

패널자료의 분석: 이론과 실제

류근관

서울대학교 경제학부 교수
ezstat.co.kr, ryu@snu.ac.kr

발표순서

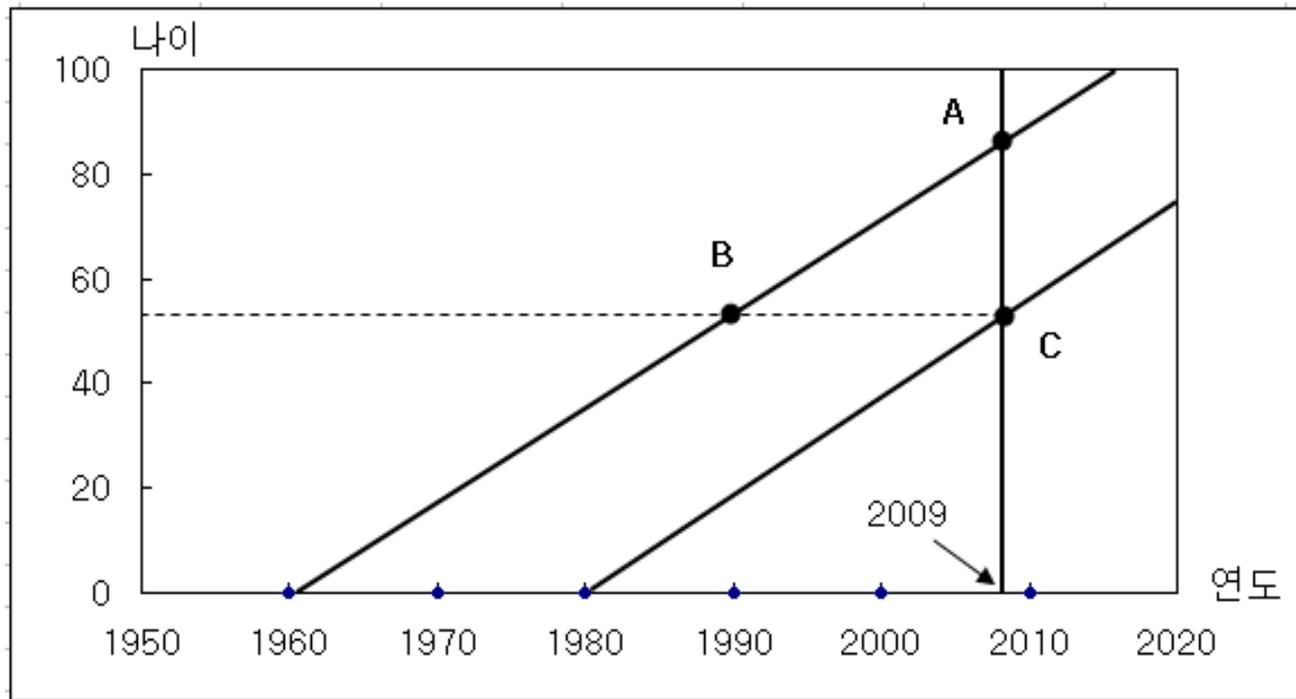
1. 패널자료 이용의 장점
2. 패널모형의 기본개념
3. 패널모형의 확장
 - 1) 내생성
 - 2) 동태적 모형
 - 3) 패널이항모형
 - 4) 동태적 패널이항모형
 - 5) 기타 이슈들
4. 패널자료 모형의 사용 사례
5. 맺음말

패널자료 이용의 장점

- $w_{it} = \alpha + \beta x_{it} + \gamma(\text{entrycohort})_{it} + \delta(\text{USage})_{it} + u_{it}$
- 횡단면 자료를 이용해서는 γ 와 δ 식별 불가능
이는 횡단면 자료의 경우
 $(\text{entrycohort})_{it} + (\text{USage})_{it} = t = (\text{const})$ 이므로
두 변수간 완벽한 선형종속 관계 성립하기 때문
- 패널자료 이용해야 코호트효과 (cohort effect)와 동화효과 (assimilation effect) 구분 가능

패널자료 이용의 장점

- 횡단면자료와 패널자료의 차이: c.f. 렉시스 도표



- 횡단면 자료: A와 C의 비교 (코호트효과, 동화효과 혼재)
- 패널자료 이용하여 A, B 비교하면 동화효과를, B, C 비교하면 코호트효과를 알 수 있음

