험조건에서 얻어지는 여러 추정량의 성능은 독립적이지 않으므로 이분산성과 함께 공분산도 같이 고려해야 한다는 의견을 제시하였다. 심사위원의 의견에 추후 차근차근 따져보겠다는 답변을 했는데, 후속 연구를 통해 추정량 사이의 공분산을 무시했을 때와 고려했을 때 결과가 많이 다르며, 이분산성과 공분산을 같이 고려해 분석하면 실험조건에 따른 추정량의 성능 차이를 더 잘 찾아낸다는 결론에 이르게 되었다. 본 연구는 이런 결론에 이르게 된 과정을 서술하였다.

¹ Corresponding author: Department of Statistics and Actuarial Science, Soongsil University, 369 Sangdo-Ro, Dongjak-Gu, Seoul 06978, Korea. E-mail: jxk61@ssu.ac.kr

본 연구는 후속연구이므로 문제를 설명하는 부분 등에서 선행연구와 중복되는데 중복을 피하기 위해 새로이 추가된 내용을 이해하는 데 필수적인 부분만 간략히 다시 설명하였다. 2절에서 공분산을 같이 고려하는 분석 방법을 일반적으로 서술하였고, 3절에서 구체적으로 한 사례에 적용하였다. 공분산을 같이 고려하면 명목신뢰수준을 보장하면서도 길이가 짧은 동시신뢰구간(simultaneous confidence intervals)을 얻을 수 있다는 것을 보이기 위해 필요한 모의실험도 실시하였다. 4절에서 결과를 요약하였다.

2. 분석 방법

하나의 모수 θ 를 추정하는 다양한 방법이 존재하고, 어떤 추정방법이 더 나은지를 모의실험을 통해 알아보고자 하는 경우를 가정한다. 추정방법 또는 추정량의 성능은 모의실험 조건에 따라 달라질 수 있다. c 개의 다양한 모의실험 조건에서 m 개 추정량 $\hat{\theta}_{ij}$, $j = 1, \dots, m$ 의 성능을 비교하기 위해 r 번 독립적으로 난수를 생성해서 모의실험을 실시했다고 하자. 추정량을 얻기 위해 어떤 자료를 썼는지는 우리의 관심 대상이 아니다. 우리의 분석대상 자료는 모의실험에서 얻어진 자료로서 추정량에 대한 성능측도이다. 성능측도는 추정량의 분산이나 편향 또는 이 둘을 종합한 제곱근평균제곱오차(root mean square error; RMSE) 등이 될 수 있다.

i 번째 모의실험 조건 하에 k 번째 이루어진 반복실험에서 m 개 추정량에 대해 얻어진 성능측도 실현값을 $y_{i1k}, y_{i2k}, \dots, y_{imk}$ 라고 할 때 $\mathbf{y}_{ik} = (y_{i1k}, y_{i2k}, \dots, y_{imk})^T$ 는 m 개 추정량의 성능측도를 나타내는 확률벡터(random vector)가 된다. 차원이 $m \times 1$ 인 벡터 \mathbf{y}_{ik} 를

$$\mathbf{y}_{ik} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ik}$$

와 같이 나타낼 수 있다. 실험조건 i 에 따라 평균이 달라질 수 있지만 동일한 실험조건에서 이루어진 반복 실험에서는 동일한 평균을 갖는다. 모의실험을 할 때마다 난수를 독립적으로 생성하므로 \mathbf{y}_{ik} 는 i 또는 k 가 달라지면 독립이다. 실험 조건에 따라 분산공분산행렬(variance-covariance matrix)이 달라질 수 있는데, \mathbf{y}_{ik} 의 분산공분산행렬 또는 $\boldsymbol{\epsilon}_{ik}$ 의 분산공분산행렬을 $\boldsymbol{\Omega}_i$ 로 표시하기로 한다. 행렬 $\boldsymbol{\Omega}_i$ 의 차원은 $m \times m$ 이다.

$\mathbf{y} = (\mathbf{y}_{11}^T, \mathbf{y}_{12}^T, \dots, \mathbf{y}_{1r}^T, \mathbf{y}_{21}^T, \mathbf{y}_{22}^T, \dots, \mathbf{y}_{2r}^T, \dots, \mathbf{y}_{c1}^T, \mathbf{y}_{c2}^T, \dots, \mathbf{y}_{cr}^T)^T$ 로서 동일한 실험조건 i 에서 r 번 반복해 얻어진 벡터 $\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{ir}$ 를 이어 붙여서 얻어지는 벡터라고 하자. (자료 벡터 \mathbf{y} 의 배열순서가 Kim과 Kim (2021)과 다르다는 점에 주의해야 한다.) 벡터 \mathbf{y} 의 차원을 $n \times 1$ 이라고 할 때 $n = cmr$ 이 된다. n 개의 성능측도 실현값에 대한 모형을 행렬로 표현하면 다음과 같다.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (2.1)$$

