

연구 자료

패널자료 분석방법

서 진 교*

Abstract

Recently, empirical research in economics has been enriched by the availability of a wealth of new sources of data; cross sections of individuals observed over time. These allow us to construct and test more realistic behavioral models that could not be identified using only a cross section or a single time series data set. Several estimation techniques for panel data analysis as well as conceptual explanation for several models - fixed and random effect models - are introduced, even the arguments is focused on the linear case.

- | | |
|-------------------------------|--|
| 1. 머리말 | 4. 임의효과모델(Random Effect Model) |
| 2. 패널자료모델의 구분 | 5. 우도추정(Maximum Likelihood Estimation) |
| 3. 고정효과모델(Fixed Effect Model) | 6. 결 론 |

1. 머리말

최근 경제현상분석에서 패널자료에 근거해 다양한 결론을 도출해 내고 있다. 패널자료란 횡단자료(Cross-section Data)와 시계열자료(Time-series Data)를 하나로 통합한 자료를 말하는 것으로서 우리 주위에서 쉽게 찾아볼 수 있는 대표적인 자료형

태 중의 하나이다¹. 예를 들어 3,000개의 표본농가에 대해 1976년부터 2000년까지 농가경제조사가 이루어졌다면 개별연도에 대해서는 3,000개의 표본수를 갖는 횡단자료이지만 1976년부터 2000년까지는 시계열자료로 볼 수 있어 전체자료는 하나의 패널

* 부연구위원

¹ 따라서 패널자료를 종종 “cross-sections over time” 또는 “pooled cross-section time-series data”로 부르기도 한다.

자료가 된다.

패널자료가 갖는 장점은 매우 다양하다. 무엇보다도 횡단분석이나 시계열분석에 비해 관측치(observations)의 수가 크기 때문에 자유도(degree of freedom) 문제가 상대적으로 줄어들며 동시에 추정값(estimates)의 효율성도 높아진다. 특히 가계소비자료인 경우 개별 경제주체인 가계간 차이가 상당히 크기 때문에 시계열분석만으로는 적절한 추정값을 얻기가 곤란한데 반해 패널자료분석은 횡단관측치간의 이질성(heterogeneity)을 감안할 수 있기 때문에 시계열자료나 횡단자료에 비해 보다 복잡한 동태적·행동적 가설검정이 가능하다. 즉 시계열자료는 시간이라는 공간에서 횡단관측치들의 개별 특성을 무시하고 이를 통합(aggregate)내지는 평균화해 놓은 것임에 반해 패널자료에서는 시간경과에 따른 횡단관측치간의 변화추이를 고려할 수 있다. 또한 패널자료를 이용함으로써 계량추정식에서 종종 분석하기가 어려웠던 잠재된(latent) 혹은 관측 불가능한(unobservable) 교란항을 보다 심도있게 분석할 수 있는 장점도 있다.²

이러한 패널자료의 계량분석은 당연히 기존 횡단분석방법이나 시계열분석방법과는 다르다. 패널자료에 횡단분석기법을 적용한다면 시간경과에 따른 표본의 동태적인 변화를 고려하지 못하는 잘못을 범하게 되며, 반대로 시계열자료 분석기법만을 적용한다면 횡단관측치간의 차이(individual heterogeneity)를 고려하지 않은 추정결과

를 얻게 된다. 외국의 유수 논문들이 패널자료의 계량분석을 이용해 다양한 연구결과를 발표하고 있지만 상당수는 계량경제패키지 이용에 그쳐, 실제 컴퓨터의 패널자료추정이 어떤 가정하에서 이루어진 것인지 또는 어떻게 계산된 것인지 모르는 경우가 많아 말 그대로 그들의 추정결과가 블랙박스(Black Box)화 되는 경향이 있다. 여기에서는 이러한 점에 착안하여 패널자료분석시 이용되는 계량경제적 추정방법을 소개하고 아울러 그에 따른 문제점을 적시함으로써 실증연구에 도움이 되고자 한다.

2. 패널자료모델의 구분

패널자료의 계량경제학적 모델은 대개 다음과 같이 가장 복잡한(일반적인) 경우에서부터 단순한 경우에 이르기까지 5가지로 분류할 수 있다.

