응용통계연구 제 **14**권 1호, *2001*년, *pp. 121–135*

패널회귀모형에서 예측량의 효율에 관한 비교*

정병철[1)] 조민화[2)] 송석헌[3)]

요 약

본 논문에서는 이원오차성분을 가지는 패널회귀모형에서 미래시점에 대한 다양한 예측량들을 유도하고, 예측량들의 효율성을 모의실험을 통하여 비교하였다. 모의실험 결과, FGLS추정량을 이용한 예측량들은 참 GLS를 이용한 예측량과 효율성에서 서로 큰 차이를 보이지 않았다. 또한 계산상 매우 복잡한 ML과 REML을 이용한 예측량과도 거의 비슷한 효율성을 보여주었다.

주요용어: 패널회귀모형, BLUP.

1. 서론

패널회귀모형을 다루는 대부분의 많은 연구들은 회귀계수와 분산성분들에 대한 추정과 분산성분의 존재유무에 대한 가설검정에 집중되고 있다(Baltagi(1995), Hsiao(1986), Maddala(1993)). 반면 회귀모형 분석의 주요한 목적 중 하나인 미래시점의 예측(Prediction)을 패널회귀모형에서 다룬 연구는 아직까지 매우 드문 실정이다. 회귀모형에서 만일 오차항이 구형성(spherial)을 만족하는 경우에는 보통최소제곱추정량(Ordinary least squared estimator, OLSE)을 이용하여 미래시점을 쉽게 예측할 수 있다. 그러나 오차항이 서로 상관되어 있는 경우 OLS추정량을 이용한 미래시점의 예측은 오차항의 상관관계를 무시함으로써 효율성이 떨어지게 될 것이다. 이러한 문제를 해결하기 위하여 Goldberger(1962)는 일반선형회귀모형(Generalized linear regression model)에서 미래시점에 대한 최량선형불편예측량(Best linear unbiased predictor, BLUP)을 제안하였다. Goldberger(1962)의 연구를 바탕으로 Wansbeek와 Kepteyn(1978)과 Taub(1979)은 일원오차성분모형을 가지는 패널회귀모형에서 미래시점에 대한 BLUP를 유도하였다. 그러나, 이와 같이 유도된 미래시점의 BLUP는 분산성분들이 알려졌다는 가정하에서 일반화최소제곱(Generalized Least Squares, GLS) 추정량을 이용하기 때문에 단지 이론적인 예측량에 지나지 않는다. 본 연구에서는 이와 같은 문제를 해결하기 위해서 실제 적용가능한 다양한 예측량들을 유도하고 각 예측량들의 효율성을 비교 연구하려한다. 더불어 본 연구에서는 패널회귀모형에서 가장 많이 적용되고있는 이원오차성분모형에 대한 적절한 예측량들을 다루고자한다.

---

* 본 연구는 2000년도 이화여자대학교 BK 학술연구비 지원에 의하여 수행되었음.

1) (120-750) 서울시 서대문구 대현동, 이화여자대학교 통계학과, Post Doc. 연구원
2) (156-704) 서울시 동작구 대방동 339-1, 에스링크(주), 연구원
3) (136-701) 서울시 성북구 안암동 5-1, 고려대학교 통계학과, 조교수
E-mail: ssong@korea.ac.kr

본 논문의 구성은 다음과 같다. 먼저 2장에서는 이원오차성분모형을 가지는 패널회귀모형을 소개하고, 3장에서는 이러한 모형에서 BLUP를 유도하고, 미래시점에 대하여 분산성분들과 회귀계수의 추정방법에 따른 여러가지 가능한 예측량들을 제안한다. 4장에서는 모의실험을 통하여 각 예측량들의 예측평균제곱오차 (Prediction Mean Square Error, PMSE)를 이용하여 각 예측량들의 효율성을 비교하며, 5장에서는 실증적인 패널자료에 적용하여 각 예측량들의 효율성을 비교한다. 마지막으로 6장에서는 결론을 다루었다.

2. 모 형

다음과 같은 패널회귀모형을 고려해 보자.

$$y_{it} = x'_{it}\beta + u_{it}, \qquad i = 1, \cdots, N, \qquad t = i, \cdots, T \tag{2.1}$$

여기서 y_{it}는 i번째 개체(개인, 가구, 국가 등)의 시점 t에서의 관측치를 나타내는 반응값이고, x_{it}는 k개의 변수로 이루어진 설명변수벡터이다. 모형 (2.1)에서 오차항 u_{it}는 다음과 같은 이원오차성분모형 (Two-way error components model)을 갖는다고 가정한다.

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \tag{2.2}$$

μ_i는 관측될 수 없는 개체효과(individual specific effect)를 나타내고, λ_t는 시간효과(time specific effect)를 나타내는 확률변수이며, ν_{it}는 나머지 오차항을 나타낸다. μ_i, λ_t 와 ν_{it}는 서로 독립이며 각각 $\mu_i \sim IID(0, \sigma_\mu^2)$, $\lambda_t \sim IID(0, \sigma_\lambda^2)$ 와 $\nu_{it} \sim IID(0, \sigma_\nu^2)$ 이라고 가정하자. 모형 (2.1)을 행렬을 이용하여 나타내면 다음과 같다.

$$y = X\beta + u \tag{2.3}$$

y는 $(NT \times 1)$인 반응변수벡터, X는 $(NT \times k)$인 독립변수행렬이며, β는 $(k \times 1)$인 회귀계수벡터이다. 또한 식 (2.2)의 오차항을 행렬로 나타내면 다음과 같다.

$$u = (I_N \otimes i_T)\mu + (i_N \otimes I_T)\lambda + \nu \tag{2.4}$$

여기서 I_N과 I_T는 각각 차원이 N과 T인 단위행렬이고, i_N과 i_T는 각각 모든 원소가 1로 이루어진 크기가 N과 T인 벡터를 나타내며, $\mu = (\mu_1, \cdots, \mu_N)'$, $\lambda = (\lambda_1, \cdots, \lambda_T)'$, $\nu = (\nu_{11}, \cdots, \nu_{NT})'$이며 $\otimes$는 크로네커곱을 나타낸다. 이와 같은 가정으로부터 오차항의 분산-공분산 행렬은 다음과 같이 구해진다.

$$\Omega = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\lambda^2(J_N \otimes I_T) + \sigma_\nu^2(I_N \otimes I_T) \tag{2.5}$$

J_N과 J_T는 모든 원소가 1인 $N \times N$과 $T \times T$인 행렬이다.