$\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{ir}$ 에 대응하는 $\boldsymbol{\epsilon}_{i1}, \boldsymbol{\epsilon}_{i2}, \dots, \boldsymbol{\epsilon}_{ir}$ 는 동일한 분산공분산행렬 $\boldsymbol{\Omega}_i$ 를 가지므로 $\boldsymbol{\epsilon}$ 의 분산공분산행렬은 다음과 같이 블록대각행렬(block diagonal matrix)이 된다.

$$\boldsymbol{\Omega}_{n \times n} = \begin{pmatrix} I_r \otimes \boldsymbol{\Omega}_1 & \mathbf{0}_{mr \times mr} & \cdots & \cdots & \mathbf{0}_{mr \times mr} \\ \mathbf{0}_{mr \times mr} & I_r \otimes \boldsymbol{\Omega}_2 & \mathbf{0}_{mr \times mr} & \cdots & \mathbf{0}_{mr \times mr} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{mr \times mr} & \cdots & \mathbf{0}_{mr \times mr} & I_r \otimes \boldsymbol{\Omega}_{c-1} & \mathbf{0}_{mr \times mr} \\ \mathbf{0}_{mr \times mr} & \cdots & \cdots & \mathbf{0}_{mr \times mr} & I_r \otimes \boldsymbol{\Omega}_c \end{pmatrix}. \quad (2.2)$$

위 식에서 I_r 은 크기 r 인 항등행렬(identity matrix)을, $\mathbf{0}_{mr \times mr}$ 은 영행렬을 나타낸다. $I_r \otimes \boldsymbol{\Omega}_i$ 는 두 행렬의 크로네커

곱(Kronecker product)으로 다음과 같다.

$$I_r \otimes \Omega_i = \begin{pmatrix} \Omega_i & \mathbf{0}_{m \times m} & \cdots & \mathbf{0}_{m \times m} \\ \mathbf{0}_{m \times m} & \Omega_i & \cdots & \mathbf{0}_{m \times m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m \times m} & \cdots & \mathbf{0}_{m \times m} & \Omega_i \end{pmatrix}$$

$\epsilon_{ik}, k = 1, \dots, r$ 는 동일한 분산공분산행렬 Ω_i 를 가지므로 잔차벡터 $\mathbf{e}_{ik}, k = 1, \dots, r$ 로부터 Ω_i 를 추정할 수 있다. 전체 잔차벡터 \mathbf{e} 는 $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ 이다. \mathbf{e} 의 일부인 $\mathbf{e}_{ik}, k = 1, \dots, r$ 로부터 얻은 표본분산공분산행렬을 $\hat{\Omega}_i$ 라고 하자. 즉,

$$\hat{\Omega}_i = \frac{1}{r-1} \sum_{k=1}^r (\mathbf{e}_{ik} - \bar{\mathbf{e}}_i)(\mathbf{e}_{ik} - \bar{\mathbf{e}}_i)^T,$$

$\bar{\mathbf{e}}_i = (1/r) \sum_{k=1}^r \mathbf{e}_{ik}$ 라고 하자. 그러면 $\hat{\Omega}_i$ 는 식 (2.1)과 식 (2.2)가 참일 때 Ω_i 의 일치추정량(consistent estimator)이 된다. 따라서 반복실험횟수 r 을 늘리면 $\Omega_i, i = 1, \dots, c$ 와 Ω 에 대한 일치추정량을 얻을 수 있다. Ω_i 의 (h, l) 칸에 있는 모수를 σ_{hl} 라고 하면, 각 모수의 추정을 위한 관측값이 r 개 있는 셈이므로 $r = 50$ 도 작은 값이 아니다. r 의 크기에 따라 결과가 어떻게 달라지는지를 다음 절에서 살펴보기로 한다.

Kim과 Kim (2021)은 \mathbf{y} 를 구성하는 n 개의 관측값이 모두 독립이되 분산은 다를 수 있다고 가정했다. 이에 따라 Ω 가 대각행렬이 된다. White (1980)는 Ω 를 직접 추정하지 않고 회귀계수 추정량 $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ 의 분산공분산행렬을 추정하는 방법을 제시하였는데, Kim과 Kim (2021)은 이 방법을 적용하였다. 만약 Ω 가 식 (2.2)와 같은 블록대각행렬이 되면 Ω 를 직접 추정하는 것이 가능해진다. 추정해야 할 모수 개수는 $c(m(m+1)/2)$ 개다. 예를 들어 $c = 8, m = 6, r = 50$ 일 때 $n = cmr = 2400$ 이 되고 $c(m(m+1)/2) = 168$ 이 된다. 본 연구에서는 Kim과 Kim (2021)에서 쓴 방법 대신 블록대각행렬인 Ω 를 직접 추정하는 것을 제안한다.