① 추정계수가 횡단관측치 및 시간경과에 모두 따라 변하는 경우

$$Y_{it} = \alpha_{it} + \sum_{k=1}^K \beta_{kit} X_{kit} + u_{it}, \\ i=1, \dots, N, \quad t=1, \dots, T$$

여기서 i 는 횡단면상의 개별관측치를 의미하며 t 는 시간변수를 나타낸다. N 은 표본에 있는 횡단관측치의 총합이고, T 는 시계열기간을 나타낸다. 따라서 표본의 총수는 NT 개가 된다. (앞에서 든 농가경제조사의 예를 적용할 경우 N 은 3,000이며 T 는 25로 총 75,000개의 표본이 된다.)

² 패널자료 이용시의 장점은 Baltagi (1995)와 Hsiao(1985, 1986, 1989)에 자세히 나와있다.

② 추정계수가 횡단관측치에 따라 변하지만 시간경과에 따라서는 일정한 경우

$$Y_{it} = \alpha_i + \sum_{k=1}^K \beta_k X_{kit} + u_{it}, \\ i=1, \dots, N, t=1, \dots, T$$

③ 기울기(slope) 계수(β)는 일정하나 절편(intercept)의(α) 값은 횡단관측치 및 시간에 따라 변하는 경우

$$Y_{it} = \alpha_{it} + \sum_{k=1}^K \beta_k X_{kit} + u_{it}, \\ i=1, \dots, N, t=1, \dots, T$$

④ 기울기(slope) 계수(β)는 일정하나 절편(intercept)의(α) 값이 횡단관측치에 따라서만 변하는 경우

$$Y_{it} = \alpha_i + \sum_{k=1}^K \beta_k X_{kit} + u_{it}, \\ i=1, \dots, N, t=1, \dots, T$$

⑤ 모든 추정치의 값이 횡단 또는 시간에 관계없이 일정한 경우

$$Y_{it} = \alpha + \sum_{k=1}^K \beta_k X_{kit} + u_{it}, \\ i=1, \dots, N, t=1, \dots, T$$

위의 5가지 경우는 다시 교란항(disturbance term)의 구조(structure)를 어떻게 가정하는가에 따라 아래의 식(i)의 경우와 같은 One-way Error Component Model과 식(ii)와 같은 Two-way Error Component Model로 나눌 수 있다.³

$$(i) \quad u_{it} = \mu_i + \varepsilon_{it}$$

$$(ii) \quad u_{it} = \mu_i + \lambda_t + \varepsilon_{it}$$

패널자료분석에서 이와 같은 교란항의 분해는 다음과 같은 이유에서이다. 추정모델을 ①~⑤와 같이 설정했다는 것은 교란항 역시 횡단관측치 및 시계열자료에 의해 영향을 받는다고 가정하는 것이므로 설정된 모델에서 생략된 혹은 관측 불가능한 변수들의 영향 역시 횡단관측치만에 의한 영향(μ_i)과 시계열만에 의한 영향(λ_t), 혹은 이 둘을 혼합한 영향(ε_{it})으로 구분해 볼 수 있다. 따라서 종종 μ_i 를 individual time-invariant variable, λ_t 를 period(or time) individual-invariant, 그리고 ε_{it} 를 individual time-varying variable로 부르기도 한다.

이러한 오차수정모델(ECM: Error Component Model)은 각각의 교란항을 고정된(fixed) 상수로 볼 것인가 또는 확률변수(random variables)로 볼 것인가에 따라 다시 고정효과모델(Fixed Effect Model)과 임의효과모델(Random Effect Model)로 구분된다.

대부분의 패널자료 경제모델에서는 세 번째와 네 번째의 경우, 즉 기울기는 일정하지만 절편이 개별관측치 또는 시간에 따라 변하는 경우를 상정하여 분석을 시도하는데, 그 이유는 이러한 모델설정의 추정계

³ 이와 같은 이유로 패널자료 모델을 종종 오차

수가 모든 개별관측치 및 시간에 따라 일정하다는 기준모델의 대안으로써 간단하면서도 합리적인 결론을 쉽게 유도 할 수 있기 때문이다. 여기서도 이러한 모델에 한해 그 추정방법에 초점을 맞추어 패널자료 분석방법을 살펴본다.