식 (2.5)와 같은 분산-공분산 행렬을 갖는 회귀모형에서 예측량을 구하기 위해서는 Ω의 역행렬을 필요로 한다. Ω^{-1}를 구하기 위하여 식 (2.5)의 J_N를 $N\bar{J}_N$, 여기서 $\bar{J}_N = J_N/N$,

I_N를 $E_N + \bar{J}_N$, 여기서 $E_N = I_N - \bar{J}_N$, J_T를 $T\bar{J}_T$, I_T를 $E_T + \bar{J}_T$로 치환하고 같은 항끼리 정리하면 식 (2.5)의 분산-공분산 행렬은 다음과 같이 다시 주어진다.

$$\Omega = \lambda_1 Q_1 + \lambda_2 Q_2 + \lambda_3 Q_3 + \lambda_4 Q_4 = \sum_{i=1}^{4} \lambda_i Q_i \tag{2.6}$$

여기서 $\lambda_1 = \sigma_\nu^2$, $\lambda_2 = T\sigma_\mu^2 + \sigma_\nu^2$, $\lambda_3 = N\sigma_\lambda^2 + \sigma_\nu^2$, $\lambda_4 = T\sigma_\mu^2 + N\sigma_\lambda^2 + \sigma_\nu^2$, $Q_1 = E_N \otimes E_T$, $Q_2 = E_N \otimes \bar{J}_T$, $Q_3 = \bar{J}_N \otimes E_T$, $Q_4 = \bar{J}_N \otimes \bar{J}_T$이다. 또한 $Q_i(i = 1, \cdots, 4)$는 멱등행렬이며 서로 직교하는 행렬이므로 이러한 Q_i행렬의 성질을 이용하면 Ω^{-1}는 다음과 같이 구해진다.

$$\Omega^{-1} = \sum_{i=1}^{4} \frac{1}{\lambda_i} Q_i \tag{2.7}$$

3. 예 측 량

Goldberger(1962)의 결과를 패널회귀모형에 확장 적용하면 이원오차성분모형에서 i번째 개체의 시점 $(T+s)$에서의 최량선형불편예측량(BLUP)은 다음과 같이 표현된다.

$$\widetilde{y}_{i,T+s} = x'_{i,T+s}\widetilde{\beta}_{GLS} + w'\Omega^{-1}\widetilde{u}_{GLS} \tag{3.1}$$

여기서 $\widetilde{\beta}_{GLS}$는 회귀계수에 대한 GLS추정량을 나타내고, $w = E(u_{i,T+s}u)$이며 $\widetilde{u}_{GLS}$는 GLS 잔차를 나타낸다. 이 경우 본 연구에서 고려된 모형에서 (3.1)의 두 번째 식을 정리하면 다음과 같은 결과를 유도할 수 있다.

$$\begin{aligned}
w'\Omega^{-1}\widetilde{u}_{GLS} &= E(u_{i,T+s}u)'\Big(\sum_{i=1}^{4}\frac{1}{\lambda_i}Q_i\Big)\widetilde{u}_{GLS} \\
&= E\Big[(\mu_i + \lambda_{T+s} + \nu_{i,T+s})u\Big]'\Big(\sum_{i=1}^{4}\frac{1}{\lambda_i}Q_i\Big)\widetilde{u}_{GLS} \\
&= \sigma_\mu^2(l_i \otimes i_T)'\Big(\sum_{i=1}^{4}\frac{1}{\lambda_i}Q_i\Big)\widetilde{u}_{GLS} \\
&= \frac{\sigma_\mu^2}{\lambda_2}(l_i'E_N \otimes i_T')\widetilde{u}_{GLS} + \frac{\sigma_\mu^2}{\lambda_4}(l_i'\bar{J}_N \otimes i_T')\widetilde{u}_{GLS} \\
&= \frac{\sigma_\mu^2}{\lambda_2}\Big[(l_i \otimes i_T)' - \frac{1}{N}(i_N \otimes i_T)'\Big]\widetilde{u}_{GLS} + \frac{\sigma_\mu^2}{N\lambda_4}(i_N \otimes i_T)'\widetilde{u}_{GLS} \\
&= \frac{\sigma_\mu^2}{\lambda_2}\Big(\sum_{t=1}^{T}\widetilde{u}_{it,GLS} - \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{u}_{it,GLS}\Big) + \frac{\sigma_\mu^2}{N\lambda_4}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{u}_{it,GLS} \\
&= \frac{T\sigma_\mu^2}{\lambda_2}\bar{\widetilde{u}}_{i.,GLS} - \frac{T\sigma_\mu^2}{\lambda_2}\bar{\widetilde{u}}_{..,GLS} + \frac{T\sigma_\mu^2}{\lambda_4}\bar{\widetilde{u}}_{..,GLS} \\
&= T\sigma_\mu^2\Big(\frac{1}{\lambda_2}\bar{\widetilde{u}}_{i.,GLS} + (\frac{1}{\lambda_4} - \frac{1}{\lambda_2})\bar{\widetilde{u}}_{..,GLS}\Big)
\end{aligned} \tag{3.2}$$

여기서 세 번째 등식의 l_i는 I_N의 i번째 열을 나타내며, 세 번째 등식은 μ_i, λ_t와 ν_{it}가 서로 독립이라는 사실을 이용하면 얻을 수 있다. 네 번째 등식은 $i_T E_T = 0$와 $i_T \bar{J}_T = i_T$라는 성질을 이용하였으며, $l_i \bar{J}_N = i_N / N$이므로 다섯 번째 등식이 성립하게 된다. 또한 마지막 등식에서 $\bar{\tilde{u}}_{i.,GLS} = \sum_{j=1}^{T} \tilde{u}_{ij,GLS}/T$와 $\bar{\tilde{u}}_{..,GLS} = \sum_{i=1}^{N}\sum_{j=1}^{T} \tilde{u}_{ij,GLS}/NT$이다. 그러므로 이원 오차성분모형에 대한 BLUP는 다음과 같이 구해진다.

$$\tilde{y}_{i,T+s} = x'_{i,T+s}\tilde{\beta}_{GLS} + T\sigma_\mu^2\Big(\frac{1}{\lambda_2}\bar{\tilde{u}}_{i.,GLS} + (\frac{1}{\lambda_4} - \frac{1}{\lambda_2})\bar{\tilde{u}}_{..,GLS}\Big) \tag{3.3}$$

그러나 식 (3.3)에서 구해진 BLUP는 σ_μ^2과 σ_λ^2이 알려지지 않을 경우 이론적인 값에 불과하다. 이와 같은 BLUP는 4장의 모의실험에서 제안된 여러 최적예측량들을 비교할 때 기준이 되는 값으로 사용된다.

3.1. OLS예측량

오차항의 분산-공분산행렬의 형태에 관계없이 사용하는 OLS추정량을 이용한 i번째 개체의 $(T+s)$시점에서의 예측량은 다음과 같이 구해진다.

$$\hat{y}_{i,T+s}(\mathrm{OLS}) = x'_{i,T+s}\hat{\beta}_{\mathrm{OLS}} \tag{3.4}$$

여기서 $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$이다. OLS예측량은 회귀계수 추정시에 분산성분을 고려하지 않을뿐 아니라 식 (3.3)의 두 번째 항 및 세 번째 항을 모두 고려하지 않기 때문에 효율성이 떨어지게 될 것이다.