우리의 중요한 관심은 m 개 추정량 중에서 더 나은 성능을 갖는 추정량을 찾는 데에 있다. 자료를 잘 설명하는 식 (2.1)을 선택하는 과정을 통해 추정량의 성능 우위가 실험조건에 따라 어떻게 달라지는지를 아는 것도 중요하지만, 각 실험조건에서 어떤 추정량이 더 나은 성능을 갖는지를 알아보는 것도 중요하다. 이를 위해서 m 개 추정량 중에서 차례로 두 추정량을 선택해 두 추정량의 성능 차이의 기대값에 대한 신뢰구간을 구하되, 두 추정량을 선택하는 방법에 $m(m-1)/2$ 개가 있으므로 신뢰구간을 구할 때 모임별오류율(family-wise error rate)을 고려하는 동시신뢰구간을 구해야 한다 (Tukey, 1953). 수식으로 표현하면 Kim과 Kim (2021)에서처럼 $E[\mathbf{D}\hat{\mathbf{y}}_s] = \mathbf{D}\mathbf{X}_s\boldsymbol{\beta}$ 에 대한 동시신뢰구간을 구하면 된다. Kim과 Kim (2021)의 식 (2.5)에 나타난 $m(m-1)/2 \times m$ 행렬 \mathbf{D} 는 m 개 추정량 중에서 차례로 두 개를 선택해 그 차이를 나타내기 위한 행렬이다. 차원이 $m \times p$ 인 행렬 \mathbf{X}_s 는 차원이 $n \times p$ 인 설계행렬(design matrix) \mathbf{X} 의 일부인데, 특정한 실험조건에서 m 개 추정량의 성능을 나타내기 위해 필요한 행렬이다. 만약 실험조건이 i 라면 \mathbf{y}_{i1} 에 대응하는 \mathbf{X} 의 행들을 나타낸다. $\mathbf{y}_{ik}, k = 1, \dots, r$ 중 어떤 것이든 대응하는 \mathbf{X}_s 는 동일하다. 구체적 예를 다음 절에서 볼 수 있다.

$E[\mathbf{D}\hat{\mathbf{y}}_s]$ 를 구성하는 $m(m-1)/2$ 개의 평균성능차이에 대한 동시신뢰구간을 구하는 방법으로 Kim과 Kim (2021)에서와 같이 Bonferroni 방법을 적용하였다. 다만 $\mathbf{D}\hat{\mathbf{y}}_s$ 의 분산공분산행렬 $\mathbf{D}\mathbf{X}_s\Omega_{\beta}\mathbf{X}_s^T\mathbf{D}^T$ 을 Kim과 Kim (2021)과 달리 추정하였다. $\Omega_{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Omega \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$ 에서 오차항의 분산공분산행렬 Ω 이 식 (2.2)와 같은 블록대각행렬임을 가정하고 잔차로부터 직접 추정하였다. Bonferroni 방법은 보수적(conservative) 검정방법이긴 하지만 VanderWeele과 Mathur (2018)은 Bonferroni 방법을 권장하였다. 6개 추정량의 모든 짝에 대해 그 평균성능차이를 동시신뢰구간으로 비교하고자 할 때, 각 신뢰구간의 유의수준은 $\alpha = 0.05$ 에서 $\alpha/15 = 0.0033$ 로 1/15배 낮아지지만, 대응하는 z 값의 기각값은 1.96에서 2.94로 1.5배 정도만 커진다. 따라서 만약 비교하고자 하는 두 추정량의 차이에 대한 표준오차가 작으면 실제 유의하게 되는 차이는 그다지 크지 않을 수 있다. 다음 절에서 보일 예에서 한 가지 경우를 미리 보면, 15개의 신뢰구간을 구할 때 각 신뢰구간에서 $\alpha = 0.05$