3. 고정효과모델(Fixed Effect Model)⁴

고정효과모델은 다음의 식으로 설명 가능하다.

$$(1) \quad Y_{it} = \alpha + \sum_{k=1}^K \beta_k X_{kit} + u_{it}, \\ u_{it} = \mu_i + \varepsilon_{it}, \quad \varepsilon_{it} \sim iid(0, \sigma^2_\varepsilon)$$

여기서 i 는 가계, 개인, 기업, 국가 등 횡단관측치를 뜻하며 t 는 시간변수이다. 따라서 i 는 횡단공간을 나타내며, t 는 시간공간을 의미하게 된다. 총 K 개의 설명변수가 있고 특정 시점에서 횡단자료의 수는 N 개이며, 총시계열수는 T 개로 전체 관측치는 NT 개가 된다. 교란항(u_{it})은 다시 관측 불가능한(unobservable) 횡단효과(μ_i)와 나머지 교란항(ε_{it})으로 구분된다. 앞에서 언급한 대로 고정효과모델은 바로 이 μ_i 를 고정된 상수로 취급한다.

⁴ 여기서는 one-way error component model에 중점을 두겠다. two-way error component model의 경우도 이의 자연스러운 연장선상에 있다.

식 (1)을 다시 쓰면 다음과 같다.

$$(1-1) \quad Y_{it} = (\alpha + \mu_i) + \sum_{k=1}^K \beta_k X_{kit} + \varepsilon_{it} \\ = \alpha_i + \sum_{k=1}^K \beta_k X_{kit} + \varepsilon_{it}, \text{ 단 } \alpha_i = \alpha + \mu_i$$

위 식을 대표적인 횡단관측치에 대해서 정리하면 다음과 같다.

$$(2) \quad y_i = \alpha_i l_T + X_i \beta + \varepsilon_i, \quad i=1, \dots, N.$$

여기서 y_i 는 $T \times 1$ 행렬이며(즉 $y_i = [Y_{i1}, Y_{i2}, \dots, Y_{iT}]'$) $l_T = [1, 1, \dots, 1]$ 는 그 원소가 모두 1인 $T \times 1$ 행렬, 그리고 X_i 는

$$X_i = \begin{bmatrix} X_{1i} & X_{2i} & \dots & X_{Ki} \\ X_{1iT} & X_{2iT} & \dots & X_{KiT} \end{bmatrix} \text{인 } T \times K \text{ 행렬, 그리고 } \varepsilon_i \text{는 } T \times 1 \text{ 행렬이다. } (\varepsilon_i = [\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT}]')$$

이제 다음과 같은 $T \times T$ 변환행렬 Q_T ($= I_T - \frac{1}{T} l_T l_T'$)를 식 (2)에 곱하면 식 (2)의 절편 α_i 는 소거되고 설명변수들은 시간에 걸친 평균과의 편차로 다음과 같이 간략히 변환된다.

$$(2-1) \quad Q_T y_i = Q_T X_i \beta + Q_T \varepsilon_i, \\ i=1, \dots, N.$$

즉 $Q_T y_i = y_i - l_T \bar{y}_i$ (단 $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$), $Q_T \alpha_i l_T = 0$, 그리고 $Q_T X_i =$

$X_i - l_T \bar{X}'_i$ 이 되어 종속변수와 설명변수들은 특정 횡단면치에 대하여 그 시계열상의 산술평균을 구해 각각의 값에서 빼준 형태의 새로운 변수로 전환된다⁵. 이제 횡단자료공간에서 식(2-1)을 행렬식 형태로 고쳐 주면

$$(3) \quad \begin{bmatrix} Q_T y_1 \\ \vdots \\ Q_T y_N \end{bmatrix} = \begin{bmatrix} Q_T X_1 \\ \vdots \\ Q_T X_N \end{bmatrix} \beta + \begin{bmatrix} Q_T \varepsilon_1 \\ \vdots \\ Q_T \varepsilon_N \end{bmatrix}$$

또는

$$(3-1) \quad Qy = QX\beta + Q\varepsilon,$$

단 $Q = I_N \otimes Q_T$ ⁶

여기서 \otimes 는 두 행렬간의 크로네커(Kronecker product) 곱을 의미한다.