3.2. FGLS추정량을 이용한 최적가능예측량

실제적으로 미래시점에 대한 최적가능예측량을 유도하기 위해서는 알려지지 않은 분산성분들과 회귀계수에 대한 추정이 선행되어야 한다. 이때 분산성분들과 회귀계수에 대한 추정방법으로는 다음과 같은 "추정가능" GLS(Feasible GLS, FGLS)추정량을 고려할 수 있다

$$\tilde{\beta}_{\mathrm{FGLS}} = (X'\tilde{\Omega}^{-1}X)^{-1}X'\tilde{\Omega}^{-1}y \tag{3.5}$$

여기서 $\tilde{\Omega}^{-1}$는 추정된 분산성분으로 이루어진 분산-공분산 행렬의 추정치이다. 본 절에서는 Amemiya (1971)와 Swamy와 Arora (1972)에 의해 제안된 방법으로 분산성분을 추정하여 회귀계수를 추정하고, 이를 바탕으로 "추정된" BLUP(Estimated BLUP, EBLUP)를 구하고자 한다.

3.2.1. Amemiya Type 예측량

Amemiya (1971)는 분산성분에 대한 추정량을 구하기 위하여 개체효과와 시간효과를 모두 제거시키는 Q_1변환(Within transformation)을 통해 얻어지는 내부변환잔차(within residuals)를 이용하였다. 이 경우 Q_1변환을 통해 얻어지는 회귀계수에 대한 추정량은 $\tilde{\beta}_{s,\mathrm{WTN}} =$

$(X_s'Q_1X_s)^{-1}X_s'Q_1y$가 된다. 이때 X_s는 설명변수행렬 X에서 $Q_1i_{NT}=0$라는 성질때문에 소거되는 상수항을 제외한 $NT\times(k-1)$행렬이며, β_s는 상수항에 해당되는 모수를 제외한 $(k-1)\times 1$ 모수벡터이다. 그러므로 회귀계수에 대한 내부변환추정량을 이용한 내부변환 잔차는 다음과 같다.

$$\widetilde{u}_{WTN}=y-\widetilde{\alpha}_{\mathrm{WTN}}i_{NT}-X_s'\widetilde{\beta}_{s,\mathrm{WTN}}, \tag{3.6}$$

여기서 $\widetilde{\alpha}_{\mathrm{WTN}}=\bar{y}_{..}-\bar{X}_{s,..}'\widetilde{\beta}_{s,\mathrm{WTN}}$이다. 식 (3.6)의 내부변환잔차를 이용한 Amemiya 형태의 분산성분에 대한 추정량은 다음과 같이 구해진다(Amemiya(1971)).

$$\widetilde{\lambda}_{i,\mathrm{AM}}=\frac{\widetilde{u}_{WTN}'Q_i\widetilde{u}_{WTN}}{tr(Q_i)},\qquad i=1,2,3 \tag{3.7}$$

이와 같이 추정된 분산성분을 식 (3.5)에 대입하면 Amemiya형태의 회귀계수에 대한 추정량을 얻을 수 있다. 식 (3.7)에 의해 추정된 분산성분과 식 (3.5)에 의해 추정되는 회귀계수를 식 (3.3)에 대입하면 다음과 같은 Amemiya형태의 EBLUP를 얻을 수 있다.

$$\widetilde{y}_{i,T+s}(\mathrm{AM})=x_{i,T+s}'\widetilde{\beta}_{\mathrm{AM}}+T\widetilde{\sigma}_{\mu,\mathrm{AM}}^2\Big(\frac{1}{\widetilde{\lambda}_{2,\mathrm{AM}}}\bar{\widetilde{u}}_{i.,\mathrm{AM}}+(\frac{1}{\widetilde{\lambda}_{4,\mathrm{AM}}}-\frac{1}{\widetilde{\lambda}_{2,\mathrm{AM}}})\bar{\widetilde{u}}_{..,\mathrm{AM}}\Big) \tag{3.8}$$

여기서 $\widetilde{u}_{\mathrm{AM}}=y-X\widetilde{\beta}_{\mathrm{AM}}$, $\bar{\widetilde{u}}_{i.,\mathrm{AM}}=\sum_{j=1}^{T}\widetilde{u}_{ij,\mathrm{AM}}/T$와 $\bar{\widetilde{u}}_{..,\mathrm{AM}}=\sum_{i=1}^{N}\sum_{j=1}^{T}\widetilde{u}_{ij,\mathrm{AM}}/NT$이며 $\widetilde{\sigma}_{\mu,\mathrm{AM}}^2=(\widetilde{\lambda}_{2,\mathrm{AM}}-\widetilde{\lambda}_{1,\mathrm{AM}})/T$, $\widetilde{\lambda}_{4,\mathrm{AM}}=\widetilde{\lambda}_{2,\mathrm{AM}}+\widetilde{\lambda}_{3,\mathrm{AM}}-\widetilde{\lambda}_{1,\mathrm{AM}}$이다. 식 (3.8)에서 얻어진 EBLUP를 AM이라 정의하자.

3.2.2. Swamy와 Arora Type 예측량

Swamy와 Arora (1972) 는 분산성분을 추정하기 위하여 내부변이, 개체변이 및 시간 변이를 이용하는 세 개의 회귀모형을 적합하였으며, 각 모형의 평균제곱 오차를 이용하여 분산성분을 다음과 같이 추정하였다.

$$\begin{aligned}\widetilde{\lambda}_{1,\mathrm{SW}}&=\frac{y'Q_1y-y'Q_1X_s(X_s'Q_1X_s)^{-1}X'Q_1y}{(N-1)(T-1)-k+1}\\ \widetilde{\lambda}_{2,\mathrm{SW}}&=\frac{y'Q_2y-y'Q_2X_s(X_s'Q_2X_s)^{-1}X'Q_2y}{N-k}\\ \widetilde{\lambda}_{3,\mathrm{SW}}&=\frac{y'Q_3y-y'Q_3X_s(X_s'Q_3X_s)^{-1}X'Q_3y}{T-k}\end{aligned} \tag{3.9}$$

이와 같이 추정된 분산성분을 식 (3.5)에 대입하면 Swamy와 Arora형태의 회귀계수에 대한 추정량을 얻을 수 있다. 그러므로 추정된 분산성분과 회귀계수를 식 (3.3)에 대입하면 다음과 같은 Swamy와 Arora형태의 EBLUP를 얻을 수 있다.

$$\widetilde{y}_{i,T+s}(\mathrm{SA})=x_{i,T+s}'\widetilde{\beta}_{\mathrm{SA}}+T\widetilde{\sigma}_{\mu,\mathrm{SA}}^2\Big(\frac{1}{\widetilde{\lambda}_{2,\mathrm{SA}}}\bar{\widetilde{u}}_{i.,\mathrm{SA}}+(\frac{1}{\widetilde{\lambda}_{4,\mathrm{SA}}}-\frac{1}{\widetilde{\lambda}_{2,\mathrm{SA}}})\bar{\widetilde{u}}_{..,\mathrm{SA}}\Big) \tag{3.10}$$

여기서 $\bar{\widetilde{u}}_{i.,\mathrm{SA}}=\sum_{j=1}^{T}\widetilde{u}_{ij,\mathrm{SA}}/T$와 $\bar{\widetilde{u}}_{..,\mathrm{SA}}=\sum_{i=1}^{N}\sum_{j=1}^{T}\widetilde{u}_{ij,\mathrm{SA}}/NT$이다. 식 (3.10)에서 얻어진 EBLUP를 SA라 정의하자.