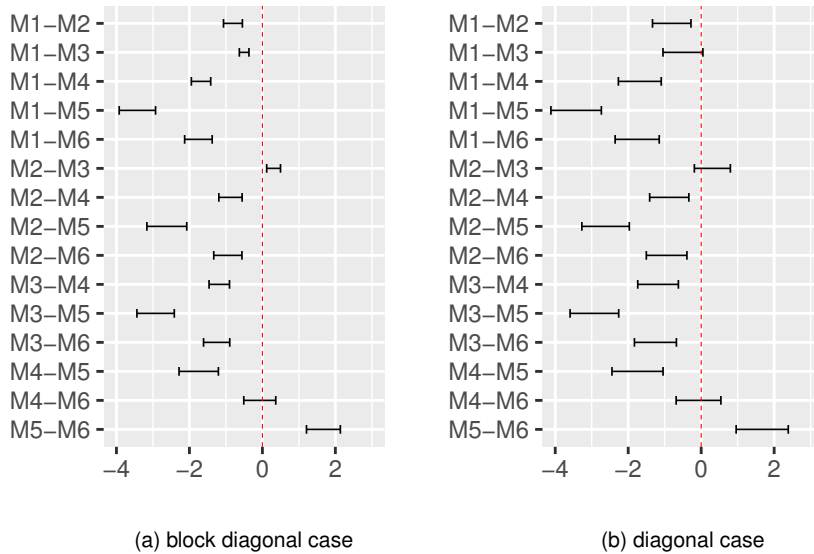


Figure 1: Simultaneous confidence intervals for the two types of variance-covariance matrix.

로 잘못 두면 유의한 차이를 보이는 신뢰구간의 수는 14개였다. 다중비교를 위해 $\alpha = 0.0033$ 으로 제대로 두어도 14개 모두 여전히 유의한 차이를 보였다. 유의한 차이를 보이는 14개 신뢰구간에 대응하는 z 값의 절대값은 4.764에서 20.179 사이에서 관측되었다. 이 자료에서는 두 추정량의 성능에 차이가 있는 경우 1.96에서 2.94 사이를 벗어나는 아주 큰 값을 갖게 되어 유의수준이 0.05에서 0.0033으로 낮아지더라도 결론이 달라지지 않게 되었다. 이 예는 Bonferroni 방법이 실용적인 관점에서 그렇게 보수적이지 않을 수 있음을 보여준다. VanderWeele와 Mathur (2018)도 비슷한 예를 들었는데, 유의수준 0.05에서 유의한 차이를 보이는 20개 중에서 17개가 Bonferroni 방법에 따른 유의수준에서도 여전히 유의한 차이를 보이는 예를 들었다.

3. 적용 사례

앞 절에서 일반적으로 기술한 방법을 구체적 사례에 적용해 보고자 한다. Kim과 Kim (2021)에 있는 것과 같은 사례에 적용해서 식 (2.1)에서 오차항의 분산공분산행렬이 대각행렬로부터 블록대각행렬로 바뀐에 따라 결과가 어떻게 달라지는지를 알아보고자 하였다.

사례에 대한 구체적 설명은 Kim과 Kim (2021)에 있으므로 여기서는 간략히 요약만 하기로 한다. 개별처리 효과(individual treatment effect)를 추정하는 비모수적 방법이 여러 개 있는데 이 중에서 유력한 6가지 방법에 의한 추정량의 성능을 비교하고자 모의실험을 실시하였다. 모의실험을 16개 서로 다른 실험조건에서 50번씩 반복 실시하였는데, 벡터 \mathbf{y}_{ik} 는 i 번째 조건에서 k 번째 반복된 실험으로부터 얻어진 6개 추정량의 성능측도를 나타낸다. 추정량의 성능측도는 제곱근평균제곱오차(RMSE)이다.

모의실험 결과에서 얻어지는 자료 \mathbf{y} 를 이용하여 $\mathbf{E}[\mathbf{y}]$ 를 제대로 설명하는 설계행렬 \mathbf{X} 의 열을 정해야 한다. 설계행렬의 열을 정하는 방법, 즉 변수선택의 방법과 선택된 모형이 갖는 의미는 Kim과 Kim (2021)을 참고하면 된다. 각 실험조건에서 어떤 추정량이 가장 나은 성능을 보이는지를 알아보려면 동시신뢰구간을 구하면 된다. 6개 추정량에서 선택할 수 있는 모든 가능한 쌍의 수는 15개인데, 두 추정량의 평균 성능 차이에 대한 15개의 동시신뢰구간을 구하였다. 오차항의 분산공분산행렬에서 이분산성만 고려하고 공분산을 무시했을 때

오차항의 분산공분산행렬은 대각행렬이 되고, 이분산성과 함께 공분산도 고려했을 때 블록대각행렬이 된다. 두 경우 동시신뢰구간이 어떻게 달라지는지를 그림으로 비교하였다. Figure 1은 16개의 실험조건 중에서 첫 번째 실험조건에 대한 결과이다. 두 그림을 비교해보면 공분산을 고려했을 때 신뢰구간의 길이가 훨씬 짧아진다는 것을 알 수 있다. 그 결과로 첫 번째와 세 번째 추정량의 성능 차이(M1-M3), 그리고 두 번째와 세 번째 추정량의 성능 차이(M2-M3)와 같이 공분산을 고려하지 않았을 때 유의한 차이를 보이지 않았던 것이 공분산을 고려하면 차이를 보이게 될 수 있다는 것도 알 수 있다.