따라서 식 (3-1)은 통상의 최소자승법(OLS) 적용이 가능하며 기울기의 추정치 $\hat{\beta}$ 는 다음과 같이 구할 수 있다.

$$(4) \quad \hat{\beta} = (X' Q' Q X)^{-1} X' Q' Q y$$

이 때 개별 절편은

$$(5) \quad \hat{\alpha}_i = \bar{y}_i - \bar{X}'_i \hat{\beta}, \quad i=1, \dots, N.$$

단 $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}$, $\bar{X}'_i = [\bar{X}_{1i}, \dots, \bar{X}_{Ki}]$,

$$\bar{X}_{ki} = \frac{1}{T} \sum_{t=1}^T X_{kit}.$$

⁵ 이와 같은 방정식 내부에서 일어난 변환을 내부변환(Within Transformation)이라고 한다.

⁶ 행렬 Q 는 자신과 그 전치행렬이 동일한 idempotent 행렬이다.

로 얻을 수 있다. 물론 당연히 위의 OLS 추정치들은 효율적(efficient)이며 불편추정치(unbiased estimator)가 된다.

아울러 $\hat{\beta}$ 의 분산은

$$(6) \quad \text{Var}(\hat{\beta}) = \sigma_\varepsilon^2 (X' Q X)^{-1},$$

$$\hat{\sigma}_\varepsilon^2 = \frac{(Qy - QX\hat{\beta})'(Qy - QX\hat{\beta})}{N(T-1)-K}$$

으로부터 구할 수 있다.

이와 같이 고정효과모델에서 변환행렬 Q 에 의한 추정방법을 최소자승 덤미변수(LSDV: Least Square Dummy Variable)에 의한 추정방법이라고도 하는데 이는 α_i 를 마치 횡단덤미변수로 취급해 추정해도 위와 동일한 결과를 얻을 수 있기 때문이다. 그러나 이러한 추정방법은 횡단관측치의 개수가 많을 경우 자유도에 영향을 줄 수 있다.

4. 임의효과모델(Random Effect Model)

임의효과모델은 고정효과모델에서 상수로 취급했던 μ_i 를 확률변수로 취급한 경우이다. 따라서 μ_i 와 ε_{it} 간의 분포에 대한 기본가정이 필요하다.

$$(7) \quad Y_{it} = \alpha + \sum_{k=1}^K \beta_k X_{kit} + \mu_i + \varepsilon_{it}$$

대개의 경우 이들 가정은 다음과 같다.

$$(8-1) \quad \text{모든 } i, t \text{ 에 대해 } E(\mu_i) =$$

$$E(\varepsilon_{it}) = 0$$

(8-2) 모든 i, j, t 에 대해 $E(\mu_i \varepsilon_{jt}) = 0$

(8-3) 모든 i, j 에 대해

$$E(\mu_i \mu_j) = \begin{bmatrix} \sigma_\mu^2 & (i=j) \\ 0 & (i \neq j) \end{bmatrix}$$

(8-4) 모든 i, j, t, s 에 대해

$$E(\varepsilon_{it} \varepsilon_{js}) = \begin{bmatrix} \sigma_\varepsilon^2 & (i=j, t=s) \\ 0 & (\text{기타}) \end{bmatrix}$$

즉 교란항 μ_i 와 ε_{it} 간에 서로 확률적으로 독립이며 각각은 자체의 동분산을 갖는다고 가정한다. 위의 식 (7)를 행렬을 이용해 표현하면 다음과 같다.

$$(9) \quad y = X\beta + (I_N \otimes I_T)\mu + \varepsilon,$$

단 $y' = [Y_{11} Y_{12}, \dots, Y_{NT}]$, $X' = [X_1 X_2, \dots, X_N]$, X_i 는 시간변수에 대해 X_{kit} 를 집약한(stacking) $T \times K$ 행렬이다. 따라서 임의효과모델은 다음의 식(10)과 같이 행렬을 이용해 간단히 표현할 수 있다.

$$(10) \quad y = X\beta + u,$$

$$u = (I_N \otimes I_T)\mu + \varepsilon$$

앞서 설정한 가정을 이용하면 교란항 u 의 분산-공분산행렬은 다음과 같이 표현된다.

