

# 노동시장정책 평가방법론 및 다부문 거시산업모형 DB 구축

윤윤규 · 홍민기 · 강창희

이진면 · 이해정 · 고영우

연구기관 : 한국노동연구원



# 목 차

## 제1부 노동시장정책 평가방법론 (홍민기 · 강창희 · 이해정 · 고영우)

제1장 서론 .....	3
제2장 용어에 대한 정의 .....	6
제1절 결과에 대한 정의 .....	6
제2절 정책에 참여하는 방식 .....	7
제3절 평균처방효과(Average Treatment Effects) .....	9
제3장 회귀분석을 이용한 분석의 문제점 .....	11
제4장 무작위 배정에 의한 정책 평가 .....	16
제1절 무작위 배정의 의미와 사례 .....	16
제2절 처방효과의 이질성: 분위회귀분석 .....	24
제5장 이중차분법 .....	31
제1절 이중차분의 의미 .....	31
제2절 이중차분법의 일반화 .....	38
제6장 성향점수 매칭법 .....	41
제1절 성향점수 매칭의 정의 .....	41
제2절 성향점수 매칭법 .....	43
제3절 성향점수 매칭의 장단점 .....	44
제7장 도구변수 추정법 .....	46
제1절 도구변수 추정법 .....	46

제2절 도구변수 추정법의 국지성과 도구변수의 어려움 .....	53
<b>제8장 회귀단절모형 .....</b>	<b>58</b>
제1절 회귀단절의 개념 .....	59
제2절 회귀단절 방법론 .....	61
1. Sharp 회귀단절 .....	62
2. Fuzzy 회귀단절 .....	64
제3절 회귀단절법을 적용하는 두 가지 구체적인 방법들 .....	66
1. 모수추정법(Parametric Methods) .....	66
2. 비모수추정법(Non-parametric Methods) .....	67
제4절 회귀단절법 적용사례 .....	67
<b>제9장 컴퓨터 프로그램을 이용한 분석 예시 .....</b>	<b>75</b>
제1절 최소자승(OLS) 추정법 .....	75
제2절 성향점수 매칭 .....	78
제3절 도구변수(Instrumental Variables : IV) 추정법 .....	82
제4절 회귀단절모형 .....	86
제5절 이중차분법(Difference in Differences : DID) .....	92
제6절 소 결 .....	95

<p><b>제2부 산업별 고용효과 분석을 위한 다부문 거시산업모형 DB 구축</b></p> <p>(윤윤규 · 이진면 · 고영우)</p>
---

제10장 서 론 .....	103
<b>제11장 다부문 거시산업모형 .....</b>	<b>105</b>
제1절 모형의 이론적 성격 .....	105
제2절 모형의 체계와 하부 모형의 구조 .....	107
1. 모형의 체계 .....	107
2. 하부 모형의 구조 .....	109

<b>제12장 다부문 거시산업모형 DB 구축</b>	116
제1절 DB의 체계	116
제2절 산업분류	117
1. 중요성	117
2. 고려사항	118
3. 한국노동연구원 모형의 분류	123
제3절 산업별 중간수요·투입 및 최종수요·공급 DB	125
1. 산업연관표와 국민계정의 연계	125
2. 시계열화	131
제4절 산업별 노동과 임금 DB	136
1. 산업별 노동통계	136
2. 산업별 임금통계	137
제5절 산업별 자본스톡과 수출입 DB	138
1. 자본스톡의 개념과 추정방법	138
2. 수출입통계의 구축	141
제6절 DB의 불변가격(실질변수)화 작업	144
 <b>제13장 다부문 거시산업모형 DB를 활용한 기초 분석</b>	 146
제1절 산업별 수요 및 공급구조 분석	146
1. 산업의 수요와 공급구조	146
2. 산업생산구조의 변화	151
3. 산업별 중간투입계수와 부가가치계수 구조 변화	154
제2절 산업별 고용구조	157
제3절 산업별 임금구조	159
 <b>제14장 결론 및 향후 발전방향</b>	 161
제1절 노동 및 임금통계의 확대	161
제2절 모형과 DB 구축의 이원화	162
제3절 정부 정책의 유형화 및 모형 내 연계방안 마련	163
 <b>참고문헌</b>	 165

## 표 목 차

〈표 5-1〉 지역에 따른 평균 집값 .....	32
〈표 7-1〉 정책참여 기회와 실제 참여 여부 .....	54
〈표 12-1〉 1980년 이후 산업연관표의 작성 현황 .....	120
〈표 12-2〉 국민계정의 통계제공 산업분류 .....	121
〈표 12-3〉 한국노동연구원의 산업분류 .....	125
〈표 12-4〉 소비지출항목의 연계 .....	128
〈표 12-5〉 민간소비의 비목별·산업별 전환행렬 추정 결과(2005) .....	129
〈표 12-6〉 고정자본 형성행렬의 추정 결과(2005) .....	130
〈표 12-7〉 불연속 시계열의 유형 .....	131
〈표 12-8〉 노동량의 측정단위 .....	137
〈표 13-1〉 27개 산업의 수요와 공급구조(1990) .....	148
〈표 13-2〉 27개 산업의 수요와 공급구조(2000) .....	149
〈표 13-3〉 27개 산업의 수요와 공급구조(2009) .....	150
〈표 13-4〉 27개 산업별 생산구조 변화 추이(총산출 기준) .....	152
〈표 13-5〉 27개 산업별 생산구조 변화 추이(부가가치 기준) .....	153
〈표 13-6〉 27개 산업별 중간투입계수 변화 추이(경상 중간투입액/총산 출액) .....	155
〈표 13-7〉 27개 산업별 부가가치계수 변화 추이(경상 부가가치액/총투 입액) .....	156
〈표 13-8〉 27개 산업별 취업자 비중과 연평균 증가율 추이 .....	158
〈표 13-9〉 27개 산업별 임금수준 및 연평균 증가율 추이 .....	160

## 그림목차

[그림 4- 1] 처방효과의 분포 .....	25
[그림 8- 1] 배정확률(Assignment probabilities) : SRD .....	63
[그림 8- 2] 잠재/관측 결과 회귀함수(Potential and observed outcome regression functions) : SRD .....	63
[그림 8- 3] 배정확률(Assignment probabilities) : FRD .....	64
[그림 8- 4] 잠재/관측 결과 회귀함수(Potential and observed outcome regression functions) : FRD .....	65
[그림 9- 1] Simple OLS Stata 프로그램 및 결과 .....	76
[그림 9- 2] Multiple OLS Stata 프로그램 및 결과 .....	77
[그림 9- 3] PSM Stata 프로그램 검색 결과 .....	79
[그림 9- 4] PSM Stata 프로그램 및 결과(attnd.ado 및 attr.ado) .....	80
[그림 9- 5] IV Stata 프로그램 및 결과(ivregress.ado) .....	83
[그림 9- 6] Weak IV 검정 Stata 프로그램 및 결과 .....	84
[그림 9- 7] IV Stata 프로그램 및 결과(ivreg2.ado) .....	85
[그림 9- 8] RDD Stata 프로그램 및 결과: 모수적 방법 .....	88
[그림 9- 9] RDD Stata 프로그램 및 결과: 비모수적 방법(rd.ado) .....	90
[그림 9-10] DID Stata 프로그램 및 결과 .....	93
[그림 9-11] DID Stata 프로그램 및 결과(패널분석) .....	95
[부도 9- 1] 시뮬레이션 데이터 생성을 위한 Stata 프로그램 (Do 파일) .....	98
[부도 9- 2] 시뮬레이션 데이터 활용한 회귀분석 Stata 프로그램 (Do 파일) .....	99
[그림 11- 1] 다부문 거시산업모형의 체계 .....	108
[그림 12- 1] 다부문 거시산업모형의 DB체계 .....	117

[그림 12- 2] 수정 RAS 방법의 절차 .....	135
[그림 13- 1] 전 산업의 수요 및 공급구조 .....	147
[그림 13- 2] 전 산업의 수요 및 공급구조 .....	147
[그림 13- 3] 산업생산구조 변화 추이 .....	151
[그림 13- 4] 산업별 중간투입계수 변화 추이 .....	154
[그림 13- 5] 산업별 부가가치계수 변화 추이 .....	154
[그림 13- 6] 산업별 취업자 수 증가 추이 .....	157
[그림 13- 7] 산업별 임금변화 추이 .....	159



**【제 1 부】**

**노동시장정책 평가방법론**



실증경제학 연구에서 가장 중요한 화두 중의 하나는 중요 변수의 내생성(endogeneity)을 어떻게 통제할 것인가 하는 문제다. 통상적인 선형회귀모형의 관점에서 내생성이란 중요 설명변수와 모형의 오차항이 통계적으로 서로 독립적이지 않고 체계적인 양(+) 또는 음(-)의 상관관계가 있는 경우를 통칭한다. 내생성이 존재할 경우에는 설명변수가 종속변수에 영향을 미치는 정도가 실제 인과효과를 과소 또는 과대 추정되게 된다. 극단적인 경우에는 내생성이 통제되지 않은 채 도출된 추정 결과와 실제 인과관계가 반대 방향이 되기도 한다. 실증연구를 통해 도출된 추정치가 인과효과를 정확히 반영하지 못하는 경우, 정부의 정책이나 개인의 의사결정에 심각한 왜곡 현상이 발생할 수 있다.

주로 설문조사 자료(observational survey data)를 이용하는 사회과학 연구에서는 전통적으로 관측 단위에 대하여 최대한 많은 정보를 취합하고 이를 회귀분석의 설명변수로 포함시키는 방법을 통하여 내생성 문제를 해결하고자 하였다. 즉, 관심 설명변수 이외에 여타의 설명변수를 가능한 한 많이 추정모형에 추가하여 내생성의 강도를 줄이고자 하였다. 이러한 방법은 통제되는 설명변수가 충분히 많아지면 관심 설명변수와 모형의 오차항 간에 독립성

\* 본고의 제1부(노동시장정책 평가방법론)는 2011년 7월부터 10월까지 한국노동연구원  
에서 개최한 「노동시장 프로그램 평가방법론」 강의(중앙대 강창희 교수)의 내용을  
바탕으로 작성되었음을 밝힌다.

(independence)이 발생한다는 매우 강한 ‘외생성(exogeneity)’ 가정에 기반하고 있다. 설명변수가 아무리 많이 포함되더라도, 설문조사 자료를 이용하는 사회과학 연구에서는 ‘외생성’ 가정을 만족시키는 경우를 찾기란 쉽지 않다. 관심 설명변수가 모형의 오차항과 독립적인 경우는 자연과학의 인위실험(artificial experiments)에서와 같이 관측대상이 처방집단(treatment group)과 통제집단(control group)으로 무작위 배정되는 경우에만 가능하기 때문이다. 의학이나 생물학 연구에서도 사회과학에서와 같이 설문조사 방법과 통계학을 결합한 실증연구가 가능하지만, 굳이 무작위 실험을 고집하는 이유는 바로 내생성을 통제하기 어렵기 때문이다. 사회과학 연구자들은 전통적으로 사회과학 분야에서는 무작위 배정(random assignment) 사례를 찾아보기 어렵다고 생각하여 수많은 설명변수를 통제하거나 또는 복잡한 통계방법을 활용하여 인과효과를 구하는 데 집중해 왔다. 그러나 최근의 많은 실증경제학 연구들은 사회 내의 각종 제도적인 요인들에 의하여 무작위 배정이 드물지 않게 일어나고 있고, 이를 사회과학 분야의 연구에 요긴하게 활용할 수 있음을 보여주고 있다. Angrist and Pischke(2009)는 최근 10여 년 동안 경제학계에 보고된 많은 자연실험(natural experiments) 사례들을 보여준다.

자연실험을 활용하는 현대 실증경제학의 방법론은 정부의 정책효과를 정확히 평가하는 방법론에 유용하게 적용될 수 있다. 실증경제학자들이 자연실험을 찾는 가장 큰 이유는 인위적인 실험(artificial experiments)을 통해 만들어진 자료가 쉽게 구해지지 않기 때문이다. 그러나 정부의 정책에서는 적절한 프로그램 설계를 통하여 성과 평가에 유용하게 사용될 인위적인 실험 자료를 얻는 것이 가능하다. 즉, 현대 실증경제학의 방법론은 정부 정책 및 노동시장 프로그램을 원래의 목적에 맞게 설계하고 그 효과를 평가하는 매우 유용한 아이디어들을 제공해 준다.

본 보고서는 최근 실증경제학에서 개발·발전되고 있는 프로그램 평가방법론을 소개하는 것을 목적으로 한다. 최근의 프로그램 평가방법론은 과거에 사용되었던 통계모형들에 비해 보다 간단명료하게 체계화되어 가고 있다. 과거의 평가방법론은 프로그램 실행 이후에 통상적인 설문조사 방법을 통하여

자료를 수집하고 통계방법을 활용하여 프로그램의 효과를 추정하는 데 집중하였다. 그러나 최근의 평가방법론은 사후적으로 프로그램의 효과를 추정하기 위해 필요한 조건들을 명확히 제시하면서, 사전적으로 프로그램을 디자인하는 많은 아이디어를 제공해 준다. 본 연구는 최근의 노동시장 평가방법론을 소개함으로써 노동시장 평가 실무 연구자들의 시야를 프로그램의 디자인 단계로까지 확장하는 것을 목표로 한다.

본 보고서에서는 현대 실증경제학에서 주로 사용되고 있는 대표적인 노동시장 프로그램 평가방법론(program evaluation)을 다룬다. 특히, 무작위 실험자료나 자연실험을 이용하는 계량분석에 초점을 맞춘다. 더불어, 비실험자료(non-experimental observational data)를 분석하는 통계방법에 대해서도 설명한다. 이 연구에서 다루어지는 통계방법은 프로그램 전후비교법(Before/After methods), 이중차분법(difference-in-differences), 도구변수법, 회귀단절모형(regression discontinuity design) 및 성향점수 매칭법 등이다. 위의 방법들이 실제 적용되어 연구에서 다루어지는 사례로는, 학급규모 효과, 직업훈련 효과, 동료 효과(peer effects), 학교 평준화 효과, 노동공급, 범죄율, 부정부패 등이 포함된다.

## 제2장

# 용어에 대한 정의

어떤 정책의 효과를 분석하고자 하는데,  $N$  명에 대한 자료가 있어서 이를 이용하고자 한다. 이 중 일부는 그 정책에 참여하였고, 나머지는 정책에 참여하지 않았다고 하자. 정책에 참여하지 않은 사람들은 애초에 정책의 대상자가 아니었을 수도 있고, 정책의 대상자였으나 개인의 선택에 의해 참여하지 않았을 수도 있다.

만약 어떤 개인이 정책에 참여하였으면  $D_i = 1$ 로 표시하고, 참여하지 않았으면  $D_i = 0$ 으로 나타낸다. 즉,  $D$ 는 정책참여 여부를 나타내는 변수다. 그리고 정책에 참여한 사람들의 집단을 정책참여집단(treatment group, 혹은 처리집단, 처방집단)이라고 하고, 참여하지 않은 사람들을 비교집단(control group, 혹은 통제집단)이라고 부른다. 정책에 참여하지 않은 사람들의 수를  $N_0$ , 참여한 사람들의 수를  $N_1$ 이라고 하자. 그리고 각 개인의 특성을 나타내는 변수들의 집합을  $X_i$ 라고 하자.

## 제1절 결과에 대한 정의

정책평가의 원리는 어떤 개인이 정책에 참여하였을 때의 결과(outcome)와

그 개인이 정책에 참여하지 않았을 때의 결과를 비교하는 것이다. 어떤 개인  $i$ 가 정책에 참여하였을 때 나타나는 결과를  $Y_i(1)$ 이라고 하고, 정책에 참여하지 않았을 때의 결과를  $Y_i(0)$ 이라고 하자. 정책평가의 원칙은  $Y_i(0)$ 와  $Y_i(1)$ 을 비교하는 것이다.

그런데 우리는 이 두 가지 결과를 동시에 볼 수 없다. 한 시기에 한 개인은 정책참여와 비참여 중 한 가지 상태에만 있게 되고 따라서  $Y_i(0)$ 와  $Y_i(1)$  중 하나만 현실에서 실현이 되고 관찰이 된다. 어떤 사람이 정책에 참여하지 않았다면  $Y_i(0)$ 만 실현되고  $Y_i(1)$ 는 실제에서는 관찰되지 않는다. 반대로 어떤 사람이 정책에 참여하였다면  $Y_i(1)$ 만 실현이 되어 관찰되고  $Y_i(0)$ 는 실현되지 않는다. 두 가지 결과 가운데 실현되지 않아서 관찰할 수 없는 결과를 가상적 결과(potential outcome, 혹은 잠재적 결과)라고 부른다.

정책평가가 어려운 이유는 정책평가에 필요한 두 가지의 결과  $Y_i(0)$ 와  $Y_i(1)$  중 하나만 실제로 실현이 되고, 나머지 하나는 실현되지 않지 않는다는 데에 있다. 이후에 소개될 정책평가 방법들은 실현되지 않아서 실제로는 관측되지 않는 가상적 결과들을 어떻게 얻을 것인가와 관련이 된다. 가상적인 결과는 관측되지 않아서 기록이 되지 않은 결측 자료(missing data)라고도 볼 수 있다. 정책평가의 핵심은 관측 가능한 자료를 잘 이용하여 관측이 되지 않은 결측 자료를 얻는 데에 있다.

## 제2절 정책에 참여하는 방식

정책평가의 핵심적 요소는 각 개인이 정책에 참여하는 방식이다. 이 방식에 따라 정책을 평가하는 방법론이 달라지게 된다. 정책참여 방식은 크게 세 가지로 구분된다.

- (1) 무작위적 (randomized) 배정 : 무작위적 배정에서 각 개인이 정책에 참

여할 확률은 가상적 결과( $Y$ )와는 무관하며 각 개인의 특성변수( $X$ )의 함수다. 무작위적 배정은 또한 크게 세 가지로 구분된다. ① 완전히 무작위적인 배정(completely random assignment)의 경우, 어떤 개인이 정책에 참여할 것인가 혹은 참여하지 않을 것인가가 완전히 (동전 던지기와 같은) 무작위적으로 이루어진다. 한편, ② 처음에 한 쌍을 고른 다음 두 번째 단계에서 무작위적으로 한 명을 골라서 정책에 참여하게 하고, 나머지 한 명은 정책에 참여하지 않게 하는 방식이 있다. 이 방식을 대진방식의 무작위 배정(pairwise randomization)이라고 부른다. ③ 대진방식 무작위 배정을 보다 일반화하여, 전체 집단을 몇 개의 하위집단(strata)로 나누고, 그 하위집단 내에서는 무작위 배정을 하는 방식이 있다. 이러한 방식을 일반적인 층화법(generalized stratified experiment)라고 한다.

(2) 두 번째 종류의 정책참여 방식에서는 정책에 참여할 확률이 가상적 결과와는 무관하지만 특성변수( $X$ )의 함수도 아니다. 이를 수학적으로 표현하면

$$D_i \perp (Y_i(0), Y_i(1)) | X_i$$

즉, 정책에 참여하게 될 확률은 특성변수를 통제하고 나면 결과와는 확률적으로 독립이다. 이러한 방식에는 매우 다양하게 지칭하는 용어가 있다. 즉, unconfounded assignment, selection on observables, exogeneity, and conditional independence 등이다.

(3) 위 두 가지 방식 외에 정책참여 방식이 결과변수( $Y$ )에 결부되어 있는 경우가 세 번째 방식이다. 이러한 상황에서 정책평가를 하는 방법은 매우 복잡해지고 상황과 자료에 따라 사용되는 정책평가 방법이 크게 달라진다.



### 제3절 평균처방효과(Average Treatment Effects)

정책의 효과를 측정하여 그 결과를 나타내고 요약하는 몇 가지 방법이 있다. 일반적으로 정책평가 방법론에서 가장 많이 사용되고 있는 것은 정책처방 평균효과(Average Treatment Effect: ATE)와 정책처방집단의 평균효과(Average Treatment Effect on the Treated: ATET)다.

정책처방 평균효과는 개인별 정책효과에 기댓값을 취한 것이다. 즉,

$$\tau_{ATE} = E[Y_i(1) - Y_i(0)]$$

만약 어떤 정책이 모든 사람들을 대상으로 하고 있고, 이 중 일부는 정책에 참여하고 나머지는 정책에 참여하지 않았다면 정책처방 평균효과를 재는 것이 적합하다.

반면, 정책처방집단의 평균효과(ATET)는 정책처방을 받은 사람들을 대상으로 정책의 효과를 측정하는 것이다. 즉,

$$\tau_{ATET} = E[Y_i(1) - Y_i(0) | D_i = 1]$$

많은 경우, 정책처방 평균효과보다는 정책처방집단의 평균효과가 관심의 대상이 된다. 직업훈련 프로그램을 예로 들어 보자. 직업훈련 프로그램은 주로 임금이 낮은 사람들을 대상으로 한다. 임금이 매우 높고 안정적인 일자리에 있는 사람들을 통제집단으로 삼아서 직업훈련 프로그램의 효과를 측정하는 것은 의미가 없다. 직업훈련 프로그램 평가를 통하여 알고 싶은 것은 직업훈련 프로그램에 참여한 사람들에게 이 프로그램이 얼마나 도움이 되었는가이다.

또한 신청자 중 일부만을 대상으로 어떤 정책이 실행되는 경우에도 정책처방 평균효과보다는 정책처방집단의 평균효과를 측정하는 것이 적합하다. 정

책의 대상이 아닌 사람 혹은 정책에 신청하지 않은 사람들을 비교집단으로 하여 그 정책의 효과를 측정하는 것은 의미가 없기 때문이다.

## 제3장

# 회귀분석을 이용한 분석의 문제점

간단히 다음과 같은 회귀식을 추정하여 정책의 효과를 분석하는 상황을 생각해 보자.

$$Y_i = \beta_0 + \beta_1 D_i + U_i \quad (3-1)$$

식 (3-1)의 OLS 추정치는

$$\hat{\beta}_1^{OLS} = \frac{\sum_{i=1}^N (D_i - \bar{D})(Y_i - \bar{Y})}{\sum_{i=1}^N (D_i - \bar{D})^2} = (\bar{Y}|D=1) - (\bar{Y}|D=0) \quad (3-2)$$

이다. 식 (3-2)에서 변수 위에 막대기가 있는 것은 평균값을 의미한다. 정책 처방집단의  $Y$ 의 평균값에서 통제집단의  $Y$ 의 평균값을 빼면 간단히 정책의 효과를 측정할 수 있다. 이러한 추정치가 불편(unbiased)추정량인지를 파악하기 위해 기댓값을 취해 보면,

$$\begin{aligned} E[\hat{\beta}_1^{OLS}] &= E\left[\frac{1}{N_1} \sum_{i \in N_1} Y_i - \frac{1}{N_0} \sum_{i \in N_0} Y_i\right] \\ &= E\left[\frac{1}{N_1} \sum_{i \in N_1} (\beta_0 + \beta_1 + U_i) - \frac{1}{N_0} \sum_{i \in N_0} (\beta_0 + U_i)\right] \\ &= \beta_1 + (E[U_i|D_i=1] - E[U_i|D_i=0]) \end{aligned} \quad (3-3)$$

이다. 식 (3-3)에서 마지막 식의 괄호 안이 0이 되면 OLS 추정치는 불편추정량이 된다. 마지막 두 항이 0이 된다는 것은 정책처방집단의 관측 불가능한 변수의 평균값이 통제집단의 평균값과 같다는 것을 의미한다. 일반적으로는 이 두 값이 같다고 볼 수 없다. 직업훈련 프로그램의 예를 들어 보자. 일반적으로 미래지향적이고 적극적인 사람이 직업훈련에 참여할 가능성이 높고, 반면 동기부여가 약한 사람은 직업훈련에 참가하지 않을 가능성이 높다. 그렇다면 두 집단 간 관측 불가능한 요소들의 평균값이 0이 되지 않을 것이다. 따라서 이 경우 OLS 추정치에 편의(bias)가 발생하게 된다.