3.3. ML추정량을 이용한 예측량

ML추정방법은 오차항의 정규성 가정하에서 우도함수를 최대로 하는 추정량을 찾는 방법으로, 분산성분들과 회귀계수를 동시에 추정할 수 있다. 먼저 $\phi_1 = \sigma_\mu^2/\sigma_\nu^2$, $\phi_2 = \sigma_\lambda^2/\sigma_\nu^2$ 및 $\Omega = \sigma_\nu^2\Sigma$라 정의하면 정규성 가정하에서 모형 (2.1)에 대한 로그 우도함수(Log likelihood function)는 다음과 같다.

$$\log L = -\frac{NT}{2}\log(2\pi) - \frac{NT}{2}\log\sigma_\nu^2 - \frac{1}{2}\log|\Sigma| - \frac{1}{2\sigma_\nu^2}(y - X\beta)'\Sigma^{-1}(y - X\beta) \tag{3.11}$$

식 (3.11)의 우도함수에서 β와 σ_ν^2에 대한 일차미분값은 ϕ_1과 ϕ_2가 주어졌을 때 다음과 같은 대수적인 해를 제공해 준다.

$$\ddot{\beta}_{\mathrm{ML}} = (X'\ddot{\Sigma}^{-1}X)^{-1}X'\ddot{\Sigma}^{-1}y \tag{3.12}$$

$$\ddot{\sigma}^2_{\nu,\mathrm{ML}} = \ddot{u}'_{\mathrm{ML}}\ddot{\Sigma}^{-1}\ddot{u}_{\mathrm{ML}}/NT \tag{3.13}$$

그러나 ϕ_1과 ϕ_2의 일차미분값은 비선형함수로 나타난다. 그러므로 반복에 의한 수치해석적인 방법이 필요하다. 본 논문에서는 Fisher의 스코어링(scoring) 방법을 사용하여 ϕ_1과 ϕ_2에 대한 ML추정량을 유도하였다. 이때 필요한 일차미분값과 정보행렬은 다음과 같이 구해진다.

$$\begin{aligned}
\frac{\partial logL}{\partial\phi_1} &= -\frac{1}{2}\Big[\frac{(N-1)T}{T\phi_1+1} + \frac{T}{(T\phi_1+N\phi_2+1)}\Big] \\
&\quad + \frac{1}{2\sigma_\nu^2}\Big[\frac{T}{(T\phi_1+1)^2}u'Q_2u + \frac{T}{(T\phi_1+N\phi_2+1)^2}u'Q_4u\Big] \\
\frac{\partial logL}{\partial\phi_2} &= -\frac{1}{2}\Big[\frac{N(T-1)}{N\phi_2+1} + \frac{N}{(T\phi_1+N\phi_2+1)}\Big] \\
&\quad + \frac{1}{2\sigma_\nu^2}\Big[\frac{N}{(N\phi_2+1)^2}u'Q_3u + \frac{N}{(T\phi_1+N\phi_2+1)^2}u'Q_4u\Big]
\end{aligned} \tag{3.14}$$

$$\begin{aligned}
E\Big[-\frac{\partial^2 logL}{\partial\phi_1^2}\Big] &= \frac{1}{2}\Big[\frac{T^2(N-1)}{(T\phi_1+1)^2} + \frac{T^2}{(T\phi_1+N\phi_2+1)^2}\Big] \\
E\Big[-\frac{\partial^2 logL}{\partial\phi_2^2}\Big] &= \frac{1}{2}\Big[\frac{N^2(T-1)}{(N\phi_2+1)^2} + \frac{N^2}{(T\phi_1+N\phi_2+1)^2}\Big] \\
E\Big[-\frac{\partial^2 logL}{\partial\phi_1\partial\phi_2}\Big] &= \frac{NT}{2(T\phi_1+N\phi_2+1)^2}
\end{aligned} \tag{3.15}$$

적당한 초기값에서 시작하여 $(n+1)$번째 단계에서 갱신되는 $\ddot{\phi}_1$과 $\ddot{\phi}_2$는 다음과 같이 구해진다.

$$\begin{bmatrix} \ddot{\phi}_1 \\ \ddot{\phi}_2 \end{bmatrix}_{n+1} = \begin{bmatrix} \ddot{\phi}_1 \\ \ddot{\phi}_2 \end{bmatrix}_n + \begin{bmatrix} E[-\frac{\partial^2 logL}{\partial\phi_1^2}] & E[-\frac{\partial^2 logL}{\partial\phi_1\partial\phi_2}] \\ E[-\frac{\partial^2 logL}{\partial\phi_1\partial\phi_2}] & E[-\frac{\partial^2 logL}{\partial\phi_2^2}] \end{bmatrix}_n^{-1} \begin{bmatrix} \frac{\partial logL}{\partial\phi_1} \\ \frac{\partial logL}{\partial\phi_2} \end{bmatrix}_n \tag{3.16}$$

각 단계에서 $\partial logL/\partial\phi_1$과 $\partial logL/\partial\phi_2$는 식 (3.14)에 의하여 구하고, 정보행렬은 식 (3.15)에 의하여 구한다. 이와 같이 구해진 회귀계수와 분산성분에 대한 ML추정량을 이용한 EBLUP는 다음과 같다.

$$\ddot{y}_{i,T+s}(\mathrm{ML}) = x'_{i,T+s}\ddot{\beta}_{\mathrm{ML}} + T\ddot{\sigma}^2_{\mu,\mathrm{ML}}\Big(\frac{1}{\ddot{\lambda}_{2,\mathrm{ML}}}\ddot{\bar{u}}_{i.,\mathrm{ML}} + (\frac{1}{\ddot{\lambda}_{4,\mathrm{ML}}} - \frac{1}{\ddot{\lambda}_{2,\mathrm{ML}}})\ddot{\bar{u}}_{..,\mathrm{ML}}\Big) \tag{3.17}$$

여기서 $\ddot{u}_{\mathrm{ML}}$는 ML잔차이며, 식 (3.17)에서 얻어진 EBLUP를 ML이라 정의하자.

