



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

碩 士 學 位 論 文

중도절단 패널회귀모형에서
이중 일반화 최대엔트로피
추정량에 관한 연구

高麗大學校 大學院

應用統計學科

李 棟 熙

2017年 6月 日

全 秀 榮 教 授 指 導
碩 士 學 位 論 文

중도절단 패널회귀모형에서
이중 일반화 최대엔트로피
추정량에 관한 연구

이 論文을 統計學 碩士學位 論文으로
提出함.

2017年 6月 日

高麗大學校大學院
應 用 統 計 學 科

李 棟 熙 (印)



李 棟 熙의 統計學 碩士學位論文

審査를 完了함.

2017年 6月 日

委員長 전 수 영



委 員 진 서 훈



委 員 최 보 승



요 약 문

패널회귀모형은 횡단면자료와 시계열자료가 병합된 패널자료를 기반으로 한다. 그러나 이 패널자료는 다수의 개체를 장기간 추적조사 하므로 조사기간 동안 자료의 일부가 누락되거나 오염되는 문제가 종종 발생한다. 특히 조사기간 중에 종속변수의 일부가 누락되는 경우를 중도절단(censoring)이라 하는데, 이를 고려하지 않을 경우 잘못된 통계적 추론을 하게 된다. 따라서 패널회귀모형 분석시 자료의 특성을 충분히 고려한 분석이 요구된다.

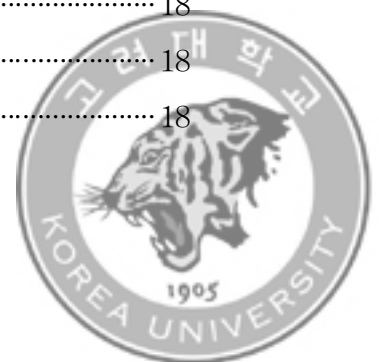
최근 일원오차성분을 갖는 중도절단 패널회귀모형에서의 이중 일반화 최대엔트로피 추정량이 Lee et al.(2016)에 의해 제안되었다. 본 연구는 Lee et al.(2016)이 제안한 추정량을 확장하여, 표본의 크기가 작고 다중공선성이 있는 중도 절단된 상태의 패널자료에서의 일원오차성분모형과 이원오차성분모형을 위한 일반화 최대엔트로피 추정량(DGME)을 제안한다. 본 연구에서 제안한 추정량의 효율성을 보기 위해, 중도절단되고 다중공선성이 있는 외생변수의 패널회귀모형에서 모의실험을 하였고, 분석결과 DGME가 안정적이고 우수한 추정량임을 보여주었다. 또한 미국 미시간 대학 사회조사연구소에서 조사된 PSID(Panel Study of Income Dynamics)의 실 자료를 분석한 결과 모의실험에서와 같이 DGME가 타 추정량들에 비해 우수한 추정량임을 보여주었다.

주요용어: 중도절단, DGME, 불완전 자료, 다중공선성, 패널자료



목 차

요 약 문	i
목 차	ii
표 목 차	iv
제 1 장 서 론	1
제 2 장 이중 일반화 최대엔트로피 추정량	3
2.1 일반화 최대엔트로피형식	3
2.1.1 일원오차성분모형의 재모수화	3
2.1.2 일반화최대엔트로피 추정량	7
2.2 이중 일반화 최대엔트로피형식	9
제 3 장 중도절단 패널회귀모형	11
3.1 중도절단 일원오차성분모형	12
3.2 중도절단 이원오차성분모형	13
제 4 장 회귀계수에 대한 고전적 추정량	15
4.1 일원오차성분모형	15
4.1.1 OLS 추정량	15
4.1.2 GLS 추정량	15
4.1.3 FGLS 추정량	17
4.2 이원오차성분모형	18
4.2.1 OLS 추정량	18
4.2.2 GLS 추정량	18



4.2.3 FGLS 추정량	20
제 5 장 회귀계수에 대한 DGME 추정량	21
5.1 중도절단 일원오차성분모형	21
5.2 중도절단 이원오차성분모형	26
5.3 DGME 추정량의 점근적 성질	31
제 6 장 모의실험	36
6.1 다중공선성을 고려한 자료	36
6.1.1 중도절단 일원오차성분모형	36
6.1.1.1 모의실험 과정	37
6.1.1.2 모의실험 결과	39
6.1.2 중도절단 이원오차성분모형	41
6.1.2.1 모의실험 과정	42
6.1.2.2 모의실험 결과	43
6.2 다중공선성과 $n < p$ 를 고려한 자료	45
6.2.1 중도절단 일원오차성분모형	46
6.2.2 중도절단 이원오차성분모형	47
제 7 장 실 자료 분석	49
7.1 자료 소개	49
7.2 중도절단 일원오차성분모형	51
7.3 중도절단 이원오차성분모형	54
제 8 장 결론	57
참 고 문 헌	58

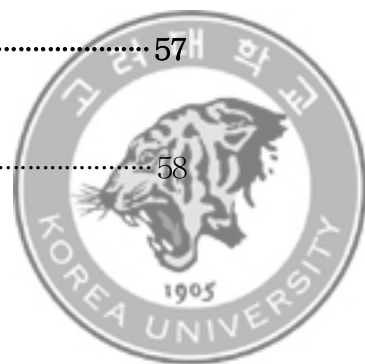


표 목 차

<표 6.1> 독립변수들 간의 다중공선성	40
<표 6.2> 일원오차성분모형에서 모의실험의 RMSE 결과	40
<표 6.3> 일원오차성분모형에서 한번 모의실험의 모수추정 결과 (추정치(표준편차))	40
<표 6.4> 이원오차성분모형에서 모의실험의 RMSE 결과	44
<표 6.5> 이원오차성분모형에서 한번 모의실험의 모수추정 결과 (추정치(표준편차))	45
<표 6.6> 독립변수들 간의 다중공선성($n < p$)	46
<표 6.7> 일원오차성분모형에서 한번 모의실험의 모수추정치와 RMSE	47
<표 6.8> 이원오차성분모형에서 한번 모의실험의 모수추정치와 RMSE	48
<표 7.1> 패널자료 생성을 위한 기준	50
<표 7.2> 변수 설명(1979년도 Codebook 기준)	50
<표 7.3> 일원오차성분모형에서 실 자료 분석결과	53
<표 7.4> 이원오차성분모형에서 실 자료 분석결과	56



제 1 장 서 론

패널자료는 동일 시점에서 관측된 여러 개체들의 관측값을 보여주는 횡단면 자료와 각 개체들이 시간의 흐름에 따라 어떠한 변화를 보여주는지를 반영한 시계열 자료가 혼합된 자료라 할 수 있다. 따라서 일반적인 횡단면 자료 혹은 시계열 자료 분석에 비해 시간에 따른 개체 특성의 변동이나 개체에 따른 시간 특성의 변동 등 복합적인 문제에 대한 해답을 제공하는 장점이 있다.

그러나 패널자료를 구성하기 위해서는 다수의 개체를 장기간 추적조사 해야 하므로, 조사기간 동안 자료의 일부가 누락되거나 오염되는 문제가 종종 발생한다. 예를 들어 연구자가 가장의 연간 근로소득에 관심이 있는 경우, 응답자가 개인적 사유로 인해 특정연도에 대해 응답을 거부하면 근로를 하고 있음에도 불구하고 해당연도에 대한 데이터는 $y=0$ 을 관측하게 된다. 이를 중도절단이라 하는데, 이러한 중도절단 문제는 패널자료 구축시 피할 수 없는 문제 중 하나이다.

최근 이러한 문제들을 해결하기 위해 자료의 생성과정에 제한이 주어졌거나 다중공선성이 있는 경우를 고려한 다양한 연구(Golan et al., 1996; Lee, Choi, 2004; Lee, Cheon, 2014, 2015)가 활발히 진행 되었고, 더 나아가 중도절단 자료에 대한 연구도 진행 중에 있다(Lee et al., 2016). 이러한 연구들은 공통적으로 불완전 자료에 대한 적절한 통계적 추론을 위해 최대엔트로피(maximum entropy, ME) 형식을 사용하고 있다.

가장 최근 연구된 Lee et al.(2016)을 살펴보면 중도절단된 패널자료에 대해 랜덤 개체효과를 고려한 일원오차성분모형에



서 DGME 추정법을 다루었다. 하지만, 실제 관측된 패널자료에서는 개체들 간의 개별적인 특성 외에도 시간의 흐름에 따른 변화도 반영될 수 있기 때문에 시간효과도 고려할 경우 좀 더 의미있는 결과를 도출할 가능성이 있다는 것은 의심의 여지가 없다. 따라서 본 연구에서는 Lee et al.(2016)에서 제안했던 중도절단 일원오차성분모형에서의 추정방법을 더 자세히 다루고, 이를 확장하여 중도절단 패널자료의 개체 및 시간 랜덤효과 오차성분을 가지는 이원오차성분모형에서의 이중 일반화 최대엔트로피(dual generalized maximum entropy, DGME) 추정량을 제안한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 DGME 추정량을 유도하기 위한 DGME 형식에 대해 서술한다. 3장에서는 본 연구에서 다루고자 하는 중도절단 패널회귀모형을 소개한다. 4장에서는 일원·이원오차성분을 가지는 패널모형에 적용 가능한 OLS, GLS, FGLS 추정량을 제시한다. 5장에서는 중도절단 패널모형에서 적용 가능한 DGME 추정량을 제시한다. 6장에서는 모의실험에서 구한 추정량별 RMSE를 통해 각 추정량의 효율성을 비교한다. 7장에서는 실제 데이터 분석에서 구한 추정량별 PRMSE를 통해 각 추정량의 효율성을 비교한다. 마지막으로 8장에서는 결론을 다룬다.



제 2 장 이중 일반화 최대엔트로피 추정량

2.1 일반화 최대엔트로피 형식

우선 Song, Cheon(2006)에 의해 소개되었던 패널모형에서의 일반화 최대엔트로피 방법에 대해 간략히 설명한다. Song, Cheon(2006)은 유한하고 이산형이고 선형적이며 오차항이 있는 일반선형모형을 고려하였으나, 여기서는 개체효과를 고려하는 일원오차성분모형을 기반으로 최대엔트로피 형식을 논한다.

$$y_{it} = X\beta + \mu_i + e_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (2.1)$$

여기서 y 는 $NT \times 1$ 인 종속변수벡터, X 는 $NT \times K$ 인 독립변수행렬, β 는 $K \times 1$ 인 회귀계수벡터, μ 는 $N \times 1$ 인 개체효과 e 는 $NT \times 1$ 인 오차항 벡터이다.

2.1.1 일원오차성분모형의 재모수화

알려지지 않은 모수 $\beta = (\beta_1, \beta_2, \dots, \beta_K)$ 와 개체효과 μ , 오차 e 가 제한된 사전정보와 비표본정보를 가졌다고 가정한다. 예를 들어 모수와 개체효과, 오차가 임의의 범위를 가지고 있다고 가정을 하면, 모수와 개체효과, 오차가 비표본정보를 가지는 사전 가중치나 유한 대응치(supports)를 가지는 이산확률변수로 구성되어있다고 할 수 있다. 따라서 다음과 같이 모수 β 와 개체효과 μ , 오차 e 를 재모수화(reparameterization)할 수 있다.

각 β_k 를 콤팩트 대응치(compact support) z 와 $M(2 \leq M < \infty)$



개의 가능한 결과치로 구성된 이산확률변수라고 가정한다. 만일 z_{k1} 과 z_{kM} 이 각각 모수 β_k 의 가능한 극한값들(상한값과 하한값)이라 하면 β_k 를 이러한 두 점의 볼록조합(convex combination)으로 표현할 수 있다. 즉, 다음과 같이 $M=2$ 에 대해 각 극한값에 확률을 할당한다.

$$\beta_k = p_k z_{k1} + (1-p_k)z_{kM}. \quad (2.2)$$

여기서 모수공간은 $\mathcal{L} = (\beta_1, \dots, \beta_K) \subset \mathbb{R}^K$ 에 의해 표현되며 $p_k \in [0, 1]$ 이다.

모수들을 일반적으로 표현하기 위하여 z_k 를 k 번째 차원까지 확장된 M 개의 점들의 집합이라고 하자. $\sum_{all\ m} p_{km} = 1$ 인 양의 가중치 M 차원 벡터가 주어졌을 때 k 번째 모수는 가중치 p_k 를 가지는 점 z_k 의 볼록 조합으로서 표현된다. 더 나아가 이러한 볼록 조합은 어떠한 $\beta \in \text{int}(\mathcal{L})$ 라도 다음과 같이 행렬로 표현할 수 있다.

$$\beta = Zp = \begin{bmatrix} z'_1 & 0 & \cdots & 0 \\ 0 & z'_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & z'_K \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{bmatrix}. \quad (2.3)$$

여기서 Z 는 $K \times KM$ 행렬, p 는 $KM \times 1$ 인 가중치벡터이다.

다음으로 개체효과 벡터 μ 를 유한한 위치모수와 척도모수들을 가지는 확률벡터라고 가정한다. 각 μ_i 를 $R(2 \leq R < \infty)$ 개의



가능한 결과치를 가지는 유한이산확률변수라고 하고, 만일 g_{i1} 과 g_{iR} 이 각각 개체효과 μ_i 의 가능한 극한값들(상한값과 하한 값)이라고 하면, μ_i 를 이러한 두 점의 볼록 조합으로 표현할 수 있다.

$$\mu_i = f_i g_{i1} + (1 - f_i) g_{iR}. \quad (2.4)$$

여기서 $f_i \in [0, 1]$ 이다.

개체효과들을 일반적으로 표현하기 위하여 다음과 같이 행렬의 형태로 표현할 수 있다.

$$\mu = Fg = \begin{bmatrix} f'_1 & 0 & \cdots & 0 \\ 0 & f'_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & f'_N \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_N \end{bmatrix}. \quad (2.5)$$

여기서 F 는 $N \times NR$ 행렬, g 는 $NR \times 1$ 인 가중치벡터이다.