내생성의 문제를 보다 일반적으로 파악하기 위해 다음 식 (3-4)와 같이 통제변수  $X$ 를 포함하여 추정을 하는 상황을 생각해 보자.

$$Y_i = \beta_0 + \beta_1 D_i + X_i \beta_2 + U_i \quad (3-4)$$

OLS를 통해  $\beta_1$ 을 추정하는 경우 만약  $Cov(D_i, U_i) \neq 0$ 이라면 편의가 발생한다. 만약  $Cov(D_i, U_i) > 0$  이면 양(+)의 편의(혹은 위쪽으로의 편의)가 발생하고,  $Cov(D_i, U_i) < 0$ 이라면 음(-)의 편의(혹은 아래쪽으로의 편의)가 발생한다.

일반적으로 내생성이 발생하는 이유는 크게 네 가지로 구분될 수 있다.

- ① 중요한 변수가 생략된 경우
- ② 반대의 인과관계(혹은 동시성)
- ③ 자기선택으로 인한 편의(예: 직업훈련 프로그램의 참가를 선택하는 경우)
- ④ 측정 오차(예를 들면 흡연과 건강과의 관계인데, 담배를 피우는 것에 대해 솔직하게 이야기하지 않을 가능성이 높음)

[예 1] 자본주의의 발달과 기독교 정신(Becker & Woessmann, 2009)

막스 베버는 자본주의의 발달이 기독교 정신과 밀접한 관련이 있다고 주장하였다. 즉, 근검절약 정신, 부의 축적을 문제시하지 않는 기독교의 논리(과거 가톨릭에서는 부의 축적을 바람직하지 않다고 보았다)가 자본주의의 사상

과 체제를 발전시키는 데 기여하였다는 것이다. 이에 대해 논쟁이 있어 왔는데, 이 논문에서는 이를 실증적으로 확인하고 있다.

이 논문에서 개별 관측치  $i$ 는 지역이고,  $D_i$ 는 그 지역이 기독교 지역인가 아니면 가톨릭 지역인가를 나타낸다. 지역은 주민들의 75% 이상이 신교가 사는 지역, 회색이 25~75%인 지역, 파란색이 25% 미만인 지역으로 나뉜다.  $X$ 는 지역의 특징(인구, 강수량, 1인당 농지 분포 등)을 보여주는 관측 가능한 특성변수이고,  $u$ 는 관측할 수 없는 오차항이다. 단순회귀분석에 따르면  $D_i$ 의 계수값이 유의미한 양수로 추정된다. 즉, 가톨릭 지역에 비해 기독교 지역에서 경제발전이 더 잘 된다는 것으로 해석된다.

그런데 이 회귀분석의 결과에 대해 내생성의 문제가 있을 것(혹은 생략변수의 문제가 있을 것)이라는 의심을 가질 수 있다. 이 논문에서 생략변수의 하나로 특별히 주목한 것이 가독률(literacy rate)이다. 가톨릭에서는 사제들이 성경을 해석해 주는 반면, 기독교에서는 신도들이 직접 성경을 읽고 해석하는 것을 중시한다. 이는 마틴 루터의 생각에 기반한 것으로, 사제의 해석에 의지하기보다는 신도들이 직접 성경을 읽고 해석하면서 직접 사제가 되는 것을 중시한다. 따라서 신교 지역에서는 성경을 읽기 위해 반드시 글자를 알아야 한다. 기독교의 전파는 가독률의 증가를 수반한다. 이제 회귀식에 가독률 변수를 추가하여 추정을 하면, 신교냐 가톨릭이냐는 나타내는 변수  $D_i$ 의 계수 추정치가 통계적으로 0이 된다. 가독률 외에 기독교와 관련된 다른 요소들은 유의미한 영향을 주지 않은 것으로 나타났다. 기독교 정신 자체가 자본주의 발전에 영향을 주었다기보다는 가독률이라는 매개를 통해 경제발전에 영향을 주었다고 할 수 있다. 이 사례는 중요한 변수를 회귀식에서 생략하여 추정을 하면 인과관계에 대한 잘못된 결론을 얻게 된다는 것을 보여준다.

[예 2]  $Y$ 가 범죄발생 건수,  $D$ 가 경찰의 숫자라고 하자. 경찰이 늘어나면 범죄가 과연 줄어들 것인가가 관심 대상이다. 지역별로 범죄발생 건수, 경찰의 숫자 및 그 지역의 특성을 나타내는 정보를 수집하여 자료로 모았다고 하

자. 이 자료를 이용하여 OLS로 추정을 하면 일반적으로  $D$ 의 계수값이 양(+)으로 추정된다. 즉 경찰의 숫자가 많을수록 오히려 범죄가 늘어나는 것으로 해석된다. 상식에 어긋나는 결과를 보고  $X$ 를 많이 포함시켜 보아도 결과는 크게 변화하지 않는다. 즉, 통제변수를 늘리는 것이 도움이 되지 않는다. 왜 이러한 결과가 나타나는가? 사실은 경찰이 많을수록 범죄가 늘어나는 것이 아니라 범죄가 많은 지역에 경찰을 많이 투입한 것이다.

[예 3] 의사가 많을수록 전염병이 줄어드는가를 알고 싶다고 하자. 이 문제에서  $Y$ 는 전염병의 빈도,  $D$ 는 의사의 수가 된다. 지역별 전염병 빈도와 의사 수, 지역특성을 보여주는 정보 등을 수집하여 자료를 만들고 이 자료로 OLS 추정을 하면 의사가 많을수록 전염병이 오히려 늘어난다는 결과를 얻게 된다. 이 결과를 보고 러시아 당국에서는 전염병을 줄이기 위해 의사를 줄이라는 정책을 내놓기도 하였다. 사실은 의사가 많아서 질병이 창궐한 것이 아니라 질병이 많아서 의사가 많이 투입된 것이다.

[예 4] 어떤 혈압약이 혈압을 낮추는 효과가 있는가를 알고 싶다. 이 문제에서  $Y$ 는 혈압,  $D$ 는 혈압약을 먹는가 먹지 않는가를 나타내는 변수다. OLS로 추정을 하면  $D$ 의 계수값이 양(+)으로 나와서, 혈압약을 먹으면 오히려 혈압이 올라간다는 결과가 나오기도 한다.

[예 5] 얼마 전 캐나다의 한 경찰관은 여성들이 성폭행 피해를 입지 않으려면 옷을 매춘부처럼 야하게 입고 다니지 말라고 언급한 바 있다. 이 말의 진위를 알아보려고 여성들의 야한 옷차림이 성폭행을 더 발생시키는지를 알고 싶다고 하자. 이 문제에서  $Y$ 는 성폭행 발생 빈도(확률)이고,  $D$ 는 여성들의 옷차림이 야한지 여부가 된다. 여성들 개인의 정보를 수집하여 자료를 구축한 후 OLS를 활용하여 추정치를 구하면  $D$ 의 계수값이 양(+)의 값을 갖는다. 결국 경찰관의 말처럼 여자들이 옷을 야하게 입을수록 성폭행이 더 발생한다는 결과를 얻게 된다. 그런데 과연 이것이 인과관계를 나타내는 것인지에 대

해서는 생각해 볼 필요가 있다. 여성들이 옷을 야하게 입는 것이 성폭행 유발에 대한 직접적인 원인이 아니라 어떤 다른 원인이 있을 것이라는 점이다. 일반적으로 성폭행 사고가 상대적으로 우범 지역에서 발생하고, 그 지역에는 매춘부도 상당수 살고 있을 수 있기 때문에 옷을 야하게 입은 사람들이 눈에 많이 보일 수 있다. 그렇다면 사람들이 상대적으로 우범 지대에 살고 있기 때문에 성폭행의 위험에 많이 노출되어 있는 것이지, 옷을 야하게 입었기 때문은 아닌 것이다.

과거 경제학, 교육학, 사회학 등에서는 이 내생성의 문제를 통제변수를 늘리는 방법으로 해결하고자 하였다. 그런데  $Y$ 에 영향을 주는 요소들이 굉장히 많은데 이 정보들을 모두 수집하여 자료로 만든다는 것이 매우 어렵고, 때로는 관측 불가능하기 때문에 회귀식에 포함시킬 수 없는 변수들도 있다. 예를 들어 인생에 대한 태도 등은 관측하여 변수로 만들기 매우 어렵다. 따라서 통제변수를 늘리는 것으로는 내생성의 문제를 근본적으로 해결하기 어렵다.

위 문제를 근본적으로 해결하는 가장 간편한 방법은 무작위적 배정방식을 취하는 것이다. 무작위적 배정은 약의 효능을 측정하기 위해 많이 사용되고 있다. 약을 두 종류, 즉 하나는 진짜 약, 다른 하나는 가짜 약을 만들고 어떤 약을 투여할 것인가를 무작위적으로 결정한다. 무작위적으로 배정을 한다는 것은  $D$ 가  $U$ 와 통계적으로 독립이라는 것을 의미한다. 따라서 내생성 때문에 생기는 문제를 쉽게 해결할 수 있다.

## 제4장

# 무작위 배정에 의한 정책 평가

### 제1절 무작위 배정의 의미와 사례

무작위 배정(random assignment)에서는 어떤 개인이 정책참여집단에 속할 것인가 아니면 통제집단에 속할 것인가가 무작위로 배정된다. 무작위 배정방식에 의해 정책참여가 이루어지면 간단히 두 집단 간 결과값의 평균을 비교하면 어떤 정책의 효과를 측정할 수 있다. 이는  $Y$ 를  $D$ 에 회귀하여  $D$ 의 계수값을 OLS로 추정하는 방식과 마찬가지로  $D$ 가 무작위로 배정되었다면 내생성의 문제가 없는 것이므로 정책의 효과를 올바르게 측정할 수 있게 된다.

무작위 배정을 이용하여 정책의 효과를 분석하는 방법은 매우 간단하면서도 결과의 신뢰성도 매우 높다는 장점이 있다. 이 방법을 이용할 때 가장 핵심적인 문제는 과연 정책참여가 무작위로 배정되었는가를 판단하는 것이다. 이를 간접적으로 판단하는 방법이 있다. 만약  $D$ 가 무작위라면  $U$ 와도 독립이고 또  $X$ 와도 통계적으로 독립이어야 한다. 따라서  $D$ 를  $X$ 에 회귀하면 그 계수값이 통계적으로 유의미하지 않아야 한다. 또한  $Y$ 를  $D$ 에 회귀하였을 때  $D$ 의 계수 추정치가,  $Y$ 를  $D$ 와  $X$ 에 회귀하였을 때  $D$ 의 계수 추정치가 통계적으로 차이가 없어야 한다. 즉  $X$ 를 추가하나 하지 않으나 그 결과가 변화하지 않아야 한다.



[예 1] 교육환경을 무작위로 배정한 경우(Gould, Lavy & Paserman, 2004)

초등학교 때의 교육환경이 고등학교에서의 성적이나 대학 진학률 등 이후 교육성파에 미치는 영향을 분석하고자 한다. 이 상황에서  $Y$ 는 고등학교 성적 혹은 대학 진학률이고,  $D$ 는 초등학교 때의 교육환경이다. 통상적으로 조사 자료를 이용하여 이 문제를 분석하곤 하였다. 고등학생을 대상으로 현재의 시험 성적, 대학진학 여부 등을 조사하고 초등학교 때 어떤 동네에서 학교를 다녔는지 등의 자료를 수집한 뒤 회귀분석을 한다. 그러면 일반적으로 초등학교 때의 환경이 이후 교육성파에 매우 긍정적인 효과가 있는 것으로 나타난다. 하지만 회귀분석에 의한 결과에 대해서는 내생성의 문제를 의심하게 된다. 어떤 초등학교에 다닐 것인가가 무작위로 배정된 것이 아니면 부모의 관심, 학생 자신의 동기부여 등의 요소들이 어떤 초등학교에 다녔는가와 이후 교육성파에 모두 영향을 주게 된다. 즉 중요한 변수를 생략해서 양(+)의 편의가 발생할 수 있다. 반대로 음(-)의 편의에 대해서도 생각해 볼 수 있다. 성적이 떨어지는 학생에 대해서는 좀 더 작은 교실로 보낸다던가, 좀 더 잘 가르치는 선생님을 배정할 수도 있고, 반대로 성적이 나은 학생에 대해서는 학교에서 신경을 덜 쓸 수 있다. 따라서 초등학교의 환경과 고등학교 성적 간의 인과관계를 밝히는 것은 쉽지 않은 일이다.

이 논문에서는 초등학교 환경이 무작위로 배정된 사례를 이용하여 분석을 하고 있다. 1991년 이디오피아에서 폭동이 일어나자 이스라엘 정부는 이디오피아에 살고 있는 유대인 1,400명을 5월 24일부터 25일 36시간 동안 비행기로 이스라엘에 데려왔다. 이들 가운데 상당수의 초등학생들이 있었는데, 이스라엘 정부는 이 학생들을 이스라엘 전역의 학교에 무작위로 배정을 하였다. 이 논문의 저자들은 이 학생들의 기록을 찾아내서 이들의 학교 기록, 성적 등의 자료를 수집하였다. 그리고 간단하게  $Y$ 의 평균값을 비교하여 초등학교 환경이 이후 교육성파에 미치는 인과관계를 측정하였다.

[예 2] 형량(prison time)의 무작위 배정(Drago, Galbiati, & Vertova, 2009)

형량을 늘리는 것이 과연 범죄를 낮추는 효과가 있을까? 이 문제를 회귀방

법을 통해 분석하는 방법은 범죄자를 대상으로 형량, 범죄율(재범 여부), 성별, 가족 등등의 자료를 수집하여 OLS 추정을 하는 것이다. 이러한 분석에서는 형량이 높을수록 범죄율이 떨어진다는 결과가 나타나기도 한다.

이 논문에서는 내생성의 문제를 극복하기 위해서는 형량이 무작위로 배정된 경우를 찾아내서 분석하였다. 2006년 7월 이탈리아에서는 감옥에 수감된 죄수들이 너무 많아서, 중범죄자를 제외한 모든 제소자의 형량을 3년 감형한다는 법이 통과되었다. 만약 5년형을 받았는데 이미 3년을 살았다면 이 제소자는 출소하게 된다. 다만 출소하는 대신에 5년 이내에 다시 범죄를 저지르면, 그 범죄에 대한 형량뿐만 아니라 예전에 남았던 형량까지도 더하게 된다는 조건이 붙었다. 10년의 형을 받고 9년을 복역한 사람은 나머지 1년을 감형받게 되고, 2년 형을 받고 6개월 감옥에 있던 사람은 1년 6개월을 감형받게 된다. 원래의 형량에 상관없이 얼마나 감형을 받게 되었는가가 무작위로 설정된 것이나 마찬가지로 되었다. 다시 범죄를 저지르면 감형된 양만큼 형량이 추가되므로 미래의 가능한 형량이 무작위로 설정된 것이나 마찬가지로 되었다. 논문에서는 약 2년이 지난 시점에 2006년 7월 시점에 감형을 받고 출소한 사람들이 범죄를 저질러 투옥되었는가를 조사하여 형량이 범죄에 미치는 효과를 분석하였다. 연구자가 의도적으로 형량을 무작위 배정하지는 않았지만 예상되지 않은 정부 정책에 따라 무작위 배정과 다른 없는 사건이 일어난 것을 잘 이용한 사례다.

#### [예 3] 동료(peer)의 무작위 배정(Sacerdote, 2001)

우리나라에서 평준화 논쟁이 있다. 나의 친구가 나에게 어떤 영향을 주는가에 대한 문제와 관련이 된 것이다. 평준화 제도하에서는 같은 반에 있는 학생들의 지적 능력이 천차만별이다. 공부 잘하는 친구도 있고, 공부 못하는 친구도 주위에 있게 된다. 반면 비평준화 제도에서는 비슷한 수준의 학생들을 함께 있게 만든다. 동료가 한 개인에 어떤 영향을 주는가는 교육학이나 경제학에서 매우 오래된 주제다. 미국에서는 특히 흑인에 대한 교육이 이슈가 되었었다. 1960년대 제임스 콜만이라는 사회학자는 흑인들의 교육수준을 증진

시킬 수 있는 방안을 연구하면서, 동료의 효과가 매우 크다는 것을 발견하였다. 이 연구에서는 어떤 개인의 성적을 그 개인을 제외한 급우들의 성적에 회귀하였을 때 그 계수값이 매우 유의미하게 양(+)의 값이 나왔기 때문에 이러한 결론을 내렸다. 하지만 이 연구에 대해서도 내생성의 문제를 의심할 수 있다. 예를 들어 강남에 있는 학교에 다니는 학생들의 부모들이 부자들이고, 다른 지역 학교에 다니는 학생들의 부모들이 상대적으로 가난하다고 해 보자. 강남에 있는 학교의 반 성적이 높고 나의 성적이 높을 것이다. 반대로 가난한 지역 학교의 반 성적이 낮고 나의 성적도 낮을 것이다. 따라서 반 성적과 나의 성적은 인과관계가 아니라 단순히 상관관계를 나타낼 뿐이다. 부모의 직업, 교육, 소득 등도 모두 다르고, 이와 더불어 반 성적과 나의 성적에 영향을 줄 수 있는 요소들이 매우 많기 때문에 회귀분석에 의해 인과관계를 밝히는 것은 매우 어렵다.

이 문제를 해결하는 방법은 동료의 배정을 무작위로 하는 것이다. 이 논문에서는 동료의 배정이 무작위로 이루어진 경우를 찾아서 동료의 효과를 분석하고 있다. 미국의 다트머스(Dartmouth) 대학 신입생에게 기숙사를 배정할 때 남자와 여자만을 구분하고, 방 배정은 추첨으로 하였다. 각 방마다 두 명씩 배정이 되었는데, 내가 어떤 방에 들어가면 나의 룸 메이트는 무작위로 배정된다는 것을 의미한다. 그 룸메이트가 공부를 열심히 하는 사람일 수도 있고, 그렇지 않고 다른 활동(오락, 여가, 음주 등)을 좋아하는 사람일 수도 있다. 이렇게 동료가 무작위로 배정되었을 때 동료의 특성이 나의 성적에 어떠한 영향을 미치는가를 간단한 자료와 방법으로 분석할 수 있게 된다.

[예 4] 이웃(neighborhoods)을 무작위 배정(Kling, Liebman & Katz, 2007)

안 좋은 동네에 사는 것이 청소년기와 성인기의 여러 가지 결과에 어떠한 영향을 주는가를 분석하는 것이 연구의 목적이다. 회귀분석을 이용한 분석은 내생성의 문제점이 있다. 어떤 동네에 살 것인가와 결과에 동시에 영향을 주지만 관측할 수 없는 가족의 특성이 있을 수 있다. 그리고 어떤 동네에 살 것인가를 선택할 수 있기 때문에 자기선택으로부터 발생하는 편의가 있을 수

있다.

이 논문에서는 이웃을 무작위로 배정한 사건에 주목하고 있다. 미국에서 도심 지역에는 가난한 흑인들이 많이 거주하고 있다. 1994-97년 다섯 개 도시 볼티모어, 보스턴, 시카고, LA, 뉴욕에서 도심 지역에 거주하고 있는 흑인 여성 단독가구주인 저소득층을 대상으로 만약 다른 곳으로 이주하면 이주비와 월세를 지원해 주는 정책(Moving to Opportunity : MTO)을 실시하였다. 이 지원을 받는 사람은 신청자 중 추첨을 통해 결정되었다. 어떤 가구가 추첨되었다면 다른 곳으로 떠날 수 있는 기회가 생기는 것으로 이웃의 배정이 무작위로 결정된 경우다.

[예 5] 인종의 무작위 배정(Bertrand & Mullainathan, 2004)

미국에서의 인종차별이 예전부터 계속되었다고 믿어지는데, 과연 인종차별의 정도가 임금과 같은 노동시장 성과에서 어느 정도 나타나는지를 파악하고자 하는 것이 연구의 목적이다. 이를 분석하기 위해 기존 연구들은 일반적으로 횡단면 자료(cross-sectional data) 또는 패널 자료(panel data)를 활용하여 다음과 같은 회귀분석 모형을 설정하게 된다.

$$y_i = \beta_0 + B_i\beta_1 + X_i\beta_2 + u_i \quad (4-1)$$

여기서  $y$ 는 임금이고,  $B$ 는 인종(흑인 여부)을 의미한다. 개인의 다양한 인적특성( $X$ )을 통제한 후에 OLS를 통해  $\beta_1$ 을 추정하면 음(-)의 값으로 흑인에 대한 노동시장 차별이 있다는 것이 이전 연구들의 주요 결과다. 하지만 이러한 결과에 대해  $\beta_1$ 이 왜 ‘차이’가 아닌 ‘차별’이라는 질문을 던질 수 있다.

그렇다면 ‘차별’적인 부분을 어떻게 분리할 것인가? 분리를 위해서는 인종( $B$ )이 무작위 배정되어야 한다. 그래야  $B$ 가 오차항  $u$ 와 독립적이 되기 때문이다. 예를 들어, 흑인이 백인에 비해 IQ가 떨어지고 근로의욕 등이 낮다는 시나리오를 생각해 본다면,  $B$ 가 내생적이라는 것을 부인하기 어렵다. 따라서  $B$ 가 무작위 배정된 경우를 고려해야 하는데, 인종의 경우에는 사람이 태어날

때 인종을 선택해서 태어나는 사람은 없으므로 자연에 의해 무작위 배정되었다고 할 수 있다.