$$(11) \quad E(uu') = (\sigma_\mu^2 + \sigma_\varepsilon^2) \left(\frac{\sigma_\mu^2}{(\sigma_\mu^2 + \sigma_\varepsilon^2)} (I_N \otimes I_T l_t') + \frac{\sigma_\varepsilon^2}{(\sigma_\mu^2 + \sigma_\varepsilon^2)} I_{NT} \right) = \sigma^2 (\rho (I_N \otimes I_T l_t') + (1 - \rho) I_{NT})$$

$$= \sigma^2 Q$$

$$\text{단 } Q = \rho (I_N \otimes I_T l_t') + (1 - \rho) I_{NT},$$

$$\sigma^2 = (\sigma_\mu^2 + \sigma_\varepsilon^2), \quad \rho = \frac{\sigma_\mu^2}{(\sigma_\mu^2 + \sigma_\varepsilon^2)}.$$

특히 Q 는 block diagonal 행렬로 주대각 원소는 다음과 같은 행렬 A 로 표현할 수 있으며, 행렬 A 는 그 주대각원소는 1이고 나머지 부분의 원소는 ρ 인 $T \times T$ 행렬이다.

$$Q = \rho (I_N \otimes I_T l_t') + (1 - \rho) I_{NT} = I_N \otimes [\rho l_t l_t' + (1 - \rho) I_T] = I_N \otimes A$$

흔히 이 ρ 를 내부상관계수(intracorrelation)라고 부르는데 그 정확한 의미는 교란항의 총변동중 횡단치로부터 발생한 변동비율을 의미하는 것으로 ρ 가 0값을 갖게 되면 이는 패널분석결과와 전체자료를 단순히 회귀분석한, 즉 관측치를 횡단 또는 시계열 구분없이 동일하게 취급하여 회귀분석한 결과와 같게 된다. 반면 $\rho=1$ 이면 패널결과는 횡단분석기법에 의한 결과와 같게 된다. 따라서 이론적으로 ρ 값은 0과 1사이에 놓이게 된다.

한편 임의효과모델의 식 (10)은 일반최소자승법(GLS)을 적용할 수 있고, 따라서 β 의 추정치는 다음과 같이 구할 수 있다.

$$(12) \quad \widehat{\beta}_{GLS} = (X' Q^{-1} X)^{-1} (X' Q^{-1} y).$$

그러나 임의효과모델의 실제응용에 있어서 일반최소자승법을 적용하는 예는 드물

다. 그 이유는 Ω^{-1} 을 계산하기 위해서는 A 의 역행렬을 먼저 구해야 하는데 일반적으로 행렬 A 의 차수(rank)가 상당히 크기 때문에 그 역행렬 A^{-1} 을 계산하기가 쉽지 않기 때문이다. 이러한 이유로 일반최소자승법 보다는 행렬 A 가 대칭행렬이고 따라서 그에 상응하는 직교행렬(orthogonal matrix)이 존재하므로 이를 이용해 원자료를 변형하여 통상의 최소자승법(OLS)을 적용한다.

이를 보다 자세히 설명하면 행렬 A 는 $T \times T$ 행렬로 그 T 개의 특성근(characteristic root)은 1개의 $\xi = 1 - \rho + T\rho$ 와 $(T-1)$ 개의 $\eta = 1 - \rho$ 로 이루어져 있다. 이 사실을 이용하여 임의효과모델을 대표적인 횡단관측치에 대해 정리한 다음 양변에 $A^{-1/2}$ 을 곱하여 모델식을 사전에 변형하는 것이다. 이를 보다 구체적으로 보면

$$\begin{aligned} y_i^* &= A^{-1/2}y_i = \xi^{-1/2} \begin{pmatrix} \bar{y}_i \\ \dots \\ \bar{y}_i \end{pmatrix} \\ &+ \eta^{-1/2} \begin{pmatrix} y_{i1} - \bar{y}_i \\ \dots \\ y_{iT} - \bar{y}_i \end{pmatrix} = \begin{pmatrix} y_{i1}^* \\ \dots \\ y_{iT}^* \end{pmatrix} \end{aligned}$$

단 $\bar{y}_i = \frac{1}{T} \sum_{t=1}^N y_{it}$ 이다.