마지막으로 오차벡터 $e = (e_{11}, \dots, e_{it}, \dots, e_{NT})$ 를 유한한 위치모수와 척도모수들을 가지는 확률벡터라고 가정한다. 각 e_{it} 를 $J(2 \leq J < \infty)$ 개의 가능한 결과치를 가지는 유한이산확률변수라고 하고, 만일 w_{it1} 과 w_{itJ} 가 각각 오차 e_{it} 의 가능한 극한값들(상한값과 하한값)이라고 하면, e_{it} 를 이러한 두 점의 볼록 조합으로 표현할 수 있다.

$$e_{it} = v_{it} w_{it1} + (1 - v_{it}) w_{itJ}. \quad (2.6)$$



여기서 $v_{it} \in [0, 1]$ 이다.

오차들을 일반적으로 표현하기 위하여 다음과 같이 행렬의 형태로 표현할 수 있다.

$$e = Vw = \begin{bmatrix} v'_{11} & 0 & \cdots & 0 \\ 0 & v'_{12} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & v'_{NT} \end{bmatrix} \begin{bmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{NT} \end{bmatrix}. \quad (2.7)$$

여기서 V 는 $NT \times NTJ$ 행렬, w 는 $NTJ \times 1$ 인 가중치벡터이다.

식 (2.3), (2.5), (2.7)에 의해 재모수화된 $\beta = Zp$, $\mu = Fg$, $e = Vw$ 를 사용함으로써 식 (2.1)을 다음과 같이 재정의 할 수 있다.

$$y = X\beta + \mu + e = XZp + Fg + Vw. \quad (2.8)$$

다음으로 Z 와 F , V 의 선택문제에 대해 알아본다. 먼저 GME 형식 하에서 Z 의 선택은 일치성과 정규화 가법성에 의해 변하지 않아야 한다. 하지만 F 와 V 는 명백히 ε 의 성질에 의존한다. 예를 들어 확률변수 x 가 평균은 0, 분산은 σ^2 을 가진다고 하자. 체비셰프 부등식(Chebychev's inequality)을 사용해서 임의의 $v > 0$ 에 대해 $\Pr[|x| < v\sigma] \geq v^{-2}$ 를 구할 수 있으며 체비셰프 오차의 하한 값과 상한 값은 각각 $v_{it1} = -v\sigma$ 와 $v_{itJ} = v\sigma$ 가 되어 V 로서 사용할 수 있다. 또한 3σ rule에 의해 V 를 선택할 수도 있다(Pukelsheim, 1994).



2.1.2 일반화최대엔트로피 추정량

일반화최대엔트로피(generalized maximum entropy, GME) 형식은 다음과 같다.

$$H(p, g, w) = -p' \ln p - g' \ln g - w' \ln w. \quad (2.9)$$

여기서 제약조건은 일치성 조건인 $y = XZp + Vw$ 와 정규화 가법성 조건인 $i_K = (I_K \otimes i'_M)p$, $i_N = (I_N \otimes i'_R)g$, $i_{NT} = (I_{NT} \otimes i'_J)w$ 이며, \otimes 는 크로네커곱(Kronecker product)을 나타낸다.

식 (2.9)을 최대화하는 \hat{p} , \hat{g} , \hat{w} 를 구하기 위해 라그랑지안 함수를 다음과 같이 정의한다.

$$\begin{aligned} L = & -p' \ln p - g' \ln g - w' \ln w + \lambda' [y - XZp - (I_N \otimes i_T)Fg - Vw] \\ & + \theta' [i_K - (I_K \otimes i'_M)p] + \zeta' [i_N - (I_N \otimes i'_R)g] \\ & + \tau' [i_{NT} - (I_{NT} \otimes i'_J)w]. \end{aligned} \quad (2.10)$$

여기서 I_K 는 차원이 K 인 단위행렬이고, i_K 는 모든 원소가 1로 이루어진 크기가 K 인 열벡터를 나타낸다.

식 (2.10)에서 p , g , w , λ , θ , ζ , τ 값에 대해 각각 1차 미분을 하면 다음의 결과들을 얻을 수 있다.

$$\nabla_p L = -\ln p - i_{KM} - Z'X'\lambda - (I_K \otimes i_M)\theta = 0, \quad (2.11)$$

$$\nabla_g L = -\ln g - i_{NR} - F'(I_N \otimes i'_T)'\lambda - (I_N \otimes i_R)\zeta = 0, \quad (2.12)$$



$$\nabla_w L = -\ln w - i_{NTJ} - V'\lambda - (I_{NT} \otimes i_J)\tau = 0, \quad (2.13)$$

$$\nabla_\lambda L = y - XZp - (I_N \otimes i_T)Fg - Vw = 0, \quad (2.14)$$

$$\nabla_\theta L = i_K - (I_K \otimes i'_M)p = 0, \quad (2.15)$$

$$\nabla_\zeta L = i_N - (I_N \otimes i'_R)g = 0, \quad (2.16)$$

$$\nabla_\tau L = i_{NT} - (I_{NT} \otimes i'_J)w = 0. \quad (2.17)$$

이 결과들을 연립하여 \hat{p} , \hat{g} , \hat{w} 추정량을 구해보면 다음과 같다.

$$\hat{p} = \exp(-Z'X'\lambda) \odot \exp\{-i_{KM} - (I_K \otimes i_M)\theta\}, \quad (2.18)$$

$$\hat{g} = \exp\{-F'(I_N \otimes i'_T)\lambda\} \odot \exp\{-i_{NR} - (I_N \otimes i_R)\zeta\}, \quad (2.19)$$

$$\hat{w} = \exp(-V'\lambda) \odot \exp\{-i_{NTJ} - (I_{NT} \otimes i_J)\tau\}. \quad (2.20)$$

또한 식 (2.11)에서 구한 p 를 식 (2.15)에 대입하여 정리하면 다음의 결과를 얻을 수 있다.

$$\exp\{-i_{KM} - (I_K \otimes i_M)\theta\} = \exp\{(I_K \otimes J_M)\exp(-Z'X'\lambda)\}^{\odot -1}. \quad (2.21)$$

따라서 \hat{p} 은 다음과 같이 재정의 될 수 있다.

$$\hat{p} = \exp(-Z'X'\lambda) \odot \exp\{(I_K \otimes J_M)\exp(-Z'X'\lambda)\}^{\odot -1}. \quad (2.22)$$

같은 원리에 의해 \hat{g} , \hat{w} 역시 다음과 같이 재정의 될 수 있다.

$$\hat{g} = \exp(-F'(I_N \otimes i'_T)\lambda) \odot \exp\{(I_N \otimes J_R)\exp(-F'(I_N \otimes i'_T)\lambda)\}^{\odot -1}. \quad (2.23)$$



$$\hat{w} = \exp(-V'\lambda) \odot \exp\{(I_{NT} \otimes J_J) \exp(-V'\lambda)\}^{\odot -1}. \quad (2.24)$$

여기서 \odot 는 하다마드곱(Hadamard product)이고 $\odot -1$ 는 하다마드 역(Hadamard inverse)이다.

마지막으로 모수 $\hat{\beta}_{GME}$ 를 추정하면 다음과 같다.

$$\begin{aligned} \hat{\beta}_{GME} &= Z\hat{p} \\ &= Z \cdot \exp(-Z'X'\lambda) \odot \exp\{(I_K \otimes J_M) \exp(-Z'X'\lambda)\}^{\odot -1}. \end{aligned} \quad (2.25)$$

2.2 이중 일반화 최대엔트로피형식

본 절에서는 Lee, Cheon(2014, 2015)에서 제안한 제한조건이 없는 이중 최대엔트로피 형식(DGME)에 대해 설명한다. 모형은 앞 절에서와 같이 일원오차성분모형을 기반으로 이중 일반화 최대엔트로피형식을 논한다.

우선 라그랑지안 함수인 식 (2.10)을 최대화 하면서, 추정된 가중치벡터 식 (2.22), (2.23), (2.24)를 만족하는 이중목적함수(dual objective function)를 다음과 같이 정의한다.

$$\begin{aligned} \mathcal{L} &= -p'\log(p) - g'\log(g) - w'\log(w) \\ &\quad + \lambda'[y - XZp - (I_N \otimes i_T)Fg - Vw] \\ &= -p'[-Z'X'\lambda - \log(\Omega(\lambda))] \\ &\quad - g'[-F'(I_N \otimes i_T)'\lambda - \log(\Phi(\lambda))] \\ &\quad - w'[-V'\lambda - \log(\Psi(\lambda))] \end{aligned}$$



$$\begin{aligned}
& + [y' - p'Z'X' - g'F'(I_N \otimes i_T)' - w'V']\lambda \\
= & y'\lambda + p'\log\{(I_N \otimes J_M)\exp(-Z'X'\lambda)\} \\
& + g'\log\{(I_N \otimes J_R)\exp(-F'(I_N \otimes i_T')\lambda)\} \\
& + w'\log\{(I_{NT} \otimes J_J)\exp(-V'\lambda)\} \\
= & y'\lambda + p'\log(\Omega(\lambda)) + g'\log(\Phi(\lambda)) + w'\log(\Psi(\lambda)). \quad (2.26)
\end{aligned}$$

식 (2.26)의 이중목적함수 \mathcal{L} 를 최소로 하는 λ (즉, $\hat{\lambda} = \min_{\lambda} \mathcal{L}$)가 p, g, w 의 최적의 해를 가짐으로, β 에 대한 DGME 추정량을 다음과 같이 구할 수 있다.

$$\hat{\beta}_{DGME} = Z\hat{p}. \quad (2.27)$$

여기서 $\hat{p} = \exp(-Z'X'\hat{\lambda}) \odot [(I_K \otimes J_M)\exp(-Z'X'\hat{\lambda})]^{\odot -1}$ 이다.



제 3 장 중도절단 패널회귀모형

횡단면단위와 시간단위가 병합된 패널자료에 대하여 다음의
패널회귀모형을 고려한다.

$$y_{it} = f_{it}(x_{it}, \beta_{it}, u_{it}), \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (3.1)$$

여기서 i 는 관측치의 횡단면 단위를 나타내고, t 는 시간을 나타내며 f_{it} 는 독립변수벡터 x_{it} , 모수벡터 β_{it} , 오차항 u_{it} 로써 종속변수 y_{it} 를 설명하는 횡단면-시간의 함수적 관계를 나타낸다. 물론 이러한 일반적인 패널모형은 관측치보다 많은 모수로 인하여 실제 적용이 어렵다. 그러므로 모형을 식별하고 효과적인 추론을 이끌어 내기 위해서는 가정 등을 통하여 위의 모형 (3.1)을 단순화하여야 한다(전수영, 2002).

단순화 작업은 일반적으로 고정된 개체효과를 제외하고 모든 개체에 있어서 회귀계수가 동일하다는 가정이 간단한 단계이다. 이는 각 개체에 대한 서로 다른 절편을 인정함으로써 이루어질 수 있다. 이러한 공분산 모형은 추정이 쉽고, 검정을 가능케 한다. 그러나 일반적인 개체효과는 직접 관측되지 않은 많은 개체효과들의 결과들일 것이다. 그러므로 고정개체효과보다도 랜덤개체효과를 취급하는 오차성분모형을 다루는 것이 더 자연스러울 것이다(Baltagi, 1995).

본 연구에서는 다음의 모형을 기반으로 중도 절단된 이원오차성분모형과 중도 절단된 이원오차성분모형 두 가지를 고려해 본다.



$$y_{it}^* = x_{it}'\beta + u_{it}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T. \quad (3.2)$$

여기서 y_{it}^* 는 i 번째 횡단면자료에 대한 t 번째 시계열자료의 종속변수 관측치이며, x_{it} 는 $K \times 1$ 의 설명변수벡터, β 는 $K \times 1$ 의 모수벡터이다.

모형 (3.2)에서 관측치 y_{it}^* 를 사용하여 새로운 중도 절단된 관측치 y_{it} 를 다음과 같이 정의한다.

$$y_{it} = \begin{cases} y_{it}^* & \text{if } y_{it}^* > 0, \\ 0 & \text{if } y_{it}^* \leq 0. \end{cases} \quad (3.3)$$

전통적인 오차가정 하에서 모형 (3.3)을 중도절단된 회귀모형(censored regression model) 또는 토빗모형(Tobit model)이라 한다(Tobin, 1958; Maddala, 1983; Amemiya, 1984).

3.1 중도절단 일원오차성분모형

모형 (3.3)에서 오차항 u_{it} 는 다음과 같은 일원오차성분모형(one-way error components model)을 갖는다고 가정하자.

$$u_{it} = \mu_i + e_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (3.4)$$

여기서 μ_i 는 관측될 수 없는 개체효과를 나타내는 확률변수이며, e_{it} 는 나머지 오차항을 나타낸다. μ_i , e_{it} 는 서로 독립이며 각각 $\mu_i \sim i.i.d.(0, \sigma_\mu^2)$, $e_{it} \sim i.i.d.(0, \sigma_e^2)$ 이라고 가정한다.



이와같이 오차항 u_{it} 가 일원오차성분을 갖는 경우 모형 (3.3)을 행렬을 이용하여 나타내면 다음과 같다.

$$\begin{aligned} y &= X\beta + u, \\ u &= (I_N \otimes i_T)\mu + e. \end{aligned} \quad (3.5)$$

모형 (3.5)에서 $u = (u_{11}, \dots, u_{1T}, \dots, u_{N1}, \dots, u_{NT})'$, $\mu = (\mu_1, \dots, \mu_N)'$, $e = (e_{11}, \dots, e_{1T}, \dots, e_{N1}, \dots, e_{NT})'$ 이고 y 는 $NT \times 1$ 인 종속변수벡터, X 는 $NT \times K$ 인 독립변수행렬, β 는 $K \times 1$ 인 회귀계수벡터, u 는 $NT \times 1$ 인 개체효과를 포함한 오차항 벡터, i_T 는 T 개의 1로 이뤄진 열벡터이다.

3.2 중도절단 이원오차성분모형

모형 (3.3)에서 오차항 u_{it} 는 다음과 같은 이원오차성분모형 (two-way error components model)을 갖는다고 가정하자.

$$u_{it} = \mu_i + \lambda_t + e_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (3.6)$$

여기서 μ_i 는 관측될 수 없는 개체효과를 나타내는 확률변수이고 λ_t 는 시간효과를 나타내는 확률변수, e_{it} 는 나머지 오차항을 나타낸다. μ_i , λ_t , e_{it} 는 서로 독립이며 각각 $\mu_i \sim i.i.d.(0, \sigma_\mu^2)$, $\lambda_t \sim i.i.d.(0, \sigma_\lambda^2)$, $e_{it} \sim i.i.d.(0, \sigma_e^2)$ 이라고 가정한다.