3.4. REML추정량을 이용한 예측량

분산성분에 대한 ML추정량은 회귀계수로 인한 자유도 손실을 고려하지 않아 일반적으로 편의 추정량이라는 단점을 가지고 있다. 이와같은 문제를 해결하기 위하여 Patterson과 Thompson (1971)은 비정칙 변환(Singular transformation)을 이용한 제한적 최대우도추정(Restricted ML, REML)방법을 제안하였다. 그동안 분산성분에 대한 REML추정량의 효율성에 대해서는 다양한 연구들이(Harville (1977), Engle (1990)) 진행되어 왔다. 반면 이를 미래시점의 예측에 적용한 연구는 Latif와 King (1993)이 있으나 패널회귀모형에 적용한 연구는 없었다.

REML추정방법은 원자료에 대하여 $y'[A \mid \Sigma^{-1}X/\sigma^2_\nu]$ 이라는 변환에 근거한 추정법으로, 여기서 행렬 A는 $A'A = I - X(X'X)^{-1}X'$과 $AA' = I$를 만족하는 계수 $(NT-k)\times NT$인 행렬이다. 이 때 $A'y$과 $X'\Sigma^{-1}y/\sigma^2_\nu$는 독립이 되며 다음과 같은 분포를 따르게 된다.

$$\begin{bmatrix} A'y \\ X'\Sigma^{-1}/\sigma^2_\nu \end{bmatrix} \sim N\Big(\begin{bmatrix} 0 \\ X'\Sigma^{-1}X\beta/\sigma^2_\nu \end{bmatrix}, \begin{bmatrix} \sigma^2_\nu A\Sigma A' & 0 \\ 0 & X'\Sigma^{-1}X/\sigma^2_\nu \end{bmatrix}\Big) \tag{3.18}$$

식(3.18)에서 $A'y$와 $X'\Sigma^{-1}y/\sigma^2_\nu$의 로그우도함수를 각각 $logL_1$과 $logL_2$라 하면 $logL_1$은 회귀계수에 의존하지 않게 된다. 그러므로 이 부분을 최대화하면 분산성분들에 대한 REML추정량을 얻게 된다. Hocking (1985)에 의하면 $A'(A\Sigma A')^{-1}A' = \Sigma^{-1}[I-X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}]$라는 사실을 알 수 있으므로 $logL_1$에 대한 σ^2_ν, ϕ_1과 ϕ_2에 대한 일차미분값은 다음과 같이 구해진다.

$$\begin{aligned}
\frac{\partial logL_1}{\partial\sigma^2_\nu} &= -\frac{NT-k}{2\sigma^2_\nu} + \frac{1}{2\sigma^4_\nu}y'\Sigma^{-1}Py \\
\frac{\partial logL_1}{\partial\phi_1} &= -\frac{1}{2}tr[(I_N\otimes J_T)\Sigma^{-1}P] + \frac{1}{2\sigma^2_\nu}y'[\Sigma^{-1}P(I_N\otimes J_T)\Sigma^{-1}P]y \\
\frac{\partial logL_1}{\partial\rho_2} &= -\frac{1}{2}tr[(J_N\otimes I_T)\Sigma^{-1}P] + \frac{1}{2\sigma^2_\nu}y'[\Sigma^{-1}P(J_N\otimes I_T)\Sigma^{-1}P]y
\end{aligned} \tag{3.19}$$

여기서 $P = I - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$ 이다. 식 (3.19)의 값이 0이 되는 해가 모수 σ^2_ν, ϕ_1과 ϕ_2에 대한 REML추정량이 된다. ϕ_1과 ϕ_2가 주어졌을 때 식 (3.19)의 첫 번째 식을 풀면 σ^2_ν에 대한 다음과 같은 추정량을 얻을 수 있다.

$$\hat{\sigma}^2_{\nu,\mathrm{REML}} = y'[\ddot{\Sigma}^{-1}P]y/(NT-k) \tag{3.20}$$

그러나, ϕ_1과 ϕ_2에 대해서는 대수적인 해를 구할 수 없으므로 반복에 의한 수치적인 해를 구해야만 한다. 여기에서도 ϕ_1과 ϕ_2를 추정하기 위하여 Fisher의 'scoring' 방법이 사용되었다. 이에 필요한 정보행렬은 Harville (1977)의 결과를 이용하면 쉽게 구할 수 있다(분산성분 모형에 대한 ML추정량과 REML추정량들에 대한 수치적인 방법들의 장단점 및 자세한 설명은 Harville (1977) 및 Searle et. al (1992)등을 참조하기바람).

갱신되는 $\ddot{\phi}_{1,\text{REML}}$과 $\ddot{\phi}_{2,\text{REML}}$의 값은 식 (3.16)과 같은 식에 의하여 얻어진다. REML추정량도 ML추정량과 마찬가지로 갱신된 값이 0보다 작게 되면, 이는 0으로 대치된다. 그러므로 회귀계수와 분산성분에 대한 REML추정량을 이용한 EBLUP는 다음과 같다.

$$\begin{aligned}\ddot{y}_{i,T+s}(\text{REML}) &= x'_{i,T+s}\ddot{\beta}_{\text{REML}} \\ &+T\ddot{\sigma}^2_{\mu,\text{REML}}\Big(\frac{1}{\ddot{\lambda}_{2,\text{REML}}}\ddot{\bar{u}}_{i.,\text{REML}} + (\frac{1}{\ddot{\lambda}_{4,\text{REML}}} - \frac{1}{\ddot{\lambda}_{2,\text{REML}}})\ddot{\bar{u}}_{..,\text{REML}}\Big) \end{aligned} \quad (3.21)$$

식 (3.21)에서 얻어진 EBLUP를 REML이라 정의하자.

4. 모의실험

4.1. 모의실험 방법

3장에서 제안된 미래시점에 대한 여러가지 EBLUP들의 효율성을 비교하기 위하여 다음과 같은 패널회귀모형을 이용하여 모의실험을 실시하였다.

$$y_{it} = \alpha + x_{it}\beta + u_{it}, \qquad i = 1, 2, \cdots, N, \qquad t = 1, 2, \cdots, T+s \quad (4.1)$$

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \quad (4.2)$$

여기서 s는 예측될 미래시점의 수를 나타낸다. 모형 (4.1)에서 설명변수 x_{it}는 Nerlove (1971)의 방법을 사용하여 발생시켰다. 즉, w_{it}가 균일 분포 (-0.5 , 0.5)를 따르는 난수라 했을 x_{it}는 다음과 같은 식을 통하여 반복적으로 생성되었다.