따라서 변형된 y_{it}^* 는 행렬 A 의 특성근의 평방근(square root)의 역수를 가중치로 하는 원자료의 횡단평균치(individual mean)

\bar{y}_i 와 횡단평균치와의 편차, $(y_{it} - \bar{y}_i)$ 의 가중평균치로 해석할 수 있다. 마찬가지로

$$X_i^* = A^{-1/2}X_i = \xi^{-1/2} \begin{pmatrix} 1 \bar{X}_{2i}, \dots, \bar{X}_{Ki} \\ \dots \\ 1 \bar{X}_{2i}, \dots, \bar{X}_{Ki} \end{pmatrix} + \eta^{-1/2} \begin{pmatrix} 0 X_{2i1} - \bar{X}_{2i}, \dots, X_{K1} - \bar{X}_{Ki} \\ \dots \\ 0 X_{2iT} - \bar{X}_{2i}, \dots, X_{KT} - \bar{X}_{Ki} \end{pmatrix}$$

이 되고 변형된 식은 $y_i^* = X_i^* \beta + u_i^*$ $i = 1, 2, \dots, N$ 이고 $E(u_i^* u_i^{*\prime}) = \sigma^2 I_T$ (단 $\sigma^2 = \sigma_\mu^2 + \sigma_\epsilon^2$)

$$(13) \quad \begin{pmatrix} y_1^* \\ \dots \\ y_N^* \end{pmatrix} = \begin{pmatrix} X_1^* \\ \dots \\ X_N^* \end{pmatrix} \beta + \begin{pmatrix} u_1^* \\ \dots \\ u_N^* \end{pmatrix}$$

또는 $Y^* = X^* \beta + u^*$.

그러므로 변형된 식 (13)은 통상의 최소자승법만을 적용해 일반최소자승법으로 구한 추정량을 구할 수 있는 장점이 있다.

한편 실제 패널자료분석시 Ω 또는 행렬 A 를 모르기 때문에 이를 추정해야 하는 어려움이 따른다. 이의 핵심은 ρ 값의 추정 문제로 최종적으로는 각각의 교란항을 어떻게 추정할 것인가에 초점이 모아지게 된다. 현재까지 다양한 추정방법이 알려져 있으며 중요한 것은 점근적으로(asymptotically) 바람직한 성질을 갖는 ρ 의 일치추정량(consistent estimator)을 구하는 것이다. 패널자료를 다루는 대부분의 계량경제학책

이 이 부분에 대해서 자세히 다루고 있지 않으며, 또한 이론적으로 문제가 있는데도 불구하고 그대로 사용하고 있는 경우도 있다. 그 대표적인 예가 Green과 Judge (1993,1995, Limdep Manual Chapter 17)의 접근방법이다. ρ 의 추정치 $\hat{\rho}$ 을 구하기 위해서 σ_{μ}^2 , σ_{ϵ}^2 의 추정치가 필요한데, Green 은 $\hat{\sigma}_{\epsilon}^2$ 을 고정효과모델에서의 잔차항 (residual)을 이용해 계산하고 $\hat{\sigma}_{\mu}^2$ 은 다음과 같이 계산하고 있다.

$$(14) \quad \hat{\sigma}_{\mu}^2 = \hat{\sigma}_{*}^2 - \frac{\hat{\sigma}_{\epsilon}^2}{T}.$$

여기서 $\hat{\sigma}_{*}^2$ 은 $e_* = \bar{Y}_i - \bar{X}_i\beta$, $i=1, 2, \dots, N$ 로서 (단 $\bar{Y}_i = \frac{1}{T} \sum Y_{it}$, $\bar{X}_i = \frac{1}{T} \sum X_{kit}$), 시간공간에서 횡단관측치의 평균 값(이것은 사실상 횡단분석기법과 유사하다)의 회귀분석(이를 종종 between group regression이라고 한다)으로부터 나온 잔차 항을 이용해 계산한다. 그러나 식 (14)에서 알 수 있듯이 Green과 Judge의 접근방법은 $\hat{\sigma}_{\mu}^2$ 의 값이 음이 될 수 있어(이는 곧 $\hat{\rho}$ 이 음의 값이 됨을 의미) 이론적으로 문제가 있다.(특히 시계열 표본수가 크지 않거나 lag 변수가 있는 경우 그 가능성이 높아진다) 대부분의 계량경제 프로그램에서 이 방법을 채택하고 있는 것도 또 다른 문제중의 하나이다. 이에 반해 Balestra-Nerlove (1966)는 instrumental variables를 이용해 $\hat{\rho}$ 를 계산하는 방법을 제시했으며, Nerlove(1971)는 다시 Monte Carlo 실험을