공분산을 고려하면 신뢰구간의 길이가 짧아지는 이유를 단순화시켜 설명해 볼 수 있다. Y_1 과 Y_2 를 두 추정량의 성능을 나타낸다고 하자.(추정량의 성능은 난수생성을 하는 모의실험에서 반복적으로 얻어지므로 확률변수로 간주할 수 있다.) 두 추정량의 성능 차이에 대한 분산을 $V(Y_1 - Y_2)$ 라고 하면, $V(Y_1 - Y_2) = V(Y_1) + V(Y_2) - 2Cov(Y_1, Y_2)$ 이고 두 추정량은 동일한 모수를 추정하므로 보통 양의 공분산 값을 갖는다. 따라서 공분산을 고려하면 독립 가정을 했을 때보다 성능차이에 대한 분산이 더 작아지게 된다.

Figure 1은 하나의 실험조건에서 본 결과인데, 16개 실험조건에서 종합적으로 봤을 때 결과에 어떤 차이가 있는지를 살펴보았다. 각 실험조건에서 15개씩의 신뢰구간이 있으므로 총 $16 \times 15 = 240$ 개의 신뢰구간이 있는데, 240개 신뢰구간 길이의 평균과 표준오차, 그리고 유의한 차이를 보이는 개수를 계산하여 아래에 나타내었다. 공분산을 고려하면 공분산을 무시했을 때와 비교했을 때 신뢰구간의 길이가 평균 $0.644/1.332 \approx 0.48$ 배로 줄어들어 0을 포함하지 않는, 즉 유의한 차이를 보이는 신뢰구간의 수는 $200/163 \approx 1.23$ 배 많아진다.

	신뢰구간 길이의 평균	평균의 표준오차	유의한 신뢰구간의 수
공분산을 고려한 경우	0.644	0.016	200
공분산을 무시한 경우	1.332	0.013	163

위 결과는 공분산을 무시하지 않고 고려하면 0을 포함하지 않는 신뢰구간의 수가 더 많아져서 두 추정량의 성능에 차이가 있다고 결론 내리는 경우가 더 많아진다는 것을 의미한다. 반복실험횟수가 많아지면 실험조건 i 에서 얻어지는 잔차의 표본분산공분산행렬 $\hat{\Omega}_i$ 가 Ω_i 의 일치추정량이 되고, $\hat{\Omega}_i$ 를 대각원소로 갖는 블록대각행렬은 식 (2.2)의 일치추정량이 된다. 따라서 이에 기반한 방법의 명목신뢰수준도 보장된다. 만약 두 추정량의 성능 차이에 대한 신뢰구간이 참값을 포함하게 되는 확률이 명목신뢰수준을 보장한다면 두 추정량의 성능에 차이가 있다고 결론 내리는 경우가 더 많은 방법이 더 나은 방법이라고 할 수 있다. 이 사례에서 관측된 50 번이라는 반복횟수는 명목신뢰수준 0.95를 가정해도 될 만큼 충분히 큰가를 확인할 필요가 있다. 이를 위해 시뮬레이션 연구를 다음과 같이 진행하였다.

1. 주어진 자료에서 추정된 모형을 참모형 식 (2.1)로 간주한다.
 - 주어진 자료로부터 추정된 회귀계수 $\hat{\beta}$ 와 오차항의 분산공분산행렬 $\hat{\Omega}$ 를 각각 참회귀계수 β 와 식 (2.2)에 있는 참분산공분산행렬 Ω 로 간주한다. 설계행렬 \mathbf{X} 는 주어진 자료와 동일하게 고정한다.
2. 참모형으로부터 모의자료 \mathbf{y} 를 생성한다.
 - 각 실험조건 $i, i = 1, \dots, 16$ 에서 차원이 6×1 인 다변량정규난수벡터 $\mathbf{y}_{ik}, k = 1, \dots, 50$ 을 반복 생성해서 결합해 차원이 $n \times 1$ 인 다변량정규난수 \mathbf{y} 를 생성한다. $n = 16 \cdot 6 \cdot 50 = 4800$ 이다.
3. 생성된 자료에 모형 식 (2.1)을 적합한다.
 - 이때 설계행렬은 고정된 것으로 간주하므로 변수 선택을 하지 않는다. 회귀계수와 회귀계수의 분산만 추정한다.
 - 회귀계수의 분산을 추정하기 위해 적합된(fitted) 모형의 잔차를 이용하여 블록대각행렬 식 (2.2)을 추정한다.