$$x_{it} = 0.1t + 0.5x_{i,t-1} + w_{it} \quad (4.3)$$

이때 초기값 x_{i0}는 $5 + 10w_{i0}$에서 구하였다. 모의실험 전체를 통하여 $\alpha = 5$, $\beta = 0.5$와 전체 분산 $\sigma^2 = \sigma^2_\mu + \sigma^2_\lambda + \sigma^2_\nu = 20$으로 고정하였다. 각 실험에서는 $\rho_1 = \sigma^2_\mu/\sigma^2$, $\rho_2 = \sigma^2_\lambda/\sigma^2$의 값을 0.0에서 0.8 사이에서 0.2 단위로 변화시켜가며 항상 $1-\rho_1-\rho_2$의 값이 0보다 크도록 실험하였다. N개의 μ_i는 $N(0, \sigma^2_\mu)$에서, $T+s$개의 λ_t는 $N(0, \sigma^2_\lambda)$에서 생성하였고, $N(T+s)$개의 ν_{it}는 $N(0, \sigma^2_\nu)$에서 생성하였다. 본 연구에서는 두 개의 미래시점 ($s = 2$)에 대하여 예측값을 구하였다. 모든 실험은 1000번 독립적으로 반복실시 하였으며, SAS/IML 프로시져를 이용하여 수행되었다. 대부분의 패널자료는 개체의 수가 크고 관측 시간이 작은 것이 일반적이다. 그러므로 실험에 사용된 표본은 T를 작은값($T = 10, 20$)으로 고정하고 N을 $10, 50, 100, 200$으로 증가시켜 보았다. 각 실험에서는 두 개의 미래시점에 대하여 3장에서 제안된 각 EBLUP를 계산하고, 이에 대하여 다음과 같은 예측평균제곱오차(Prediction Mean

Square Error, PMSE)를 계산하였다.

$$PMSE = \frac{1}{NR}\sum_{r=1}^{R}\sum_{i=1}^{N}(y_{i,T+s} - \widehat{y}_{i,T+s})^2, \qquad s = 1, 2 \tag{4.4}$$

여기서 $y_{i,T+s}$는 i번째 개체의 $(T+s)$번째 생성된 실제 관측치이고, $\widehat{y}_{i,T+s}$는 각 예측방법에 의해서 적합된 EBLUP이다. 이때 r는 반복수를 나타내며 본 실험에서는 $R = 1000$이다. 이와 같이 얻어진 각 예측량들의 PMSE값을 식 (3.3)를 이용한 참 GLS의 PMSE에 대한 상대효율(Relative Efficiency, RE)로 나타내면 다음과 같다.

$$RE = \frac{PMSE(\widetilde{\beta})}{PMSE(\widehat{\beta}_{GLS})} \tag{4.5}$$

여기서 $PMSE(\widetilde{\beta})$는 각 예측방법에 의해서 계산된 EBLUP의 PMSE를 나타내고, $PMSE$ $(\widehat{\beta}_{GLS})$는 실험에서 실제 사용된 모수값을 사용하여 얻어진 참 BLUP에 대한 PMSE이다. 따라서 상대효율 RE값이 1에 가까워질수록 효율성이 좋은 예측방법이라고 할 수 있다.

4.2. 모의실험 결과

표 4.1에서 표 4.3까지는 각각 (N, T)가 (10,10), (20,10)과 (100,10)인 경우에 2개의 미래 시점에 대하여 3장에 제시되어있는 여러가지 예측방법들의 상대효율을 나타낸다 (지면관계상 제시하지 않은 모의실험 결과는 저자들에게 요구할 수 있음). 표 4.1에서 표 4.3까지의 결과를 요약하면 다음과 같다.

표 4.1: 각 예측량들의 상대효율(N=10, T=10)

		$s=1$					$s=2$				
ρ_1	ρ_2	OLS	AM	SA	ML	REML	OLS	AM	SA	ML	REML
0.0	0.0	1.012	1.030	1.024	1.018	1.020	1.014	1.029	1.026	1.019	1.021
0.0	0.2	1.032	1.045	1.040	1.037	1.038	1.032	1.037	1.039	1.034	1.035
0.0	0.4	1.046	1.057	1.052	1.051	1.052	1.048	1.052	1.055	1.051	1.051
0.0	0.6	1.065	1.066	1.066	1.064	1.064	1.066	1.058	1.054	1.056	1.056
0.0	0.8	1.099	1.089	1.086	1.088	1.088	1.099	1.089	1.085	1.087	1.087
0.2	0.0	1.144	1.015	1.009	1.008	1.007	1.154	1.022	1.016	1.016	1.015
0.2	0.2	1.193	1.026	1.022	1.023	1.022	1.194	1.041	1.035	1.037	1.037
0.2	0.4	1.230	1.048	1.045	1.047	1.047	1.221	1.038	1.034	1.036	1.036
0.2	0.6	1.245	1.052	1.048	1.053	1.053	1.269	1.063	1.057	1.062	1.062
0.4	0.0	1.465	1.014	1.009	1.009	1.008	1.476	1.018	1.011	1.011	1.010
0.4	0.2	1.494	1.025	1.017	1.019	1.019	1.533	1.026	1.022	1.023	1.023
0.4	0.4	1.570	1.036	1.032	1.035	1.035	1.555	1.038	1.034	1.037	1.037
0.6	0.0	2.155	1.012	1.006	1.005	1.005	2.182	1.010	1.006	1.006	1.005
0.6	0.2	2.303	1.033	1.027	1.030	1.030	2.243	1.030	1.026	1.028	1.028
0.8	0.0	4.204	1.011	1.005	1.002	1.002	4.321	1.011	1.003	1.003	1.003

표 4.2: 각 예측량들의 상대효율(N=20, T=10)

		$s=1$					$s=2$				
ρ_1	ρ_2	OLS	AM	SA	ML	REML	OLS	AM	SA	ML	REML
0.0	0.8	1.092	1.090	1.091	1.091	1.091	1.097	1.085	1.085	1.085	1.085
0.0	0.6	1.059	1.062	1.058	1.060	1.060	1.082	1.049	1.076	1.077	1.077
0.0	0.4	1.049	1.046	1.044	1.045	1.045	1.048	1.059	1.048	1.048	1.049
0.0	0.2	1.021	1.021	1.021	1.021	1.021	1.024	1.029	1.025	1.024	1.024
0.0	0.0	1.002	1.005	1.008	1.003	1.003	1.001	1.005	1.010	1.004	1.004
0.2	0.6	1.287	1.068	1.066	1.067	1.067	1.303	1.082	1.081	1.082	1.082
0.2	0.4	1.275	1.067	1.065	1.066	1.066	1.254	1.052	1.050	1.052	1.052
0.2	0.2	1.197	1.020	1.018	1.019	1.019	1.194	1.018	1.016	1.017	1.017
0.2	0.0	1.164	1.003	1.003	1.002	1.002	1.170	1.004	1.003	1.003	1.003
0.4	0.4	1.660	1.046	1.046	1.047	1.047	1.642	1.051	1.051	1.052	1.052
0.4	0.2	1.576	1.017	1.016	1.017	1.017	1.590	1.027	1.024	1.026	1.026
0.4	0.0	1.512	1.003	1.001	1.001	1.001	1.513	1.002	1.001	1.001	1.001
0.6	0.2	2.386	1.033	1.032	1.033	1.033	2.320	1.015	1.014	1.015	1.015
0.6	0.0	2.257	1.002	1.001	1.000	1.000	2.268	1.002	1.001	1.001	1.001
0.8	0.0	4.454	1.002	1.000	1.000	1.000	4.476	1.002	1.000	1.000	1.000