이와 같이 오차항 u_{it} 가 이원오차성분을 갖는 경우 모형



(3.3)을 행렬을 이용하여 나타내면 다음과 같다.

$$\begin{aligned} y &= X\beta + u, \\ u &= (I_N \otimes i_T)\mu + (i_N \otimes I_T)\lambda + e. \end{aligned} \quad (3.7)$$

모형 (3.7)에서 $u = (u_{11}, \dots, u_{1T}, \dots, u_{N1}, \dots, u_{NT})'$, $\mu = (\mu_1, \dots, \mu_N)'$, $\lambda = (\lambda_1, \dots, \lambda_T)'$, $e = (e_{11}, \dots, e_{1T}, \dots, e_{N1}, \dots, e_{NT})'$ 이고 y 는 $NT \times 1$ 인 종속변수벡터, X 는 $NT \times K$ 인 독립변수행렬, β 는 $K \times 1$ 인 회귀계수벡터, u 는 $NT \times 1$ 인 개체효과와 시간효과를 포함한 오차항 벡터, i_T 와 i_N 은 각각 T 개와 N 개의 1로 이뤄진 열벡터이다.



제 4 장 회귀계수에 대한 고전적 추정량

4.1 일원오차성분모형

본 장에서는 다양한 회귀계수 추정량들의 이해를 돕기 위해 전수영(2002), 박성현(2007, p.178, p.321), Teimour(2012)를 참조하여 간략히 정리한다.

4.1.1 OLS 추정량

일원오차성분모형 (3.5)에서 회귀계수 추정을 위한 보통최소제곱(ordinary least squares, OLS) 추정량은 다음과 같다.

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y. \quad (4.1)$$

이는 오로지 최소제곱법에 의해 오차제곱합을 최소화시키기 위한 과정에서 유도된 식으로, 패널모형의 회귀계수 추정시 효율성이 떨어질 수 있다.

4.1.2 GLS 추정량

일원오차성분모형 (3.5)에서 회귀계수 추정을 위한 일반화최소제곱(generalized least squares, GLS) 추정량은 다음과 같다.

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y. \quad (4.2)$$



이는 오차들이 이분산($\text{Var}(\epsilon_i) = \sigma_i^2$)일 경우 최량선형불편추정량(best linear unbiased estimator, BLUE)인 최소제곱추정량을 구한 식이다. 여기서 오차항 $u = (I_N \otimes i_T)\mu + e$ 에 대한 분산 공분산행렬은 $\Omega = E(uu') = \sigma_\mu^2(I_N \otimes J_T) + \sigma_e^2(I_N \otimes I_T)$ 이다.

식 (4.2)를 구하기 위해서는 Ω 를 다음과 같이 분광분해(spectral decomposition)하여 Ω^{-1} 을 구해야 한다.

$$\begin{aligned}\Omega &= T\sigma_\mu^2(I_N \otimes \overline{J_T}) + \sigma_e^2(I_N \otimes E_T) + \sigma_e^2(I_N \otimes \overline{J_T}) \\ &= (T\sigma_\mu^2 + \sigma_e^2)(I_N \otimes \overline{J_T}) + \sigma_e^2(I_N \otimes E_T) \\ &= \sigma_1^2 P + \sigma_e^2 Q.\end{aligned}$$

여기서 $\overline{J_T} = J_T/T$, $E_T = I_T - \overline{J_T}$, $\sigma_1^2 = T\sigma_\mu^2 + \sigma_e^2$, $P = I_N \otimes \overline{J_T}$, $Q = I_N \otimes E_T$ 이다. 이를 이용하면 다음과 같이 Ω^{-1} 을 구할 수 있다.

$$\Omega^{-1} = \frac{1}{\sigma_1^2} P + \frac{1}{\sigma_e^2} Q. \quad (4.3)$$

이때 $\Omega = (\Omega^{1/2})(\Omega^{1/2}) = (\Omega^{1/2})(\Omega^{1/2})'$ 이 성립하므로 $\Omega^{-1/2}$ 을 식 (3.5)의 양변에 곱하면 $(\Omega^{-1/2})y = z$, $(\Omega^{-1/2})X\beta = W\beta$, $(\Omega^{-1/2})u = v$ 를 만족하는 다음의 식이 정의된다.

$$z = W\beta + v. \quad (4.4)$$

식 (4.4)의 벡터 v 의 분산-공분산 행렬은 $\text{Var}(v) = I\sigma^2$ 을 만족하므로 일원오차성분모형에 대한 GLS 추정량은 식 (4.3)을



이용해 다음과 같이 구해진다.

$$\begin{aligned}\hat{\beta}_{GLS} &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \\ &= \left[X' \left(\frac{1}{\sigma_1^2} P + \frac{1}{\sigma_e^2} Q \right) X \right]^{-1} X' \left(\frac{1}{\sigma_1^2} P + \frac{1}{\sigma_e^2} Q \right) y.\end{aligned}\quad (4.5)$$

그러나 식 (4.5)는 σ_μ^2 와 σ_e^2 이 알려지지 않은 경우 이론적인 값에 불과하다. 따라서 GLS 추정량은 6장 모의실험에서 다른 추정량들과 함께 비교를 위해 사용되지만, 7장 실 자료 분석에서는 사용되지 않는다.

4.1.3 FGLS 추정량

실 자료 분석 시에는 분산성분들이 추정되어야 GLS 추정량을 구할 수 있다. 이렇게 추정된 분산성분들을 통해 추정되는 GLS 추정량을 실현가능한 일반화최소제곱(Feasible GLS) 추정량이라 한다.

분산성분을 추정하는 방법에는 여러 가지가 있는데, 여기서는 Song, Cheon(2006)에서 소개된 Amemiya(1971)의 분산성분 추정방법을 사용하여 FGLS(AM) 추정량을 구하고자 한다.

Amemiya(1971)는 분산성분에 대한 추정량을 구하기 위해 개체효과를 제거시킨 최소제곱더미변수(least square dummy variables)잔차를 사용하였다. 이 경우 얻어지는 회귀계수에 대한 추정량은 다음과 같다.



$$\begin{aligned}\hat{\beta}_{AM} &= (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y \\ &= \left[X' \left(\frac{1}{\hat{\sigma}_1^2} P + \frac{1}{\hat{\sigma}_e^2} Q \right) X \right]^{-1} X' \left(\frac{1}{\hat{\sigma}_1^2} P + \frac{1}{\hat{\sigma}_e^2} Q \right) y.\end{aligned}\quad (4.6)$$

여기서 추정된 분산성분 $\hat{\sigma}_1^2$ 과 $\hat{\sigma}_e^2$ 는 다음과 같다.

$$\begin{aligned}\hat{\sigma}_1^2 &= \frac{\hat{u}' P \hat{u}}{tr(P)} = T \sum_{i=1}^N \frac{\hat{u}_i}{N}. \\ \hat{\sigma}_e^2 &= \frac{\hat{u}' Q \hat{u}}{tr(Q)} = \sum_{i=1}^N \sum_{t=1}^T \frac{(u - \hat{u}_i)}{N(T-1)}, \quad \hat{u} = \hat{u}_{LSDV}.\end{aligned}$$

여기서 소개된 AM 추정량 식 (4.6)은 6장의 모의실험 및 7장 실 자료 분석에서 사용된다.

4.2 이원오차성분모형

4.2.1 OLS 추정량

이원오차성분모형 (3.7)에서 회귀계수 추정을 위한 OLS 추정량은 다음과 같다.

$$\hat{\beta}_{OLS} = (X' X)^{-1} X' y. \quad (4.7)$$

4.2.2 GLS 추정량

이원오차성분모형 (3.7)에서 회귀계수 추정을 위한 GLS 추정량은 다음과 같다.



$$\hat{\beta}_{GLS} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y. \quad (4.8)$$

여기서 $\Omega = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\lambda^2(J_N \otimes I_T) + \sigma_e^2(I_N \otimes I_T)$ 이다.

식 (4.8)을 구하기 위해서는 Ω^{-1} 을 계산해야 하는데, Ω 에서 J_N 을 $N\overline{J_N}$ 으로, I_N 을 $E_N + \overline{J_N}$ 으로, J_T 를 $T\overline{J_T}$ 으로, I_T 를 $E_T + \overline{J_T}$ 으로($\overline{J_N} = J_N/N$, $E_N = I_N - \overline{J_N}$) 치환하고 같은 항끼리 정리하면 식 (4.8)의 분산-공분산 행렬은 다음과 같이 재정의 된다.

$$\begin{aligned} \Omega &= \sigma_e^2(E_N \otimes E_T) + (T\sigma_\mu^2 + \sigma_e^2)(E_N \otimes \overline{J_T}) \\ &\quad + (N\sigma_\lambda^2 + \sigma_e^2)(\overline{J_N} \otimes E_T) + (T\sigma_\mu^2 + N\sigma_\lambda^2 + \sigma_e^2)(\overline{J_N} \otimes \overline{J_T}) \\ &= \lambda_1 Q_1 + \lambda_2 Q_2 + \lambda_3 Q_3 + \lambda_4 Q_4 = \sum_{i=1}^4 \lambda_i Q_i. \end{aligned} \quad (4.9)$$

또한 $Q_i (i=1,2,3,4)$ 는 멱등행렬($Q_i Q_i = Q_i$)이며 서로 직교하는 행렬이므로 이러한 Q_i 행렬의 성질을 이용하면 Ω^{-1} 는 다음과 같이 구해진다.

$$\Omega^{-1} = \sum_{i=1}^4 \frac{1}{\lambda_i} Q_i. \quad (4.10)$$

따라서 이원오차성분모형에 대한 GLS 추정량은 다음과 같이 구해진다.

$$\hat{\beta}_{GLS} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y. \quad (4.11)$$



4.1.2절에서 서술한 바와 같이 식 (4.11)역시 σ_μ^2 와 σ_λ^2 , σ_ϵ^2 이 알려지지 않은 경우 이론적인 값에 불과하다. 따라서 GLS 추정량은 6장 모의실험에서 다른 추정량들과 함께 비교를 위해 사용되지만, 7장 실 자료 분석에서는 사용되지 않는다.

4.2.3 FGLS 추정량

이원오차성분모형의 분산성분 추정 역시 Song, Cheon(2006)에서 소개된 Amemiya(1971)의 분산성분 추정방법을 사용하여 FGLS(AM) 추정량을 구하고자 한다.

Amemiya(1971)는 분산성분에 대한 추정량을 구하기 위해 개체효과와 시간효과를 모두 제거시키는 Q_1 변환을 통해 얻어지는 내부변환잔차를 사용하였다. 이 경우 Q_1 변환을 통해 얻어지는 회귀계수에 대한 AM 추정량은 다음과 같다.

$$\hat{\beta}_{AM} = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y. \quad (4.12)$$

여기서 $\hat{\Omega}^{-1} = \sum_{i=1}^4 \hat{\lambda}_i^{-1} Q_i$, $\hat{\lambda}_i = \frac{\hat{u}' Q_i \hat{u}}{tr(Q_i)}$, $\hat{u} = y - X\hat{\beta}_{OLS}$ 이며

$\hat{\lambda}_4 = \hat{\lambda}_2 + \hat{\lambda}_3 - \hat{\lambda}_1$ 이다.

여기서 소개된 AM 추정량은 6장의 모의실험 및 7장 실 자료 분석에서 사용된다.



제 5 장 회귀계수에 대한 DGME 추정량

5.1 중도절단 일원오차성분모형

DGME 추정량을 구하기 위해 모수(β)와 개체효과(μ), 그리고 오차항(e)을 재모수화 한다.

우선, 모수를 재모수화하면 다음과 같다.

$$\beta = Zp = \begin{bmatrix} z'_1 & 0 & \cdots & 0 \\ 0 & z'_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & z'_K \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{bmatrix}. \quad (5.1)$$

여기서 Z 는 $K \times KM$ 행렬, p 는 $KM \times 1$ 벡터, $z_k = (z_{k1}, \dots, z_{kM})'$, $p_k = (p_{k1}, \dots, p_{kM})'$, 모든 $k(k=1, \dots, K)$ 에 대해 $\beta_k = \sum_{m=1}^M z_{km} p_{km}$ 이다.

다음으로 개체효과를 재모수화하면 다음과 같다.

$$\mu = Fg = \begin{bmatrix} f'_1 & 0 & \cdots & 0 \\ 0 & f'_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & f'_N \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_N \end{bmatrix}. \quad (5.2)$$

여기서 F 는 $N \times NR$ 행렬, g 는 $NR \times 1$ 벡터, $f_i = (f_{i1}, \dots, f_{iR})'$, $g_i = (g_{i1}, \dots, g_{iR})'$, 모든 $i(i=1, \dots, N)$ 에 대해 $\mu_i = \sum_{r=1}^R f_{ir} g_{ir}$ 이다.

마지막으로 나머지 오차항을 재모수화하면 다음과 같다.



$$e = Vw = \begin{bmatrix} v'_{11} & 0 & \cdots & 0 \\ 0 & v'_{12} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & v'_{NT} \end{bmatrix} \begin{bmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{NT} \end{bmatrix}. \quad (5.3)$$

여기서 V 는 $NT \times NTJ$ 행렬, w 는 $NTJ \times 1$ 벡터, $v_{it} = (v_{it1}, \dots, v_{itJ})'$, $w_{it} = (w_{it1}, \dots, w_{itJ})'$, 모든 $i (i = 1, \dots, N)$ 와 $t (t = 1, \dots, T)$ 에 대해 $e_{it} = \sum_{j=1}^J v_{itj} w_{itj}$ 이다.

따라서 모형 (3.5)를 다음과 같이 재모수화 할 수 있다.

$$y = X\beta + u = XZp + (I_N \otimes i_T)Fg + Vw. \quad (5.4)$$

모형 (5.4)는 중도절단된 자료에 대한 패널회귀모형이므로, 정상적으로 관측된 관측치를 y_1 으로, 중도 절단된 관측치를 y_2 로 표현하면 다음과 같이 재표현이 가능하다.