그렇다면  $B$ 는 더 이상 내생적이지 않다는 의미인가? 이 연구의 저자들은 노동시장 성과의 인종차별을 분석함에 있어서 인종 자체보다 인종을 인식하는 사람들에 의한 차별에 초점을 맞추었다. 예를 들어, 어떤 개인이 흑인이라서 차별을 받는다면, 그것은 그가 흑인이어서가 아니라 차별하는 기업이 그를 흑인으로 인식했기 때문이라는 것이다. 조금 극단적인 예로 어떤 개인이 실제로는 백인인데 까맣게 분장을 해서 회사에서는 그를 흑인으로 취급하여 발생하는 차이가 바로 ‘차별’이라는 것이다. 다시 말해서, 인종차별에 있어서 어떤 사람이 백인 또는 흑인으로 태어났느냐 자체가 중요한 것이 아니라, 다른 사람들이 그를 백인 혹은 흑인으로 ‘인식’하는 것이 중요한 요인이라는 것이다. 따라서 저자들은 노동시장 성과에 대한 인종차별을 연구할 때, 어떤 사람에 대한 인식을 무작위 배정하는 방법을 사용하였다. 동일한 인적특성을 갖는 이력서를 2부 작성하는데, 하나에는 흑인들이 자주 사용하는 이름을 기재하고, 다른 하나에는 백인들이 자주 사용하는 이름을 기재하는 것이다. 이렇게 무작위로 작성된 두 셋(set)의 이력서( $B$ )를 보스턴과 시카고의 구인광고 업체에 보내서 인터뷰 요청이 오는지, 즉 서류전형이 통과되었는지( $y$ )를 관찰하여 분석을 수행하였다.

[예 6] 고등학교(high school)를 무작위 배정(Lee & Kang, 2011)

한국 고등학교 졸업생들에 대해서 학연(school tie)의 영향을 추정하겠다는 것이 연구의 주요 목적이다. 어떤 사람이 좋은 학교를 졸업해서 그 학교의 학연 때문에 부당하게 다른 사람보다 좋은 대우를 받느냐에 대한 문제인식이다. 다시 말해서, 학연이라는 것이 한국 사회에서 기능을 하느냐에 대한 기본적인 물음이다. 예를 들어, A학교를 졸업한 학생이 B학교를 졸업한 학생보다 더 좋은 노동시장 성과를 보이는 것이 과연 모두 A학교와 B학교의 학연의 차이로 설명이 될 것인지에 대한 부분에 있어서는 그렇지 않을 수도 있다는 것이다. 일반적으로 학연의 효과를 추정하기 위해서 다음과 같은 실증모형을

고려해 보자.

$$y_i = \beta_0 + ST_i\beta_1 + X_i\beta_2 + u_i \quad (4-2)$$

여기서  $y$ 는 개인의 노동시장 성과지표로 임금, 대기업 취업 여부 등의 변수가 사용될 수 있다.  $ST$ 는 학연(school tie)의 크기(또는 질)를 나타내는 변수이고,  $X$ 에는 개인의 인적특성(성별, 연령, 출생지, 부모의 학력 등)이 포함된다.

이러한 실증분석 모형에 대해서 조사 자료를 이용해 OLS로  $\beta_1$ 을 추정하면, 강한 양(+)의 값이 나오며, 이로부터 학연의 크기(또는 질)가 좋아지면 노동시장 성과도 좋아진다는 인과관계 해석이 이루어진다.

하지만 학연의 크기(또는 질)를 의미하는  $ST$ 에 대해서 내생성이 존재하기 때문에 위의 OLS 결과는 인과관계가 아닌 단순한 상관관계를 의미할 수 있음에 주의해야 한다. 그렇다면 학연의 내생성은 왜 발생하는가? 우선, 임금과 같은 노동시장 성과인  $y$ 에 영향을 미치는 요인들을 살펴보면, 최소한 다음의 세 가지를 고려할 수 있다. 첫 번째는 개인의 능력이다. 이는 학교와 관련 없이 주어지는 개인의 인적특성인데, 상당 부분이 관찰되지 않거나 측정하기가 어려워  $X$ 에 포함되지 않는다. 두 번째는 학교의 교육환경이다. 이는 학연과는 구별되는 개념으로, 학교별로 교육의 질적 수준 및 학생들의 학구열 등 학교 전체의 교육환경 차이를 의미한다. 세 번째는 졸업 이후 형성되는 졸업자들의 네트워크, 바로 학연이다. 따라서 학연의 순수한 효과를 추정하기 위해서는 나머지 두 가지 요소에 대해서 완벽한 통제가 이루어져야 한다. 하지만 관찰되지 않거나 측정하기 어려운 요소들을 통제한다는 것은 실증적으로 쉬운 일이 아니다.

그렇다면 어떻게 실증분석을 수행해야 하는가? 결국 학연이 무작위로 배정된 경우를 고려해야 한다. 한국의 경우에는 1974년부터 순차적으로 이루어진 고등학교 평준화 제도를 활용할 수 있다. 평준화 제도는 추첨을 통해서 학교 내 고등학교에 학생을 임의로 배정하는 제도로서, 학생들에게 학연(고등학교)

이 무작위로 배정된 사례라 할 수 있다. 다시 말해서, 평준화 제도로 인해 어떤 학생이 입학하는 학교가 그에게 무작위로 배정되었는데, 이는 그가 졸업한 이후에 참석하게 될 동문회가 무작위로 배정되었음을 의미한다. 결국 선배(졸업생)들의 네트워크(동문회)가 무작위로 배정된 것이므로, 이를 바탕으로 이루어지는 학연 역시 무작위로 배정된 사례인 것이다. 따라서 각 고등학교 별로 학연의 크기(또는 질)인  $ST$ 를 적절히 측정하면, 앞에 제시한 모형을 OLS로 추정한  $\hat{\beta}_1$ 이 학연이 개인의 노동시장 성과에 미치는 인과적인 효과인 것이다.

이 연구에서 학연의 크기(또는 질)를 의미하는  $ST$ 는 한국거래소(구 증권거래소)에서 제공하는 2004년 「경영인명록」을 활용하여, 등재되어 있는 약 6,800명 중 평준화 이전에 고등학교를 입학한 사람들을 대상으로 학교별 인원 수를 측정하여 사용하였다. 따라서  $\hat{\beta}_1$ 은 평준화 제도로 고등학교에 입학한 어떤 사람의 선배가 「경영인명록」에 1명 증가할 때 그의 노동시장 성과가 증가하는 정도를 의미한다.

다만, 고등학교의 무작위 배정을 의미하는 평준화 제도를 활용하더라도 다음과 같은 경우에 분석 결과를 해석함에 있어 주의해야 한다. 평준화 이전의 졸업생(선배)들이 평준화 이후의 졸업생(후배)들과 네트워크(동문회)를 공유하지 않는 경우를 생각해 보자. 이러한 경우에는 고등학교가 무작위로 배정된 것이 곧바로 학연의 무작위 배정으로 이어지지 않으므로, 학연의 크기(또는 질)가 미치는 영향이 정확하게 추정되었다고 주장할 수 없는 부분이 존재한다.

그렇다면 과연 학연이 무작위로 배정되었는지를 검정할 수는 없는가? 학연이 무작위로 배정되었는지를 직접적으로 검정하기 위해서는  $ST$ 가 오차항  $u$ 와 서로 독립적(공분산이 0)임을 보여야 하지만, 실증적으로 관찰되지 않는 오차항  $u$ 와의 독립성을 검정하는 것은 불가능하다. 따라서  $ST$ 가 다른 설명변수  $X$ 와 독립적인지를 간접적으로 검정한다. 구체적으로,  $ST$ 를  $X$ 에 대해 회귀분석하여 그 추정계수가 유의한지를 F-검정하는 것이다. 이러한 검정이

가능한 이유는  $ST$ 가 만약 무작위로 배정되었다면, 관찰되는 개인의 인적특성들과도 서로 상관관계가 없을 것이기 때문이다. 이외에도 간접적인 검정방법이 한 가지 더 있는데, 그것은  $ST$ 가 무작위로 배정되었기 때문에, 회귀모형에 개인의 인적특성  $X$ 가 설명변수로 추가되어도  $ST$ 에 대한 추정계수  $\hat{\beta}_1$ 은 비슷한 값을 가져야 한다는 점을 이용하는 방법이다.

이상의 논의를 바탕으로 종합해 보면, 고등학교 평준화 제도를 활용하여 추정한 학연의 효과  $\hat{\beta}_1$ 은 과거(평준화 이전)의 졸업생(선배)들이 평준화 이후의 후배들을 인정하고 네트워크(동문회)를 공유할 때, 그리고 그것이 실질적으로 작용할 때, 어떤 사람의 노동시장 성과가 학연으로 인해서 좋아진다고 한정적인 해석을 할 수 있다.

그렇다면 만약에  $\beta_1$ 이 0이면 무엇을 의미하는가? 두 가지로 해석이 가능하는데, 하나는 학연 자체가 작용하지 않는다는 것이고, 다른 하나는 작용하더라도 비평준화 안에서만 이루어지는, 즉 평준화 이전 졸업생(선배)들과 평준화 이후의 후배들 사이의 학연은 작용하지 않는다는 것이다.

## 제2절 처방효과의 이질성 : 분위회귀분석

지금까지 어떤 정책의 효과성을 평가하기 위해서 평균처방효과(ATE)를 추정함에 있어서 처방(treatment)에 대한 무작위 배정(randomization)이 중요하다는 것을 살펴보았다. 다시 말해서, 처방에 대한 무작위 배정이 담보되면, OLS 추정을 통해서 간단하게 평균처방효과(ATE)를 구할 수 있다. 이를 수식으로 표현하면 다음과 같다.

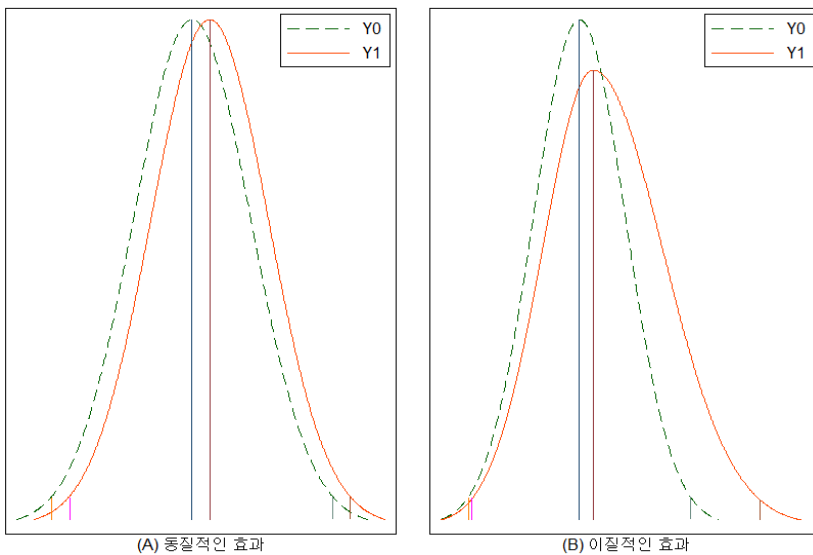
$$\begin{aligned} E(\hat{\beta}_1^{OLS}) &= E(Y_i \mid D_i = 1) - E(Y_i \mid D_i = 0) \\ &= E(Y_{1i}) - E(Y_{0i}) = E(Y_{1i} - Y_{0i}) \end{aligned} \quad (4-3)$$



여기서  $Y_{1i}$ 는 어떤 개인  $i$ 가 처방을 받았을 때( $D_i=1$ )의 성과이고,  $Y_{0i}$ 는 처방을 받지 않았을 때( $D_i=0$ )의 성과다. 따라서 처방( $D$ )이 무작위 배정된 경우의 OLS 추정치  $\widehat{\beta_1^{OLS}}$ 은 각 개인의 처방효과( $Y_{1i} - Y_{0i}$ )에 대한 기댓값(평균)을 의미한다. 이는 어떤 개인  $i$ 가 평균적인 사람일 때 그가 처방을 받을 경우와 받지 않을 경우에 얻을 수 있는 성과의 차이로도 해석될 수 있다.

하지만 어떤 처방에 대한 효과를 추정함에 있어서, 평균처방효과(ATE)가 그 효과를 모두 설명할 수 있는지에 대해서는 논란의 여지가 있다. 즉, 처방의 효과가 과연 동질적(homogeneous)인지에 대한 논란이 있을 수 있다. 이러한 논란을 아래 [그림 4-1]을 통해 살펴보면 보다 쉽게 이해할 수 있다. [그림 4-1]은 처방을 받지 않은 집단과 받은 집단의 성과에 대한 분포를 나타낸다. (A)는 처방에 대한 효과가 모든 개인에 대해 동일하게 나타나는 동질적인 경우이고, (B)는 처방에 대한 효과가 개인별로 다르게 나타나는 이질적인 경우를 보여주고 있다. 처방효과가 이질적으로 나타나는 경우(B)에 평균처방효과(ATE)가 상위 수준 개인의 처방효과와는 차이가 있음을 쉽게 발견할 수 있다.

[그림 4-1] 처방효과의 분포



앞선 학연의 효과를 추정하는 예를 통해서 이러한 문제를 살펴보면, 이질적인 처방효과에 대해서 보다 명확한 인식을 가질 수 있다. 일반적으로 학연이라는 것이 최상위의 학교 일부, 또는 최상위 수준의 학생 일부에게 적용되는 개념이라면, 그 효과를 추정함에 있어서 평균적인 수준의 학생들에 대한 추정치를 제공하는 것이 많은 의미를 갖지 못할 수 있다. 예를 들어서, OLS를 통해서 추정한 평균처방효과(ATE)  $\hat{\beta}_1^{OLS}$ 는 1,000명 중 500등인 학생에 대한 학연의 효과인데, 이것이 1,000명 중에 10등인 학생의 학연효과와는 다를 수 있다는 것이다.

그렇다면 처방효과가 이질적인 것이라 생각되는 경우에 각각의 처방효과를 어떻게 추정할 것인가? 가장 편한 방법으로 이질적인 효과가 예측되는 집단을 구분하여 해당 집단의 평균처방효과(ATE)를 추정하는 방법을 생각할 수 있다. 하지만 집단을 구분하여 평균처방효과를 추정할 경우에는 구분된 표본을 사용하기 때문에 표본선택편의(sample selection bias)가 발생하게 된다.

처방이 무작위 배정된 상황에서 이질적인 처방효과를 분석하기 위해서, 최소자승법(OLS) 대신 분위회귀법(Quantile Regression)을 고려할 수 있다. 분위회귀분석을 활용하면, [그림 4-1]의 (B)와 같이 평균 수준에서의 처방효과와 상위 수준(high tail)에서의 처방효과가 다른 경우에 각각의 효과들에 대한 정보를 간접적으로 추정할 수 있다. 즉, 어떠한 처방(정책)의 수혜 유무에 따른 상위 수준(high tail) 사람들의 차이, 중위 수준(median or mean) 사람들의 차이, 하위 수준(low tail) 사람들의 차이를 각각 다르게 추정할 수 있다.

분위회귀 추정치 도출을 수학적으로 살펴보기 위해서 다음의 모형을 고려해 보자.

$$Y_i = (\beta_0 + D_i\beta_1 + X_i\beta_2) + U_i \equiv \rho(D_i, X_i) + U_i \quad (4-4)$$

식 (4-4)는 OLS를 분석하는 데에 있어서 설정된 모형과도 동일하다. OLS는 평균에 대한 효과를 추정하는 방법인데, 다음과 같이 오차의 제곱값을 최소화하는 추정치  $\hat{\beta}_k$ 을 구하는 것이다.

$$\min_{\beta} \sum U_i^2 = \sum [Y_i - \rho(D_i, X_i)]^2 \quad (4-5)$$

그런데 OLS는 특이 관측치(outlier)에 대해서 매우 민감하게 반응한다는 문제를 내포하고 있다. 이러한 문제를 해소하기 위해서, 즉 특이 관측치에 덜 민감하고 강건(robust)한 추정치를 구하기 위해서 고안된 방법이 중위회귀법(median regression)이다. 식 (4-5)에서 극소화하는 목적함수의 제곱값을 다음 식 (4-6)과 같이 절댓값으로 바꾸면 중위회귀 추정치를 구할 수 있다.

$$\min_{\beta} \sum |U_i| = \sum |Y_i - \rho(D_i, X_i)| \quad (4-6)$$

이와 같이 평균을 대신하여 중위수를 대푯값으로 하는 추정치를 고려하면, 중위수뿐만 아니라 각 백분위수(percentile)에서의 추정치를 도출할 수 있는데, 위 극소화 문제의 목적함수를 다음과 같이 변형하여 쉽게 구할 수 있다.

$$\min_{\beta^{\theta}} \left\{ \begin{array}{l} \sum_{i: Y_i \geq \rho(D_i, X_i)} \theta \cdot |Y_i - \rho(D_i, X_i)| + \\ \sum_{i: Y_i < \rho(D_i, X_i)} (1 - \theta) \cdot |Y_i - \rho(D_i, X_i)| \end{array} \right\} \quad (4-7)$$

여기서  $\theta \in (0, 1)$ 는 관심 있는 분위수로, 예를 들어 90분위수(90 percentile)에 관심이 있으면 0.9가  $\theta$ 값이 된다. 이와 같이 관심 있는 분위수  $\theta$ 만을 결정하면, 해당 분위에 대한 처방효과 추정치  $\hat{\beta}_1^{\theta}$ 을 구할 수 있다.

OLS 추정치와 달리 분위회귀 추정치를 해석함에 있어서는 상당한 주의를 기울여야 한다. 앞에서도 언급하였듯이, OLS 추정치  $\hat{\beta}_1^{OLS}$ 은 처방집단(treatment group)과 통제집단(control group) 각각의 평균 성과의 차이로, 이는 어떤 평균 수준의 개인  $i$ 에게 처방이 주어졌을 때와 주어지지 않았을 때의 차이라고 해석이 가능하다. 즉, 평균 수준 개인  $i$ 가 처방을 받음으로 해서 얻게 되는 성과인 처방효과라고 해석이 가능함을 의미한다. 하지만 분위회귀 추정치  $\hat{\beta}_1^{\theta}$ 은 개인의 입장에서의 처방효과 수준이라고 해석할 수 없다. 다시

말해서,  $\hat{\beta}_1^\theta$ 은 처방집단의  $\theta$ 분위 사람과 통제집단의  $\theta$ 분위 사람 사이의 성과 차이를 의미하는 것이지, 어떤  $\theta$ 분위의 개인  $i$ 가 처방을 받음에 따라서 얻게 되는 성과인 처방효과로는 해석할 수 없다는 것이다. 이는 분위회귀 추정치가 OLS 추정치의 기댓값 연산자와 달리 수학적으로 선형 연산자(linear operator)가 아니기 때문이다. 이를 수식으로 표현하면 다음과 같은데, 위에서 살펴본 OLS 추정치와 달리 연산자 사이의 결합이 불가능하기 때문에  $\theta$ 분위 사람에 대한 처방효과로 추정치를 해석하는 것에 주의해야 한다.

$$\begin{aligned} Quant_\theta(\hat{\beta}_1^\theta) &= Quant_\theta(Y_i \mid D_i=1) - Quant_\theta(Y_i \mid D_i=0) \\ &= Quant_\theta(Y_{1i}) - Quant_\theta(Y_{0i}) \neq Quant_\theta(Y_{1i} - Y_{0i}) \end{aligned} \quad (4-8)$$

이와 같이 분위회귀 추정법을 통해서 어떤 다른 수준의 개인 각각에 대한 처방효과를 추정할 수 있는 것은 아니다.<sup>1)</sup> 그렇다면 분위회귀 추정치로부터 얻을 수 있는 이질적인 처방효과에 대한 함의는 무엇인가? 상위 수준 사람들의 성과 차이와 하위 수준 사람들의 성과 차이의 크기를 비교함으로써, 상위 수준 또는 하위 수준 사람들의 처방효과 크기가 차이날 수 있으며 어느 수준에서 처방효과가 더 좋을 수 있는지를 예측하는 데 분위회귀 추정치가 함의를 제공할 수 있다.

뿐만 아니라, 어떤 정책의 효과를 평가함에 있어서도 유용한 정보를 제공할 수 있다. 예를 들어, 어떤 정책이 저소득층에 얼마나 영향을 미치고 있는지를 평가하고자 한다고 해보자. 만약에 해당 정책의 효과가 소득 수준에 따라서 이질적이라면, 평균적인 소득 수준의 사람에 대한 정책효과를 제시하는 OLS 추정치로는 저소득층에게 미치는 영향을 파악할 수 없다. 이러한 경우에 분위회귀 추정치를 활용하여 소득 수준별로 상이한 정책의 효과를 간접적으로 유추할 수 있다. 더욱이, 만약에 정책의 주요 목적이 저소득층의 보조에 있음에도 불구하고 저소득층보다는 고소득층에서 정책의 효과가 클 것으로

1) 아직까지 각 분위에 있는 사람에게 미치는 분위별 처방효과(quantile treatment effect)를 정확히 추정하는 일반적인 계량경제학 방법은 전무한 상태다.

유추된다면, 해당 정책에 대한 평가의 방향을 아주 달라질 수 있게 하는 함의를 분위회귀 추정치가 제시할 수 있다.

[예 1] 고등학교(high school) 학연효과에 대한 이질성 고려(Lee & Kang, 2011)

앞에서 언급한 바 있는 한국의 고등학교 평준화 제도에 따른 고등학교 무작위 배정을 이용한 학연효과 추정의 예에서, 노동시장 성과 수준별로 효과의 이질성이 나타나는지를 분위회귀 추정을 통해서 살펴보는 것이 연구의 추가적인 목적이다. 연구 결과의 요지를 정리하면, 학연에 대해서 일반적으로 생각할 수 있듯이, 노동시장 성과의 수준( $\theta$ )이 증가할수록 분위회귀 추정치  $\hat{\beta}_1^\theta$ 의 크기가 커지고 통계적으로 유의하다는 것이다. 이러한 결과로부터, 한국에서의 학연의 크기(또는 질)가 개인의 노동시장 성과에 미치는 영향은 상위 수준의 학생들 일부에게 작용된다는 것을 유추할 수 있다.

이 예에서도 마찬가지로 분위회귀 추정치의 해석에 있어서는 상당한 주의를 기울여야 한다. 이 예에서 분위회귀 추정치  $\hat{\beta}_1^\theta$ 는 다음과 같이 해석되어야 한다. 2004년 「경영인명록」에 등재되어 있는 평준화 이전 졸업 선배의 수( $ST$ )가 1명 차이 나는 두 학교의  $\theta$ 분위 학생들 사이에 나타나는 노동시장 성과의 차이가 바로  $\hat{\beta}_1^\theta$ 인 것이다. 어떤 학교의 평준화 이전 졸업 선배가 「경영인명록」에 1명 증가할 때, 그 학교  $\theta$ 분위 학생의 노동시장 성과가  $\hat{\beta}_1^\theta$ 만큼 증가함을 의미하는 것이 절대 아니라는 점을 반드시 기억해야 한다.