통해 Balestra-Nerlove의 방법으로 $\hat{\rho}$ 가 불일치추정량이 됨을 밝히면서 그 대안으로서 고정효과모델과 pooled모델(횡단 또는 시계열에 구분없이 모든 관측치를 이용해 회귀분석하는 모델)에서의 잔차항을 이용해 계산하는 방법을 제시하고 있다.

$$(15) \quad \hat{\sigma}_{\mu}^2 = \frac{1}{N} \sum (\bar{Y}_i - \bar{X}_i\hat{\beta})^2,$$

$$\hat{\sigma}_{\epsilon}^2 = \frac{1}{T-1} \sum (\epsilon_{it} - \bar{\epsilon}_i)^2,$$

$$\bar{\epsilon}_i = \frac{1}{T} \sum \epsilon_{it}$$

사실 이를 3가지 방법은 모두 고정효과모델의 추정결과를 이용하고 있고 여기에서 결과되는 $\hat{\beta}$ 의 값이 불일치추정량이기 때문에 어느 방법을 사용하더라도 $\hat{\rho}$ 값이 불일치 추정량이 될 수밖에 없다. 문제는 자료의 특성을 파악, 즉 횡단관측치의 변동과 시계열상의 변동을 잘 관측하여 적절한 방법을 선택해야 하고 이는 전적으로 연구자의 의도에 따라 결정될 수밖에 없다. 그러나 적어도 이론적으로 음의 값을 갖게 될 가능성성이 있는 또는 매우 0에 가까운 값이 나올 수 있는 Green-Judge의 방법은 시계열표본수가 많지 않는 한 사용하지 않는 것이 최근의 패널분석연구의 추세이다.

5. 우도추정(Maximum Likelihood Estimation)

교란항이 정규분포를 한다고 가정하면 우도함수(likelihood function)는 어렵지 않

계 도출할 수 있다. 즉 앞서의 임의효과모델로부터, $y = X\beta + u$, $u = (I_N \otimes l_T) \mu + \varepsilon$ 로부터 (단 $u \sim N(0, \sigma^2 \Omega)$)

$$(16) \quad L(\beta, \sigma_\mu^2, \sigma_\varepsilon^2 | Y_{11}, \dots, Y_{1T}, Y_{21}, \dots, Y_{N1}, \dots, Y_{NT}) \\ = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{NT} |\Omega|^{-1/2} \\ \exp \left(-\frac{1}{2\sigma^2} (y - X\beta)' \Omega^{-1} (y - X\beta) \right)$$

실제 이용에 있어서는 앞서의 임의효과모델에서 본 바와 같이 원자료의 변형을 이용한다. 즉 $\Omega = I_N \otimes A$ 이고 $|\Omega| = |I_N|^N$ $|A|^N = |A|^N$ 이며, A 의 특성방정식의 근은 $1 - \rho + T\rho$ 와 $1 - \rho$ 이다. 또한 임의 행렬의 결정식(determinants)은 그 행렬의 특성방정식의 근의 곱이기 때문에

$|A| = [(1 - \rho + T\rho)(1 - \rho)^{T-1}]^N$ 이고 따라서 $(y - X\beta)' \Omega^{-1} (y - X\beta)$ 은 다음과 같이 표현된다.

$$(y - X\beta)' \Omega^{-1} (y - X\beta) \\ = \sum_{i=1}^N (y_i - X_i \beta)' A^{-1} (y_i - X_i \beta) \\ = \sum_{i=1}^N (A^{-1/2} y_i - A^{-1/2} X_i \beta)' \\ (A^{-1/2} y_i - A^{-1/2} X_i \beta)$$