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 = 0 \end{bmatrix} &= \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\ &= \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} Zp + \begin{bmatrix} (I_N \otimes i_{T_1}) \\ (I_N \otimes i_{T_2}) \end{bmatrix} Fg + \begin{bmatrix} V_1 w_1 \\ V_2 w_2 \end{bmatrix}. \end{aligned} \quad (5.5)$$

여기서 y_1 과 y_2 는 각각 $NT_1 \times 1$ 과 $NT_2 \times 1$ 인 종속변수벡터, X_1 과 X_2 는 각각 $NT_1 \times K$ 와 $NT_2 \times K$ 인 독립변수행렬, β 는 $K \times 1$ 인 회귀계수벡터, V_1 과 V_2 는 각각 $NT_1 \times NT_1J$ 와 $NT_2 \times NT_2J$



인 행렬, w_1 과 w_2 는 각각 $NT_1J \times 1$ 과 $NT_2J \times 1$ 인 벡터이다. 또한 $T = T_1 + T_2$ 라 한다.

모형 (5.5)에 대한 일반화 최대 엔트로피 형식은

$$H(p, g, w_1, w_2) = -p' \log(p) - g' \log(g) - w_1' \log(w_1) - w_2' \log(w_2). \quad (5.6)$$

을 최대로 하는 p, g, w_1, w_2 를 산출하는 문제로 귀결된다. 여기서 다음과 같은 두 가지 제약조건을 갖는다.

첫 번째 제약조건은 일치성으로 다음과 같다.

$$\begin{aligned} y_1 &= X_1 Z p + (I_N \otimes i_{T_1}) F g + V_1 w_1, \\ y_2 &= X_2 Z p + (I_N \otimes i_{T_2}) F g + V_2 w_2. \end{aligned}$$

두 번째 제약조건은 정규화 가법성으로 다음과 같다.

$$\begin{aligned} i_K &= (I_K \otimes i_M') p, \quad i_N = (I_N \otimes i_R') g, \\ i_{NT_1} &= (I_{NT_1} \otimes i_J') w_1, \quad i_{NT_2} = (I_{NT_2} \otimes i_J') w_2. \end{aligned}$$

식 (5.6)에서 두 가지 제약조건을 만족하는 p, g, w_1, w_2 의 해를 구하기 위해, 라그랑지안 함수를 다음과 같이 정의한다.

$$\begin{aligned} L &= -p' \log(p) - g' \log(g) - w_1' \log(w_1) - w_2' \log(w_2) \\ &\quad + \lambda_1' [y_1 - X_1 Z p - (I_N \otimes i_{T_1}) F g - V_1 w_1] \end{aligned}$$



$$\begin{aligned}
& + \lambda'_2 [y_2 - X_2 Z p - (I_N \otimes i_{T_2}) F g - V_2 w_2] \\
& + \theta' [i_K - (I_K \otimes i'_M) p] + \zeta' [i_N - (I_K \otimes i'_R) g] \\
& + \gamma' [i_{NT_1} - (I_{NT_1} \otimes i'_J) w_1] + \tau' [i_{NT_2} - (I_{NT_2} \otimes i'_J) w_2].
\end{aligned}$$

라그랑지안 함수 L 를 최대화 하는 모수를 구하기 위해 $p, g, w_1, w_2, \lambda_1, \lambda_2, \theta, \zeta, \gamma, \tau$ 에 대하여 각각 1차 미분한 뒤 0이라 하여 방정식들을 풀면 p, g, w_1, w_2 의 해를 다음과 같이 구할 수 있다.

$$p = \exp(-Z X'_1 \lambda_1 - Z X'_2 \lambda_2) \odot \exp[-i_{KM} - (I_K \otimes i_M) \theta], \quad (5.7)$$

$$g = \exp(-F'(I_N \otimes i_{T_1})' \lambda_1 - F'(I_N \otimes i_{T_2})' \lambda_2) \odot \exp[-i_{NR} - (I_N \otimes i_R) \zeta], \quad (5.8)$$

$$w_1 = \exp(-V'_1 \lambda_1) \odot \exp[-i_{NT_1 J} - (I_{NT_1} \otimes i_J) \gamma], \quad (5.9)$$

$$w_2 = \exp(-V'_2 \lambda_2) \odot \exp[-i_{NT_2 J} - (I_{NT_2} \otimes i_J) \tau]. \quad (5.10)$$

여기서 \odot 는 하다마드 곱(Hadamard product)이다. 또한 식 (5.7)에서 $\exp[-i_{KM} - (I_K \otimes i_M) \theta] = [(I_K \otimes J_M) \exp(-Z X'_1 \lambda_1 - Z X'_2 \lambda_2)]^{\odot(-1)}$ 이므로 다음과 같이 재표현이 가능하다.

$$\begin{aligned}
p &= \exp(-Z X'_1 \lambda_1 - Z X'_2 \lambda_2) \odot [(I_K \otimes J_M) \exp(-Z X'_1 \lambda_1 - Z X'_2 \lambda_2)]^{\odot(-1)} \\
&= \exp(-Z X' \lambda) \odot [(I_K \otimes J_M) \exp(-Z X' \lambda)]^{\odot(-1)}. \quad (5.11)
\end{aligned}$$

여기서 $\lambda = (\lambda_1, \lambda_2)'$ 이고 $\odot(-1)$ 는 하다마드 역(Hadamard inverse)이다. 같은 방식으로 식 (5.8), (5.9), (5.10)은 각각 다음과 같이 재표현 된다.

$$g = \exp(-F'(I_N \otimes i_T)' \lambda) \odot [(I_N \otimes J_R) \exp(-F'(I_N \otimes i_T)' \lambda)]^{\odot(-1)}, \quad (5.12)$$



$$w_1 = \exp(-V_1\lambda_1) \odot \exp[(I_{NT_1} \otimes J_J) \exp(-V_1\lambda_1)]^{\odot(-1)}, \quad (5.13)$$

$$w_2 = \exp(-V_2\lambda_2) \odot \exp[(I_{NT_2} \otimes J_J) \exp(-V_2\lambda_2)]^{\odot(-1)}. \quad (5.14)$$

라그랑지안 함수 L 과 p, g, w_1, w_2 의 해인 식 (5.11), (5.12), (5.13), (5.14)는 정규화가법성을 만족하므로, λ 의 함수로 이중 목적함수 \mathcal{L} 을 정의할 수 있다.

$$\begin{aligned} \mathcal{L} &= -p' \log(p) - g' \log(g) - w_1' \log(w_1) - w_2' \log(w_2) \\ &\quad + \lambda_1' [y_1 - X_1 Z p - (I_N \otimes i_{T_1}) F g - V_1 w_1] \\ &\quad + \lambda_2' [y_2 - X_2 Z p - (I_N \otimes i_{T_2}) F g - V_2 w_2] \\ &= -p' [-Z' X' \lambda - \log(\Omega(\lambda))] - g' [-F' (I_N \otimes i_T)' \lambda - \log(\Phi(\lambda))] \\ &\quad - w_1' [-V_1' \lambda_1 - \log(\Psi(\lambda_1))] - w_2' [-V_2' \lambda_2 - \log(\Psi(\lambda_2))] \\ &\quad + [y_1' - p' Z' X_1' - g' F' (I_N \otimes i_{T_1})' - w_1' V_1'] \lambda_1 \\ &\quad + [y_2' - p' Z' X_2' - g' F' (I_N \otimes i_{T_2})' - w_2' V_2'] \lambda_2 \\ &= y' \lambda + p' \log(\Omega(\lambda)) + g' \log(\Phi(\lambda)) + w' \log(\Psi(\lambda)). \end{aligned} \quad (5.15)$$

여기서 $\omega = (\omega_1, \omega_2)'$ 이고, $\Omega(\lambda) = (I_K \otimes J_M) \exp(-Z' X' \lambda)$, $\Phi(\lambda) = (I_N \otimes J_R) \exp(-F' (I_N \otimes i_T)' \lambda)$, $\Psi(\lambda_1) = (I_{NT_1} \otimes J_J) \exp(-V_1' \lambda_1)$, $\Psi(\lambda_2) = (I_{NT_2} \otimes J_J) \exp(-V_2' \lambda_2)$ 이다.

식 (5.15)의 이중목적함수 \mathcal{L} 를 최소로 하는 λ (즉, $\hat{\lambda} = \min_{\lambda} \mathcal{L}$)가 p, g, w_1, w_2 의 최적의 해를 가짐으로, β 에 대한 DGME 추정량을 다음과 같이 구할 수 있다.



$$\hat{\beta}_{DGME} = Z\hat{p}.$$

여기서 $\hat{p} = \exp(-Z'X'\hat{\lambda}) \odot [(I_K \otimes J_M) \exp(-Z'X'\hat{\lambda})]^{\odot(-1)}$ 이다.

5.2 중도절단 이원오차성분모형

DGME 추정량을 구하기 위해 모수(β)와 개체효과(μ), 시간 효과(λ) 그리고 오차항(e)을 재모수화 한다.

우선, 모수를 재모수화하면 다음과 같다.

$$\beta = Zp = \begin{bmatrix} z'_1 & 0 & \cdots & 0 \\ 0 & z'_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & z'_K \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{bmatrix}. \quad (5.16)$$

여기서 Z 는 $K \times KM$ 행렬, p 는 $KM \times 1$ 벡터, $z_k = (z_{k1}, \dots, z_{kM})'$,

$p_k = (p_{k1}, \dots, p_{kM})'$, 모든 $k(k=1, \dots, K)$ 에 대해 $\beta_k = \sum_{m=1}^M z_{km} p_{km}$

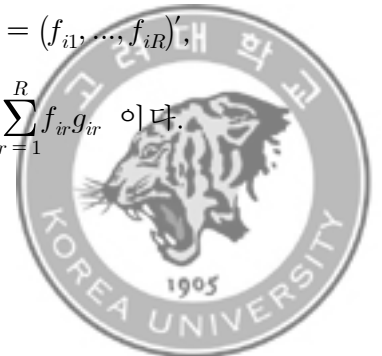
이다.

다음으로 개체효과를 재모수화하면 다음과 같다.

$$\mu = Fg = \begin{bmatrix} f'_1 & 0 & \cdots & 0 \\ 0 & f'_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & f'_N \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_N \end{bmatrix}. \quad (5.17)$$

여기서 F 는 $N \times NR$ 행렬, g 는 $NR \times 1$ 벡터, $f_i = (f_{i1}, \dots, f_{iR})'$,

$g_i = (g_{i1}, \dots, g_{iR})'$, 모든 $i(i=1, \dots, N)$ 에 대해 $\mu_i = \sum_{r=1}^R f_{ir} g_{ir}$ 이다.



다음으로 시간효과를 재모수화하면 다음과 같다.

$$\lambda = Ab = \begin{bmatrix} a'_1 & 0 & \cdots & 0 \\ 0 & a'_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a'_T \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_T \end{bmatrix}. \quad (5.18)$$

여기서 A 는 $T \times TD$ 행렬, b 는 $TD \times 1$ 벡터, $a_t = (a_{t1}, \dots, a_{tD})'$, $b_t = (b_{t1}, \dots, b_{tD})'$, 모든 $t(t=1, \dots, T)$ 에 대해 $\lambda_t = \sum_{d=1}^D a_{td} b_{td}$ 이다.

마지막으로 나머지 오차항을 재모수화하면 다음과 같다.

$$e = Vw = \begin{bmatrix} v'_{11} & 0 & \cdots & 0 \\ 0 & v'_{12} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & v'_{NT} \end{bmatrix} \begin{bmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{NT} \end{bmatrix}. \quad (5.19)$$

여기서 V 는 $NT \times NTJ$ 행렬, w 는 $NTJ \times 1$ 벡터, $v_{it} = (f_{it1}, \dots, f_{itJ})'$, $w_{it} = (w_{it1}, \dots, w_{itJ})'$, 모든 $i(i=1, \dots, N)$ 와 $t(t=1, \dots, T)$ 에 대해 $e_{it} = \sum_{j=1}^J v_{itj} w_{itj}$ 이다.

따라서 모형 (3.7)을 다음과 같이 재모수화 할 수 있다.

$$y = X\beta + u = XZp + (I_N \otimes i_T)Fg + (i_N \otimes I_T)Ab + Vw. \quad (5.20)$$

모형 (5.20)은 중도절단된 자료에 대한 패널회귀모형이므로, 정상적으로 관측된 관측치를 y_1 으로, 중도 절단된 관측치를 y_2 로 표현하면 다음과 같이 재표현이 가능하다.



$$\begin{bmatrix} y_1 \\ y_2 = 0 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} =$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} Zp + \begin{bmatrix} (I_N \otimes i_{T_1}) \\ (I_N \otimes i_{T_2}) \end{bmatrix} Fg + \begin{bmatrix} (i_N \otimes I_{T_1}) \\ (i_N \otimes I_{T_2}) \end{bmatrix} Ab + \begin{bmatrix} V_1 w_1 \\ V_2 w_2 \end{bmatrix}. \quad (5.21)$$

여기서 y_1 과 y_2 는 각각 $NT_1 \times 1$ 과 $NT_2 \times 1$ 인 종속변수벡터, X_1 과 X_2 는 각각 $NT_1 \times K$ 와 $NT_2 \times K$ 인 독립변수행렬, β 는 $K \times 1$ 인 회귀계수벡터, V_1 과 V_2 는 각각 $NT_1 \times NT_1 J$ 와 $NT_2 \times NT_2 J$ 인 행렬, w_1 과 w_2 는 각각 $NT_1 J \times 1$ 과 $NT_2 J \times 1$ 인 벡터이다. 또한 $T = T_1 + T_2$ 라 한다.

모형 (5.21)에 대한 일반화 최대 엔트로피 형식은

$$H(p, g, b, w_1, w_2) = -p' \log(p) - g' \log(g) - b' \log(b)$$

$$- w_1' \log(w_1) - w_2' \log(w_2). \quad (5.22)$$

을 최대로 하는 p, g, b, w_1, w_2 를 산출하는 문제로归结된다.

여기서 다음과 같은 두 가지 제약조건을 갖는다.