지금까지 무작위 배정에 의한 정책평가의 필요성 및 중요성, 그리고 편리성에 대해서 살펴보았다. 어떤 정책의 수혜 대상과 비수혜 대상을 구성함에 있어서 무작위 배정이 이루어진다면, 보다 간단한 계량적 추정기법을 활용하여 명확한 평가 결과를 제시할 수 있다. 이러한 논의로부터 정부 정책의 성과를 올바르게 평가하기 위해서는 정책을 계획하는 단계에서부터 이러한 방법으로 정책을 설정할 필요가 있다는 주장을 이끌어 낼 수 있다. 왜냐하면, 이러

한 방법으로 정책이 설정된다면, 정책의 평가도 제대로 이루어질 뿐만 아니라 지금까지의 문제는 무엇이었고, 해결을 위해 앞으로는 무엇을 해야 하는지도 비교적 상당히 명확하게 제시할 수 있는 등 여러 가지 장점이 있기 때문이다.

미국의 경우에는 직업훈련제휴법(Job Training Partnership Act : JTPA)을 통해서 수행되는 직업훈련 프로그램의 신청자들에 대해서 그 수혜 여부를 무작위로 배정한다. 그리고 프로그램의 평가를 위해서는 무작위로 선정된 처리 집단은 물론이거니와 무작위 탈락한 통제집단에 대한 정보 역시 필요하다. 따라서 JTPA에 의한 직업훈련 프로그램에서는 신청 후 탈락한 사람들에 대해서도 사후 관리를 한다. 이러한 미국의 예로부터, 정책평가에 있어서 무작위 배정도 중요하지만, 무작위 배정 이후 탈락자에 대한 정보를 추적 조사할 수 있는 여건을 마련하는 것 역시 상당히 중요하다는 함의를 제시할 수 있다.

## 제1절 이중차분의 의미

어떠한 정책에 참여한 사람들의 집단(treatment group, 처리(처방 혹은 처치)집단)과 참여하지 않은 사람들의 집단(control group, 통제집단)의 결과를 비교하여 정책 효과를 볼 때 처리되기 전의 집단들 상태가 서로 균형 잡혀 있어야 한다. 즉, 처리집단과 통제집단이 처리 이외의 다른 특징들(관측 가능할 수도 불가능할 수도 있음)이 동일하여야 한다는 의미다. 그렇지 않다면 동일하지 않은 상태를 제거해 주어야 처리에 대한 순수한 효과를 추정할 수 있는 것이다. 그래서 고안한 방법이 이중차분법이다. 이중차분법(Difference-in-Difference estimator : 이하 DID라고 함)은 차분을 두 번 한다는 의미다.

가장 간단한 구조인 2개의 지역(Near, Far)과 2개의 시간(1978, 1981)을 가지고 예를 들어 보자. 쓰레기 소각장의 설치가 부동산 가격에 영향을 미치는 지에 대해 관심을 가지게 되므로 다음과 같이 실험을 할 수 있다.<sup>2)</sup> 어느 날 시장(市長)이 동전을 던져서 앞면이 나온 지역에 쓰레기 소각장을 설치한다고 해 보자. 쓰레기 소각장이 들어오게 될 지역의 선정은 무작위 배정방식이라는 것을 알 수 있다. 실제로 1979년에 Near 지역으로 쓰레기 소각장이 들

2) Wooldridge(2008)의 제13장 제2절에 소개된 내용이다.

〈표 5-1〉 지역에 따른 평균 집값

(단위 : 미국달러)

	Near(1)	Far(2)	(1)-(2)
1978(Ⅰ)	63,693	82,517	-18,824
1981(Ⅱ)	70,620	101,308	-30,688
(Ⅱ)-(Ⅰ)	6,927	18,791	-11,864

어가게 되었다. 쓰레기 소각장이 들어온 Near 지역은 이로 인해 부동산 가격이 떨어지게 되어 아파 사람들이 반대를 많이 했을 것이다.

〈표 5-1〉을 보면, 쓰레기 소각장이 있기 전인 1978년 Near 지역과 Far 지역의 집값 차이는 -18,824달러이고, 1981년 두 지역의 집값의 차이는 -30,688달러다. 그런데 -30,688달러의 값은 쓰레기 소각장이 설치되기 전의 집값의 차이와 쓰레기 소각장이 설치된 후의 집값의 하락분이 포함되어 있어서 순수한 쓰레기 소각장의 영향이라고 할 수 없다. -30,688달러에서 1978년 두 지역의 집값 차이(-18,824달러)를 빼준 값 -11,864달러가 쓰레기 소각장 설치에 따른 순수한 효과라고 할 수 있다.

이를 수식으로 표현하여 설명해 보면 다음과 같다.

$$rprice_{i,t} = \gamma_0 - \gamma_1 \text{nnearinc}_{i,t} \quad (5-1)$$

여기서, rprice는 각 연도에 대한 지역의 집값이고, nnearinc는 지역의 쓰레기 소각장 설치 여부를 나타내 주는 더미변수다. 1981년과 1978년 각각에 대한 회귀식을 추정하면 아래와 같다.

$$\widehat{rprice} = 101,307.5 - 30,688.3 \text{nnearinc} \quad (5-2)$$

$$\widehat{rprice} = 82,517.2 - 18,824.4 \text{nnearinc} \quad (5-3)$$

식 (5-2)를 보면 쓰레기 소각장이 설치되면 집값이 30,688.3 달러가 하락한



다. 이 값은 <표 5-1>를 보면 1981년의 두 지역에 대한 집값의 차이와 같은 것을 알 수 있다. 그리고 절편값, 즉 101,307.5 달러는 쓰레기 소각장이 없는 지역에 대한 1981년도의 집값이다. 식 (5-3)을 보면 쓰레기 소각장이 설치 되면 집값이 18,824.4 달러가 하락한다는 것을 의미하고, <표 5-1>에서 1978년의 두 지역의 집값의 차이를 나타내고 있다. 그리고 절편값, 즉 82,517.2 달러는 쓰레기 소각장이 없는 지역에 대한 1978년도의 집값이다.

여기서 알고 싶은 것은 쓰레기 소각장이 설치된 후의 집값의 변동 ( $\widehat{\gamma}_1^{81} - \widehat{\gamma}_1^{78}$ )이고, 식 (5-4)와 같이 표현할 수 있다.

$$\begin{aligned} \text{추정하고자 하는 효과} &= \widehat{\gamma}_1^{81} - \widehat{\gamma}_1^{78} & (5-4) \\ &(\text{여기서, } \widehat{\gamma}_1 = \overline{rprice}_{nr} - \overline{rprice}_{fr}) \\ &= \overline{rprice}_{81,nr} - \overline{rprice}_{81,fr} - [\overline{rprice}_{78,nr} - \overline{rprice}_{78,fr}] \\ &= \overline{rprice}_{81,nr} - \overline{rprice}_{78,nr} - [\overline{rprice}_{81,fr} - \overline{rprice}_{78,fr}] \end{aligned}$$

식 (5-2)와 (5-3)은 차분을 한 번 한 식이다. 여기서 다시 1981년 차분식에서 1978년 차분식을 빼주어 두 번 차분한 것이 DID 추정량이며, 이것이 쓰레기 소각장 설치(처리)에 대한 효과의 값(-11,864)이다. 이는 식 (5-4)처럼 나타낼 수 있으며, 앞에서 <표 5-1>을 이용하여 설명한 값과도 일치한다는 것을 알 수 있다.

하나의 회귀모형으로 쓰레기 소각장 설치에 대한 효과를 추정하기 위해서는 데이터 구조를 1978년과 1981년 자료를 수직으로 붙여 한 개의 데이터로 구성하면 된다. 그러면 1978년과 1981년을 나타내는 연도 더미변수(year dummy( $y_{81}$ ))가 필요하고, 집 주변에 쓰레기 소각장이 있는지 여부를 나타내는 더미변수( $nnearinc$ )도 필요하다. 그리고 이 두 변수들을 곱한 상호작용(interaction) 변수를 포함한 후 집값을 종속변수로 놓고 회귀분석을 하면 된다. 이것을 수식으로 표현하면 다음과 같다.

$$rprice = \beta_0 + \delta_0 y_{81} + \beta_1 nnearinc + \delta_1 y_{81} * nnearinc + u \quad (5-5)$$

그렇다면 식 (5-5)에서 쓰레기 소각장 설치 효과를 알려 주는 계수가 무엇인지 알기 위해 조건부 기댓값(conditional expectation)을 계산하면 아래와 같다.

$$E_{11} \equiv E(rprice|y_{81} = 1, nnearinc = 1) = \beta_0 + \delta_0 + \beta_1 + \delta_1 \quad (5-6)$$

$$E_{10} \equiv E(rprice|y_{81} = 1, nnearinc = 0) = \beta_0 + \delta_0$$

$$E_{01} \equiv E(rprice|y_{81} = 0, nnearinc = 1) = \beta_0 + \beta_1$$

$$E_{00} \equiv E(rprice|y_{81} = 0, nnearinc = 0) = \beta_0$$

식 (5-6)에서  $(E_{11} - E_{10}) - (E_{01} - E_{00}) = (\beta_1 + \delta_1) - \beta_1 = \delta_1$ 이라는 것을 알 수 있으며, DID 추정량은 상호작용 변수의 계수인  $\delta_1$ 이라는 것을 알 수 있다.  $\beta_1$ 은 소각장이 들어온 지역( $nnearinc = 1$ )과 없던 지역( $nnearinc = 0$ )의 1978년( $y_{81} = 0$ ) 평균 가격의 차이이고,  $\beta_0$ 는 소각장이 없는 지역( $nnearinc = 0$ )의 1978년( $y_{81} = 0$ ) 평균 가격 수준이며,  $\delta_0$ 는 쓰레기 소각장이 없던 지역( $nnearinc = 0$ )의 1981년( $y_{81} = 1$ )과 1978년( $y_{81} = 0$ ) 평균 가격의 차이를 의미한다.

그리고 이중차분 추정 시 제한적 가정은 시간이 지나면서 처리가 없는 경우에는 결과변수의 변화량은 처리집단과 통제집단이 정확하게 같아야 한다는 것이다. 이는 변화의 추세(회귀식에서 기울기)가 같다는 것을 의미한다.

식 (5-5)는 단순한 모형인데, 여기에 다른 설명변수들( $x$ )을 추가할 수 있다. 예를 들면 개별 집들의 특성(방의 개수나 층수 등)을 통제하여 분석할 수 있으며 이는 식 (5-7)로 표현할 수 있다.

$$rprice = \beta_0 + \delta_0 y_{81} + \beta_1 nnearinc + \delta_1 y_{81} * nnearinc + x_i' \gamma + u \quad (5-7)$$

또한 식 (5-7)은 식 (5-5)와의 비교를 통해 이중차분법이 제대로 적용되는지를 확인할 수 있는 방법으로도 사용할 수 있다. 설명변수들( $x$ )을 추가하거

나 빼더라도  $\delta_1$ 의 해석(의미)이 달라지지 않아야 한다. 식 (5-5)와 식 (5-7)을 따로 추정했을 때  $\delta_1$ 의 계수 추정값이 크게 다르지 않으면 이중차분법이 잘 적용되었다고 판단할 수 있다.

앞의 예(쓰레기 소각장의 설치에 대한 집값의 효과)를 추정할 때 사용한 데이터는 반복 횡단면(repeated cross-sectional) 자료다. 만약 자료가 균형패널(balanced panel)인 경우에는 모형을 다음과 같이 설정할 수도 있다.

$$\Delta rprice_i = \tau_0 + \tau_1 nnearinc_i + x_i' \gamma + u_i \quad (5-8)$$

균형패널 자료이므로 각각의 집이 달라진 것이 없다면 1981년과 1978년에 대한 집값의 차이 값( $\Delta rprice_i$ )을 종속변수로 할 수 있다. 이렇게 하면 한번 차분을 한 것과 같다. 이때 설명변수는 쓰레기 소각장의 설치 여부 변수( $nnearinc_i$ )만 넣어야 한다. 종속변수에서 이미 한 번 차분되었고, 이  $\Delta rprice_i$ 는 소각장이 있었던 지역과 없었던 지역 차이이고,  $\tau_1$ 은 종속변수의 일계 차분값이므로 이중차분이 되는 것을 알 수 있다. 그러므로 시간변수( $y81$ )를 넣으면 안 되는 것이며, 식 (5-5)와 다른 점이다. 다른 설명변수는 추가할 수 있으며, 이중차분 추정량은  $\tau_1$ 이다.

#### [예 1] 최저임금 상승에 따른 고용효과(Card & Krueger, 1994)

이 논문은 이중차분법을 크게 유행시켰다. 연구 목적은 최저임금이 오르면 고용에 어떠한 영향이 있는지에 대해 살펴본 것이다. 1992년 4월 1일에 뉴저지의 최저임금법이 통과되면서 시간당 최저임금이 4.25달러에서 5.05달러로 오르게 되었다. 그리고 뉴저지의 경계 지역에는 펜실베이니아가 있는데 뉴저지의 최저임금은 올라갔지만 펜실베이니아는 변동이 없었다.

그래서 저자들이 생각한 것은 두 지역이 붙어 있는데 한 지역의 최저임금이 올랐을 경우에 접경 지역들을 비교해 보자는 것이었다. 최저임금을 받는 근로자를 주로 채용하는 기업은 패스트푸드점이다. 뉴저지 전체 패스트푸드점과 펜실베이니아의 접경 10마일 내 지역의 패스트푸드점을 비교해 보았다.

펜실베이니아의 전체를 보지 않은 이유는 접경 지역이어야 상당히 비슷할 것이기 때문이다. 4월 1일에 제도가 바뀌므로 1, 2월에 각 패스트푸드점에 근로자가 몇 명(정책 변화 전)인지 조사하고 최저임금이 오른 후 11, 12월에 동일한 패스트푸드점의 근로자 수(정책 변화 후)를 조사하였다. 최저임금이 달라졌다고 해서 패스트푸드점이 이사를 할 경우는 굉장히 낮을 것이기 때문에 가능하였다.

패스트푸드점의 근로자 수를 종속변수로 하였다. 지역이 뉴저지인지 펜실베이니아인지에 대한 지역 더미변수가 있으며, 정책 변화 전인지 후인지에 대한 시간 더미변수가 있고, 시간과 지역 간의 상호작용 더미변수, 그리고 레스토랑의 특성(빈민가에 있는지, 고속도로 주변에 있는지 또는 인구 등)에 대한 변수들도 함께 통제하여 회귀모형을 추정하였다. 여기서 이중차분 추정량은 시간과 지역 간의 교호 더미변수에 대한 계수라는 것을 알 수 있다. 또한 이 자료는 패널(panel) 구조이므로 종속변수를 정책 변화 전후에 대한 근로자 수의 차이로 놓고 추정할 수 있다. 이때 이중차분 추정량은 지역 더미변수의 계수다.

[예 2] 나라별 비평준화 제도에 따른 교육 불평등의 효과(Hanushek & Woessmann, 2006)

비평준화 제도를 실시하게 되면 학생들의 교육 불평등 정도에 얼마나 영향을 미치는지에 대해 살펴본 논문이다. 보통 학생들의 성적을 종속변수로 비평준화를 받았는지 받지 않았는지에 대한 여부를 더미변수로 하여 회귀분석을 실시한다. 그러나 비평준화에 대한 여부가 외생적이지 않은 경우에 계수값은 편의되어 있으므로 올바른 추정이 아닌 것이다. 저자들은 이 문제를 제기하면서 이중차분법을 제안하였다.

전 세계를 보면 어떤 나라들은 비평준화를 아주 일찍 시작하는 나라가 있는데 독일·오스트리아·헝가리와 네덜란드가 10세에 시작하여 초등학교 3학년 때 우열학교가 생긴다. 반면에 캐나다·일본·영국·한국은 늦게 시작하는 편이다. 이렇게 나라마다 비평준화를 시작하는 시점이 모두 다르다. 그

러나 초등교육은 똑같이 중등교육(secondary education)에서 달라지므로 중등교육 단계에서 교육성과의 차이를 초등교육 단계에서의 차이로 빼주는 것, 즉 이중차분 아이디어인 것이다. 초등학교와 중학교 사이의 차이를 종속변수로 하고 나라를 통제하여 회귀분석하였다. 여기서 이중차분 추정량은 각 나라가 중학교에서 비평준화를 받았는지 받지 않았는지에 대한 여부인 것이다.

[예 3] 평준화 제도에 따른 임금차이 효과분석(Kang, Park & Lee, 2007)

우리나라의 평준화 제도에 따른 임금의 차이를 분석한 논문이다. 우리나라에서는 평준화 제도가 1974년부터 이루어져 왔다. 1974년에는 서울·부산에, 1975년에는 대구·인천·광주에 인문계 고등학교 평준화 제도가 실시되었다. 평준화 제도 때문에 졸업 후 노동시장에서 나타나는 결과가 체계적으로 차이가 있다고 생각할 수 있다. 이를 확인하기 위해 1973년 그룹과 1974년 그룹 간의 차이는 평준화 제도 이외에는 없을 것이라는 가정을 한 후, 다음과 같이 모형을 설정하였다.

한 지역에 태어나서 그 지역의 고등학교에 1974년 전후로 하여 들어간 사람들을 추출한 후 1969~1973년 이전에 다닌 비평준화 제도 그룹(1973년 그룹)과 1974~1978년 이후에 다닌 평준화 제도 그룹(1974년 그룹)으로 구분하였다. 고등학교 입학이 1974년 이후 또는 1973년 이전인지를 알려주는 처리지시(treatment indicator) 변수와 2000년대 중반의 임금을 종속변수로 하여 회귀분석하였다. 1974년의 평준화 제도는 인문계 고등학교에만 적용되고 실업계 고등학교에는 적용되지 않았다. 즉 인문계 고등학교 졸업자를 처리집단으로, 실업계 고등학교 졸업자를 통제집단이라고 설정할 수 있다. 각 집단에 대하여 1973년 그룹과 1974년 그룹 간 임금차이를 구하여 '인문계-실업계'의 차분을 구하면 이는 이중차분을 적용한 평준화 효과의 추정치로 해석된다.

## 제2절 이중차분법의 일반화

다수의 그룹이고 시점이 여러 개인 경우에는 일반화 DID를 이용하여 분석하면 되는데 다음 3개의 예를 가지고 설명하겠다.

[예 1] 기회비용의 변화에 따른 종교 활동 영향 분석(Gruber & Hungerman, 2008)

이 논문은 종교 활동 참가의 기회비용이 변화함에 따라 종교 활동 참가 정도에 어떠한 영향이 있는지에 대해 분석한다.

먼저, 역사적 배경을 설명하자면 미국에서 일요일에는 상업 활동을 못하게 하는 청교도적 법률(blue law)이 예전 1930년대 이전부터 존재하였다. 그런데 이 법률을 폐지하는 주들이 생기기 시작하였다. 미국 모든 주가 같은 해에 된 것은 아니고 각 주별로 1950년대, 1960년대 또는 1970년대에 점진적으로 없어지게 되었다. 상업 활동을 일요일에 한다는 것은 종교적 관점에서 보면 기회비용이 커지는 것이다. 왜냐하면 아무 할 일이 없어서 교회 가는 기회비용이 0이라고 하면 법률 폐지로 인해 다른 활동을 할 수 있게 되면서 교회를 가는 기회비용이 0보다 커지게 되기 때문이다. 그래서 종교 활동에 대한 기회비용이 사람들의 종교 행동에 미치는 영향을 테스트해 볼 수 있는 좋은 연구라고 볼 수 있다.

여기서는 지역(state)과 시간(time)이 여러 개로 다른 경우이고, 시간과 지역의 정보를 알면 청교도적 법률이 언제 폐지되었는지 여부를 알 수 있다. 따라서 시간과 지역의 교호(interaction)인 것이고, 구해진 계수는 알고자 하는 일반화 DID의 추정값이 되는 것이다. 그리고 시간과 지역을 통제하고 각 주별로 평균 교회를 가는 사람들의 값을 종속변수로 놓고 일반화 DID 모형을 구하면 청교도적 법률의 효과를 추정할 수 있게 된다.

[예 2] 핸드폰 서비스 도입으로 인한 시장성과 효과 분석(Jensen, 2007)

이 논문의 기본적인 아이디어는 정보기술(informational technology)의 발달이 시장에서 가격분산(price dispersion)이나 균일가격제(one price) 등의 형성에 어떠한 영향을 미치는가에 대한 것이다.

인디아의 서쪽 Kerala 반도의 해변 지역에는 지역별 생선시장(fisher market)이 있는데 이들 생선시장의 운영 방식은 다음과 같다. 밤새도록 사람들이 나가서 고기를 잡고 집으로 돌아오는데, 그 당시에는 냉장고가 없었으므로 잡은 고기를 모두 팔고 와야 한다. 이때 어부들은 고기를 비싸게 팔고 싶어 할 것이다. 그러나 각 지역마다 생선 가격이 다르며 가격 정보도 알 수가 없는 상황이었다. 만약 어부들에게 핸드폰이 생겨서 각 지역의 생선 가격을 물어 볼 수 있게 된다면 균일가격제가 성립하게 될 가능성이 높을 것이다. 그래서 핸드폰 서비스가 1997년부터 2001년 사이에 각 지역에 들어오게 되면서 시장가격의 지역 간 분산에 어떠한 결과를 나타내는지 분석한 것이다. 이 예는 앞에서 설명한 제1절의 [예 1]과 같은 구조라는 것을 알 수 있다.

처리(treatment)가 각각 다른 시점으로 들어오기 때문에 통제그룹의 처리가 언제 들어오는지 그 시점은 모른다. 이렇게 각 지역마다 핸드폰 서비스 도입 시점이 다르므로 일반화 DID를 적용하고, 회귀모형은 각 지역에서 핸드폰 서비스가 있었는지 없었는지에 대한 변수와 시간과 지역변수를 통제해 주면 된다. 핸드폰 서비스가 들어오는 것은 지역과 시간의 교호(interaction)이고 여기서 구해진 계수는 알고자 하는 일반화 DID의 추정값인 것이다. 그러므로 핸드폰 서비스로 인하여 발생하는 시장성과에 대한 일반화 DID 추정량을 구할 수 있는 것이다.

[예 3] 평준화 교육의 성적효과 분석(Pekkarinen, Uusitalo, & Kerr, 2009)

본 논문은 핀란드에서 1972~1977년 사이에 확장된 초·중등학교 평준화 제도를 이용하여 평준화 교육의 성적효과를 검토한다. 1970년대 이전 핀란드의 교육 제도는 비평준화 제도로 규정되나, 1972년부터 북부 지역부터 평준화 제도가 도입되어 1977년 헬싱키 지역에 마지막으로 도입되었다. 본 논문

은 4년 동안 시차를 가지고 도입된 핀란드의 평준화 제도를 자연실험으로 활용한다.

분석에 사용된 자료는 Finnish Army Basic Skills Test로서 남자들이 군입대 시에 치르는 언어, 수학, 논리 과목의 시험 성적이다. 1962~1966년 사이에 태어난 남자 징집병들이 분석의 대상이 된다. 성적 자료를 센서스 자료와 결합하여 개인들의 생일, 거주지, 형제 관계 및 부모의 교육 정도와 소득 자료를 확보하였다.