표 4.3: 각 예측량들의 상대효율(N=100, T=10)

		$s=1$					$s=2$				
ρ_1	ρ_2	OLS	AM	SA	ML	REML	OLS	AM	SA	ML	REML
0.0	0.8	1.091	1.083	1.084	1.084	1.084	1.072	1.066	1.066	1.066	1.066
0.0	0.6	1.062	1.060	1.060	1.060	1.060	1.071	1.065	1.064	1.064	1.064
0.0	0.4	1.039	1.035	1.036	1.035	1.035	1.049	1.044	1.044	1.044	1.044
0.0	0.2	1.026	1.023	1.025	1.024	1.024	1.031	1.029	1.031	1.029	1.029
0.0	0.0	1.001	1.002	1.007	1.002	1.002	1.001	1.002	1.008	1.002	1.002
0.2	0.6	1.286	1.068	1.067	1.068	1.068	1.323	1.077	1.076	1.077	1.077
0.2	0.4	1.234	1.037	1.036	1.037	1.037	1.246	1.055	1.054	1.055	1.055
0.2	0.2	1.207	1.021	1.020	1.021	1.021	1.210	1.019	1.017	1.018	1.018
0.2	0.0	1.166	1.001	1.001	1.000	1.000	1.165	1.001	1.001	1.000	1.000
0.4	0.4	1.690	1.054	1.053	1.054	1.054	1.685	1.060	1.060	1.060	1.060
0.4	0.2	1.597	1.023	1.022	1.023	1.023	1.595	1.035	1.034	1.035	1.035
0.4	0.0	1.527	1.001	1.000	1.000	1.000	1.535	1.001	1.000	1.000	1.000
0.6	0.2	2.420	1.039	1.038	1.039	1.039	2.387	1.026	1.026	1.026	1.026
0.6	0.0	2.277	1.001	1.000	1.000	1.000	2.275	1.001	1.000	1.000	1.000
0.8	0.0	4.522	1.001	1.000	1.000	1.000	4.519	1.001	1.000	1.000	1.000

먼저 $\rho_1 = 0$인 경우에는 OLS를 이용한 예측방법이 본 연구에서 고려된 모든 예측방법들과 비교하여 효율성에 큰 차이를 보이지 않았다. OLS방법의 상대효율이 오차항들의 자기상관을 고려한 SA, AM, ML과 REML방법과 거의 같은 효율을 보이고 있음은 특이한 결과인데, 이는 다음과 같이 설명될 수 있다. 즉, $\rho_1 = 0$인 경우에는 이론적인 σ_μ^2의 값이 0이 되어 식 (3.3)에 제시되어 있는 BLUP의 두 번째 항과 세 번째 항의 값이 0이 된다. 이와 같은 경우 각 방법에 의해 추정된 EBLUP의 두 번째 항과 세 번째 항의 값도 0에 가까운 값을 갖게 된다. 그러므로 실제 추정되는 EBLUP들은 회귀계수에 대한 추정량을 제외한 나머지 부분이 OLS를 이용한 예측량과 거의 동일하게 되어, 이러한 효과가 OLS추정량의 효율성에 영향을 주었을것으로 판단된다.

ρ_1이 0보다 큰 경우에는 이론적인 σ_μ^2의 값이 커지게 되고 이로 인하여 추정되는 σ_μ^2의 값도 0보다 커지게 된다. 그러므로 식 (3.3)에 나타난 EBLUP의 두 번째 항과 세 번째 항의 값이 존재하게 되므로, 이 부분을 무시한 OLS예측량의 상대효율이 다른 예측방법(SA, AM, ML, REML)에 비해서 떨어지고 있음을 알 수 있다. 이와 같은 OLS예측량의 비효율성은 ρ_1과 ρ_2가 커질수록 또는 N과 T가 커질수록 더 심각해지는 경향이 존재한다 (표 4.1 - 표 4.3 참조). 특히, $N = 10, T = 10$인 경우 $\rho_1 = 0.8$, $\rho_2 = 0.0$에서 OLS의 상대효율은 1차의 미래시점($s = 1$)에 대하여 4.204로 나타나고 2차의 미래시점($s = 2$)에 대하여 4.321로 나타나 참 GLS를 이용한 BLUP, FGLS를 이용한 EBLUP 및 ML추정을 이용한 EBLUP들보다 4배 이상 높은 PMSE를 보여주고 있다.

반면 SA, AM, ML 및 REML방법을 통해 얻어지는 EBLUP들의 PMSE는 서로 큰 차이를 보이지 않고 있으며 참 GLS를 이용한 BLUP의 PMSE와도 큰 차이를 보이지 않고 있다. 실제로 $N = 10, T = 10$인 경우 이들의 상대효율은 1차의 미래시점($s = 1$)에 대하여 최대 1.089(AM, $\rho_1 = 0.0$, $\rho_2 = 0.8$)로 BLUP에 비하여 8.9 % 큰 PMSE를 보이고 있다. 이와 같은 상대효율은 N과 T가 커질수록 감소하여 $N = 100, T = 10$인 경우에는 거의 BLUP와 같은 효율을 나타내고 있다.

이상의 결과를 통하여 ML방법(ML, REML)과 FGLS방법(AM, SA)을 이용한 예측은 그 효율이 거의 동일하다는 것을 보았다. ML방법(ML, REML)을 이용한 예측은 정규분포의 가정이 필요하고 수치적인 방법을 이용하여 모수를 추정해야하기 때문에 계산상 어려움이 존재한다. 반면, FGLS방법(AM, SA)을 이용한 예측은 ML방법(ML, REML)에 비하여 계산이 용이하고 정규성의 가정이 필요없다는 장점이 있다. 또한 효율성에서도 FGLS방법이 ML방법에 비하여 크게 떨어지지 않으므로 패널회귀모형에서 미래시점의 예측시 ML이나 REML예측량 대신 FGLS방법을 이용한 예측량들(AM, SA)의 사용을 권장할 만하다.