첫 번째 제약조건은 일치성으로 다음과 같다.

$$y_1 = X_1 Zp + (I_N \otimes i_{T_1}) Fg + (i_N \otimes I_{T_1}) Ab + V_1 w_1,$$

$$y_2 = X_2 Zp + (I_N \otimes i_{T_2}) Fg + (i_N \otimes I_{T_2}) Ab + V_2 w_2.$$

두 번째 제약조건은 정규화 가법성으로 다음과 같다.



$$i_K = (I_K \otimes i'_M)p, \quad i_N = (I_N \otimes i'_R)g, \quad i_T = (I_T \otimes i'_D)b,$$

$$i_{NT_1} = (I_{NT_1} \otimes i'_J)w_1, \quad i_{NT_2} = (I_{NT_2} \otimes i'_J)w_2.$$

식 (5.22)에서 두 가지 제약조건을 만족하는 p, g, b, w_1, w_2 의 해를 구하기 위해, 라그랑지안 함수를 다음과 같이 정의한다.

$$\begin{aligned} L = & -p' \log(p) - g' \log(g) - b' \log(b) - w_1' \log(w_1) - w_2' \log(w_2) \\ & + \lambda'_1 [y_1 - X_1 Z p - (I_N \otimes i_{T_1}) F g - (i_N \otimes I_{T_1}) A b - V_1 w_1] \\ & + \lambda'_2 [y_2 - X_2 Z p - (I_N \otimes i_{T_2}) F g - (i_N \otimes I_{T_2}) A b - V_2 w_2] \\ & + \theta' [i_K - (I_K \otimes i'_M)p] + \zeta' [i_N - (I_K \otimes i'_R)g] \\ & + \eta' [i_T - (I_T \otimes i'_D)b] + \gamma' [i_{NT_1} - (I_{NT_1} \otimes i'_J)w_1] \\ & + \tau' [i_{NT_2} - (I_{NT_2} \otimes i'_J)w_2]. \end{aligned}$$

라그랑지안 함수 L 를 최대화 하는 모수를 구하기 위해 $p, g, b, w_1, w_2, \lambda_1, \lambda_2, \theta, \zeta, \eta, \gamma, \tau$ 에 대하여 각각 1차 미분한 뒤 0이라 하여 방정식들을 풀면 p, g, b, w_1, w_2 의 해를 다음과 같이 구할 수 있다.

$$p = \exp(-Z X'_1 \lambda_1 - Z X'_2 \lambda_2) \odot \exp[-i_{KM} - (I_K \otimes i'_M)\theta], \quad (5.23)$$

$$g = \exp(-F'(I_N \otimes i_{T_1})' \lambda_1 - F'(I_N \otimes i_{T_2})' \lambda_2) \odot \exp[-i_{NR} - (I_N \otimes i'_R)\zeta], \quad (5.24)$$

$$b = \exp(-A'(i_N \otimes I_{T_1})' \lambda_1 - A'(i_N \otimes I_{T_2})' \lambda_2) \odot \exp[-i_{TD} - (I_T \otimes i'_D)\eta], \quad (5.25)$$

$$w_1 = \exp(-V'_1 \lambda_1) \odot \exp[-i_{NT_1J} - (I_{NT_1} \otimes i'_J)\gamma]. \quad (5.26)$$

$$w_2 = \exp(-V'_2 \lambda_2) \odot \exp[-i_{NT_2J} - (I_{NT_2} \otimes i'_J)\tau]. \quad (5.27)$$



여기서 \odot 는 하다마드 곱(Hadamard product)이다. 또한 식 (5.23)에서 $\exp[-i_{KM} - (I_K \otimes i_M)\theta] = [(I_K \otimes J_M)\exp(-Z X'_1 \lambda_1 - Z X'_2 \lambda_2)]^{\odot(-1)}$ 이므로 다음과 같이 재표현이 가능하다.

$$\begin{aligned} p &= \exp(-Z X'_1 \lambda_1 - Z X'_2 \lambda_2) \odot [(I_K \otimes J_M)\exp(-Z X'_1 \lambda_1 - Z X'_2 \lambda_2)]^{\odot(-1)} \\ &= \exp(-Z X' \lambda) \odot [(I_K \otimes J_M)\exp(-Z X' \lambda)]^{\odot(-1)}. \end{aligned} \quad (5.28)$$

여기서 $\lambda = (\lambda_1, \lambda_2)'$ 이고 $\odot(-1)$ 는 하다마드 역(Hadamard inverse)이다. 같은 방식으로 식 (5.24), (5.25), (5.26), (5.27)은 각각 다음과 같이 된다.

$$g = \exp(-F'(I_N \otimes i_T)' \lambda) \odot [(I_N \otimes J_R)\exp(-F'(I_N \otimes i_T)' \lambda)]^{\odot(-1)}, \quad (5.29)$$

$$b = \exp(-A'(i_N \otimes I_T)' \lambda) \odot [(I_T \otimes J_D)\exp(-A'(i_N \otimes I_T)' \lambda)]^{\odot(-1)}, \quad (5.30)$$

$$w_1 = \exp(-V'_1 \lambda_1) \odot \exp[(I_{NT_1} \otimes J_J)\exp(-V'_1 \lambda_1)]^{\odot(-1)}, \quad (5.31)$$

$$w_2 = \exp(-V'_2 \lambda_2) \odot \exp[(I_{NT_2} \otimes J_J)\exp(-V'_2 \lambda_2)]^{\odot(-1)}. \quad (5.32)$$

라그랑지안 함수 L 과 p, g, b, w_1, w_2 의 해인 식 (5.28), (5.29), (5.30), (5.31), (5.32)는 정규화가법성을 만족하므로, λ 의 함수로 이중목적함수 \mathcal{L} 을 정의할 수 있다.

$$\begin{aligned} \mathcal{L} &= -p' \log(p) - g' \log(g) - b' \log(b) - w'_1 \log(w_1) - w'_2 \log(w_2) \\ &\quad + \lambda'_1 [y_1 - X_1 Z p - (I_N \otimes i_{T_1}) F g - (i_N \otimes I_{T_1}) A b - V_1 w_1] \\ &\quad + \lambda'_2 [y_2 - X_2 Z p - (I_N \otimes i_{T_2}) F g - (i_N \otimes I_{T_2}) A b - V_2 w_2] \\ &= -p' [-Z X' \lambda - \log(\Omega(\lambda))] - g' [-F'(I_N \otimes i_T)' \lambda - \log(\Phi(\lambda))] \\ &\quad - b' [-A'(i_N \otimes I_T)' \lambda - \log(\Gamma(\lambda))] \end{aligned}$$



$$\begin{aligned}
& -w_1'[-V_1'\lambda_1 - \log(\Psi(\lambda_1))] - w_2'[-V_2'\lambda_2 - \log(\Psi(\lambda_2))] \\
& + [y_1' - p'Z'X_1' - g'F'(I_N \otimes i_{T_1})' - b'A'(i_N \otimes I_{T_1}) - w_1'V_1']\lambda_1 \\
& + [y_2' - p'Z'X_2' - g'F'(I_N \otimes i_{T_2})' - b'A'(i_N \otimes I_{T_2}) - w_2'V_2']\lambda_2 \\
& = y'\lambda + p'\log(\Omega(\lambda)) + g'\log(\Phi(\lambda)) + b'\log(\Gamma(\lambda)) + w'\log(\Psi(\lambda)). \quad (5.33)
\end{aligned}$$

여기서 $\omega = (\omega_1, \omega_2)'$ 이고, $\Omega(\lambda) = (I_K \otimes J_M)\exp(-Z'X'\lambda)$,
 $\Phi(\lambda) = (I_N \otimes J_R)\exp(-F'(I_N \otimes i_T')\lambda)$, $\Gamma(\lambda) = (I_T \otimes J_D)\exp(-A'(i_N \otimes I_T')\lambda)$,
 $\Psi(\lambda_1) = (I_{T_{N_1}} \otimes J_J)\exp(-V_1'\lambda_1)$, $\Psi(\lambda_2) = (I_{T_{N_2}} \otimes J_J)\exp(-V_2'\lambda_2)$ 이다.

식 (5.33)의 이중목적함수 \mathcal{L} 를 최소로 하는 λ (즉, $\hat{\lambda} = \min_{\lambda} \mathcal{L}$)
가 p, g, b, w_1, w_2 의 최적의 해를 가짐으로, β 에 대한 DGME
추정량을 다음과 같이 구할 수 있다.

$$\hat{\beta}_{DGME} = Z\hat{p}.$$

여기서 $\hat{p} = \exp(-Z'X'\hat{\lambda}) \odot [(I_K \otimes J_M)\exp(-Z'X'\hat{\lambda})]^{\odot(-1)}$ 이다.

5.3 DGME 추정량의 점근적 성질

본 연구에서 제안하는 DGME 추정량의 점근적 성질을 알아보
기 위해 Golan et al.(1996, p.96)을 참조한다. DGME 추정량
이 가지는 특징 중 하나는 표본의 크기가 충분히 커짐에 따라
추정량의 분포가 근사적으로 정규분포를 따른다는 것이다.

$\hat{\beta}_{DGME} = Zp(\hat{\lambda})$ 함수는 $\hat{\lambda}$ 에 대한 연속함수이므로 $\hat{\beta}_{DGME}$ 의 분
포는 $\hat{\lambda}$ 의 분포로부터 유도될 수 있는데, 이는 δ -method



(Spanos, 1986)에 의해서 $\hat{\beta}_{DGME}$ 의 분포로 근사시킬 수 있다.

이에 대한 논의에 앞서, 우선 $X'y = X'X\beta + X'e$ 를 표본의 크기 T 으로 나눈 다음의 가중평균 벡터를 고려해 보도록 하자.

$$\begin{aligned}\frac{X'y}{T} &= \frac{X'X\beta}{T} + \frac{X'e}{T} \\ &= \frac{X'XZp}{T} + Vw.\end{aligned}\quad (5.34)$$

식 (5.34)는 포괄적 일반선형모형(generic generalized linear model, GLM)의 한 형태(normed-moment formulation)로 y 는 $T \times 1$ 인 종속변수벡터, X 는 $T \times K$ 인 독립변수행렬, β 는 $K \times 1$ 인 회귀계수벡터, e 는 $T \times 1$ 인 종속변수벡터이다. 또한 모형의 재모수화 시에 V 와 w 는 각각 $K \times KJ$ 인 행렬과 KJ 차원의 가중치 벡터로 재정의 된다.

식 (5.34)를 기반으로 한 이중목적함수를 $M(\lambda)$ 라 할 때 이를 최소화 시키는 $\hat{\lambda}$ 을 통해 $\hat{\beta}_{DGME} = Zp(\hat{\lambda})$ 을 구할 수 있다. 여기까지의 과정은 5.1절과 5.2절에서 다룬 DGME 추정량을 구하는 과정과 동일하다.

이제 식 (5.34)를 기반으로 구한 $\hat{\beta}_{DGME}$ 의 극한분포(limiting distribution)가 다음의 네 가지 가정을 기반으로 정규분포로 근사함을 증명한다.

가정 1: 다음과 같은 유한하고, 양정치 행렬 Q 가 존재한다.

$$\lim_T \left(\frac{X'X}{T} \right) = Q.$$



가정 2: $E(e) = 0$, $Var(e) = \Sigma_e$ 이고 $F(e)$ 는 다음의
린드버그 조건(Lindeberg condition; Billingsley
1986, equation 27.8)을 만족한다.

$$T^{-1} \sum_{t=1}^T \int_{\varsigma} \|e\|^2 dF(e) \rightarrow 0,$$

$$\varsigma = \{e : \|e\| > \sqrt[T]{T}\} \text{ for } \varepsilon > 0.$$

가정 3: $\varepsilon = X'e/\sqrt{T}$ 의 분산-공분산 행렬은 다음과 같이
유한하고 양정치인 행렬로 수렴한다.

$$\lim_T \left(\frac{X' \Sigma_e X}{T} \right) = \Sigma^*.$$

가정 4: 실제 베타(β_0)의 영역은 다음과 같이 정의된다.

$$\beta_0 \in \text{int}(\mathcal{L}).$$

우선, 앞서 구한 이중목적함수 $M(\lambda)$ 의 1계도조건(first order
condition, FOC)에 대한 1차 미분 테일러급수는 다음과 같다.

$$\nabla_{\lambda} M(\hat{\lambda}) = \nabla_{\lambda} M(\lambda_0) + \nabla_{\lambda\lambda'} M(\lambda^*)(\hat{\lambda} - \lambda_0). \quad (5.35)$$

여기서 λ^* 는 평균값정리(mean value theorem)에 의해 $\hat{\lambda}$ 과 λ_0
사이의 값을 갖으며, 식 (5.35)의 좌변은 FOC를 구하기 위해
0이 되어야 한다. 이를 전개하면 식 (5.35)는 다음과 같이 재



정의 될 수 있다.

$$\sqrt{T}(\hat{\lambda} - \lambda_0) = - [\nabla_{\lambda\lambda'} M(\lambda^*)]^{-1} [\sqrt{T} \nabla_{\lambda} M(\lambda_0)]. \quad (5.36)$$

이때 식 (5.36) 우변의 $[\sqrt{T} \nabla_{\lambda} M(\lambda_0)]$ 는 가정 2와 가정 3에 의해 $N(0, \Sigma^*)$ 로 수렴한다. 이에 대한 전개는 다음과 같다.

$$\begin{aligned} \sqrt{T} \nabla_{\lambda} M(\lambda_0) &= \sqrt{T} \left[\frac{X'y}{T} - \left(\frac{X'X}{T} \right) Zp(\lambda_0) - Vw(\lambda_0) \right] \\ &= \sqrt{T} \left[\frac{X'y(y - X\beta_0)}{T} + O(T^{-1}) \right] \\ &= \left(\frac{X'e}{\sqrt{T}} \right) + \frac{O}{\sqrt{T}} \sim N(0, \Sigma^*). \end{aligned}$$

더 나아가, 가정 1을 만족하고 Σ_Z 가 p 의 분산-공분산행렬인 경우 $\lim_T \nabla_{\lambda\lambda'} M(\lambda_0) = Q \Sigma_Z Q'$ 가 유도된다. 이에 대한 전개는 다음과 같다.

$$\begin{aligned} \lim_T \nabla_{\lambda\lambda'} M(\lambda_0) &= \lim_T \left[\left(\frac{X'X}{T} \right) \Sigma_Z(\lambda_0) \left(\frac{X'X}{T} \right)' + \Sigma_V \right] \\ &= Q \Sigma_Z Q', \quad (V = O(T^{-1})). \end{aligned}$$

여기서 $\hat{\lambda}$ 의 극한분포는 다음과 같다.

$$\sqrt{T}(\hat{\lambda} - \lambda_0) \rightarrow N(0, (Q \Sigma_Z Q')^{-1} \Sigma^* (Q \Sigma_Z Q')^{-1}).$$



따라서 네 가지 가정을 기반으로 $\hat{\beta}_{DGME} = Zp(\hat{\lambda})$ 의 연속성과 δ -method에 의해 $\hat{\beta}_{DGME}$ 의 극한분포는 다음과 같이 정규분포로 근사됨을 알 수 있다.

$$\sqrt{T}(\hat{\beta}_{DGME} - \beta_0) \rightarrow N(0, Q^{-1}\Sigma^*Q^{-1}). \quad (5.37)$$



제 6 장 모의실험

본 논문에서 제안하는 DGME의 우수성을 보기 위해 간단한 모의실험을 실시하여 보통최소제곱(ordinary least squares, OLS)과 일반화최소제곱(generalized least squares, GLS), 실현가능한 일반화최소제곱(feasible GLS, FGLS) 추정량의 평균제곱근오차(root mean squared error, RMSE)를 비교하고자 한다.