패널자료 이용의 장점

- $y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + \varepsilon_{it}$
 - study of dynamics (αy_{it-1}) possible
 - control of unobserved heterogeneity (η_i) possible
 - in cross-section, η_i and ε_{it} can neither be separately specified nor separately identified

패널자료 이용의 장점

- 패널자료 이용하면 미관측 이질성 통제 가능
- 예시 1: 소유구조와 기업가치간의 관계

$$y_{it} = \alpha + \beta x_{it} + \gamma(\text{ownership})_{it} + \eta_i + \varepsilon_{it}$$

- 소유구조는 한 기업 내에서도 시간에 따라 변화
- 기업별로 미관측 이질성 η_i 존재
- 소유구조의 내생성 Demsetz and Lehn (JPE 1985)
- 정보 비대칭성 하 소유구조의 history 고려해야
Lee and Ryu (mimeo 2008)
Ryu and Yoo (mimeo 2009)

패널자료 이용의 장점

- 예시 2: 금융발전과 경제성장의 관계

Rajan and Zingales (NBER 1996)

- 국가간 금융발전의 정도와 경제성장률을 직접 비교하는 cross-sectional analysis는 국가간 차이를 적절히 통제하지 못함
- 이때 (i =국가, t =산업)의 형태로 패널자료 분석방법을 사용하면 국가간 차이 및 산업간 차이를 통제하기 훨씬 용이함

패널자료 이용의 장점

참고: 패널자료 없는 경우의 단순비교

	금융선진국	금융후진국
중공업	A	B
경공업	C	D

국가간 비교 : A-B (국가간 차이 통제 안됨)

산업간 비교 : A-C (산업간 차이 통제 안됨)

이상은 단순비교의 문제점

패널자료 이용하여 이중비교하면 유리함

패널자료 이용의 장점

(i =국가, t =산업) 형태의 패널자료 이용한 이중비교

	금융선진국	금융후진국
중공업	A	B
경공업	C	D

- 이중차분 (difference in difference) 방법 :
 $\tau = (A-B)-(C-D)$ 또는 $(A-C)-(B-D)$
- $A-B$ =금융효과+국가간 차이 ($\tau + \gamma$)
 $A-C$ =금융효과+산업간 차이 ($\tau + \delta$)
 $C-D$ =국가간 차이 (γ), $B-D$ =산업간 차이 (δ)
- $y_{it} = \alpha + \beta x_{it} + \gamma d_i + \delta d_t + \tau(d_i d_t) + u_{it}$

패널자료 이용의 장점

패널자료 이용한 이중비교 사용 사례

- 군납 담합으로 인한 손해액 산정
 y =리터당 낙찰가 또는 로그 낙찰가
 i =군납, 비군납, t =담합시기, 비담합기
- 광고허용이 제품가격에 미치는 효과
 y =리터당 주류판매가격
 i =RI 주, MA 주, t =주류광고 허용 이전, 허용 이후
- 상해보상이 유급휴가일수를 증가시키는 효과
 y =상해로 인한 연간 유급휴가일수
 i =고임금근로자, 저임금근로자, t =보상한도 인상 전, 후

패널모형의 기본개념

- $y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + \varepsilon_{it}$
- static ($\alpha = 0$) vs. dynamic ($\alpha \neq 0$)
- x_{it} and η_i are correlated or not
- x_{it} strictly exogenous, currently endogenous, or predetermined w.r.t. ε_{it}
- Data is balanced or not (panel attrition issue)

패널모형의 기본개념

- 고정효과 대 확률효과 (η_i 처리의 문제)
- incidental parameters problem :
 η_i 를 추정대상으로 파악하는 경우 too many parameters to estimate! Bias for small T.
- N과 T의 상대적 크기
 - typical micro data (PSID, KLIPS) : $N \gg T$
 - typical macro data (Penn world table) : $N \sim T$

패널모형의 기본개념

- $y_{it} = \alpha + \beta x_{it} + \eta_i + \varepsilon_{it}$
 - **Pooled OLS** : η_i 빼고 전체자료 이용하여 OLS
 - **Fixed effect** : η_i 를 N-1개의 개인별더미변수로 처리하고 모형 추정. 이는 mean-deviated data 에 OLS 적용하는 것과 동일함
 - **Random effect** : η_i 를 x_{it} 와 독립인 확률변수로 보고 GLS 적용. 이는 “fractional mean-deviated” data에 OLS 적용하는 것과 동일함.