5. 실증적 자료 분석

본 장에서는 3장에서 유도한 이원오차성분모형에 대한 여러가지 예측량들의 예측값들을 Baltagi와 Griffin(1983)이 제시한 실증적인 패널자료에 적용하여 구하고, 각 예측방법들의 결과를 비교해보고자 한다. 이 자료는 OECD에 가입된 18개국에 대하여 1960에서 1978년 동안의 가솔린 수요에 대한 자료이다(Baltagi(1995), p 227-234). 3개년의 미래시

점($s = 1, 2, 3$)을 고려하여 각 예측방법들에 의한 EBLUP을 구했으며 이를 실제 자료와 비교하였다. 개체와 시간의 수는 각각 $N = 18, T = 16$이며 사용된 모형은 다음과 같다.

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \tag{5.1}$$

여기서 y는 자동차당 소비되는 가솔린 양이고, x_1은 실제 국가당 수입 x_2는 실제 자동차 가솔린가격이고 x_3는 국가당 자동차 보유수이다. 모든 값들은 log단위로 얻어졌다. 이 때 실제 자료에 대한 각 예측값들의 비교를 위해서 사용된 PMSE는 다음과 같이 얻어진다.

$$PMSE(\widehat{\beta}) = \frac{1}{18} \sum_{i=1}^{18} (y_{i,T+s} - \widehat{y}_{i,T+s})^2, \qquad s = 1, 2, 3 \tag{5.2}$$

여기서 $y_{i,T+s}$는 i번째 개체에 대한 $(T+s)$번째의 실제 관측값이고, $\widehat{y}_{i,T+s}$는 예측방법에 의해서 적합된 EBLUP이다. 표 5.1은 실제자료를 이용하여 얻은 3개년간의 실제 각 예측치의 PMSE값을 ML방법을 사용하여 얻은 PMSE 값에 대한 상대효율로 나타낸 것이다. 표 5.1에 나타난 AM과 SA는 각각 식 (3.7)과 식(3.9)의 분산성분 추정방법을 이용하여 식 (3.8)과 식 (3.10)에서 얻어진 미래시점에 대한 예측량들을 나타내며 ML과 REML은 분산성분에 대한 각각 ML추정량과 REML추정량을 이용하여 식 (3.17)과 식 (3.21)에 의해서 얻어진 예측량을 나타낸다.

표 5.1: 가솔린 자료에 대하여 각 예측량들의 ML예측량에 대한 상대효율

YEAR	ML	OLS	AM	SA	REML
1976	1.000	4.910	1.001	1.010	1.029
1977	1.000	4.296	1.001	1.037	1.040
1978	1.000	3.300	1.003	1.063	1.065
평균	1.000	4.169	1.002	1.037	1.045

표 5.1의 결과는 앞 장에서 고려한 모의실험의 결과와 거의 동일한 결과를 제공해 주고 있다. 이를 자세히 살펴보면, 오차항의 상관을 고려하지 않은 OLS방법은 ML방법에비해 최대 4.91(1976년)배 큰 PMSE를 보이고 있어 비효율적인 예측량임을 알 수 있다. 반면 FGLS 방법(AM, SA) 및 REML방법의 PMSE는 ML방법과 큰 차이를 보이지 않고 있음을 볼 수 있다. 특히 본 연구에서 고려한 가솔린 수요자료에 대해서 AM방법은 ML방법과 거의 같은 PMSE를 갖는 것으로 나타났다.

6. 결 론

본 연구에서는 패널회귀모형에서 미래시점에 대한 여러가지 예측량들 을 제안하고 그들의 효율성을 모의실험과 실증적 패널자료를 통하여 비교하였다. 모의실험 결과, 미래시점의 예측에 대하여 오차항의 상관을 고려하지 않는 OLS방법은 다른 예측방법들에 비하

여 효율성이 떨어지는 결과를 보여주었다. 반면, 오차항의 상관을 고려한 FGLS방법(AM, SA)과 ML방법(ML, REML)에 의한 예측량들은 BLUP와 비교할 때 상대효율이 크게 떨어지지 않았다. 특히, FGLS방법을 이용한 예측량(AM, SA)은 ML방법(ML, REML)에서 필요로하는 오차항의 정규성 가정이 필요없고, 계산이 간편하다는 장점이 존재하므로 패널회귀모형의 예측에 사용을 권장한다.

참고문헌

[1] Amemiya, T. (1971). The estimation of the variances in a variance components model, *International Econometric Review*, Vol. 12, 1-13.

[2] Baltagi, B.H. (1995). *Econometric Analysis of Panel Data*, John Wiley, New York.

[3] Baltagi, B.H. and J.M. Griffin (1983). Gasoline demand in the OECD: An application of pooling and testing procedures, *European Economic Review*, Vol. 22, 117-137.

[4] Engle, B. (1990). The analysis of unbalanced linear models with variance components, *Statistica Neerlandica*, Vol. 44, 195-219.

[5] Goldberger, A.S. (1962). Best linear unbiased prediction in the generalized linear regression model, *Journal of the American Statistical Association*, Vol. 57, 369-375.

[6] Hsiao, C. (1986). *Analysis of Panel Data*, Cambridge University Press, Cambridge.

[7] Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and related to problems, *Journal of the American Statistical Association*, Vol. 72, 320-340.

[8] Hocking, R.R. (1985). *The Analysis of Linear Models*, Brooks/Cole company, Monterey, California.

[9] Latif, A. and L. King (1993). Linear regression forecasting in the presence of AR(1) disturbances, *Journal of Forecasting*, Vol. 12, 513-524.

[10] Maddala, G.S. (1993). *The Econometrics of Panel Data, I, II*, Cambridge University Press, Cambridge.

[11] Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections, *Econometrica*. Vol. 39, 359-382.

[12] Patterson, H.D. and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal, *Biometrica*, Vol. 58, 545-554.

[13] Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*,Wiley, New

York.

[14] Swamy, P.A.V.B. and S.S. Arora (1972). The exact finite sample properties of the estimators of coefficients in the error components regression model, *Econometrica*, Vol. 40, 261-175.

[15] Taub, A.J. (1979). Prediction in the context of the variance-components model data, *Journal of Econometrics*, Vol. 13, 203-223.

[16] Wansbeek, T.J. and A. Kapteyn (1978). The separation of individual variation and systematic change in the analysis of panel data, *Annales de I'INSEE*, vol. 30-31, 659-680.

[2000년 9월 접수, 2001년 2월 채택]

A Comparison of Predictors in a Panel Data Regression Model*

Byoung-Cheol Jung[1)] Min-Wha Cho[2)] Seuck-Heun Song[3)]

ABSTRACT

This paper derives the BLUP in a panel data regression model with two way error components and investigates the performance of various predictors. Through simulation study and real data anaysis some of basic finding is following: the computationally simple FGLS(AM, SA) predictors perform reasonably well when compared with the computationally involved MLE and RMLE predictors.

Keywords: Panel Data Regression; BLUP.

* This work was supported by Ewha Womans University BK21 Grant.
1) Post Doctorial Researcher, Dept of Statistics, Ewha University.
2) Researcher, S-Link Co..
3) Assistance Professor, Dept of Statistics, Korea University. E-mail: ssong@korea.ac.kr