그리므로 로그우도함수(log likelihood function)는

$$(17) \quad \log L(\beta, \sigma_\mu^2, \sigma_\varepsilon^2 | Y_{11}, \dots, Y_{1T}, Y_{21}, \dots, Y_{N1}, \dots, Y_{NT})$$

$$= -\frac{NT}{2} \log 2\pi - \frac{NT}{2} \log \sigma^2 \\ - \frac{N}{2} \log (1 - \rho + T\rho) \\ - \frac{N(T-1)}{2} \log (1 - \rho) \\ - \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - X_{kit} \beta)^2$$

한편 $\beta, \sigma_\mu^2, \sigma_\varepsilon^2$ 의 우도추정량은 위의 로그우도함수의 값을 최대로 만드는 1차 조건을 연립하여 풀어 구할 수도 있으나 그 결과가 매우 복잡하기 때문에 보통은 Grid-search 방식을 사용한다. 즉 먼저 σ^2 를 중심으로 우도함수를 단순화시킨다 (흔히 이때의 단순화된 우도함수를 concentrated log likelihood function이라고 한다) 즉 위의 식(17)을 σ^2 에 대해서 미분하여 σ^2 의 추정치를 구한 후 그 추정치를 다시 식(17)에 대입하여 정리하면

$$(18) \quad \log L(\beta, \rho | \dots) \\ = -\frac{NT}{2} (\log 2\pi + 1) \\ - \frac{NT}{2} \log \hat{\sigma}^2 - \frac{N}{2} \log (1 - \rho \\ + T\rho) - \frac{N(T-1)}{2} \log (1 - \rho) \\ \text{단, } \hat{\sigma}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (Y_{it} - X_{kit} \beta)^2}{NT}.$$

따라서 식 (18)에서 미지의 모수는 (17)과는 달리 ρ 와 β 로 줄어들게 되고 또한 ρ 값이 이론적으로 0과 1사이에 위치하고 있다는 것을 활용하여 ρ 값을 0과 1사이에

서 적절히 설정해 놓으면 (예를 들면 처음에 ρ 를 0.1로 고정시켜 놓고) 결국 구하는 모수는 β 뿐이며 우도함수를 최대로 만드는 $\hat{\beta}$ 를 추정하기가 그만큼 용이해 진다. 이러한 일련의 과정을 ρ 값을 0에서 1까지 변화시키면서 반복하여 우도함수의 값을 최대로 만드는 $\hat{\beta}$ 값을 찾는 것이다.

6. 결 론

근래에 들어 경제분석시 다양하게 쓰이는 패널자료분석에 대하여 그 장점과 모델링방법 및 그에 따른 추정방법을 비교·검토해 보았다. 앞서 언급한대로 패널자료분석은 기존의 횡단면분석이나 시계열분석만으로는 곤란한 특정 횡단면자료치(혹은 시계열자료치)의 특성을 감안하여 보다 정확한 계량모형의 설정 및 추정을 위한 도구로서 이미 경제학 여러 분야에서 다양하게 이용되고 있다.

여기서 언급하지 못한 동태적인 패널자료모델과 기타 교란항 자체에 자기상관 및 이질성이 있는 경우, 그리고 연립방정식체계에 있어서의 패널 자료의 분석 등 패널자료분석에 대한 다양한 계량경제학적인 연구가 아직도 진행중이며 이와 관련해서는 또다

른 기회를 이용하여 소개할 예정이다.

참 고 문 헌

- Balestra, P. and M. Nerlove. 1966. "Pooling Cross-Sectional and Time-Series Data in the Estimation of a Dynamic Economic Model: The Demand for Natural Gas." *Econometrica* 34: 585-612.
- Baltagi B., H. 1995. *Econometric Analysis of Panel Data* New York: John Wiley & Sons.
- Hsiao Cheng. 1996. *Analysis of Panel Data* 2nd. eds. Cambridge University Press.
- Judge, G., R.C. Hill, W. Griffiths, H. Lutkepohl, and T-C. Lee. 1988. *Introduction to the Theory and Practice of Econometrics* 2nd. eds. New York: John Wiley & Sons.
- Maddala, G.S. 1971. "The Use of Variance Component Models in Pooling Cross Section and Time Series Data." *Econometrica* 39: 341-3585.
- Mundlak, Y. 1978. "On the Pooling of Cross Section and Time Series Data." *Econometrica* 46: 69-86
- Nerlove, M. 1971. "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections." *Econometrica* 39: 359-382.