Subtract $(1 - \theta) \cdot (\text{mean})$, $\theta = \sqrt{\sigma_\varepsilon^2 / (T\sigma_\eta^2 + \sigma_\varepsilon^2)}$

패널모형의 확장: 내생성

- $y_{it} = \alpha + \beta x_{it} + \eta_i + \varepsilon_{it}$

- Random effect : η_i 를 x_{it} 와 독립이 아닌 확률 변수로 보고 IV/GLS 적용. Hausman and Taylor (Econometrica, 1981).

이는 “fractional mean-deviated” data (GLS 변환)에 IV 또는 GMM적용하는 것과 동일함.

이때 설명변수 중 외생변수 일체, time-varying하는 변수에 대해서는 이들 변수의 mean-deviations 일체를 도구변수로 사용함. time-constant 내생성만 취급

패널모형의 확장: 동태적 모형

- $y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + \varepsilon_{it}$
- Within (DVLS) vs. Differencing

$$y_{it} - \bar{y}_i = \alpha(y_{it-1} - \bar{y}_{i(-1)}) + \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

$$y_{it} - y_{it-1} = \alpha(y_{it-1} - y_{it-2}) + \beta(x_{it} - x_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1})$$

-Applying OLS yields biased estimation

-Easier to find IVs in the case of differencing transformation. Think why. Hint: Lagged y's.

패널모형의 확장: 동태적 모형

- $y_{it} - y_{it-1} = \alpha(y_{it-1} - y_{it-2}) + \beta(x_{it} - x_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1})$
 - Anderson and Hsiao (JASA 1981, JOE 1982)
Use y_{it-2} as IV.
 - Arellano and Bond (RES 1991)
Use $y_{it-2}, y_{it-3}, y_{it-4}, \dots, y_{i1}$ as IVs.
 - Depending on assumptions on the initial value y_{i1} and also on the relation between x_{it} and ε_{is} (StExog, PreDet, or CurrEnd), available moment conditions change.

패널모형의 확장: 패널이항모형

- $y_{it} = \beta x_{it} + \eta_i + \varepsilon_{it}$, $y=0, 1$ (binary dep. var.)
- Linear probability model (nothing new!)
- Panel probit
- Panel Logit (easier than panel probit)
- Index models (semi-parametric)
- Parameter or Partial effect (PE)

패널모형의 확장: 패널이항모형

- Panel probit model

$$y_{it} = \beta x_{it} + \eta_i + \varepsilon_{it}$$

$$P(y_{it} = 1 \mid x_{it} = x, \eta_i = \eta) = \Phi(\beta x + \eta)$$

- Partial effect at (x, η)

$$\theta_j(x, \eta) = \partial P(y_{it} = 1 \mid x_{it} = x, \eta_i = \eta) / \partial x_{jit}$$

$$\theta_j(x, \eta) = \beta_j \phi(\beta x + \eta)$$

패널모형의 확장: 패널이항모형

- Problem is :

We do not have information on η .

- PEA : $\theta_j(x, \eta = 0) = \beta_j \phi(\beta x)$, $E\eta = 0$ used
- APE :

$$E\theta_j(x, \eta) = \partial E\Phi(\beta x + \eta) / \partial x_j = \beta_\eta \phi(\beta_\eta x)$$

where $\eta \sim N(0, \sigma_\eta^2)$ is assumed,

$$\text{and } \beta_\eta = \beta / \sqrt{1 + \sigma_\eta^2}$$

패널모형의 확장: 패널이항모형

- From

$$P(y_{it} = 1 \mid x_{it} = x) = E\Phi(\beta x + \eta) = \Phi(\beta_{\eta} x)$$

- β_{η} and APE are identified whereas β and PEA are not.
- Focus on parameter identification is wrong headed. Focus on identifying APE
- In the case of local APE, use $f(\eta \mid x)$.
Altonji and Matzkin (Econometrica 2005)

패널모형의 확장: 패널이항모형

$$P(y_{it} = 1 | x_{it}, \eta_i) = \Phi(\beta x_{it} + \eta_i)$$

- Treating unobserved heterogeneity η_i as well as β as parameters to estimate, result in bias (incidental parameters problem).
- Hahn and Newey (Econometrica 2004) proposed bias reduction through jackknife techniques

$$\hat{\beta}_{HN} = T\hat{\beta} - (T-1) \cdot \text{average } \hat{\beta}_{(t)}$$