분석 결과에 의하면, 수학과 논리 과목에서는 평준화의 효과가 유의하게 나타나지 않는다. 그러나 언어 과목에서는 양(+)의 효과가 있으나 그리 크지는 않다. 부모의 배경에 따라 평준화의 효과가 다르게 나타난다. 저학력의 부모를 가진 학생들은 평준화 제도하에서 더 높은 학업 성취도를 보인다. 그러나 고학력의 부모를 가진 학생들에게는 평준화 제도의 유의한 효과가 나타나지 않는다.



## 성향점수 매칭법

### 제1절 성향점수 매칭의 정의

새로운 정책을 도입하여 실시하면 그 정책이 잘 수행되고 있는지에 대해 성과평가를 하고자 한다. 그 정책에 대한 인과관계(causality), 즉 처리효과를 알아보고자 하는 것이다. 처리는 직업훈련과 같은 교육프로그램으로 한 사람에게 참여하거나 참여하지 않는 경우인 두 가지 상태로 나타나며, 처리효과는 정책에 참여했을 때 얻는 성과(Factual)와 동일한 참여자가 정책에 참여하지 않았더라면 얻었을 성과(Counterfactual)의 차이인 것이다. 그러나 정책에 참여한 대상이 참여하지 않아서 얻는 성과는 관측할 수 없는 값이어서 구할 수 없으므로 아주 유사한 특성을 가지면서 정책에 참여하지 않은 대상을 찾아서 매칭(matching)하여 차이를 구한다면 정책에 대한 처리효과를 추정할 수 있을 것이다. 이 과정이 성향점수 매칭법(Propensity Score Matching : PSM)이다.

수식을 가지고 설명하면 다음과 같다.  $y_1$ 은 처리를 받은 경우의 결과,  $y_0$ 는 처리를 받지 않은 경우의 결과이고,  $d$ 는 처리 여부에 대한 더미변수(Dummy Variable)다. 처리를 받은 대상에 대한 평균효과(average effect of treatment on the treated : ATT)가 알고자 하는 값이며, 식 (6-1)과 같이 표현

된다. 그런데  $E(y_0|d=1)$ 의 값은 실제로는 처리를 받을 사람이 만약에 처리를 받지 않았다면 갖게 될 값으로 이 값은 직접적으로 관찰이 불가능한 값이다.

$$ATT=\tau = E(y_1 - y_0|d=1) = E(y_1|d=1) - E(y_0|d=1) \quad (6-1)$$

그래서 ATT를 추정하기 위해서 식 (6-2)로 표현되는 조건부 독립성 가정(Conditional Independence Assumption : CIA)이 도입된다. 이 가정의 의미는  $X$ 가 주어지면 처리효과( $d$ )와 잠재적인 결과값( $y_0, y_1$ )은 독립이라는 것이다.

$$(y_0, y_1) \perp\!\!\!\perp d | X \quad (6-2)$$

$X$ 는 관측할 수 있는 값(학력, 성별, 연령, 직업 등)으로 조건이 주어지면 처리 여부와 상관없이  $y_0$ 의 결과를 구할 수 있고 Counterfactual로 사용할 수 있는 것이다. 그래서 CI 가정에 의하여  $E_X(E(y_0|d=1, X)|d=1)$ 은  $E_X(E(y_0|d=0, X)|d=1)$ 으로 표현이 가능하여 ATT는 식 (6-3)과 같이 표현할 수 있다.

$$\begin{aligned} ATT=\tau &= E(y_1|d=1) - E_X(E(y_0|d=1, X)|d=1) \\ &= E(y_1|d=1) - E_X(E(y_0|d=0, X)|d=1) \end{aligned} \quad (6-3)$$

예를 들어  $X$ 를 성별이라고 한다면  $E_X(E(y_0|d=0, X)|d=1)$ 는 남성들 중에서 처리를 안 받은 사람들의  $y_0$ 의 평균과 여성들 중에서 처리를 안 받은 사람의  $y_0$ 의 평균을 구한 후, 이 둘에 대한 가중평균을 구하면 되는 것이다. 이 예는 성별이 하나인 경우였지만 학력·연령 또는 출신 지역 등으로  $X$ 의 차원(dimension)이 많아지게 되면 계산이 복잡하고 정확한 매칭도 이루어지기 어려우므로 이 방법으로는 해결하기 어렵다. 그래서  $X$ 의 특성을 하나의

값으로 나타내는 성향점수(Propensity Score)  $p(X)$ 를 사용하여 ATT를 구하고 식 (6-4)와 같이 표현할 수 있다.

$$ATT = \tau = E(y_1 | d = 1) - E_{p(X)}(E(y_0 | d = 0, p(X)) | d = 1) \quad (6-4)$$

## 제2절 성향점수 매칭법

PSM 과정을 매칭할 때 사용하는 방법 중의 하나인 Nearest neighbor matching을 가지고 설명해 보겠다.

(1)  $d$ 를 종속변수로,  $X$ 를 설명변수로 놓고 프로빗 모형(Probit Model) 또는 로짓 모형(Logit Model)에서 구한 추정치의 예측 확률값(Predicted Probability Value)인 성향점수( $\hat{p}(X_i)$ )를 구한다.

(2) 처리집단 부표본(treatment subsample)의 각  $i$ 를 선택하고, 그 부표본에서  $i$ 를 지운다. 다음은 통제집단 부표본(control subsample)에서 관측값  $j$ 를 찾는다. 여기서  $\hat{p}(X_j)$ 와  $\hat{p}(X_i)$ 의 값은 가장 가까워야 하며, 관측치  $j$ 는 다음  $\hat{p}(X_j)$ 를 찾을 때 다시 사용한다. 즉, 복원추출(with replacement) 방식으로 진행되는 것이다. 처리집단 부표본의 모든 관측값에 대응하는  $\hat{p}(X_j)$ 를 찾을 때까지 이 과정을 반복한다. 그러면  $\hat{E}(y_0 | d = 1)$ 에 해당되는  $\hat{E}(y_0 | d = 0)$ 의 표본 평균을 계산할 수 있다.

(3) 이를 통해 구해진  $\hat{E}(y_0 | d = 0)$ 와  $\hat{E}(y_1 | d = 1)$ 의 차이인 처리효과(ATT,  $E(y_1 | d = 1) - E(y_0 | d = 0)$ )를 구하게 되고, 표준오차(standard error)는 부스트랩(bootstrap) 방법을 이용하여 구할 수 있다.

매칭의 질을 높이기 위해서 Common Support가 고려되는데, 이는 처리집단 부표본과 통제집단 부표본의 성향점수가 동시에 관측되는  $\hat{p}$ 의 구간에 속

한  $i$ 들만을 매칭에 사용하고 그렇지 않은 관측치들은 제외하는 방법이다. 예를 들어, 처리받은 사람들의 가장 작은 확률값이 0.3이라면, 처리집단의 Counterfactual의 확률값이 0.3 이하이면 처리집단과 너무 멀어져서 매칭하는데 좋은 후보가 될 수 없으므로 제외하자는 것이다.

위의 방법 외에도 매칭할 때 사용하는 방법은 Radius 매칭(propensity score를 범위로 설정한 후 그 안에 포함되는 관측치들의 평균값으로 매칭)과 Kernel 매칭(모든 처리 단위를 비교단위의 가중평균으로 매칭하거나 비모수적 방법으로 추정하여 매칭) 등이 있다.

### 제3절 성향점수 매칭의 장단점

PSM의 장점은 처리집단과 비교집단의 쌍을 비교함으로써 선택편의의 문제를 해결할 수 있다. 그리고 처리집단과 비교집단에서의  $X$  분포는 무작위 표본보다 짝짓기 표본(Matched Sample)에서 더 비슷하기 때문에 ATT 추정치의 분산 역시 무작위 표본보다 짝짓기 표본에서 더 적은 값을 가진다. 또한 모형에 기반을 조정하는 경우에도 무작위 표본을 대상으로 하는 경우보다 짝짓기 표본을 대상으로 하는 경우가 통계적으로 더욱 강건(robust)하다(Heckman, Ichimura, & Todd, 1997).<sup>3)</sup>

단점은 분석적으로 주어지지 않기 때문에 좋은지 나쁜지 판단하기 어려우며, 비모수적 방법을 이용할 경우에는 실제 값보다 크게 나오는 경향이 있다. 요즘에는 PSM을 잘 사용하지 않는 편이며 PSM의 단점을 지적하는 논문들이 많이 나오고 있다. Angrist(1998)는 CI 가정하에서는 회귀(regression)와 매칭은 단지 가중치를 어떻게 주는지에 따른 차이일 뿐임을 증명하였다.

3) 최강식(2007), 「고용영향 분석평가 방법론 연구」 참조.

매칭에서의 ATT는

$$E[Y_{1i} - Y_{0i} | d_i = 1] = \frac{\sum_X \delta_X P(d_i = 1 | X_i = x) \cdot P(X_i = x)}{\sum_X P(d_i = 1 | X_i = x) \cdot P(X_i = x)} \text{이고}$$

$$(\delta_X \equiv E(Y_i | X_i, d_i = 1) - E(Y_i | X_i, d_i = 0)),$$

OLS 추정량은

$$\widehat{\delta_R} = \frac{\sum_X \delta_X P(d_i = 1 | X_i = x) [1 - P(d_i = 1 | X_i = x)] \cdot P(X_i = x)}{\sum_X P(d_i = 1 | X_i = x) [1 - P(d_i = 1 | X_i = x)] \cdot P(X_i = x)}$$

으로서

(회귀모형은  $Y_i = \sum_X M_{iX} \alpha_X + \delta_R d_i + e_i$ , 여기서  $M_{iX} = 1[X_i = x]$ 임)

수식으로도 확인되는 것을 알 수 있다.

## 제7장

# 도구변수 추정법

도구변수를 이용한 추정 이론은 이미 1940년대 초부터 알려져 있었다. 이론적으로는 이미 오래 전에 알려졌지만 신뢰성 있는 도구변수를 찾을 가능성이 별로 없다는 생각 때문에 실제 연구에서는 도구변수 추정법이 폭넓게 사용되지는 않았다. 그런데 Joshua Angrist(1991)의 논문(QJE)에서 도구변수 추정법이 유용하게 사용될 수 있음을 보여주면서 이후 15여 년 동안 유행을 하게 된다. 이때 이후로 도구변수 추정법의 이론과 해석 방법이 발전을 하게 된다.

## 제1절 도구변수 추정법

횡단면 자료로써 다음과 같은 방정식을 추정하려고 한다.

$$y_i = \beta_0 + d_i\beta_1 + X_i\beta_2 + u_i$$

연구의 관심사는 평균처방효과인  $\beta_1$ 을 추정하는 것이다. 그런데 위 식에서  $d_i$ 가 내생변수라고 하면 위 식을 OLS로 추정하는 것은 바람직하지 않다.

도구변수 추정법의 아이디어는  $u_i$ 와는 독립적이면서  $d_i$ 의 변화를 일으키

는 변수  $z_i$ 를 찾는 것이다. 즉 도구변수  $z$ 란  $d$ 와는 상관관계가 매우 높아서  $z$ 가 변화할 때  $d$ 도 같이 변화하면서도  $u$ 와는 상관관계가 없는 변수를 말한다. 만약에 이러한 특성을 만족하는 도구변수가 있다면 평균처방효과  $\beta_1$ 에 대한 일치추정량을 얻을 수 있다.

도구변수 추정량은 다음과 같다.

$$\widehat{\beta}_{IV} = \left( \sum_i Z_i D_i' \right)^{-1} \left( \sum_i Z_i y_i \right) = (Z'D)^{-1} (Z'y)$$

이 추정량은 다음과 같은 2단계 추정법과 수학적으로 동일하다.

1단계 (축약형 방정식)

$$d_i = \gamma_0 + z_i \gamma_1 + X_i \gamma_2 + e_i = Z_i' \gamma + e_i$$

2단계 (구조방정식)

$$y_i = \beta_0 + d_i \beta_1 + X_i \beta_2 + u_i = D_i' \beta + u_i$$

2단계 최소자승 추정량은 구조방정식의  $d_i$ 를  $\widehat{d}_i^{OLS}$  으로 대체하여서 얻을 수도 있다. OLS 추정량은  $X$ 변수 스스로가 도구변수인 도구변수 추정량의 일종이라고 볼 수 있다.

도구변수법을 직관적으로 이해하기 위해 도구변수  $z_i$ 가 더미변수이고 다른 설명변수는 없는 상황을 생각해 보자. 즉 추정식이 다음과 같다고 하자.

$$y_i = \beta_0 + d_i \beta_1 + u_i \tag{7-1}$$

$d$ 가 내생변수인 경우, 위의 식 (7-1)을 OLS 추정해 얻은  $\widehat{\beta}_1^{OLS}$ 는  $d$ 의 변화가  $y$ 에 미치는 인과효과뿐만 아니라 편의(bias)도 포함한다. 그러므로 OLS를 사용해서는 진정한 인과효과의 크기를 구할 수 없다. 이러한 상황에서 도구변수가 있다면 도구변수 추정량  $\widehat{\beta}_1^{IV}$ 는 다음과 같은 Wald 추정량이 된다.

$$\hat{\beta}_1^V = \frac{\hat{E}[y_i|z_i = 1] - \hat{E}[y_i|z_i = 0]}{\hat{E}[d_i|z_i = 1] - \hat{E}[d_i|z_i = 0]} \quad (7-2)$$

식 (7-2)에서 분모는 도구변수가 1에서 0으로 변했을 때 내생변수  $d$ 가 변화하는 정도를 의미한다. 이는 수학적으로  $\partial d/\partial z$ 로 표현할 수 있다. 한편 분자는 도구변수  $z$ 가 1에서 0으로 변했을 때 종속변수  $y$ 가 변화한 정도를 의미하고, 이는  $\partial y/\partial z$ 로 표현된다. 분자를 분모로 나누어 준 값은  $(\partial y/\partial z)/(\partial d/\partial z)$ 이고 이는  $\partial y/\partial d$ 로서 우리가 구하고자 하는 정책 한계효과와 동일하다. 도구변수의 핵심 조건 중 하나가  $u$ 와 무관하다는 것이므로 위 도구변수 추정법은  $u$ 와 관계가 없는  $z$ 의 변동을 이용하여 내생성의 문제를 해결하면서 우리가 구하고자 하는 정책효과를 추정할 수 있는 방법이다.

도구변수의 두 번째 조건, 즉  $d$ 와의 상관관계가 커야 한다는 조건은 통계적으로 검증할 수 있다. 이 조건을 검증하는 방법은 도구변수 추정법의 제1단계 추정에서  $\gamma_1$ 이 통계적으로 0이라는 가설을 검증하는 것이다. 이 가설을 기각하여야 도구변수의 두 번째 조건이 충족된다.

만약  $d$ 와 상관관계가 커야 한다는 도구변수의 조건을 만족하지 않는 도구변수를 사용한다면 어떻게 될까?  $z$ 와  $d$ 가 전혀 상관관계가 없다면 위 도구변수 추정량 식 (7-2)의 분모가 0이 되어서 추정이 불가능하다.  $z$ 와  $d$ 가 상관관계가 있더라도 그 정도가 매우 약해서 거의 0에 가깝다면 식 (7-2)의 분모가 거의 0에 가깝기 때문에 도구변수 추정치가 굉장히 커진다. 이렇게  $d$ 와 상관관계가 아주 작은 도구변수를 ‘약한 도구변수(weak instruments)’라고 부른다.

$z$ 와  $d$ 와의 상관관계가 작으면 도구변수의 추정치가 매우 커진다는 문제가 발생하므로  $z$ 와  $d$ 와의 상관관계가 상당히 커야 보다 바람직하다. Stock et al.(2002)에 의하면 도구변수의 계수 추정치에 대한 F-검정 통계량이 10을 넘으면 ‘강한 도구변수’, 10 미만이면 ‘약한 도구변수’로 분류된다. F값이 매우 작으면(예 : 2이하) 도구변수라고 보기 어렵다.  $t$ 값을 제공하면 F값이 되므로  $t$ 값 기준으로는 약 3.2 정도가 된다. 도구변수의 두 번째 조건을 검증하는 방



법은 최근 통계 프로그램에서는 대부분 제공하고 있다.

최소자승법에서는 변수의 설명력을 판단하는 기준으로  $R^2$ 를 사용하고 있다. 마찬가지로 도구변수의 두 번째 조건을 검증하는 기준으로  $R^2$ 를 사용할 수도 있다. Stock et al.(2002)에 의하면,  $R^2$ 값과 F값은 동일한 기준이어서 하나로부터 다른 하나를 수학적으로 도출할 수 있다. 따라서 굳이  $R^2$ 를 보기보다는 F값을 사용하는 것이 편리하다.

도구변수의 첫 번째 조건, 즉  $u$ 와 상관관계가 0이라는 조건은 테스트가 불가능하다.  $u$ 가 관측되지 않기 때문이다. 통계적 검증방법이 없기 때문에 이 조건에 대해서는 이론적으로 정당화하고 설득하는 과정을 거칠 수밖에 없다. 그런데 그동안 사람들이 매우 설득력 있다고 받아들일 수 있는 도구변수가 없었기 때문에 1990년 전까지는 도구변수 추정법이 거의 사용되지 않았다.

Joshua Angrist 이후 외생성 조건을 만족한다고 받아들여졌던 도구변수들은 무작위적으로 발생한 사건을 이용한 것들이다. 즉 동전 던기기와 비슷한 특성을 가진 것들이다. 결국 도구변수 추정에도 무작위 배정 개념이 들어 있다고 할 수 있다.

[예 1] 자녀의 수가 어머니의 노동공급에 미치는 효과(Angrist & Evans, 1998)  
다음과 같은 식을 추정하고자 한다.

$$y_i = \beta_0 + N_i\beta_1 + X_i\beta_2 + u_i \quad (7-3)$$

식 (7-3)에서  $N_i$ 는 자녀의 수,  $y_i$ 는 어머니의 노동공급 지표를 가리킨다. 이론적으로 자녀 수가 여성의 노동공급에 미치는 영향은 불분명하다. 자녀 수가 늘어나면 양육비가 늘어나기 때문에 여성이 노동공급을 늘릴 가능성이 있는 반면, 자녀를 돌볼 시간이 필요하기 때문에 여성의 노동공급이 줄어들 가능성도 있다.

식 (7-3)을 OLS로 추정하는 것은 바람직하지 않다. 그 이유는 첫째, 출산과 여성의 노동공급은 동시에 이루어지는 경향이 있기 때문이고, 둘째, 미래의

사회경력을 보다 중요시할수록 자녀 출산이 적어질 가능성이 높기 때문이다. 첫 번째 문제는 동시성의 문제이고, 두 번째 문제는 변수 누락의 경우다.

Angrist and Evans는 위 내생성 문제를 해결하기 위해 부모들이 자녀 구상에 대해 가지는 선호를 이용하였다. 미국에서는 자녀들의 성이 서로 다른 것을 선호하여서, 기존에 태어난 자녀들의 성이 같을 때, 성이 다른 자녀를 갖기 위해서 아이를 더 낳을 가능성이 높다고 한다. 만약 태아의 성별을 감식하여 낙태를 할 가능성이 없다면, 기존에 태어난 자녀들의 성별은 무작위적으로 배정된 것이라고 할 수 있다. 여기서 기존에 태어난 자녀들이 같은 성을 가졌는가 그렇지 않은가가 무작위적이라면 도구변수가 외생적이어야 한다는 조건을 만족한다. 그리고 기존 자녀들의 성은 이후 아이 출산에 영향을 미치기 때문에 도구변수의 두 번째 조건도 만족한다. 이 두 번째 조건은 다음과 같은 1단계 추정식을 통해 검증할 수 있다.

$$N_i = \gamma_0 + z_i\gamma_1 + X_i\gamma_2 + e_i \quad (7-4)$$

이 식 (7-4)에서 처음 두 자녀의 성이 같으면  $z_i = 1$ 이고, 같지 않다면  $z_i = 0$ 이다. 이처럼 처음 두 자녀의 성을 도구변수로 하여 자녀의 수가 어머니의 노동공급에 미치는 영향을 추정할 수 있다.

#### [예 2] 교육이 임금에 미치는 효과(Angrist & Krueger, 1991)

교육이 임금에 미치는 효과를 추정하기 위해 Angrist and Krueger는 의무교육 제도를 이용하였다. 미국 학교는 보통 9월에 새 학기가 시작하는데, 당해년 12월 31일 기준으로 만 6세가 되어야 입학할 수 있다. 이러한 제도에 따라서 출생월에 따라 입학하는 연도가 달라진다. 예를 들어, 2001년 1월에 태어난 아이는 2000년 12월에 태어난 아이보다 1년 늦게 학교를 시작하게 된다. 또한 의무교육 제도에서는 만 16세(혹은 주에 따라 만 17세)가 되기 전까지는 반드시 학교에 다녀야 한다고 규정하고 있다. 만 16세가 지나야 자퇴할 수 있고, 그 전까지는 자퇴가 허용되지 않는다. 16세 이전에 학생이 자퇴를

하면 부모가 처벌을 받는다. 따라서 4분기에 태어난 학생은 1~3분기에 태어난 학생보다 학교에 오래 다녀야 한다. 실제 통계에서도 태어난 분기에 따라 교육연수가 상당히 차이가 난다는 것을 보여주고 있다.

이처럼 의무교육 제도 때문에 태어날 분기가 교육연수에 영향을 주는 반면 태어난 달은—일부러 부모님이 특정한 달에 아이를 낳기 위해 임신을 통제하지 않았다면—거의 무작위적으로 배정되었다고 할 수 있다. 따라서 태어난 분기를 교육연수의 도구변수로 사용할 수 있다.

[예 3] 자녀의 수가 자녀에 대한 교육투자에 미치는 영향(Kang, 2011)

식 (7-3)에서  $y_i$ 가 사교육비 지출액이라고 하자. 식 (7-3)을 OLS로 추정하는 것은 바람직하지 않다. 예를 들어, 변수 누락의 문제가 있어서 부모님의 학력이 높을수록 소득이 높고, Becker의 양과 질 간의 선택 논리에 따르면 자녀의 수가 적을 수 있다. Kang은 처음에 태어난 자녀의 성을 자녀 수의 도구변수로 사용하였다. 한국에서는 남아 선호가 남아 있어서 첫째 아이가 딸 일 때 자녀의 수가 커지는 경향이 있다. 첫 번째 자녀의 성이 무작위적으로 배정되었다고 볼 수 있기 때문에 첫째 아이의 성이 자녀 수에 대한 도구변수가 된다. 도구변수의 두 번째 조건은 다음과 같은 1단계 추정을 통해 검증할 수 있다.

$$N_i = \gamma_0 + G_i\gamma_1 + X_i\gamma_2 + e_i$$

이 식에서  $i$ 는 두 번째 아이를 가리킨다. 첫째 아이가 딸이면  $G_i = 1$ 이고, 아들이면  $G_i = 0$ 이다.