4. 각 실험조건에서 $E[\mathbf{D}\hat{\mathbf{y}}_s] = (\mathbf{D}\mathbf{X}_s\boldsymbol{\beta})_{15 \times 1}$ 에 대한 95% 동시신뢰구간을 구해서 15개 신뢰구간의 평균길이와 제1종오류 발생 여부를 저장한다.
 - 15개의 신뢰구간 중에서 하나라도 참값을 포함하지 않으면 제1종오류가 생긴 것으로 간주한다.
 - 단계 4의 작업을 16개 실험조건에서 차례로 수행한다.
5. 앞 단계 중에서 2부터 4까지 단계를 2000번 반복해서 평균적인 성능을 구한다.
 - 16번의 실험조건에서 각각 15개 동시신뢰구간의 평균길이와 제1종오류 발생 여부가 2000개 얻어진다.
 - 2000개 평균길이의 평균과 실제유의수준을 각 실험조건에서 구하여 출력한다. 실제유의수준(의 추정값)은 2000번 중에서 제1종오류가 발생한 비율이 된다.

위에서 설명한 모의실험 단계 1과 3에서 $\hat{\boldsymbol{\beta}}$ 는 보통최소제곱(Ordinary Least Squares; OLS)추정량이다. OLS 추정량은 등분산 가정이 성립할 때, 즉 $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$ 일 때 최량선형비편향추정량(best linear unbiased estimator)이 되지만 본 연구에서와 같이 등분산 가정이 성립하지 않을 때는 최적추정량이 아닐 수 있다. 오차항의 분산 공분산행렬 $\boldsymbol{\Omega}$ 가 보다 일반적인 형태일 때 다음과 같이 정의되는 일반화최소제곱(Generalized Least Squares; GLS)추정량을 고려할 수 있다.

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = \underset{\mathbf{b}}{\operatorname{argmin}} (\mathbf{y} - \mathbf{X}\mathbf{b})^T \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\mathbf{b}).$$

OLS 추정량은 $\hat{\boldsymbol{\beta}}_{\text{OLS}} = \underset{\mathbf{b}}{\operatorname{argmin}} (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b})$ 와 같이 정의된다. $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$ 인 경우 두 추정량의 정의가 일치하므로, OLS 추정량을 GLS 추정량의 특수한 경우로 볼 수 있다. $\boldsymbol{\Omega}$ 를 알고 있다면 GLS 추정량 $\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y}$ 이 최량선형비편향추정량(best linear unbiased estimator)이 된다. 하지만 $\boldsymbol{\Omega}$ 를 자료로부터 추정해야 하는 경우 GLS 추정량의 분산이 OLS 추정량 $\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ 의 분산보다 더 커질 수도 있다 (Rao와 Griliches, 1969). 모의실험을 통해 GLS 추정량의 성능을 OLS 추정량과 비교해 보았다. OLS 추정량을 쓸 때도 추정량의 분산 추정을 달리 할 수 있으므로 여러 가지 추정방법이 가능하다. 총 4가지 추정방법의 성능을 비교하였다.

- **OLS1**으로 표기할 첫 번째 방법은 앞에서 서술한 사례연구에서 쓴 방법이다. $\boldsymbol{\beta}$ 의 점추정량으로 OLS 추정량을 쓰고 추정량의 분산을 $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Omega}} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$ 으로 추정한다. 이때 $\hat{\boldsymbol{\Omega}}$ 는 블록대각행렬로서 i 번째 블록에 해당하는 행렬 $\hat{\boldsymbol{\Omega}}_i$ 는 간차벡터 $\mathbf{e}_{ik}, k = 1, \dots, r$ 의 표본분산공분산행렬로부터 얻어진다.
- **OLS2**로 표기하는 두 번째 방법은 OLS 추정량을 쓰되 추정량의 분산에 대한 추정량으로 HC3 추정량을 쓰는 방법이다 (MacKinnon과 White, 1985). Kim과 Kim (2021)에서 쓴 방법인데 공분산을 무시하고 이분산성만 고려한 경우이다. 지난 연구 결과와 비교하기 위해 고려하였다.
- **OLS3**로 표기하는 세 번째 방법은 OLS 추정량을 쓰고 추정량의 분산을 $\hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}$ 로 추정하는 방법이다. 이때 $\hat{\sigma}^2$ 은 y_1, \dots, y_n 이 독립이고 등분산성을 가정했을 때 얻어지는 분산의 추정값이다. 적절하지 않은 방법이지만 공분산과 이분산성을 모두 무시할 때 성능이 얼마나 나빠지는지 알아보고자 고려하였다.
- **GLS**로 표기하는 네 번째 방법은 GLS 추정량을 이용하는 방법인데 추정량의 분산을 $(\mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1}$ 으로 추정한다. $\hat{\boldsymbol{\Omega}}$ 은 첫 번째 방법에서 구한 블록대각행렬을 나타낸다. (GLS 추정량을 얻기 위해서 $\hat{\boldsymbol{\Omega}}$ 가 필요하고 $\hat{\boldsymbol{\Omega}}$ 를 얻기 위해서 다시 GLS 추정량이 필요하다. 이 문제를 해결하는 방법으로 OLS 추정량에 기반한 간차를 이용해 $\hat{\boldsymbol{\Omega}}$ 를 먼저 얻는다. 이 분산공분산행렬을 이용해 GLS 추정량을 구하고 GLS 추정량에 기반한 간차를 다시 구해 분산공분산행렬을 재추정한다. 재추정된 분산공분산행렬에 기반한 GLS 추정량을 다시 구하고, 이 작업을 수렴된 GLS 추정량을 얻을 때까지 반복하는 방법을 생각해볼 수 있으나, 이 반복적인 방법은