패널모형의 확장: 동태적 패널이항모형

- Chay and Hyslop (UCB WP 2001)

$$P(y_{it} = 1 | x_{it}, y_{it-1}, \eta_i) = \Phi(\alpha y_{it-1} + \beta x_{it} + \eta_i)$$

- Model

$$\eta_i = \gamma + \delta y_{i1} + \tau \bar{x}_i + \varepsilon_i \text{ and } \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

- Then, we have $P(y_{it} = 1 | x_{it}, y_{it-1}, \bar{x}_i, y_{i1})$

$$= \Phi([\alpha y_{it-1} + \beta x_{it} + \gamma + \delta y_{i1} + \tau \bar{x}_i] / \sqrt{1 + \sigma_\varepsilon^2})$$

패널모형의 확장: 동태적 패널이항모형

- Running pooled probit of y_{it} on y_{it-1}, x_{it} , constant, y_{i1}, \bar{x}_i will identify all the parameters up to a common scale in the form of the original parameters divided by $\sqrt{1 + \sigma_\varepsilon^2}$

Denote these estimated parameters as

$$\alpha_\varepsilon, \beta_\varepsilon, \gamma_\varepsilon, \delta_\varepsilon, \tau_\varepsilon.$$

- We can compute APE of previous LFP status (y_{it-1}) on the current LFP (y_{it})

패널모형의 확장: 동태적 패널이항모형

- APE : $\Phi(\alpha_\varepsilon \cdot 1 + \beta_\varepsilon x_{it} + \gamma_\varepsilon + \delta_\varepsilon y_{i1} + \tau_\varepsilon \bar{x}_i)$
 $-\Phi(\alpha_\varepsilon \cdot 0 + \beta_\varepsilon x_{it} + \gamma_\varepsilon + \delta_\varepsilon y_{i1} + \tau_\varepsilon \bar{x}_i)$
- APE averaged over (i, t) in the sample = 0.26
- The same APE without accounting for unobserved heterogeneity $\eta_i = 0.84$.
- Much of the observed persistency in women's LFP is explained by unobserved heterogeneity than by genuine state dependency.
- Still, some evidence of state dependency

패널모형의 확장: 기타 이슈들

- Sample selection in panel data : panel attrition
Wooldridge (JOE 1995), Ryu (Cambridge Univ Press 2001)
- Panel count data models (Poisson model)

$$E(y_{it} | x_{it}, y_{it-1}, \eta_i) = \eta_i \exp(\alpha y_{it-1} + \beta x_{it} + \gamma)$$

$$E(y_{it} | x_{it}, y_{it-1}, \bar{x}_i, y_{i1}, \varepsilon_i) = \varepsilon_i \exp(\alpha y_{it-1} + \beta x_{it} + \gamma + \delta y_{i1} + \tau \bar{x}_i)$$

Use Gamma distribution for unobserved heterogeneity (η_i, ε_i) with mean equal to one.

- Duration analysis (Fallick & Ryu, REStat 2007)
- Event studies

패널자료 모형의 사용 사례

국내외 패널자료 예시

- KLIPS (한국노동패널)
제안: 3년에 한번씩 가계 저축, 금융자산 실태 조사 추가
- PSID, NLS, NLSY
- BHPS, ECHP
- JPSC: Japanese panel survey on consumers
- Compustat, CRSP
- Scanner data (마케팅 분야)
- 신용카드 보유자의 월별 업종별 지출액 패널 자료

패널자료 모형의 사용 사례

국내외 패널자료 예시

- 통신회사(전화, 인터넷)가 보유한 개인별 통신 사용자료
- 전기 등 각종 공과금의 월별 지출액 자료
- 은행의 기업/가계별 대출 및 사후 관리 자료
- 기업의 신용등급 변화 자료
- 결혼정보회사의 개인별 matching history 자료

패널자료 모형의 사용 사례

국내외 패널자료 예시

- Multicountry analysis of growth and convergence
 - Penn world table
 - Barro (QJE 1991) initiated growth empirics
 - Islam (QJE 1995) panel approach
 - Temple (JEL 1998) new evidences

패널자료 모형의 사용 사례

- Granger causality between savings and growth

$$g_{it} = \alpha g_{it-1} + \beta s_{it-1} + \eta_i + \varepsilon_{it}$$

$$s_{it} = \alpha s_{it-1} + \beta g_{it-1} + \eta_i + \varepsilon_{it}$$

- The former equation tests whether savings Granger causes growth, whereas the latter whether growth Granger causes savings
- One can use similar models to test whether inflation rate Granger causes wage increase (노조) or the other way around (전경련).