[예 4] 공부 시간이 교육성과에 미치는 영향(Stinebrickner & Stinebrickner, 2008)

공부 시간이 길면 학점이 올라가는지를 파악하기 위해 다음과 같은 식을 추정하고자 한다.

$$y_i = \beta_0 + S_i\beta_1 + X_i\beta_2 + u_i$$

이 식에서  $S_i$ 는 공부 시간을,  $y_i$ 는 학점을 가리킨다. OLS로 추정을 하면 변수 누락의 문제, 동시성의 문제 때문에 편의가 발생한다. 관측되지 않는 능력이나 동기부여가 있을 수 있고, 자습 시간과 학점과 동시에 변화할 수 있기 때문이다.

이 논문에서는 베뢰아 대학 기숙사 배정이 무작위적으로 이루어진다는 것을 이용하였다. 신입생의 경우, 기숙사 룸메이트는 무작위적으로 배정되기 때문에 도구변수의 한 조건이 충족된다. 이 논문에서 도구변수는 룸메이트가 비디오나 컴퓨터 게임을 가지고 있는가를 나타내는 변수다. 룸메이트가 비디오 게임을 가지고 있으면 공부 시간이 줄어들 것이라고 예상되므로 도구변수의 다른 조건도 충족할 것으로 예상된다. 이 조건이 만족하는가를 검증하기 위해 다음과 같은 식을 1단계에서 추정을 한다.

$$S_i = \gamma_0 + V_i\gamma_1 + X_i\gamma_2 + e_i$$

룸메이트가 비디오나 컴퓨터 게임을 가지고 있으면  $V_i = 1$ 이고, 그렇지 않으면  $V_i = 0$ 이다.

[예 5] 종교 활동이 행복에 미치는 영향(Cohen-Zada & Sander, 2010)

종교 활동이 행복에 미치는 영향을 분석하기 위해 OLS 추정을 한다면 마찬가지로 변수 누락과 동시성의 문제 때문에 편의가 발생한다. 종교 활동 참여와 행복에 모두 영향을 주지만 측정할 수 없어서 추정식에 포함될 수 없는 요소, 예를 들어 인생관, 태도 등이 있을 수 있다.

이 논문에서는 이 문제를 해결하기 위해 종교 활동 참여에 대한 도구변수로 안식일 엄수법(blue laws)의 폐지라는 자연실험을 이용하였다. 이 말은 18 세기에 'blue'라는 말이 경멸적인 의미인 '도덕적으로 경직된'을 뜻하던 데서 유래했다고 한다. 성서(聖書) 중심 공동체인 청교도들이 가장 엄격하게 지켰던 안식일 엄수법은 일요일에 일상적인 일뿐만 아니라 구매 행위, 판매 행위,

여행, 공개적인 오락, 운동 시합 등을 금지했다. 식민지 시대 미국에서는 어디에서나 어느 정도 이와 유사한 법률이 존재했으며, 대체로 미국 독립전쟁 이후 사라졌고 공식적으로는 1940~50년대에 폐지되었다. 하지만 안식일 엄수법의 영향으로 금지되었던 행위가 주마다 약간씩 달랐고, 폐지된 시점도 약간씩 달랐다.

이중차분법에서 논의된 Gruber and Hungerman(2008)에 따르면, 안식일 엄수법의 폐지로 인하여 사람들의 종교 활동이 영향을 받는다. 이 사실에 근거하여, 이 연구에서는 안식일 엄수법이 폐지되는 상황을 종교 활동 정도에 대한 도구변수로 활용한다.

## 제2절 도구변수 추정법의 국지성과 도구변수의 어려움

도구변수 추정법이 어떠한 방식으로 평균처방효과를 식별(identify)하는가를 이해하는 것이 해석을 위해서도 매우 중요하다. 설명을 간단히 하기 위해 정책처방변수와 도구변수가 모두 더미변수인 상황을 생각해 보자. 예를 들어, 직업훈련 프로그램의 효과를 측정하려고 한다. 가상적으로 훈련 프로그램의 참여를 동전 던지기로 결정하기로 하자. 만약 동전의 앞면이 나오면 ( $z=1$ ) 훈련 프로그램에 참여하기로 하고, 뒷면이 나오면 ( $z=0$ ) 참여할 기회를 아예 주지 않는다고 하자. 직업훈련 프로그램에 참여하였으면  $d_i=1$ 이고, 그렇지 않았으면  $d_i=0$ 로 정의한다. 즉  $d_i$ 는 선택에 의해 결정되는 선택변수다. 동전 던지기로 결정된 기회와 실제의 참여 여부에 따라 다음 <표 7-1>과 같이 네 가지 경우가 생길 수 있다.

동전의 앞면이 나오나 뒷면이 나오나 참여하지 않는 집단을 절대불참집단(never-taker)이라고 하고, 앞면이 나오나 뒷면이 나오나 훈련 프로그램에 참여하는 사람들을 항상수용(always taker)집단이라고 하자. 동전의 뒷면이 나

〈표 7-1〉 정책참여 기회와 실제 참여 여부

$z_i = 1$	$z_i = 0$	
	$d_i = 0$	$d_i = 1$
$d_i = 0$	절대 불참 (never-taker)	반항 (defier)
$d_i = 1$	순응 (complier)	항상 수용 (always-taker)

와 기회가 부여되면 훈련에 참여하고 앞면이 나와 기회가 없어진다면 훈련에 참여하지 않는 사람들을 순응집단(complier)이라고 하자. 그리고 기회가 부여되면 참여하지 않고, 반대로 기회가 부여되지 않으면 참여하는 사람들을 반항집단(defier)이라고 하자.

도구변수 추정법에서는 도구변수가 변화할 때 이에 맞추어 정책변수가 변화를 하는 집단에 대해서만 정책효과를 식별할 수 있다. 위의 예로 든 상황에서는 순응집단만이 도구변수의 변화에 맞추어 변화를 보여준다. 따라서 이 예에서 도구변수 추정량은 오로지 순응집단의 평균처방효과를 재는 것으로 해석된다. 도구변수 추정법으로는 불응, 반항, 항상수용집단의 경우에는 정책효과를 측정할 수 없다. 도구변수 추정법이 전체 집단을 대상으로 정책의 효과를 재는 것이 아니라 도구변수의 영향을 받는 일부 집단만을 대상으로 정책효과를 측정한다는 의미에서 도구변수를 이용한 추정치를 ‘국지적 평균효과(local average treatment effect)’ 라고 부르기도 한다.

만약 정책의 효과가 개인별로 혹은 집단별로 크게 다르지 않다면 도구변수 추정치가 전체 집단의 효과를 대표한다고 할 수 있다. 하지만 만약에 정책의 효과가 매우 이질적이라면 도구변수 추정치에 대한 해석은 매우 제한적일 수밖에 없다. 앞에서 예로 든 Angrist and Evans(1998)의 논문에서는 자녀의 수가 어머니의 노동공급에 미치는 효과를 분석하기 위해 처음 두 자녀의 성을 도구변수로 사용하고 있다. 이 도구변수에 영향을 받는 가구는 처음 두 자녀의 성이 같아서 세 번째 아이를 갖지 않은 가구와 처음 두 자녀의 성이 같지 않아서 세 번째 아이를 갖은 가구다. 따라서 이 도구변수는 자녀의 수가 2명

인 가구와 3명인 가구를 대상으로 자녀의 수가 여성 노동공급에 미치는 효과를 식별하고 있다. 반면 자녀의 수가 없거나, 1명뿐이거나 4명 이상인 가구에 대해서는 자녀 수의 효과를 식별하지 못한다. 만약 자녀 수가 2명에서 3명으로 늘 때나 기타의 경우 모두 자녀 수 증가가 동일한 효과(homogeneous effect)를 가진다면 처음 두 자녀의 성을 도구변수로 사용하여 추정한 결과가 전체 집단을 대표할 수 있다. 하지만 자녀 양육에서 규모의 경제나 학습의 효과 등이 작용하기 때문에 자녀의 수가 2명에서 3명으로 늘 때와 5명에서 6명으로 늘 때 늘어난 자녀 1명의 효과가 동일하다고 사전적으로 단정하기는 매우 어렵다.

특정 도구변수가 특정 집단에 대해서만 정책효과를 식별하기 때문에, 다른 집단에서의 정책효과를 측정하려면 해당 집단에 대해 작용을 하는 다른 도구변수를 찾아야 한다. 전체를 A, B, C, D 집단으로 나누었을 때, 어떤 도구변수  $Z_A$ 로 집단 A에서의 정책효과를 추정하였고, 다른 도구변수  $Z_B$ 로 집단 B에서의 정책효과를 추정하였다면, 이 두 추정치가 아주 다를 수 있다.

도구변수 추정량은 OLS 추정량과 다른 경우가 많은데, 보통은 OLS 추정의 편의 때문인 것으로 해석한다. 하지만 도구변수 추정량의 국지성을 고려하면 다른 해석도 가능하다. OLS 추정에서는 정책효과가 동질적이라고 가정한다. 도구변수 추정법은 특정 집단에 대해서만 작용을 한다. 만약 정책효과가 매우 이질적이라면 OLS 추정에서 편의가 발생하지 않더라도 OLS 추정 결과와 도구변수 추정 결과가 아주 다를 수 있다. 이렇게 되면 OLS 추정과 도구변수 추정 결과의 차이가 편의 때문에 발생한 것인지, 아니면 정책효과가 매우 이질적이어서 발생한 것인지 구분하기 어려워진다.

교육이 임금에 미치는 영향과 관련하여 한 논문에서 OLS와 도구변수 추정법을 비교한 바 있다. 미국의 경우 OLS로 교육의 수익률을 추정하면 6~8% 정도가 나오고, 도구변수로 추정을 하면 10~15% 정도로 다소 높게 나온다. 이를 두고 실제로 교육수익률이 매우 높는데 OLS에서는 적게 추정을 하고 있다고 해석할 수 있다. 반면 도구변수 추정은 국지적 효과를 측정하는 것이므로 도구변수에 반응을 보이는 사람들에 대해서만 교육수익률을 측정하는 것

이라고 볼 수도 있다. 도구변수에 대해 반응을 보이는 집단은 그렇지 않은 집단에 비해 교육에 대한 수익률이 잠재적으로 높은 집단일 가능성이 높다고 한다면, 도구변수 추정으로 교육수익률이 높게 나온 것은 전체의 특성을 측정한 것이 아니라 도구변수에 반응을 보인 집단의 특성을 국지적으로 반영한 것이라고 해석할 수 있다.

보다 극단적으로 어떤 도구변수가 어떤 집단에 대해 작용을 하는지 분명히 파악하기 어렵다면 이 도구변수로 추정된 결과가 어떤 집단에 대해서만 해당 되는지도 해석하기 어렵게 된다.

이러한 이유들 때문에 최근 도구변수 추정법을 둘러싸고 논쟁이 벌어지기도 하였다. 도구변수 추정법에 의해 식별되는 영역이 지나치게 협소하고, 그에 비해 정책효과는 매우 이질적일 가능성이 높다고 생각하는 일부 학자들은 도구변수 추정법에 대해 매우 비판적이다. 이에 대해 도구변수 추정법 옹호자들은 아무것도 안 하는 것보다는 그래도 적절한 도구변수를 찾아서 추정을 하는 것이 낫다는 주장을 하고 있다.

만약 정책효과의 이질성이 크지 않다면 도구변수 추정법의 결과가 상당한 대표성을 가지게 되며, 정책효과가 매우 이질적이라면 그렇지 않게 된다. 따라서 도구변수 추정법의 용도는 적절한 도구변수를 찾는 것과 더불어 정책효과가 이질적인가 아니면 상당히 동질적인가에 달려 있다. 이에 따라 최근에는 정책효과의 이질성을 판단하는 데 도움을 줄 수 있는 방법들이 제안되기도 한다.

도구변수의 조건 중 하나는 외생성이다. 도구변수가 과연  $u$ 와 상관관계가 없는가는 검증할 수 없기도 하거니와 실제로 외생성 조건을 만족한다고 설득력을 가진 도구변수를 찾는 것은 매우 어렵다. 적절한 도구변수를 찾는 일은 매우 참신한 아이디어와 더불어 그를 뒷받침할 만한 논리를 필요로 한다. 처음에 매우 괜찮은 것으로 받아들여졌던 도구변수들 대부분은 외생성에서 의심을 받으면서 적절한 도구변수가 아닌 것으로 판명되었다. Angrist and Krueger(1991)의 논문에서는 학력에 대한 도구변수로 출생 분기를 사용하였다. 이 논문이 처음 나왔을 때에는 이 아이디어가 매우 참신하고 획기적인 것



으로 받아들여졌다. 하지만 점차 출생 분기가 외생적이지 않다는 여러 가지 주장과 증거들이 나타나면서 도구변수로서의 장점이 상당 부분 약화되었다. 상당히 외생적이라고 간주되는 도구변수의 경우에도  $d$ 와의 관련성이 매우 약한 도구변수인 경우가 흔히 발생하였다.

## 회귀단절모형

자료를 이용한 실증분석에서 주로 문제가 되는 것은 관심변수에 내포된 내생성(endogeneity)이다. 이 내생성을 해결하기 위해 전통적인 사회과학에서는 통제되는 설명변수들의 수를 늘리는 방법을 사용하였다. 그러나 1990년대 초반 이후로 실증경제학자들은 정책 처방이 무작위 배정되는 상황을 찾고자 노력해 왔고, 그 이후로 무작위 배정 원칙을 응용한 여러 가지 방법들이 도입되었다. 첫 번째 방법은 자연에서 일어난 자연실험(natural experiments)을 이용하는 무작위 배정을 통한 처치-통제집단 비교법이다. 이 방법은 우리가 살고 있는 사회에서 제도 혹은 다른 요인에 의해서 인위적으로 의도되지 않은 채 무작위 배정이 일어난 특정 상황을 활용하는 연구 방법이다. 이와 같이 자연실험을 이용한 실증분석 결과는 상당히 높은 설득력을 가지지만, 무작위 배정이 일어나는 상황 자체가 흔하지 않기 때문에 실제 연구에서는 폭넓게 활용되지 못하고 있다. 두 번째 방법은 차선의 대안으로서 여전히 무작위 배정에 기본적인 아이디어를 두고 있는 방법들로 이중차분법, 도구변수법, 회귀단절모형 등이 있다. 이 장에서는 회귀단절모형(regression discontinuity design)에 대해 설명하고자 한다.

## 제1절 회귀단절의 개념

회귀단절(regression discontinuity design, 이하 RDD)의 개념을 개략적으로 설명하면 다음과 같다. 예를 들어 “학생이 장학금을 받는 경우 그 결과로 그 학생의 학업 성적이 더 오르는가” 하는 주제에 대해 연구를 한다고 하자. 여기서, 학생들의 장학금 수혜 여부는 무작위적(randomly)으로 결정되는 것이 아니라 선정 시점 이전에 결정된 변수(예를 들어, 직전 학기 평점 평균)를 기준으로 이번 학기 장학금 수혜 여부가 결정된다. 이러한 제도적 배경하에서 장학금 수혜 여부( $T_i$ )가 학업 성적( $Y_i$ )에 미치는 인과효과를 추정하는 연구를 생각해 보자. 이와 같은 연구에서 흔히 사용되는 실증모형은 아래와 같다.

$$\text{추정모형: } Y_i = \beta + \alpha * T_i + u_i \quad (8-1)$$

$$\text{장학금 수혜 결정식: } T_i = 1[S_i > \bar{S}] \quad (8-2)$$

여기서,  $Y_i$  = 학생  $i$ 의 최종 학업 성적

$T_i$  = 장학금 수혜 여부를 표시하는 더미변수

$S_i$  = 직전 학기 평점 평균

통상적으로 개인에 대한 설문조사 자료(individual survey data)를 사용하여 이 문제를 분석할 경우, 연구자는 이번 학기 장학금 수혜 여부를 조사하여  $T_i$ (treatment variable)로 사용하고 이번 학기말의 평점을 최종 성적변수  $Y_i$ 로 사용하는 횡단면 자료를 구축하여 위의 식 (8-1)을 OLS로 추정한다. 그러나 앞서 설명했듯이 장학금 수혜 여부는 무작위적으로 결정된 것이 아니기 때문에 ‘ $cov(T_i, u_i) \neq 0$ ’가 성립하고 그 결과로  $\alpha$ 의 추정치에 편의가 발생하게 된다. 따라서 위의 방법으로는 장학금 수혜가 학업 성적에 미치는 진정한 인과효과를 파악할 수 없다.

이제 회귀단절 방법을 고려해 보기로 한다. 가령 장학금 수혜 여부를 결정하는 기준이 지난 학기 평점이고 그 경계값이 3.0이라고 가정하자. 그리고 학생 A는 지난 학기에 평점 3.01을 받았고 학생 B는 2.99를 받았다고 가정하자. 이때 A는 이번 학기에 장학금을 받게 되고 B는 받지 못하게 된다. 이 경우 경계값 주위에서 매우 근소한 차이로 두 사람의 장학금 수혜 여부가 달라지게 된다. 이와 같이 근소한 차이로 A와 B의 장학금 수혜(status) 여부가 결정된다면 우리는 이 결정이 거의 무작위적으로 이루어진 것이라고 생각할 수 있다.

현실적인 예를 통하여 장학금 수혜 결정을 이해하기 위해 다음과 같은 상황을 가정해 보자. A와 B는 서로 친하고 지난 학기에 같은 과목을 수강하였다. B는 시험 당일 등교 시 지하철이 고장 나서 시험장에 30분쯤 늦게 들어왔고 A는 제 시각에 도착해서 시험을 보았다. 똑같이 공부하고 똑같은 능력을 가지고 있어서 최종 성적이 동일할 가능성이 높았으나, B는 늦게 들어오는 바람에 A보다 약간 낮은 점수를 받게 되었다. 그 결과로 A의 지난 학기 평점은 3.01이 되었고 B의 평점은 2.99가 되었다. 이 경우에는 통제할 수 없는 어떤 무작위적인 요인 때문에 두 사람의 성적이 3.01과 2.99로 나뉘게 된 것이므로 두 사람은 무작위적으로 장학금 수혜 여부가 결정되었다고 생각할 수 있다.

반면에 지난 학기 평점이 4.0인 C와 2.0인 D, 두 학생을 비교해 보자. C는 경계값보다 평점 1.0이 높고 D는 경계값보다 평점 1.0이 낮다. 과연 이 두 학생의 장학금 수혜 여부가 무작위적으로 결정되었다고 생각할 수 있을까? 그렇지 않다. C는 D에 비해서 원래부터 공부를 잘할 가능성이 높고 D는 C에 비해서 공부를 못할 가능성이 높다. 그러므로 장학금 수혜 여부가 C와 D 사이에서는 무작위적으로 결정되었다고 생각하기 어렵다. 또한 이 두 학생을 한 자료에 놓고 장학금 수혜 여부에 따라 이번 학기 최종 성적을 비교하는 것은 무의미하다.

위의 두 가지 서로 다른 비교방법에서 유추할 수 있는 바와 같이, 회귀단절법은 장학금 수혜 여부를 결정하는 경계값으로부터 일정한 범위(boundary)를

정해서 그 구간 안에 들어오는 학생들 간에 성적을 비교하여 그 성적의 차이를 장학금 수혜의 인과효과로 파악하는 실증분석 방법이다. 위의 예를 사용할 때 그 범위가 경계값 상하 0.01이라면 A와 B는 그 구간 내에 속하고, 이번 학기말 두 학생의 성적 차이는 장학금 수혜 여부의 차이라고 해석되는 것이다.

회귀단절 방법은 내생성을 야기하는 관측치들을 분석에서 제외시키기 때문에 일치성(consistency)과 불편성(unbiasedness)이 보장되는 인과효과의 추정치를 구하도록 만드는 장점이 있다. 하지만, 경계값(cutoff) 주변으로 분석 샘플이 한정되기 때문에 실증분석에 사용되는 관측치의 수가 작을 가능성이 높다. 관측치의 수가 작으면 추정치의 표준오차가 커지고 그 정확도(precision)가 낮아지게 된다. 즉 분석대상이 되는 구간의 크기를 좁히면 좁힐수록 내생성은 해소되는 반면 관측치의 수가 줄어들면서 추정치의 표준오차가 커져서 정확도가 떨어진다. 그러나 분석대상 구간을 넓히면 넓힐수록 관측치 수가 늘면서 표준오차가 작아지고 정확도가 좋아지는 반면 추정치의 일치성과 불편성이 보장되지 않는다. 어느 정도의 구간이 인과효과의 추정에 가장 적합한가에 대한 이론적인 해답은 회귀단절 방법에서 아직 제시되고 있지 않다. 그러므로 분석대상의 범위를 조금씩 변화시켜 가면서 추정치가 변화하는 패턴을 확인하는 방법이 통상적으로 사용된다.

## 제2절 회귀단절 방법론

본절에서는 두 종류의 회귀단절 방법이 소개된다. 첫 번째 방법은 경계값 주위에서 정책처방 여부가 뚜렷하게 분리되는 Sharp 회귀단절 방법이고, 두 번째 방법은 정책처방 여부가 Sharp 회귀단절의 경우만큼 뚜렷하지는 않으나 처방을 받을 확률에 불연속성(discontinuity)이 존재하는 Fuzzy 회귀단절 방법이다.

## 1. Sharp 회귀단절

Sharp 회귀단절(Sharp Regression Discontinuity, SRD) 방법을 통하여 추정되는 인과효과는 다음과 같이 표현된다.

$$\alpha = \lim_{s \downarrow \bar{s}} E[Y|S] - \lim_{s \uparrow \bar{s}} E[Y|S]$$

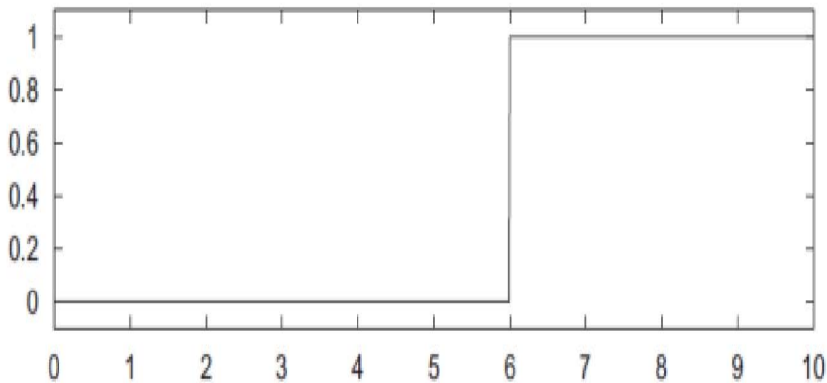
이 식에서 ' $\lim_{s \downarrow \bar{s}} E[Y|S]$ '는 S(지난학기 평점)가 점점 줄어들면서  $\bar{S}(=3.0)$ 이라는 경계값으로 수렴할 때 이번 학기 성적의 기댓값  $E[Y|S]$ 의 극한값을 표시한다. 비슷한 방식으로, ' $\lim_{s \uparrow \bar{s}} E[Y|S]$ '는 S가 점점 증가하면서  $\bar{S}$ 로 수렴할 때의 성적의 기댓값  $E[Y|S]$ 의 극한값이다. 이 두 극한 기댓값의 차이가 장학금 수혜의 인과효과인  $\alpha$ 가 된다.