Table 1: Simulation results for the comparison of four kinds of simultaneous confidence intervals

Number of replications	Experimental condition	Actual confidence level				Average length of confidence interval			
		OLS1	OLS2	OLS3	GLS	OLS1	OLS2	OLS3	GLS
$r = 50$	1	0.964	0.999	0.999	0.939	0.740	1.200	1.341	0.677
	2	0.968	1.000	0.999	0.943	0.755	1.446	1.341	0.690
	3	0.963	1.000	1.000	0.944	0.659	1.149	1.341	0.553
	4	0.967	0.999	1.000	0.944	0.701	1.126	1.341	0.635
	5	0.965	1.000	0.994	0.939	0.859	1.416	1.341	0.756
	6	0.960	1.000	0.990	0.943	0.883	1.722	1.341	0.737
	7	0.963	1.000	1.000	0.935	0.714	1.155	1.341	0.648
	8	0.963	1.000	0.999	0.935	0.769	1.263	1.341	0.686
$r = 20$	1	0.960	1.000	0.997	0.856	1.114	1.756	1.793	0.897
	2	0.964	1.000	0.998	0.863	1.066	1.841	1.793	0.866
	3	0.958	1.000	1.000	0.846	0.908	1.466	1.793	0.673
	4	0.959	1.000	1.000	0.850	0.838	1.455	1.793	0.652
	5	0.950	0.999	0.995	0.815	1.172	2.106	1.793	0.894
	6	0.947	1.000	0.990	0.828	1.207	2.180	1.793	0.908
	7	0.964	1.000	1.000	0.841	0.914	1.623	1.793	0.664
	8	0.969	1.000	1.000	0.851	0.960	1.617	1.793	0.706

시뮬레이션에서 구현하지 않았다. 그리고, GLS 추정량과 분산을 구할 때 $\hat{\Omega}$ 의 역행렬을 구해야 하는데, 블록대각행렬의 역행렬은 각 블록을 구성하는 행렬의 역행렬들로 이루어진다는 성질을 이용하였다.)

네 가지 방법에 대한 모의실험 결과를 Table 1에 정리하였다. Table 1에서 OLS1, OLS2, OLS3, GLS는 네 가지 방법을 각각 나타낸다. 16가지 실험조건에서 얻은 결과 중에서 처음 8가지 실험조건에 대한 결과만 보고하였다.(나머지 8개 실험조건에서 얻어진 결과도 비슷하였다.) 대신에 각 실험조건에서 반복 실험하는 횟수 r 을 50에서 20으로 줄였을 때 얻어진 결과를 Table 1에 같이 보고하였다.

OLS2와 OLS3는 실제신뢰수준이 명목신뢰수준인 0.95를 넘어 0.999 또는 1이라는 값을 가지므로 불필요하게 보수적임을 알 수 있다. 반면에 OLS1은 모든 실험조건에서 신뢰구간의 길이가 OLS2나 OLS3보다 짧음에도 불구하고 명목신뢰수준을 보장한다.(신뢰구간 길이의 평균에 대한 표준오차를 Table 1에 나타내지 않았으나 0.001에서 0.004 사이의 값으로서 방법별 평균길이의 차이는 모두 유의하다. 그리고 OLS3의 신뢰구간의 길이가 모든 실험조건에서 같은데, 그 이유는 $E[\mathbf{D}\hat{\mathbf{y}}_s]$ 에 대한 신뢰구간에서 $\hat{\mathbf{y}}_s$ 가 모든 s 에 대해 등분산성을 갖기 때문이다.) GLS는 OLS1보다 길이가 조금 짧지만 명목신뢰수준을 보장하지 못하는 경우가 있어 장점이라고 보기 어렵다. 특히 $r = 20$ 인 경우 실제신뢰수준이 너무 낮아 짧은 길이는 전혀 장점이 되지 못한다.