본 논문에서는 제시된 모든 모의실험은 R(version 3.3.2)을 사용하였고, DGME 추정치를 구하기 위해 `optimx` 함수에서 `nlm`(nonlinear minimization; Dennis, Schnabel, 1983; Schnabel et al., 1985)방법을 사용하였다.

6.1 다중공선성을 고려한 자료

6.1.1 중도절단 일원오차성분모형

모의실험 모형은 다음과 같은 패널회귀모형이다.

$$y = X\beta + \mu_i + e_{it}, \quad i = 1, \dots, 4; \quad t = 1, \dots, 5.$$

모의실험 자료는 Lee et al.(2016)과 같이 많은 저자들(Nerlove, 1972; Arora, 1973; Baltagi, 1981; Heckman, 1981; Wansbeek, Kapteyn, 1989)에 의해 소개된 방법에 의해 생성시켰다. 즉, 외생변수는 $x_{it} = 0.1t + 0.5x_{i,t-1} + w_{it}$, $x_{i0} = 5 + 10w_{i0}$, $w_{it} \sim Unif(-0.5, 0.5)$ 에 의해 독립적으로 생성시키고, 개체효과와 오차항 μ_i , e_{it} 는 각각 독립적으로 $N(0, \sigma_\mu^2)$, $N(0, \sigma_e^2)$ 으로부터 생성시킨다. 여기서 $\sigma_\mu^2 = \rho\sigma^2$, $\sigma_e^2 = (1-\rho)\sigma^2$, $\sigma^2 = \sigma_\mu^2 + \sigma_e^2$ 이라



정한다. DGME 추정량의 모수 대응치(z_k), 개체 대응치(f_i), 오차 대응치(v_{it})는 각각 다음과 같다.

$$\begin{aligned} z_k &= [-3 * \max(\widehat{\beta_{ols}}), -1.5 * \max(\widehat{\beta_{ols}}), 0, 1.5 * \max(\widehat{\beta_{ols}}), 3 * \max(\widehat{\beta_{ols}})], \\ f_i &= [-3, -1.5, 0, 1.5, 3], \\ v_{it} &= [-3, -1.5, 0, 1.5, 3]. \end{aligned}$$

본 연구의 모의실험은 하나의 절편과 열 개의 독립변수를 고려하여 $\beta = [-4.5, 1, 1, 1, 1, 1, 1, 1, 1, 1]$ 이라 하고, $N=4$, $T=5$, $\sigma^2=10$, $\rho=(0, 0.2, 0.4, 0.6, 0.8)$ 로 정한다. 또한 공선성문제를 고려하기 위해 Belsley(1991)와 같이 세 번째 독립변수 X_3 를 다음과 같이 생성시킨다.

$$X_3 = 0.2X_2 + \epsilon, \epsilon \sim N(0, 10^{-2} \text{Var}(0.2X_2)I).$$

전체 표본수는 $220(=N \times T \times K=4 \times 5 \times 11)$ 개로 정하였으며, ρ 값에 따른 각 실험의 반복횟수는 100회로 하였다.

6.1.1.1 모의실험 과정

단계 1: 자료 X 를 6.1.1절에 따라 랜덤 추출한다.

단계 2: 전체분산 $\sigma^2 = \sigma_\mu^2 + \sigma_e^2 = 10$ 으로 고정한다.

ρ	0.0	0.2	0.4	0.6	0.8
σ_μ^2	0	2	4	6	8
σ_e^2	10	8	6	4	2



여기서 개체효과 μ 는 $N(0, \sigma_\mu^2)$ 에서 랜덤추출하고,
나머지 오차항 e 는 $N(0, \sigma_e^2)$ 에서 랜덤추출한다.

단계 3: $y \leq 0$ 인 경우 $y=0$ 으로 둔다.

단계 4: y 와 X 를 다음과 같이 재정렬 한다.

(중도절단된 관측치들을 아래쪽으로 이동)

$$y = X \Leftrightarrow \begin{bmatrix} y_1 \\ y_2 = 0 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

단계 5: OLS, GLS, AM, DGME 추정량을 구한다.

단계 6: 단계 2에서 5까지의 과정을 100번 반복하여 계산된
추정량별 RMSE를 구하고, 효율성을 비교한다.

여기서 RMSE의 식은 다음과 같다.

$$RMSE = \frac{1}{R} \sum_{r=1}^R \sqrt{(\sum (\hat{\beta} - \beta)^2)},$$

$\beta = K \times 1$ 의 회귀계수 벡터,

$\hat{\beta} = K \times 1$ 의 추정된 회귀계수 벡터,

K = 절편을 포함한 회귀계수의 수,

R = 모의실험 반복 횟수.



6.1.1.2 모의실험 결과

<표 6.1>은 모의실험 자료 X 에서 독립변수들 간의 다중공선성을 계산한 결과이고, <표 6.2>는 일원오차성분모형에서 σ_μ^2 과 σ_e^2 의 값에 따라 계산된 추정량별 RMSE를 비교한 결과이다.

아래의 표에 담긴 결과들은 모의실험 진행시 각 효과나 오차항이 갖는 분산 값에 따라 동일한 과정을 100번씩 반복한 결과로, 각 추정량이 갖는 평균적인 추세를 반영하고 있다.

우선 <표 6.1>을 살펴보면 독립변수들 간의 다중공선성이 상당히 심한 것을 알 수 있다. 다음으로 <표 6.2>를 살펴보면 모든 경우에 있어 평균적으로 13% 내외의 중도절단을 보이고 있다. 이러한 경우 자료의 형태에 대한 고려 없이 OLS나 GLS, AM을 사용하면 잘못된 회귀계수를 추정할 수 있다. 예를 들어, <표 6.2>에서 $\sigma_\mu^2 = 2$, $\sigma_e^2 = 8$ 인 경우 RMSE 값으로 OLS는 28.463, GLS는 28.906, AM은 28.273을 갖는 반면 DGME는 상대적으로 작은 14.856값을 갖는다. 이러한 추세는 각 효과나 오차항의 분산값에 상관없이 동일하다. 결론적으로 중도절단이나 다중공선성으로 인해 자료의 왜곡상태가 심하고 자료의 크기가 작은 경우 DGME 추정량이 다른 추정량들에 비해서 강건한 추정량이 될 수 있음을 알 수 있다.

모의실험에 대한 추가적인 결과는 <표 6.3>을 참고하면 된다. 여기서는 $\sigma_\mu^2 = 2$, $\sigma_e^2 = 8$ 인 경우 각 추정량 별로 계산된 회귀계수와 표준편차 값이 계산되어 있다. 대부분의 경우 DGME로 추정된 회귀계수의 표준편차 값이 다른 추정량들에 비해 작으며, 이 차이는 다중공선성이 심할수록 더 큰 것을 알 수 있다. 이는 제시된 추정량간의 편향(bias)은 비슷하지만 분산은 DGME가 가장 작은 것을 뜻한다. 이러한 성질로 인해 DGME의 RMSE가 다른 추정량들에 비해 작은 것을 알 수 있다.



<표 6.1> 독립변수들 간의 다중공선성

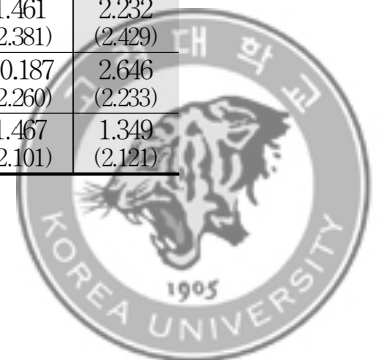
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
VIF	32.38	121.43	101.30	10.02	44.07	12.68	10.10	7.77	9.58	14.41

<표 6.2> 일원오차성분모형에서 모의실험의 RMSE 결과

σ_μ^2	σ_e^2	중도절단율	OLS	GLS	AM	DGME
0	10	13.850 %	27.071	27.071	26.725	13.647
2	8	13.100 %	28.463	28.906	28.273	14.856
4	6	13.700 %	23.902	24.825	24.281	10.574
6	4	15.350 %	24.537	23.708	25.833	11.081
8	2	14.950 %	23.661	19.597	25.038	9.065

<표 6.3> 일원오차성분모형에서 한번 모의실험의 모수추정 결과 (추정치(표준편차))

평균 중도절단율 = 13.100%							
σ_μ^2	σ_e^2		TRUE	OLS	GLS	AM	DGME
2	8	β_0	-4.5	-3.520 (2.455)	-5.433 (2.478)	-5.639 (2.449)	-4.146 (2.320)
		β_1	1	3.562 (2.738)	-1.105 (2.763)	-0.816 (2.755)	0.718 (2.666)
		β_2	1	0.684 (6.783)	0.101 (6.915)	0.421 (6.804)	0.555 (4.211)
		β_3	1	2.354 (32.226)	1.945 (32.623)	-0.262 (32.358)	0.294 (18.084)
		β_4	1	-1.231 (1.762)	2.070 (1.707)	1.836 (1.844)	0.820 (1.704)
		β_5	1	1.966 (2.960)	1.609 (3.023)	2.603 (2.942)	0.540 (2.725)
		β_6	1	2.800 (2.578)	1.569 (2.585)	4.559 (2.573)	1.171 (2.379)
		β_7	1	0.432 (2.075)	-0.380 (2.125)	0.222 (2.080)	0.009 (2.097)
		β_8	1	1.875 (2.436)	2.575 (2.547)	1.461 (2.381)	2.232 (2.429)
		β_9	1	-1.646 (2.258)	3.308 (2.309)	-0.187 (2.260)	2.646 (2.233)
		β_{10}	1	-0.984 (2.105)	0.915 (2.115)	1.467 (2.101)	1.349 (2.121)



6.1.2 중도절단 이원오차성분모형

모의실험 모형은 다음과 같은 패널회귀모형이다.

$$y = X\beta + \mu_i + \lambda_t + e_{it}, \quad i = 1, \dots, 4; \quad t = 1, \dots, 5.$$

모의실험 자료 X 는 본 논문의 6.1.1절과 동일하게 생성하고, 개체효과와 시간효과 그리고 오차항 μ_i , λ_t , e_{it} 는 각각 독립적으로 $N(0, \sigma_\mu^2)$, $N(0, \sigma_\lambda^2)$, $N(0, \sigma_e^2)$ 으로부터 생성시킨다. 여기서 $\sigma^2 = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_e^2$, $\rho = \sigma_\mu^2 / \sigma^2$, $\eta = \sigma_\lambda^2 / \sigma^2$ 이라 정한다.

DGME 추정량의 모수 대응치(z_k), 개체 대응치(f_i), 시간 대응치(a_t), 오차 대응치(v_{it})는 각각 다음과 같다.

$$z_k = [-3 * \max(\widehat{\beta_{0ls}}), -1.5 * \max(\widehat{\beta_{0ls}}), 0, 1.5 * \max(\widehat{\beta_{0ls}}), 3 * \max(\widehat{\beta_{0ls}})],$$

$$f_i = [-3, -1.5, 0, 1.5, 3],$$

$$a_t = [-3, -1.5, 0, 1.5, 3],$$

$$v_{it} = [-3, -1.5, 0, 1.5, 3].$$

회귀계수와 개체, 시간의 크기는 6.1.1절에서와 동일하게 $\beta = [-4.5, 1, 1, 1, 1, 1, 1, 1, 1, 1]$, $N=4$, $T=5$ 이라 두고, $\sigma^2 = 10$, $\rho = (0, 0.2, 0.4, 0.6, 0.8)$, $\eta = (0, 0.2, 0.4, 0.6, 0.8)$, $1 - \rho - \eta > 0$ 로 정한다. 또한 공선성문제 역시 6.1.1절과 동일하게 설정한다.

전체 표본수는 $220(= N \times T \times K = 4 \times 5 \times 11)$ 개로 정하였으며,



ρ 와 η 값에 따른 각 실험의 반복횟수는 100회로 하였다.

6.1.2.1 모의실험 과정

단계 1: 자료 X 를 6.1.2절에 따라 랜덤 추출한다.

단계 2: 전체분산 $\sigma^2 = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_e^2 = 10$ 으로 고정한다.

ρ	η	σ_μ^2	σ_λ^2	σ_e^2
0.0	0.0	0	0	10
	0.2	0	2	8
	0.4	0	4	6
	0.6	0	6	4
	0.8	0	8	2
0.2	0.0	2	0	8
	0.2	2	2	6
	0.4	2	4	4
	0.6	2	6	2
0.4	0.0	4	0	6
	0.2	4	2	4
	0.4	4	4	2
0.6	0.0	6	0	4
	0.2	6	2	2
0.8	0.0	8	0	2

여기서 개체효과 μ 는 $N(0, \sigma_\mu^2)$ 에서 랜덤추출하고,

시간효과 λ 는 $N(0, \sigma_\lambda^2)$ 에서 랜덤추출하고,

나머지 오차항 e 는 $N(0, \sigma_e^2)$ 에서 랜덤추출한다.

단계 3: $y \leq 0$ 인 경우 $y=0$ 으로 둔다.