[그림 8-1]에는 위에서 설명한 내용이 그림으로 요약되어 있다. [그림 8-1]의 x축에는 장학금 수혜 여부를 결정하는 변수(forcing variable)인 S가, y축에는 장학금 수혜 확률이 표시되어 있다. [그림 8-1]에서의 경계값은 6이다. 한 학생의 지난 학기 평점이 경계값인 6 이상이면 장학금을 받을 확률이 1이고 경계값보다 조금이라도 낮으면 장학금을 받을 확률이 0이다. 장학금을 받을 확률은 'S<6'인 경우에는 계속 0으로 유지되다가 경계값을 넘는 순간 1로 바뀌면서 불연속성(discontinuity)이 발생한다. 이러한 이유로 회귀단절에 discontinuity라는 단어가 포함되고, 회귀단절 방법에서는 반드시 정책처방을 받을 확률에 일정한 크기의 불연속성이 존재해야 한다. [그림 8-1]에서는 장학금 수혜 여부가 경계값 아래에서는 0으로 유지되다가 경계값에 도달하는 순간 1로 뚜렷하게 바뀌는 상황을 표시하고 있다. 이와 같이 경계값 주위에서 처방의 확률이 0에서 1로 바뀌는 상황을 분석하는 회귀단절을 'Sharp 회귀단절'이라고 부른다.

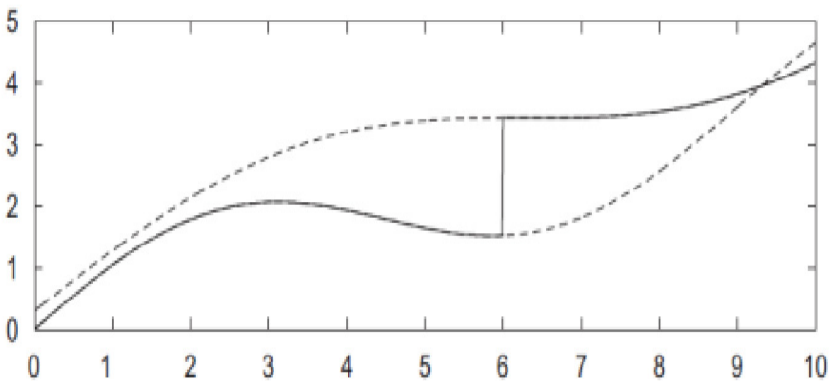
[그림 8-2]의 x축은 S(forcing variable)이고, y 축은 성과변수인 이번 학기의 최종 성적을 나타낸다. 그림 아래쪽의 점선은 장학금을 못 받았을 경우에

나타날 잠재적인 성적의 스케줄이고, 위쪽의 점선은 장학금을 받았을 경우에 나타날 잠재적인 평점의 스케줄이다. 즉, 장학금 수혜의 인과효과인  $\alpha$ 는 경계값인 6에서 평가된 두 점선의 차이다.  $S$ 가 오른쪽에서 경계값인 6에 가까워졌을 때의 최종 성적(즉,  $\lim_{s \downarrow 6} E[Y|S]$ )과  $S$ 가 왼쪽에서 경계값인 6에 가까워졌을 때의 최종 성적(즉,  $\lim_{s \uparrow 6} E[Y|S]$ ) 간의 차이가 장학금을 받음으로써 발생하는 최종 성적의 상승분으로 해석된다.

[그림 8-1] 배정확률(Assignment probabilities): SRD



[그림 8-2] 잠재/관측 결과 회귀함수(Potential and observed outcome regression functions): SRD



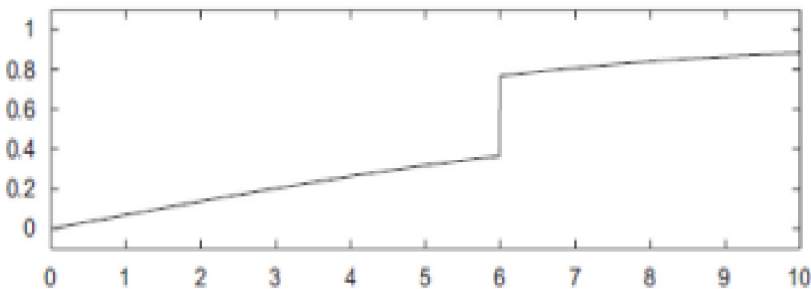
## 2. Fuzzy 회귀단절

Fuzzy 회귀단절(Fuzzy Regression Discontinuity, FRD)방법을 통하여 추정되는 인과효과는 다음과 같이 표현된다.

$$\alpha = \frac{\lim_{s \downarrow \bar{s}} E[Y|S] - \lim_{s \uparrow \bar{s}} E[Y|S]}{\lim_{s \downarrow \bar{s}} E[T|S] - \lim_{s \uparrow \bar{s}} E[T|S]}$$

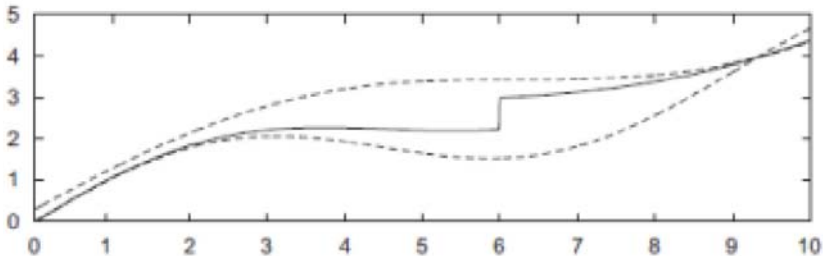
예를 들어, 장학금 수혜 여부를 선정할 때 직전 학기 평점이라는 하나의 기준만을 사용하지 않고 가정형편 및 학생에 관한 여러 사정들을 추가적으로 고려하는 상황을 생각해 보자. 이러한 상황에서 가정형편이 좋은 학생의 평점이 3.01이라면 그 학생이 장학금을 받을 확률은 높긴 하지만 1보다는 작을 수 있다. 반대로, 가정형편이 좋지 않은 학생의 평점이 2.99라도 그 학생이 장학금을 받을 확률은 0보다는 높을 수 있다. 즉 아래의 [그림 8-3]에 의하면, 지난 학기 평점이 6보다 낮을 경우 장학금 받을 확률이 0보다는 조금 크고, 6보다 높은 경우 장학금 받을 확률이 1보다 조금 낮게 되는 상황이 발생한다. 이러한 상황을 Sharp 회귀단절과 구분하여 Fuzzy 회귀단절라고 부른다. 장학금의 수혜 여부가 경계값 주위에서 뚜렷하게 구분되지 않고 약간 불분명하게 구분된다는 의미에서 Fuzzy라는 용어를 사용한다. 그러나 이 경우에도 여전히 장학금 수혜 확률에는 경계값에서 불연속성이 존재해야 한다.

[그림 8-3] 배정확률(Assignment probabilities): FRD





[그림 8-4] 잠재/관측 결과 회귀함수(Potential and observed outcome regression functions): FRD



Fuzzy 회귀단절에서 정의되는 인과효과  $\alpha$ 의 분자는  $\lim_{s \downarrow \bar{s}} E[Y|S] - \lim_{s \uparrow \bar{s}} E[Y|S]$ 로서 Sharp 회귀단절의 인과효과와 동일하다. 그러나 분모는 Sharp 회귀단절의 경우와 달리  $\lim_{s \downarrow \bar{s}} E[T|S] - \lim_{s \uparrow \bar{s}} E[T|S]$ 로서 경계값 주위에서 발생하는 장학금 수혜 확률의 차이가 반영된다. Sharp 회귀단절은  $\lim_{s \downarrow \bar{s}} E[T|S] - \lim_{s \uparrow \bar{s}} E[T|S] = 1$ 인 경우이기 때문에 Fuzzy 회귀단절이 더 포괄적인 정의이고, Sharp 회귀단절은 Fuzzy 회귀단절의 특수한 경우라고 생각할 수 있다.

회귀단절 방법을 적용하는 경우에는 다음의 두 가지 조건을 먼저 확인해야 한다.

첫째, 경계값인  $\bar{S}$ 에서 일정한 범위의 주변 구간을 정했을 때  $\bar{S}$ 의 주위에서  $S$ 의 확률밀도함수(probability density function: pdf)  $f(S)$ 에 점프나 불연속성이 존재하지 않아야 한다(McGrady(2008), *Journal of Econometrics*). 만약 학생들이 장학금의 경계값 점수가 6점인 사실을 알면 전략적인 행동을 통하여 성적을 받을 가능성이 있기 때문이다. 만약 6점보다 약간 높은 점수(예를 들어, 6.1)에는 정상보다 많은 수의 학생들이 분포해 있고, 6점보다 약간 낮은 점수(예를 들어, 5.9)에는 정상보다 적은 수의 학생들이 분포해 있을 수 있다. 이것은 학생들이 6점에 대해 전략적인 행동을 한 증거이기 때문에, 6점이라는 경계값에 의하여 장학금이 무작위로 배정되었다고 판단하기 어렵다. 그러나 경계값의 주변에서  $f(S)$ 의 점프나 불연속성이 없다면 무작위 배정이 잘 이루어졌다고 판단할 수 있다.

둘째, 정책처방 이외의 관측변수들(observable variables)이 경제값 주변에서 처방-통제그룹 간에 서로 균등화(balanced)되어 있는지 확인하여야 한다 (Lee(2008), *Journal of Econometrics*). 경제값 주변에서 정책처방 배정이 무작위로 이루어졌다면, 처방 이외의 다른 관측변수들의 평균값은 처치-통제그룹 간에 서로 비슷할 것으로 예상되기 때문이다.

### 제3절 회귀단절법을 적용하는 두 가지 구체적인 방법들

아래와 같은 추정모형이 주어져 있을 때 처방의 인과효과  $\alpha$ 를 추정하는 방법에는 다음과 같은 두 가지가 있다.

$$Y_i = \beta + \alpha \cdot 1[S_i > \bar{S}] + f(S_i) + u_i \quad (8-3)$$

첫 번째 방법은  $f(S)$ 에 대하여 비선형함수 형태를 가정하고 계수들을 OLS를 통하여 추정하는 방법으로서 모수추정법(parametric methods)라고 부른다. 두 번째 방법은 인과효과를 구성하는 네 기댓값의 극한값(즉,  $\lim_{s \downarrow \bar{s}} E[Y|S]$ ,  $\lim_{s \uparrow \bar{s}} E[Y|S]$ ,  $\lim_{s \downarrow \bar{s}} E[T|S]$ ,  $\lim_{s \uparrow \bar{s}} E[T|S]$ ) 각각을 비모수 추정법(nonparametric estimation methods)을 통하여 추정하고  $\alpha$ 의 식에 대입하는 방법이다.

#### 1. 모수추정법(Parametric Methods)

모수추정법에서는  $f(S)$ 를 ' $f(S_i) = \gamma_1 S_i + \gamma_2 S_i^2 + \gamma_3 S_i^3 + \text{etc}$ '와 같은 비선형의 형태로 가정하고 식 (8-3)에 OLS를 적용하여  $\alpha$ 를 추정하는 방법이다. 연구자의 판단에 따라 경제값 주변으로 분석 샘플을 제한하여 식 (8-3)을 추정하기도 하고, 혹은 이러한 샘플 제약 없이 주어진 전체 표본에 대해서 식

(8-3)을 추정하기도 한다. 샘플을 경계값 주변으로 제한하는 전자의 방법이 회귀단절의 원래 취지에 더 부합한다.

## 2. 비모수추정법(Non-parametric Methods)

비모수추정법에서는 통계학의 비모수추정방법을 사용하여 인과효과를 구성하는 네 기댓값의 극한값을 직접 구하여  $\alpha$ 의 식에 직접 대입한다. 모수추정법에서 분석 샘플을 경계값의 일정한 주변으로 제한하는 방법은 비모수법의 관점에서는 경계값 주위에 일정한 크기의 대역 폭(bandwidth)를 정하고 사각형 핵(kernel)을 이용하여 극한 기댓값을 추정하는 방법이다. 통계학의 비모수법을 사용하면 대역 폭의 크기와 핵의 형태에 따라 여러 종류의 극한 기댓값을 구할 수 있다. 여러 종류의 기댓값을 이용하여  $\alpha$ 를 추정하고  $\alpha$ 의 값에 큰 변화가 없는지를 검토함으로써 추정치가 강건한지를 검증할 수 있다.

### 제4절 회귀단절법 적용사례

[예 1] 노동조합이 기업성과에 미치는 영향 분석(DiNardo & Lee, 2004)

이 논문의 목적은 한 기업에 노조가 생기면 그 노조는 기업의 성과(예를 들어 수익성, 생산성, 생존 가능성 등)에 어떠한 영향을 미치는지를 탐구하는 것이다.

본 논문에서 사용된 추정모형은 아래와 같다.

$$Y_i = \alpha + X_i' \beta + D_i \delta + \epsilon_i \quad (8-4)$$

여기에서  $i$ 는 기업을 표시하고,  $Y$ 는 그 기업의 성과(판매량, 1인당 판매

량, 이윤),  $D$ 는 노조 존재 여부를 나타내는 더미변수,  $X$ 는 그 기업의 관측가능한 특성들을 모아 놓은 벡터(지역, 종업원 구성, 남녀 비율, 55세 비율 등)이다.

통상적으로 1980년대에 이 주제를 다룬 많은 논문들은 보통의 횡단면 데이터를 이용하여 OLS 회귀분석 방법을 적용하였다. 그러나 일반적으로 성과가 좋은 기업에서 노조가 조직되는 경향이 강하므로, 이와 같은 실증분석에는 노조 유무에 내생성이 존재할 가능성이 높다. 그러므로 횡단면 자료를 이용하여 식 (8-4)를 OLS 추정하여 얻은  $\hat{\delta}$ 은 노조의 인과효과와 무관하게 유의한 양(+)의 값을 가질 가능성이 높다. 따라서 이 결과가 노조가 있으면 기업의 수익성이 좋아진다고 단순히 해석될 수는 없다.

이 실증분석 문제에 존재하는 내생성을 해결하기 위해서 본 논문에서는 미국의 독특한 제도적 시스템을 이용하였다. 미국에서는 기업 내에서 교섭력을 가지는 노조로 인정받으려면 종업원들 전부가 노조 승인(union recognition) 투표를 실시하여, 그 투표에서 노조에 대한 찬성률이 50%가 넘어야 한다. 만약 이 투표에서 종업원 50% 이상의 지지를 얻지 못하면 그 노조에게는 교섭권이 인정되지 않는다.

본 논문에서는 이 노조 승인 투표에서의 지지율이 노조 성립 여부( $D$ )를 결정하는 변수  $S$ 의 역할을 한다. 다양한 투표 결과가 존재할 수 있기 때문에 원자료에는 90% 이상의 찬성으로 교섭권을 가진 노조도 있을 것이고, 10%의 낮은 지지를 받아 교섭권이 인정되지 않고 노조 설립이 부결된 노조도 있을 것이다. 회귀단절의 원칙에 의하면 이 두 종류의 기업을 분석 샘플에 포함시켜 비교하는 방법은 노조 설립의 인과효과를 보여주지 못한다. 왜냐하면, 노조의 유무와 관계없이 이 두 기업은 이미 대단히 이질적인 기업으로 간주될 수 있기 때문이다.

그러나 투표 결과에 의해 득표율이 50% 주변에서 형성된 두 기업을 생각해 보자. 예를 들어, 득표율 51%로 교섭권을 갖게 된 노조의 기업과 49%로 노조 설립이 기각된 기업을 상정해 보자. 이 두 기업에서 노조 성립 여부는 거의 무작위 배정과 유사하게 결정되었다고 가정할 수 있다. 그러므로 50%라

는 경계값 주변으로 득표율이 결정된 기업들을 추출하여 투표가 있는 후 약 5년 또는 10년 후에 그 기업의 성과가 어떻게 변화하였는지를 분석하면 노조의 인과효과를 정확히 추정할 수 있다. 본 논문은 이러한 아이디어를 이용하여 노조 유무의 내생성을 해결하고 회귀단절 방법을 적용한 결과 과거의 연구 결과와는 다르게 노조가 기업의 성과에 큰 영향이 없음을 밝혔다.

#### [예 2] 현역 의원이 재선에서의 당선 가능성(Lee, 2008)

이 논문은 현역 의원이 다음에 다시 의원이 될 확률이 얼마나 더 높은지에 대한 연구다. 현재 국회의원인 경우에 신생 국회의원보다 현역이라는 장점이 재선에서 얼마나 작용하는지 알아볼 수 있다는 것이다.

일반적인 단순회귀분석(OLS) 방법을 이용하여 분석하면, 과거에 의원을 지 지한 사람은 다음에도 그 의원을 지지하기 때문에 재선에서 당선될 수 있다는 내생성이 있어 과대추정하게 되어 좋은 방법이라고 할 수 없다. 그래서 연구자는 회귀단절 방법을 적용하여 분석해 보았다.

득표율이 90%로 당선된 사람과 10%로 떨어진 사람을 비교하는 것은 차이가 너무 커서 의미가 없다. 그래서 지난번 선거 때 득표율이 근소한 차이로 당락이 결정된 구역을 찾은 다음 근소한 차이로 당선된 사람과 낙선된 사람들을 추출하여 다음 선거 때 어떻게 되는지를 본 것이다. 이때 경계값은 50%로 하였는데, 그 이유는 득표율이 최소한 50% 이상은 되어야 당선되는 것이기 때문이다. 그리고 처리 전에 결정된 변수들의 분포는 경계값을 기준으로 양쪽은 거의 동일해야 하며 이는 무작위성을 만족하는 것이다. 지난번 선거가 다음번 선거에 영향을 주지 않는다는 결과를 알 수 있다.

#### [예 3] 학급의 학생 수에 따른 성적의 효과(Angrist & Lavy, 1999)

이 논문은 처음에는 도구변수를 사용한 예로 알려졌다가 후에 회귀단절 방법도 사용되었다는 연구로서 재발견되었다. 학급의 학생 수가 많거나 적음에 따라 성적에 어떠한 영향을 미치는지에 대한 연구다.

기존에는 시험 점수를 종속변수로 한 교실의 학생 수를 나타내는 변수와

학생의 다른 특성을 설명변수로 놓고 단순회귀분석(OLS) 방법을 실시하였다. 학교의 질은 학생과 부모와 관련된 가족 특성과 선택이 연관된다는 점 때문에 내생성을 가지게 된다. 그래서 연구자는 도구변수를 이용하여 분석하였고 도구변수에 회귀단절 방법이 포함되어 있는 것이다.

이스라엘의 학교에서는 14세기 유대인 랍비 Maimonides가 제안한 학급 구성 원칙에 의하여 한 학급에 최대 학생의 수는 40명을 넘지 못한다. 예를 들어, 41명이라면 2개의 학급을 만들어야 한다. 학급당 학생 수의 평균은 전교 학생 수 40명까지는 45도 직선으로 올라가고, 전교 학생 수가 41명이 되었을 때 20.5명으로 떨어진 후 다시 상승하다가 80명이 되었을 때 다시 40명으로 떨어진 후 올라가는 톱니바퀴 형태로 움직인다는 것을 알 수 있다. 그래서 이 제도를 학급 크기의 도구변수로 이용한 것이다. 실제로 학교들이 이 제도에 따라 움직인다는 것도 확인되었다. 여기서 회귀단절의 적용은 경계값이 나타나는 40, 80, 120을 경계값으로 설정한 후 경계값 주변에 있는 학교들을 추출하여 분석하면 되는 것이다. 모형 식은 다음과 같다.

$$2SLS \text{의 1단계} : n_{sc} = X_s' \pi_0 + f_{sc} \pi_1 + \eta_{sc}$$

$$(where \ f_{sc} = \frac{e_s}{\int ((e_s - 1)/40) + 1})$$

$$2SLS \text{의 2단계} : y_{sc} = \beta_0 + n_{sc} \beta_1 + X_s \beta_2 + u_{sc}$$

$y_{sc}$ 는 학생의 시험 성적이고,  $n_{sc}$ 는 학교 s에 있는 학급 c의 학생 수다.  $X_s$ 는 학교 특성변수들이다.  $e_s$ 는 연초 학교 s의 학년에 등록한 학생 수이고,  $f_{sc}$ 는 학교 s의 각 학년에 학급 c에 할당된 학생 수로 이론적인 학급의 학생 수를 의미한다. 2SLS (2단계 추정) 방법을 이용하는데, 1단계에서는  $n_{sc}$ 를 종속변수로 하고,  $X_s$ 를 설명변수와 도구변수로  $f_{sc}$ 를 놓고 OLS 방법으로 분석하여  $\hat{n}_{sc}$ 의 추정값을 구했다. 2단계는 앞에서 구한  $\hat{n}_{sc}$ 를 가지고 OLS 방법을 실시하여 알고자 하는  $\hat{\beta}_1$ 을 추정하였다.

[예 4] 초등학교의 질이 집값에 미치는 영향(Black, 1999)

이 논문은 초등학교의 질이 그 동네의 집값에 어떤 영향을 미치는지를 분석한 것으로서 우리나라에서도 많이 연구하고 있는 관심도 높은 주제 중의 하나다. 어떤 학교의 평균 성적이 50에서 60으로 바뀌었을 때 그 이유로 부동산 가격이 얼마나 올라가는지에 대한 인과효과를 추정하고자 하는 연구다.

이 논문에서는 미국 보스턴의 집값(평당 가격이나 평당 매매 가격)을 종속 변수로, 그 동네에 사는 초등학교 학생의 평균 성적을 설명변수로 놓고 OLS 방법을 실시한다. 그런데 시험 점수가 좋은 학교는 좋은 지역에 있는 학교일 수 있다. 좋은 지역이란 주변 환경이나 교통 시설 등의 제반 시설이 좋은 곳인데, 그런 살기 좋은 환경이 반영되어서 부동산 가격이 높은 것이지 시험 점수가 좋아서 그런 것이 아니라고 주장할 수 있기 때문에 내생성이 있다고 볼 수 있다. 그래서 연구자는 보스턴 전체를 다 보는 것이 아니라, 어느 한 구획을 보자라는 아이디어를 제안하였다. 여기서 구획이란 초등학교의 학군을 의미하며, 이 구획의 경계선을 기준으로 좁혀서 살펴본 것이다. 보스턴의 시내가 아닌 외곽의 동네는 2차선 도로를 중심으로 마주보고 있는 집들이 있다. 그 길을 기준(학군의 경계선)으로 왼편에 사는 학생은 왼편으로, 오른편에 사는 학생은 오른편의 학교를 다닐 것이다. 학군을 제외하고 나머지의 특성들(주변 시설, 대중교통이나 공원 등의 환경)은 상당히 비슷할 것이다. 그러므로 아이들이 다니는 학교에 따라서 두 집의 가격이 차이가 난다면 그것은 학교 때문이라고 볼 수 있는 것이다. 그래서 학군의 경계선을 따라 양쪽으로 0.35, 0.20 및 0.15마일(거리를 점점 줄여 가면서 살펴봄)을 기준으로 하여 집을 추출하여 OLS 방법을 실시하였다.