두 추정량의 성능 차이에 대한 신뢰구간의 길이가 짧아지면 0을 포함하지 않아 유의한 차이를 탐지하게 되는 신뢰구간의 비율이 높아진다. Table 1에 나타내지 않았는데, $r = 50$ 일 때 총 $16 \times 15 \times 2000 = 480000$ 개의 신뢰구간 중에서 0을 포함하지 않는 신뢰구간의 비율은 네 가지 방법순으로 0.840, 0.675, 0.653, 0.853이었다. 역시 OLS1이 명목신뢰수준을 보장하면서 OLS2나 OLS3보다 유의한 차이를 더 잘 찾아내는 것을 알 수 있다.

4. 결론

여러 추정량의 성능을 모의실험을 통해 비교하고자 할 때, 모의실험 결과에 회귀모형을 적합해서 분석하면 좋다는 제안을 Kim과 Kim (2021)에서 하였다. 하지만 그 연구에서 제안한 방법은 비교하고자 하는 추정량

성능측도들 사이에 존재하는 공분산을 고려하지 않고 이분산성만 고려했다는 한계가 있었다. 본 연구는 Kim 과 Kim (2021)에 대한 보완연구로서 공분산을 같이 고려하면 결과가 어떻게 달라지는지를 알아보았다.

일반적으로 공분산이 존재한다면 이를 고려해야 한다. 동일한 목적을 갖는 추정량들의 성능측도 사이에는 공분산이 존재하게 되고, 이 사실을 고려하여 비교하면 성능에 차이가 있는 추정량의 짝을 더 잘 찾아낼 수 있다는 것을 사례 연구를 통해 알 수 있었다. 사례에 대한 모의실험을 통해 명목신뢰수준을 보장하면서 차이를 더 찾아낸다는 사실도 확인할 수 있었다. 물론 이 모의실험 결과를 일반화하는 데엔 주의가 필요하지만, 공분산이 존재할 때 이를 고려해야 한다는 것은 당연하며, 다만 무시했을 때와 비교해서 결과가 달라지는 정도는 사례에 따라 달라질 것이다.

결론적으로, 모의실험 결과로 얻어진 자료에 대한 회귀모형에서 이분산성과 공분산을 모두 고려해야 하며, 오차항의 분산공분산행렬은 대각행렬이 아닌 블록대각행렬이 된다. 회귀계수의 점추정량으로 보통최소 제곱(OLS) 추정량을 쓰고, OLS 추정량으로 구한 잔차를 이용해 블록대각행렬인 분산공분산행렬을 추정하면 된다.

References

- Kim J and Kim B (2021). Analysis of simulation results using statistical models. *The Korean Journal of Applied Statistics*, **34**, 1–11.
- MacKinnon JG and White H (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, **29**, 305–325.
- Rao P and Griliches Z (1969). Small-sample properties of several two-stage regression methods in the context of auto-correlated errors. *Journal of the American Statistical Association*, **64**, 253–272.
- Tukey JW (1953). The problem of multiple comparisons. Unpublished report, Princeton University, Princeton, N.J.
- VanderWeele TJ and Mathur MB (2018). Some desirable properties of the Bonferroni correction. *American Journal of Epidemiology*, **188**, 617–618.
- White H (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

Received May 10, 2022; Revised June 15, 2022; Accepted June 25, 2022

통계모형을 이용하여 모의실험 결과 분석하기에 대한 보완연구

김지현^{1,a}, 김봉성^a

^a송실대학교 정보통계보험수리학과

요 약

비모수적 추정량의 성능을 이론적으로 비교하기 힘들 때 흔히 모의실험을 실시한다. 다양한 실험조건에서 여러 추정량에 대해 얻어진 모의실험 결과를 회귀모형을 이용해 분석하면 보다 체계적이고 정확한 비교를 할 수 있다는 것을 Kim과 Kim (2021)에서 보였다. 이 연구는 Kim과 Kim (2021)에 대한 후속연구이자 보완연구이다. 회귀모형의 오차항에 대한 분산공분산행렬에서 이분산성만 고려하고 공분산을 선행연구에서 무시했는데, 공분산을 고려하게 되면 분산공분산행렬은 블록대각행렬이 된다. 본 연구에서 블록대각행렬인 분산공분산행렬을 추정하여 분석에 이용하는 방법을 제시하였다. 이렇게 하면 명목신뢰수준을 보장하면서 유의하게 성능 차이가 나는 추정량 짝을 더 잘 찾을 수 있다는 것도 보였다.

주요용어: 분산공분산행렬, 블록대각행렬, 이분산성, 동시신뢰구간

¹교신저자: (06978) 서울시 동작구 상도로 369, 송실대학교 정보통계보험수리학과. E-mail: jxk61@ssu.ac.kr