단계 4: y 와 X 를 다음과 같이 재정렬 한다.

(중도절단된 관측치들을 아래쪽으로 이동)

$$y = X \Rightarrow \begin{bmatrix} y_1 \\ y_2 = 0 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

단계 5: OLS, GLS, AM, DGME 추정량을 구한다.

단계 6: 단계 2에서 5까지의 과정을 100번 반복하여 계산된
추정량별 RMSE를 구하고, 효율성을 비교한다.

6.1.2.2 모의실험 결과

<표 6.4>는 이원오차성분모형에서 σ_μ^2 , σ_λ^2 , σ_e^2 의 값에 따라
계산된 추정량별 RMSE를 비교한 결과이다.

아래의 표에 담긴 결과들은 모의실험 진행시 각 효과나 오
차항이 갖는 분산 값에 따라 동일한 과정을 100번씩 반복한
결과로, 각 추정량이 갖는 평균적인 추세를 반영하고 있다.

이원오차성분모형에서 사용하는 모의실험 자료 X 는 일원오
차성분모형에서 사용하는 자료와 동일하므로 <표 6.1>을 통해
독립변수들 간의 다중공선성이 상당히 심한 것을 알 수 있다.

<표 6.4>를 살펴보면 모든 경우에 있어 평균적으로 15% 내
외의 중도절단율을 보이고 있다. 이러한 경우 자료의 형태에
대한 고려 없이 OLS나 GLS, AM을 사용하면 잘못된 회귀계
수를 추정할 수 있다. 예를 들어, <표 6.4>에서 $\sigma_\mu^2 = 2$, $\sigma_\lambda^2 = 2$
 $\sigma_e^2 = 6$ 인 경우 RMSE 값으로 OLS는 26.258, GLS는 25.865,
AM은 26.854을 갖는 반면 DGME는 상대적으로 작은 12.305

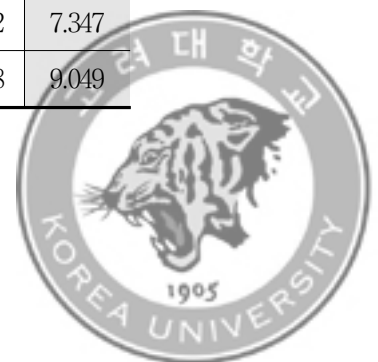


값을 갖는다. 이러한 추세는 각 효과나 오차항의 분산값에 상관없이 동일하다. 결론적으로 이원오차성분 모형에서도 중도절단이나 다중공선성으로 인해 자료의 왜곡상태가 심하고 자료의 크기가 작은 경우 DGME 추정량이 다른 추정량들에 비해서 강건한 추정량이 될 수 있음을 알 수 있다.

6.1.2절의 모의실험에 대한 추가적인 결과는 <표 6.5>를 참고하면 된다. 이에 대한 해석은 6.1.1.2절과 동일하다.

<표 6.4> 이원오차성분모형에서 모의실험의 RMSE 결과

σ_μ^2	σ_λ^2	σ_e^2	중도절단율	OLS	GLS	AM	DGME
0	0	10	14.550 %	27.879	27.879	28.027	12.534
	2	8	14.000 %	23.528	22.566	23.143	11.069
	4	6	15.150 %	25.157	22.617	22.611	11.890
	6	4	14.450 %	22.028	18.683	19.372	9.374
	8	2	16.350 %	21.709	15.537	16.663	9.413
2	0	8	14.800 %	26.840	27.326	27.833	11.967
	2	6	15.700 %	26.258	25.865	26.854	12.305
	4	4	15.350 %	23.558	21.921	21.658	10.230
	6	2	14.900 %	20.085	17.754	18.875	8.107
4	0	6	14.500 %	28.132	28.874	29.223	13.613
	2	4	14.500 %	24.706	22.331	24.889	10.132
	4	2	15.600 %	21.409	17.940	20.879	8.474
6	0	4	14.950 %	26.745	27.842	26.949	11.776
	2	2	16.450 %	21.225	18.244	20.742	7.347
8	0	2	16.100 %	22.970	19.514	23.378	9.049



<표 6.5> 이원오차성분모형에서 한번 모의실험의 모수추정
결과 (추정치(표준편차))

평균 중도절단을 = 15.700%							
σ_μ^2	σ_e^2		TRUE	OLS	GLS	AM	DGME
2	8	β_0	-4.5	-5.306 (1.876)	-4.341 (1.873)	-2.861 (1.900)	-3.957 (2.142)
		β_1	1	1.793 (2.378)	0.283 (2.324)	0.742 (2.380)	0.564 (2.259)
		β_2	1	2.160 (6.616)	2.136 (6.502)	0.027 (6.578)	0.944 (3.629)
		β_3	1	0.252 (30.720)	2.803 (30.198)	1.868 (31.180)	0.264 (15.796)
		β_4	1	0.914 (1.940)	0.959 (1.840)	2.722 (1.942)	1.680 (1.825)
		β_5	1	0.792 (2.385)	-0.149 (2.425)	1.037 (2.439)	1.800 (2.315)
		β_6	1	1.982 (2.154)	-1.784 (2.204)	-1.657 (2.224)	1.420 (2.239)
		β_7	1	-1.680 (2.348)	1.596 (2.318)	0.501 (2.357)	1.020 (2.202)
		β_8	1	0.418 (2.290)	3.242 (2.176)	0.652 (2.274)	0.057 (2.209)
		β_9	1	0.900 (2.023)	1.741 (1.980)	3.185 (2.065)	0.519 (2.000)
		β_{10}	1	3.678 (2.275)	1.067 (2.098)	1.216 (2.154)	0.666 (2.175)

6.2 다중공선성과 ($n < p$)을 고려한 자료

이번 절에서는 독립변수들 간에 강한 다중공선성을 보이고 자료의 크기가 $n(=NT) < p(=K)$ 인 경우 즉, 관측치의 수(n)가 모수의 수(p)보다 적은 경우를 추가적으로 고려해 보도록 한다.

OLS, GLS, FGLS의 경우 추정량을 계산하는 과정에서 $(X'X)^{-1}$, $(X'\Omega^{-1}X)^{-1}$ 등과 같은 역행렬 계산과정이 필요한데, 자료의 크기가 $n < p$ 인 경우 역행렬을 계산하지 못하여 추정량을 계산하지 못한다. 하지만 DGME는 제시된 추정방법들



중에 유일하게 추정량의 계산이 가능하다.

전반적인 모의실험 과정은 앞 절과 동일하며, 자료의 크기만 바뀌었기 때문에 모형별로 DGME 추정결과에 대해서만 간단히 다루도록 하겠다.

6.2.1 중도절단 일원오차성분모형

앞의 6.1.1.1 절과 동일한 과정을 통해 자료 X 를 생성한다. 단, 여기서는 $n < p$ 인 경우를 고려해야 함으로 $N = 3$, $T = 3$, $K = 11$ 으로 놓아 $n = 9(= NT)$, $p = 11(= K)$ 크기의 자료 X 를 만들어 DGME 추정량을 추정해보았다.

독립변수들 간의 다중공선성은 다음의 <표 6.6>과 같이 모두 무한대에 가까운 값이 나왔다. 즉 독립변수들 간의 다중공선성은 매우 심각함을 알 수 있다.

<표 6.6> 독립변수들 간의 다중공선성($n < p$)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
VIF	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf

전체 분산도 앞의 6.1.1.1 절과 동일과 동일하게 $\sigma^2 = 10$ 으로 놓고 σ_μ^2 과 σ_e^2 의 값을 조정해 가며 모의실험 하였다. 그 결과 주어진 σ_μ^2 과 σ_e^2 의 값에 상관없이 DGME 추정량의 전반적인 추세가 동일하게 나왔다. 그중 한 결과인 $\sigma_\mu^2 = 2$, $\sigma_e^2 = 8$ 인 경우의 추정결과를 살펴보면 <표 6.7>과 같다.



<표 6.7> 일원오차성분모형에서 한번 모의실험의 모수추정치
와 RMSE

평균 중도절단을 = 15.667% ($\sigma_{\mu}^2 = 2, \sigma_e^2 = 8$)												
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	RMSE
TRUE	-7.5	1	1	1	1	1	1	1	1	1	1	14.7719
DGME	-8.42	1.73	1.65	0.11	-0.52	0.55	0.01	2.73	-0.57	1.92	1.19	

<표 6.7>를 살펴보면 100번의 DGME 추정과정을 거친 결과, 추정량의 RMSE값은 14.7719로 계산되었고, 그중 한 결과인 DGME 추정의 결과를 보여준다. 참값으로 주어진 모수 값들과 완전히 일치하지는 않지만 매우 심각한 다중공선성과, $n < p$ 인 자료형태를 고려했을 때 일원오차성분모형을 고려할 경우 DGME가 상당히 우수한 추정량임을 알 수 있다.

6.2.2 중도절단 이원오차성분모형

앞의 6.1.2.1 절과 동일하게 이원오차성분모형에서 사용하는 모의실험 자료 X 는 일원오차성분모형에서 사용하는 자료와 동일하므로 <표 6.6>를 통해 독립변수들 간의 다중공선성이 매우 심한 것을 알 수 있다.

전체 분산도 앞의 6.1.2.1 절과 동일과 동일하게 $\sigma^2 = 10$ 으로 놓고 σ_μ^2 과 σ_λ^2 , σ_e^2 의 값을 조정해 가며 모의실험 하였다. 그 결과 주어진 σ_μ^2 과 σ_λ^2 , σ_e^2 의 값에 상관없이 DGME 추정량의 전반적인 추세가 동일하게 나왔다. 그중 한 결과인 $\sigma_\mu^2 = 2$, $\sigma_\lambda^2 = 2$, $\sigma_e^2 = 6$ 인 경우의 추정결과를 살펴보면 <표 6.8>과 같다.



<표 6.8> 이원오차성분모형에서 한번 모의실험의 모수추정치
와 RMSE

평균 중도절단율 =16.444 % ($\sigma_{\mu}^2 = 2, \sigma_{\lambda}^2 = 2\sigma_e^2 = 6$)												
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	RMSE
TRUE	-7.5	1	1	1	1	1	1	1	1	1	1	13.4887
DGME	-7.96	0.18	1.46	0.49	2.62	0.73	0.10	1.44	4.28	-1.38	0.61	

<표 6.8>를 살펴보면 100번의 DGME 추정과정을 거친 결과, 추정량의 RMSE값은 13.4887로 계산되었고, 그중 한 결과인 DGME 추정의 결과를 보여준다. 실제 값으로 주어진 모수 값들과 완전히 일치하지는 않지만 매우 심각한 다중공선성과, $n < p$ 인 자료형태를 고려했을 때 DGME가 상당히 우수한 추정량임을 알 수 있다.



제 7 장 실 자료 분석

실자료 분석을 위해 미국 미시간 대학 사회조사연구소에서 조사된 PSID(Panel Study of Income Dynamics)의 자료를 분석하였다. Heckman, Macurdy(1980)는 1968년에 30-65세 사이의 백인여성을 대상으로 연간 근로시간과 경제활동참가율을 조사하여 여성 노동 공급의 수명주기모형을 연구하였다. 이 연구에서 만일 여성이 고용이 되었는데 일을 하지 않았다면 근로시간을 0으로 보고가 된다는 점을 고려하여 이를 중도절단된 자료로 보았고, 이를 분석하기 위해 토빗모형을 고려하였다(Maddala, 1983). 비록 Heckman, Macurdy(1980)는 개체효과를 고정효과로 보았지만, 본 논문에서는 Chamberlain(1984)과 Heckman, Willis(1975)에서처럼 이를 랜덤효과로 간주하여 분석을 하였다.

7.1 자료 소개

분석에 사용될 자료는 PSID 자료에서 1979년에 남편(가장)의 나이가 20-70세 사이에 속하는 가구들을 6년간 추적 조사한 자료로, 본 연구를 위해 <표 7.1>에서 나타난 6가지 기준에 따라 각 연도별 원 자료를 분류하고 개체별, 연도별로 순차적으로 결합하는 과정을 거쳤다.

이렇게 만들어진 6년간의 패널자료에서 1979년도를 제외한 나머지 5년간(1980년 - 1984년)의 자료만 분석에 사용된다. 이는 당해 연도에 조사된 연간 근로소득이 전년도를 기준으로 조사된 결과이기 때문이다.



<표 7.1> 패널자료 생성을 위한 기준

분류	조건	기준년도
가장의 성별	Male	매년
가장의 고용상태	Working now	매년
아내의 고용상태	Working now	매년
결혼 여부	Married	매년
가장의 나이	20-70세	1979년
재산 유무	유	매년

또한 아내의 연소득이나 아내의 연간근로시간, 남편의 연간근로시간이 0으로 입력되어 있거나(결측값), 갑자기 연소득이 수배가 증가한 경우(특이값) 혹은 5년 동안 남편의 연소득이 모두 0으로 입력되어 있는 가구들을 자료에서 제거해주면 최종적으로 분석을 위한 패널자료($N=201, T=5$)가 완성된다.

변수에 대한 간략한 설명은 <표 7.2>와 같다.

<표 7.2> 변수 설명(1979년도 Codebook 기준)

변수명			설명
종속 독립	Y	HEAD 1978 WAGES	가장의 연소득(1978년)
	X ₁	AGE OF HEAD	가장의 나이
	X ₂	AGE OF WIFE	아내의 나이
	X ₃	1979 FAMILY SIZE	가족 구성원 수(1979년)
	X ₄	# CHILDREN IN FU	17세 이하 자녀의 수
	X ₅	WIFE 1978 LABOR/WAGE	아내의 연소득(1978년)
	X ₆	WF ANN WRK HRS IN 78	아내의 연간근로시간(1978년)
	X ₇	HD ANN WRK HRS IN 78	가장의 연간근로시간(1978년)
	X ₈	ANN PROP TAX 79	연간 재산세(1979년)
	X ₉	ANN \$ EATING OUT	연간 외식비용
	X ₁₀	1979 HOUSE VALUE	집가격(1979년)



7.2 중도절단 일원오차성분모형

실 자료 분석 시에는 실제 분산성분 Ω 를 알 수 없기 때문에 $(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ 로 계산되는 GLS는 구할 수 없다. 따라서 OLS와 Amemylia(1971)에 의해 제안된 FGLS(AM), R의 plm package(version 1.6-4)에서 제공하는 PLM추정량을 구해서 DGME 추정량과의 효율을 비교해보았다. 비교는 추정량별 PRMSE를 통해 이루어졌는데, 이는 모의실험과 달리 실제 β 값을 알 수가 없어 추정량의 RMSE를 구할 수 없기 때문이다.