패널자료 모형의 사용 사례

- “퍼주기” 사용에 있어서 국내 일간지간의 선도-추종관계
김재준, 김병철, 류근관 (한국언론학보 2009)
 - KINDS의 기사 검색 기능 활용
 - 일간지별로 ‘퍼주기’라는 미디어 유행어의 사용
빈도를 2000년 4월부터 2007년 12월에 걸쳐 조사
 - 그랜저 인과관계 검증 방법 사용
 - 분석결과, 조선, 중앙, 동아일보 등 3대 일간지가 다른
일간지 및 경제지의 퍼주기 사용을 초래하는 경향 존재
 - 특히 조선일보의 선도가 두드러짐
 - 퍼주기 언급은 한 달 정도의 시차를 두고 대북정책 지지
율에 부정적 영향 미침

패널자료 모형의 사용 사례

- firm's investment behavior : genuine dynamics or unobserved heterogeneity?

$$(I / K)_{it} = \alpha(I / K)_{it-1} + \beta x_{it} + \eta_i + \varepsilon_{it}$$

- Estimates of α
 - 0.27 by OLS (upward bias)
 - 0.01 by fixed effect (downward bias)
 - 0.16 by IV (GMM) to differenced equation

패널자료 모형의 사용 사례

- 노동조합이 임금에 미치는 효과
조동훈 (노동경제논집 2008)
 - KLIPS 자료 사용
 - static panel data model
 - 횡단면 분석 4.6% 임금인상 효과
 - FE panel analysis 2.1% 임금인상 효과
 - 제안: *try dynamic panel data model*

패널자료 모형의 사용 사례

- **재벌과 비재벌간 장기간에 걸친 생산효율성 비교**
Choo, Lee, Ryu, and Yoon (EDCC 2009)

- 기업별 패널자료 사용

- 재벌 계열사와 그렇지 않은 회사간 장기간에 걸친 생산상의 효율성 비교

- 1997년의 아시아 금융/외환위기 이후 재벌의 효율성이 그 이전에 비해 증가

- 1997년 이전에는 비효율적 투자가 비효율성의 주요 원인이었음

- c.f. 윤용 (SNU Ph.D thesis, 지도교수 류근관):

- 진입/퇴출, 회사규모 감소한 상태에서 장기간에 걸친 재벌과 비재벌의 성과 비교

패널자료 모형의 사용 사례

• 성형수술의 효과 Lee and Ryu (mimeo 2009)

- 실험에 참가한 서울대생이 주어진 사진 자료 각각에 대해 평가한 자료 (i =평가자, j =대상 사진)
- unbalanced panel 자료로 실험설계 (의도적)
- 대상 사진에는 보통사진, 성형 전/후 사진 등 공존
- $y_{ij} = \alpha_i + \beta_j + \gamma(\text{성형수술})_j + \varepsilon_{it}$
- 성형의 효과 γ 는 남성보다 여성에게서 높게 나타나는 것으로 추정됨
- 성형수술 투자 => 용모 개선 => 결혼시장에서 보상

패널자료 모형의 사용 사례

- **카드보유자의 월별 업종별 지출액 패널자료**

홍설희, 홍정민, SNU Ph.D Thesis (진행중, 류근관 지도)

- 경기국면에 따라 월별지출액 변화가 소득/직업간 업종별로 차별적 패턴 보이는지 분석
- 이를 경기예측 지표로 활용
- 이를 sector portfolio 투자전략에 활용
- 개인카드보유자의 월별 업종별 지출액 패널자료를 이들의 대출자료와 연계하여 개인별 지출액 변화 패턴이 대출연체 여부 예측에 도움 주는지 분석

맺음말

- “All Models are Wrong. Some are Useful.”
(George Box). Panel data are very useful.
- “Panel Data is like Red Wine: The Longer, the Better.”
Length of KLIPS=11 by year 2008.
- STATA offers many useful panel estimation routines.
Other panel programs written in Gauss, SAS etc.
available on internet
- Be smart panel analysts, not stupid panel discussants.

맺음말 : 실존주의 vs. 실용주의

- To do is to be (Descartes)
- To be is to do (Voltaire)
- Do be do be do (Frank Sinatra)

- It is better to solve the right problem approximately than the wrong problem exactly (John Tukey)