또한 본 논문에서는 소표본인 경우에서 DGME의 효율성을 살펴보고자 함으로, 앞에서 생성한 패널자료를 모두 사용하지 않고 특정 기준에 부합되는 가구들을 대상(소표본)으로 분석을 실시하였다. 여기서는 미국 상무부 산하 경제분석국에서 경제 분석을 위해 만든 지역구분 중 하나인 New England Region에서 거주하는 가구들을 분석 대상으로 하였다.

분석 과정은 다음과 같다.

단계 1: 다음의 기준에 속하는 가구를 골라낸다.

주(STATE)	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
----------	---

단계 2: 다음의 기준에 속하는 가구를 제외한다.

- ① 5년간 모든 y 값이 0인 가구
- ② 전반부 3년간 모든 y 값이 0인 가구



③ 후반부 2년 중에 $y=0$ 이 있는 가구

분석에 사용되는 자료의 크기가 다음과 같이 변한다.

$$(N=201, T=5) \Rightarrow (N=10, T=5)$$

단계 3: 독립변수들을 표준화한다.

단계 4: 자료를 분석(추정)용과 모형 평가용으로 분할한다.

분석용	$N=10, T=3$ (전반부 3년)
평가용	$N=10, T=2$ (후반부 2년)

단계 5: 분석용 자료의 y 와 X 를 다음과 같이 재정렬 한다.

(중도절단된 관측치들을 아래쪽으로 이동)

$$y = X \Rightarrow \begin{bmatrix} y_1 \\ y_2 = 0 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

단계 6: 분석용 데이터로 OLS, AM, PLM, DGME

추정량을 구한다.

단계 7: 단계 6에서 구한 추정량별로 패널모형을 만들고

평가용 데이터로 PRMSE를 구한다.

여기서 PRMSE의 식은 다음과 같다.

$$\text{PRMSE} = \sqrt{\left(\frac{1}{NT} \sum (\hat{Y} - Y)^2 \right)},$$

$Y = NT \times 1$ 의 실제 종속변수 벡터,



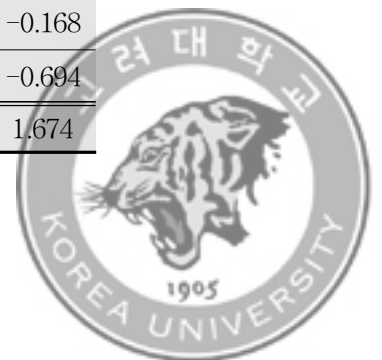
$\hat{Y} = NT \times 1$ 의 추정된 종속변수 벡터,
 $N =$ 개체 수, $T =$ 연도 수.

분석 결과는 다음과 같다.

<표 7.3>을 보면 분석용 데이터는 독립변수들 간의 다중공선성이 상당히 심각하게 존재하고 6.667%의 중도절단율을 보이고 있다. 그럼에도 불구하고 DGME의 PRMSE값이 다른 추정량들에 비해 가장 작은 것으로 보아 소표본에서 자료의 왜곡 상태가 심할 경우 DGME가 상당히 훌륭한 추정량이 될 수 있음을 알 수 있다.

<표 7.3> 일원오차성분모형에서 실 자료 분석결과

분석용 데이터의 중도절단율 = 6.667%					
	VIF	OLS	AM	PLM	DGME
β_0	-	2.236	2.187	2.370	2.184
β_1	49.386	4.051	4.084	-17.035	0.486
β_2	28.300	-2.955	-2.946	20.519	-0.382
β_3	37.676	-0.385	-0.500	-3.852	-0.153
β_4	38.060	0.550	0.611	1.188	0.264
β_5	7.597	0.202	0.080	1.045	-0.253
β_6	4.462	-1.129	-1.057	-0.555	0.576
β_7	2.259	-0.389	-0.494	0.109	0.765
β_8	8.182	-1.803	-1.603	-1.939	0.904
β_9	1.883	-0.376	-0.457	0.006	-0.168
β_{10}	5.802	1.791	1.643	-0.006	-0.694
PRMSE		1.811	1.947	8.602	1.674



7.3 중도절단 이원오차성분모형

전반적인 분석과정은 7.2.1절과 동일하며, 분석을 위해 사용되는 자료만 다르다. 여기서 사용되는 자료는 도·소매업에 종사하는 가구를 대상으로 하며 이중에서 중도절단 데이터를 포함하고 있는 주(state)만 포함한다.

분석 과정은 다음과 같다.

단계 1: 다음의 기준에 속하는 가구를 골라낸다.

주(STATE)	Massachusetts, Oregon, Missouri
가장이 속한 산업군	Wholesale and Retail Trade(도·소매업)

단계 2: 다음의 기준에 속하는 가구를 제외한다.

- ① 5년간 모든 y 값이 0인 가구
- ② 전반부 3년간 모든 y 값이 0인 가구
- ③ 후반부 2년 중에 $y=0$ 이 있는 가구

분석에 사용되는 자료의 크기가 다음과 같이 변한다.

$$(N=201, T=5) \Rightarrow (N=4, T=5)$$

단계 3: 독립변수들을 표준화한다.



단계 4: 자료를 분석(추정)용과 모형 평가용으로 분할한다.

분석용	$N=4, T=3$ (전반부 3년)
평가용	$N=4, T=2$ (후반부 2년)

단계 5: 분석용 자료의 y 와 X 를 다음과 같이 재정렬 한다.

(중도절단된 관측치들을 아래쪽으로 이동)

$$y = X \Rightarrow \begin{bmatrix} y_1 \\ y_2 = 0 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

단계 6: 분석용 데이터로 OLS, AM, PLM, DGME 추정량을 구한다.

단계 7: 단계 6에서 구한 추정량별로 패널모형을 만들고 평가용 데이터로 PRMSE를 구한다.

분석 결과는 다음과 같다.

<표 7.4>를 보면 분석용 데이터는 독립변수들 간의 다중공선성이 상당히 심각하게 존재하고 20%의 중도절단율을 보이고 있다. PRMSE는 일원오차성분 모형에서와 마찬가지로 OLS보다 DGME가 작음을 알 수 있다.

이번 결과에서 눈여겨 볼 점으로는 PLM 추정량은 회귀계수의 추정값 자체를 구하지 못했다는 점이다. 이는 주어진 데이터의 구조상 회귀계수를 구하는 과정에서 필요한 역행렬을 구하지 못하기 때문이다. 이러한 문제는 AM 추정량 계산 시에도 종종 발생하는 문제이다. 결론적으로는 이원오차성분모형에



서도 자료의 왜곡 상태가 심할 경우 DGME가 다른 추정량들에 비해 우수한 추정량이 될 수 있음을 알 수 있다.

<표 7.4> 이원오차성분모형에서 실 자료 분석결과

분석용 데이터의 중도절단을 = 20%				
	VIF	OLS	AM	DGME
β_0	-	0.609	0.541	1.111
β_1	168.541	0.403	0.651	0.812
β_2	203.949	-1.009	1.368	-1.031
β_3	39.080	-3.353	-3.304	-0.420
β_4	22.644	0.226	0.250	-0.535
β_5	3.512	1.651	1.650	0.164
β_6	9.095	-2.966	-2.952	-1.013
β_7	7.553	-0.274	-0.389	0.269
β_8	73.847	-5.410	-5.578	1.427
β_9	53.428	-3.282	-3.426	-0.596
β_{10}	25.915	1.340	1.575	-1.870
PRMSE		4.924	5.035	3.637



제 8 장 결론

패널자료는 다수의 개체를 장기간 추적조사 하므로 패널자료의 생성과정에서 자료의 일부가 누락되거나 오염되는 문제가 종종 발생한다. 특히 조사목적에 따라 개인적 사유, 경제적 혹은 건강 등의 문제들로 인해 조사 대상자가 조사를 거부할 경우 중도절단된 자료가 발생하기도 한다. 이러한 불완전 자료에 대한 올바른 통계적 추론을 위해, 본 논문에서는 Lee et al.(2016)의 연구를 확장하여 일원오차성분 혹은 이원오차성분을 가지는 패널회귀모형에서의 DGME 추정량을 제안한다.

DGME 추정량은 과거에 Song, Cheon(2006)이 제안했던 일반화최대엔트로피와는 달리 일치성과 정규화 가법성이라는 두 가지 제약조건에 구애받지 않고 쉽고 빠르게 추정량을 구할 수 있다.

특히 기존에 제안되어 있는 OLS 추정량은 자료의 이분산 문제를 고려하지 않고, FGLS 추정량은 점근적으로 GLS 추정량과 동일한 성질을 갖지만 일반적으로 소표본인 경우에는 정확한 성질이 알려져 있지 않다(Judge et al., 1988). 따라서 자료의 크기가 작고 중도절단, 다중공선성의 문제가 있는 경우 DGME 추정량을 구하여 OLS, GLS, FGLS(AM), PLM 추정량들과의 효율성을 비교하였다.

모의실험과 실 자료 분석 결과에서 알 수 있듯이 DGME 추정량이 다른 추정량들에 비해 강건한 성질 갖고 있음을 확인할 수 있었다. 이러한 결과를 통해 중도절단된 패널자료에서 개체효과나 시간효과를 고려할 경우 DGME 추정량을 사용하는 것이 적절할 것으로 생각된다.

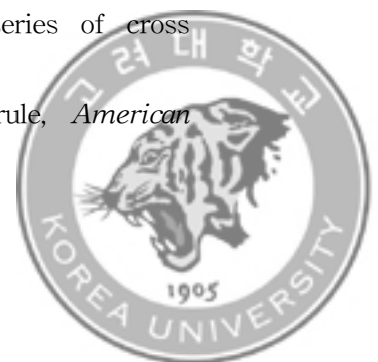


참 고 문 헌

- [1] 박성현 (2007). 제 3판 회귀분석, 민영사.
- [2] 진수영 (2002). 패널회귀모형에서 Generalized Maximum Entropy 추정량에 관한 연구, 고려대학교 대학원 석사학위논문.
- [3] Amemiya, T. (1971). The estimation of the variances in a variance-components model, *International Econometric Review*, 12, 1-13.
- [4] Amemiya, T. (1984). Tobit models: a survey, *Journal of Econometrics*, 24, 3-61.
- [5] Arora, S. (1973). Error components regression models and their applications, *Annals of Economic and Social Measurement*, 2, 451-462.
- [6] Baltagi, B. H. (1981). Pooling: an experimental study of alternative testing and estimation procedures in a two-way error component model, *Journal of Econometrics*, 17, 21-49.
- [7] Baltagi, B. H. (1995). *Econometrics Analysis of Panel Data*, John Wiley, New York.
- [8] Belsley, D. A. (1991). *Conditioning diagnostics: Collinearity and weak data in regression*, John Wiley & Sons, New York.
- [9] Billingsley, P. (1986). *Probability and Measure*, John Wiley, New York.
- [10] Chamberlain, G. (1984). Panel data, in Griliches Z. and Intrilligator M., eds., *Handbook of Econometrics*, 2, 1247-1318.
- [11] Dennis, J. E., Schnabel, R. B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice-Hall, Englewood Cliffs, New Jersey.
- [12] Golan, A., Judge, G., Miller, D. (1996). *Maximum entropy econometrics: robust estimation with limited data*, John Wiley & Sons, Chichester.
- [13] Heckman, J. J. (1981). *The incidental parameters problem and the problem of initial conditions in estimating a*



- discrete time-discrete data stochastic process*, in Manski M C., McFadden, D. eds., *Structural analysis of discrete data with econometric application*, MIT Press, Cambridge, MA.
- [14] Heckman, J. J., Macurdy, T. E. (1980). A life-cycle model of female labour supply, *Review of Economic Studies*, 47, 47-74.
- [15] Heckman, J. J., Willis, R. (1975). *Estimation of a stochastic model of reproduction: an econometric approach*, in Terleckyi, N, ed., 99-138, Household Production and Consumption, National Bureau of Economic Research, New York.
- [16] Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H., Lee, T. C. (1988). *Introduction to the Theory and Practice of Econometrics*, John Wiley, New York.
- [17] Lee, J., Cheon, S. (2014). Dual generalized maximum entropy estimation for panel data regression models, *Communications for Statistical Applications and Methods*, 21, 395-409.
- [18] Lee, J., Cheon, S. (2015). Dual generalized maximum entropy in the unbalanced panel regression model, *Journal of the Korean Data Analysis Society*, 17(3B), 1285-1295.
- [19] Lee, H., Choi, P. (2004). Multicollinearity in regression graphics, *Journal of the Korean Data Analysis Society*, 6(3), 849-860.
- [20] Lee, J., Lee, D., Cheon, S. (2016). Dual Generalized Maximum Entropy in the Censored Panel Regression Model, *Journal of the Korean Data Analysis Society*, 18(3B), 1271-1280.
- [21] Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*, Cambridge University Press, Cambridge.
- [22] Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross sections, *Econometrica*, 39, 359-382.
- [23] Pukelsheim, F. (1994). The three sigma rule, *American Statistician*, 48, 88-91.



- [24] Schnabel, R. B., Koontz, J. E., Weiss, B. E. (1985). A modular system of algorithms for unconstrained minimization, *ACM Transactions on Mathematical Software*, 11, 419-440.
- [25] Song, S. H., Cheon, S. Y. (2006). A study of generalized maximum entropy estimator for the panel regression model, *The Korean Journal of Applied Statistics*, 19, 521-534.
- [26] Spanos. A. (1986). *Statistical Foundations of Econometric Modelling*, Cambridge University Press.
- [27] Teimour, M. (2012). Relative performance of components variance estimators in random effects models, *Iranian Journal of Economic Research*, 17(50), 83-98.
- [28] Tobin, J. (1958). Estimation of relationships for limited dependent variables, *Econometrica*, 26, 24-36.
- [29] Wansbeek, T., Kapteyn A. (1989). Estimation of the error-components model with incomplete panels, *Journal of Econometrics*, 41, 341-361.