그 결과 시험 점수의 계수값이 모든 집을 대상으로 하였을 때는 0.035이었지만 0.15 마일을 기준으로 선택한 집의 경우에는 0.015로 계수값이 절반 가까이 줄어들었다. 이를 통해 그 전에 분석한 결과는 과대 추정(overestimate) 되었다고 볼 수 있다. 그리고 분석할 때 고려할 것은 어느 구획에서 추출된 집인지를 알려 주는 변수를 넣어 주어야 한다. 이처럼 구역 고정효과(boundary fixed effect)를 고려하지 않으면 경계선을 기준으로 거리를 좁혀

분석하더라도 모든 집을 대상으로 한 결과와 비슷하게 나오므로 의미가 없게 되기 때문이다. 이 연구를 통해 부모들은 자녀들의 시험 점수 5% 증가를 위해 집값으로 2.5%를 더 지불한다는 사실을 알아냈다.

[예 5] 고용보험의 사업주 직업능력개발지원사업이 기업의 훈련투자 결정에 미치는 영향(강창희 · 유경준, 2009)

우리나라 정부에서는 고용보험의 ‘사업주 직업능력개발지원사업’을 통하여 개별 기업들의 근로자 직업훈련 활동에 영향을 미치고 있다. 이 사업은 기본적으로 분담금 · 보조금 제도로서 정부는 훈련 발생 이전에 기업에 대하여 일정한 법정 분담금을 부과하고, 훈련 발생 이후에 그 분담금의 전부 또는 그것을 초과하는 보조금을 지급한다. 이 논문은 현재 실시되고 있는 고용보험의 사업주 직업능력개발지원사업(이하 ‘지원사업’)이 기업의 교육훈련투자 수준과 그 기업의 성과변수에 어떠한 영향을 미치는지에 대한 것이다.

선행연구들에서는 횡단면 자료에 대해 단순 OLS를 적용하여 추정하였는데 이는 지원사업의 요율체계로부터 유발되는 내생성을 적절히 통제했다고 판단하기 힘들며, 기업 훈련량과 성과 사이에 존재할 수 있는 내생성 또한 적절히 고려되지 못한 한계가 있다. 왜냐하면 근로자 1인당 매출액이 높은 기업일수록 근로자에 대한 직업훈련에 보다 많은 투자를 할 성향이 높은 경우 훈련이 기업성가에 미치는 영향을 과대 평가할 수 있기 때문이다. 그래서 연구자는 지원사업의 요율체계가 기업의 훈련 및 성과에 미치는 인과효과를 분석하기 위해 지원사업의 요율체계의 불연속적인 변화로부터 야기되는 무작위성(randomness)을 활용하였다.

우선, 지원사업의 분담률과 지원율의 체계는 산업군(3개)과 기업규모(각 산업군별 4개)에 따라 상이하게 적용되고 있다. 각 산업군에 따라 다른 4개의 기업규모 그룹이 있기 때문에, 4그룹을 나누는 3개의 기업규모 경계값이 각 산업군에 존재한다. 이 경계값(3개)을 기준으로 기업규모가 크거나 작음에 따라서 각 기업에 적용되는 근로자 1인당 직업훈련분담률, 최대 지원율, 최대 순지원율이 달라진다. 여기서 회귀단절을 적용하는 방법은 각 산업군에 존재



하는 기업규모 경계값들을 활용하며, 이때 분석 표본은 불연속성이 유발되는 경계값 주변의 관측치로 한정하는 것이다. 모형 식은 다음과 같다.

$$y_i = \beta_0 + \beta_1 [b_i * (1 - h_{3i})] + \delta (l_i * h_{3i}) + \beta_2 X_i + \gamma_2 h_{2i} + \gamma_3 h_{3i} + u_i$$

$$(if |\ln(\frac{s_i}{\tau_k})| \leq \epsilon, \forall k(=1, 2, 3))$$

여기서  $b_i$ 는 기업  $i$ 에 적용되는 최대 순지원율의 크기이고,  $\tau_k(k=1, 2, 3)$ 는 기업  $i$ 가 어느 산업군의 경계값 I, II, III에 속하는지를 알려 주는 변수이다.  $h_{ki} = 1(|\ln(\frac{s_i}{\tau_k})| \leq \epsilon), (k=2, 3)$ 는 기업  $i$ 의 기업규모  $s_i$ 가 경계값 I, II, III 중 어느 경계값 근처에 위치하고 있는가를 나타낸다. 이때  $h_{2i}$ 와  $h_{3i}$ 가 설명변수로 추가되어야만 각 경계값의 상하 간에 위치한 기업들을 비교할 수 있다. 마지막으로 산업분류, 노조 유무 및 기업규모와 같은 기업의 제반 특성변수들( $X_i$ )이 포함된다. 그런데 경계값 III 근처에서는 동일한 분담률과 최대 지원율이 모든 산업에 공통적으로 적용되어 최대 순지원율이 0으로서 변화하지 않는다. 즉, 모든 산업의 경계값 III 근처에서는 공통적으로  $b_i$ 값의 변화가 없다. 그래서 III은 I과 II와 구분해서 보기 위해 교차항( $l_i \cdot h_{3i}$ )을 넣어 준 것이다.  $l_i$ 는 기업  $i$ 에 적용되는 분담률(최대지원율)의 크기를 나타내는 변수다.

따라서 알고자 하는  $\beta_i$ 는 경계값 I과 II의 근처에서  $b_i$ 값의 차이가 유발하는  $y_i$ 의 평균값의 차이를,  $\delta$ 는 경계값 III 근처에서  $l_i$ 값의 차이가 유발하는  $y_i$ 의 평균값의 차이를 각각 보여준다. 여기서 조건식  $|\ln(\frac{s_i}{\tau_k})| \leq \epsilon$ 은 분석에 사용되는 기업의 규모변수  $s_i$ 는 경계값  $\tau$ 로부터 상하 대칭적으로 약  $100 \times \epsilon\%$  반경 내에 위치해 있어야 한다는 RDD의 표본 선정 조건을 나타내는 것이다.

RDD분석 전에 경계값 주변의 기업들이 자신의 의도와는 무관하게 외생적으로(마치 동전 던지기와 비슷한 방식) 경계값의 상하에 위치해야 한다는

RDD의 타당성을 검증하기 위해 두 가지 방식을 통하여 간접적으로 검증하였다. 첫 번째 방법은 임계치 주변 상하에 위치한 기업들 사이에 사전적인 기업 특성들(pre-treatment characteristics)의 체계적인 차이(특히, 평균값의 차이)가 있는지 확인하는 방법이다. 두 번째 방법은 기업의 규모  $s_i$ 의 분포가 경계값을 중심으로 상하 대칭적인지를 확인하는 방법이다. 그 결과 경계값 7% 반경 근처의 상하에 위치한 기업들 간에 체계적인 차이가 있다는 증거는 발견되지 않으므로 RDD를 실행할 수 있다는 것을 확인하였다.

분석 결과에 의하면, 법정 분담금·지원금의 격차(즉, 순지원율)는 ① 기업의 직업훈련 여부 ② 기업의 직업훈련 실시 근로자 비율 ③ 근로자 1인당 총 훈련비용 및 ④ 기업의 매출에 유의한 영향을 미친다. 따라서 지원사업의 현 효율체계는 기업의 근로자 훈련 결정과 기업성장에 유의미한 영향을 미치고 있다고 볼 수 있다.

## 제9장

# 컴퓨터 프로그램을 이용한 분석 예시

본장에서는 정책평가를 위한 자료가 주어져 있을 때, 앞에서 다루어진 다양한 노동시장 프로그램 평가방법론을 실제로 어떻게 적용할 수 있는지 제시하고자 한다. 그리고 실제로 구해지는 추정 결과로부터 각 방법론이 갖고 있는 장점 및 단점에 대해서 비교·분석하고자 한다. 우리가 주어진 자료를 활용하여 실증분석을 수행함에 있어서는 다양한 통계분석 프로그램을 활용할 수 있겠으나, 본장에서는 최근 계량경제 분석 프로그램으로 폭넓게 활용되고 있는 Stata를 활용한 분석방법을 제시하도록 한다.<sup>4)</sup>

## 제1절 최소자승(OLS) 추정법

앞에서도 언급했듯이, 우리는 일반적으로 어떠한 정책처방이 개인의 성과에 미치는 효과를 추정하기 위해서 다음과 같은 모형을 설정하게 된다.

---

4) 본장에서 제시하는 Stata 명령어들은 해당 방법론의 추정을 위한 최소한의 내용을 포함하는 것이므로, Stata 명령어에 대한 보다 자세한 사항은 Stata 도움말(help) 및 Stata Reference Book을 참고하기 바란다. 아울러, Stata 버전에 따른 명령어의 차이가 있을 수 있음을 상기하기 바란다. 본장에서는 Stata version 11을 기본으로 사용한다.

$$y_i = \beta_0 + \beta_1 d_i + u_i$$

여기서  $d_i$ 는 개인  $i$ 가 어떤 정책의 수혜를 받았는지( $d_i = 1$ ), 받지 않았는지( $d_i = 0$ )를 나타내는 가변수(dummy variable)이고,  $y_i$ 는 성과지표다. 예를 들면, 직업훈련사업에의 참여가 개인의 임금에 미치는 영향을 보고자 할 경우,  $d_i$ 는 직업훈련사업에 참여했는지 여부이고,  $y_i$ 는 임금수준이 된다.

이제 우리에게  $y$ 와  $d$ 의 정보를 가진 개인 10만 명의 데이터가 주어졌다고 하자. 10만 명의 관측치를 가진 데이터는 표본의 크기 측면에서는 특별히 문제를 삼을 수 없을 것이기 때문에, 위에 제시한 추정식을 이용하여 다음과 같이 Stata에서 OLS 추정을 수행할 수 있다.

추정 결과를 살펴보면,  $d$ 에 대한 추정계수가 약 1.15 수준이며, 통계적으로 상당히 유의적이어서 1% 유의수준에서도 통계적으로 유의한 것을 확인할 수 있다. 이로부터 처방을 받은 사람( $d_i = 1$ )이 받지 않은 사람( $d_i = 0$ )에 비해서 성과( $y_i$ )가 평균적으로 약 1.15 단위 정도 높다고 결론지을 수 있다. 이를 예를 들어 다시 표현하면, 평균적인 (소득)수준의 개인  $i$ 가 직업훈련사업

[그림 9-1] Simple OLS Stata 프로그램 및 결과

. regress y d						
Source	SS	df	MS	Number of obs = 100000		
Model	33273.4968	1	33273.4968	F( 1, 99998) = 13496.83		
Residual	246523.301	99998	2.46528232	Prob > F = 0.0000		
Total	279796.798	99999	2.79799596	R-squared = 0.1189		
				Adj R-squared = 0.1189		
				Root MSE = 1.5701		
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d	1.153669	.0099304	116.18	0.000	1.134206	1.173132
_cons	-6.579582	.0070324	-935.61	0.000	-6.593366	-6.565799

에 참여했을 때( $d_i = 1$ ), 참여하지 않은 경우( $d_i = 0$ )보다 임금수준이 약 1.15 단위 증가한다고 생각할 수 있다.

하지만 이러한 결과가 정확한 결과인지에 대해서는 의문이 존재한다. 왜냐하면, 위의 추정모형에서  $d$ 가 과연 오차항  $u$ 와 상관이 없는 외생적인 변수인지에 대한 확신이 없기 때문이다. 예를 들어,  $d$ 와  $y$ 에 동시에 영향을 미치는 어떤 변수를 모형에 포함시키지 않음으로써, 즉 중요한 변수의 누락(omitted variables)으로 인한 편의(bias)가 발생할 수 있다. 그래서 일반적으로 실증분석을 수행할 경우에 개인의 다양한 인적 특성을 설명변수로 포함하게 된다. 앞서 우리에게 주어진 10만 명의 데이터에 개인의 인적 특성으로  $x_1$ 과  $x_2$ 가 함께 주어졌다고 하자. 그렇다면 이러한 인적 특성 변수들을 활용하여 다음과 같이 추정모형을 설정할 수 있다.

$$y_i = \beta_0 + \beta_1 d_i + \beta_2 x_{1i} + \beta_3 x_{2i} + u_i$$

이 추정모형을 Stata를 활용하여 OLS 추정한 결과는 [그림 9-2]와 같다.

[그림 9-2] Multiple OLS Stata 프로그램 및 결과

. regress y d x1 x2

Source	SS	df	MS	Number of obs	=	100000
Model	121062.749	3	40354.2498	F( 3, 99996)	=	25421.54
Residual	158734.048	99996	1.58740398	Prob > F	=	0.0000
				R-squared	=	0.4327
				Adj R-squared	=	0.4327
Total	279796.798	99999	2.79799596	Root MSE	=	1.2599

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d	1.299217	.0080856	160.68	0.000	1.28337	1.315065
x1	1.081933	.0046415	233.10	0.000	1.072835	1.09103
x2	-.4330302	.0046877	-92.38	0.000	-.442218	-.4238425
_cons	-6.650108	.0056827	-1170.24	0.000	-6.661246	-6.63897

인적 특성을 고려한 추정 결과를 통해서도 처방을 받은 사람이 받지 않은 사람에 비해서 평균적으로 성과가 약 1.30 단위 정도 높다는 결론을 내릴 수 있다. 다시 말해서, 인적 특성을 고려할 경우에 추정계수의 크기는 다소 변하지만 여전히 통계적으로는 유의한 영향을 미치고 있다는 결론을 도출할 수 있는 것이다.

하지만 과연 이것이 정책 수혜 여부가 개인의 성과에 미치는 인과효과(causal effect)인지에 대해서는 논란의 여지가 있다. 왜냐하면, 개인의 다양한 인적 특성을 모형에서 고려했음에도 불구하고, 여전히 정책 수혜 변수인  $d$ 가 외생적인지에 대해서는 의구심이 들기 때문이다. 아무리 많은 개인의 인적 특성을 고려한다고 하더라도, 개인의 성과( $y$ )에 미치는 모든 요소를 고려하는 것이 불가능하기 때문에, 정책 수혜 변수  $d$ 는 여전히 오차항  $u$ 와 상관관계가 유지되는 내생변수일 가능성이 높다. 만약  $d$ 가 내생변수라면 OLS를 통해서 추정한  $\hat{\beta}_1$ 에는 편의가 발생하게 되어  $\hat{\beta}_1$ 은 우리가 관심 있는 정책 수혜 여부가 개인의 성과에 미치는 인과효과가 아니라 단순히 둘 사이의 상관성을 보여주는 수치일 수 있는 것이다.

그렇기 때문에  $d$ 가 보유하고 있을지도 모르는 내생성을 통제하여 정책 수혜에 대한 진정한 인과효과를 추정하기 위한 다양한 프로그램 평가방법론을 앞선 절에서 논의하였다. 지금부터 우리에게 데이터가 주어졌을 때, 앞에서 논의된 다양한 평가방법론들을 Stata를 활용하여 어떻게 추정할 수 있는지에 대해서 살펴보도록 한다.

## 제2절 성향점수 매칭

성향점수 짝짓기(PSM)는 일반적으로 다차원 하에서 수행되던 짝짓기(matching)를 일차원의 성향점수(propensity score)로 변환하여 보다 편리하

게 짝짓기를 수행하여 정책처방집단의 평균효과(Average Treatment Effect on the Treated : ATET)를 추정하는 방법이다. 과거에는 정책 수혜의 효과를 추정하려 할 때 정책 수혜 집단에 대한 자료만이 주어지는 경우가 종종 발생했기 때문에, 정책 비수혜 집단(통제집단, control group)에 대한 정보를 별도의 자료로부터 추출해야 했으며, 이때 사용된 방법이 짝짓기 방법이다. 따라서 짝짓기 방법은 전통적으로 ATET 추정에 많이 사용되었으며, 짝짓기를 통해서 정책변수( $d$ )의 내생성 통제가 가능하다는 믿음이 한동안 유지되어 왔다.

성향점수 짝짓기(PSM)를 통한 정책처방집단의 평균효과(ATET)를 Stata에서 비교적 손쉽게 구할 수 있다. 하지만 PSM 추정을 위한 명령어들은 Stata 설치 시 내재되어 있지 않으므로, Findit 또는 Search 명령어를 통해서 검색을 실시한 후, 해당 프로그램을 설치하고 작업을 수행해야 한다. PSM을 활용하여 ATET를 추정할 경우에는 짝짓기의 방법에 따라서 크게 네 가지로 나눌

[그림 9-3] PSM Stata 프로그램 검색 결과

```
. findit attnd
. search attnd

Keyword search

Keywords: attnd
Search: (1) Official help files, FAQs, Examples, SJs, and STBs

Search of official help files, FAQs, Examples, SJs, and STBs

SJ-5-3 st0026_2 . . . . . Software update for pscore suite
      (help attk, attnd, attnw, attr, atts, and pscore if installed)
      Q3/05 SJ 5(3):470
      bug fix for pscore

SJ-4-2 st0026_1 . . . . . Software update for pscore suite
      (help attk, attnd, attnw, attr, atts, and pscore if installed)
      Q2/04 SJ 4(2):224
      pscore now allows for bootstrap options in Stata 8 and for
      identification of matched pairs in nearest neighbor matching

SJ-2-4 st0026 . Est. of average treatment effects based on propensity scores
      (help pscore, attnd, attr, ... if installed) S. O. Becker & A. Ichino
      Q4/02 SJ 2(4):358-377
      propensity score matching estimators

(end of search)
```

수 있으며, 그에 따라서 각각의 Stata 명령어가 존재한다. attk, attnd, attnw, attr, atts 등의 프로그램이 바로 그것들이며, 구체적으로 다음과 같이 구별할 수 있다. attk는 막대 짝짓기(Kernel Matching)를 할 때 사용하는 명령어이며, attnd 및 attnw는 가장 가까운 이웃 짝짓기(Nearest Neighbor Matching) 방법이고, attr은 반지름 짝짓기(Radius Matching)를 수행할 때, 마지막으로 atts는 계층 짝짓기(Stratification Matching)를 활용할 때 사용하는 명령어이다.

위의 [그림 9-3]과 같이 findit 또는 search를 통해서 PSM 추정을 위해서 attk, attnd, attnw, attr, atts 등의 프로그램들을 설치하였다면, 다음 [그림 9-4]와 같이 간단히 PSM 명령어를 Stata에 입력함으로써 ATET를 추정할 수 있다.<sup>5)</sup>

[그림 9-4] PSM Stata 프로그램 및 결과(attnd,ado 및 attr,ado)

**. attnd y d x1 x2**

The program is searching the nearest neighbor of each treated unit.  
This operation may take a while.

ATT estimation with Nearest Neighbor Matching method  
(random draw version)  
Analytical standard errors

n. treat.	n. contr.	ATT	Std. Err.	t
50151	26041	<b>1.313</b>	<b>0.014</b>	<b>94.965</b>

Note: the numbers of treated and controls refer to actual nearest neighbour matches

**. attr y d x1 x2**

The program is searching for matches of treated units within radius.  
This operation may take a while.

ATT estimation with the Radius Matching method  
Analytical standard errors

n. treat.	n. contr.	ATT	Std. Err.	t
50151	49849	<b>1.194</b>	<b>0.010</b>	<b>117.426</b>

Note: the numbers of treated and controls refer to actual matches within radius

- 5) 본절에서는 PSM 중 상대적으로 계산 과정이 용이하여 광범위하게 사용되고 있는 가장 가까운 이웃 짝짓기(attnd)와 반지름 짝짓기(attr)의 결과를 제시하도록 한다. 이외에도 막대 짝짓기(attk) 역시 비교적 자주 활용되는 방법 중 하나이지만, 다른 방법들에 비해서 상대적으로 긴 계산 시간을 요구한다는 단점이 있다.



[그림 9-4]의 PSM 추정 결과를 살펴보면, 짝짓기의 방법에 따라 다소 차이가 있지만, 전반적으로 정책처방집단의 평균효과(ATET)가 통계적으로 유의한 수준인 것으로 나타났다. 그 크기 역시 약 1.19~1.31 수준으로 앞서 분석한 최소자승(OLS) 결과와도 유사함을 발견할 수 있다. 이상의 결과를 종합하여 유추하면, 우리에게 주어진 10만 명의 자료에서는 정책처방( $d$ )이 개인의 성과( $y$ )에 상당히 강한 양(+)의 영향을 미치는 것으로 결론지을 수 있다.

하지만 이러한 OLS 및 PSM 추정 결과를 통해 일정한 결론을 내리는 것에는 상당한 주의를 기울여야 한다. PSM 추정에 있어서 가장 중요한 가정 중의 하나는 조건부 독립성 가정(Conditional Independence Assumption)이다. 이 가정은 수식으로  $[(y_0, y_1) \perp d \mid X]$ 와 같이 표현될 수 있으며, 이는 개인의 인적 특성( $X$ )이 적절하게 고려되면 그 하위 집단에서 개인의 성과( $y$ )는 정책 수혜 여부( $d$ )와 상관이 없다는 것을 의미한다. 그런데 이것은 앞선 OLS 추정식에서 다양한 개인의 인적 특성을 설명변수로 추가하여 정책 수혜 변수( $d$ )의 내생성을 통제한다는 아이디어와 크게 다르지 않다. 따라서 OLS 추정 결과와 PSM 추정 결과가 유사하게 나타나는 것이 크게 놀라운 일은 아닐 뿐만 아니라, 두 추정 결과가 유사하다는 것만으로 정책처방( $d$ )의 효과를 추정함에 있어서 내생성이 통제된 채 모형의 강건성(robustness)이 만족된다고 주장하기에는 여전히 무리가 있다. 결국 PSM을 통해서 추정된 정책의 효과는 OLS를 활용한 추정치와 도출상의 가정이 크게 다르지 않기 때문에, 여전히 정책 수혜( $d$ )에 대한 내생성이 존재할 수 있으며, 이로 인해 진정한 정책의 인과효과를 보여주지 못할 수도 있다는 단점을 내포하고 있다.

그렇다면 어떻게 해야 하는가? 앞서 정책처방( $d$ )에 대한 내생성을 통제하여 정책과 그로 인한 개인의 성과( $y$ ) 사이의 단순한 상관관계가 아닌 정책이 성과에 미치는 인과효과를 추정할 수 있는 방법으로서 최근 실증경제학(Empirical Economics)에서 각광받고 있는 도구변수 추정법, 회귀단절모형, 이중차분법에 대해서 논의를 하였다. 지금부터 순서대로 이 세 가지 방법을 Stata에서 어떻게 활용할 수 있는지에 대해서 살펴보도록 한다.

### 제3절 도구변수(Instrumental Variables : IV)

#### 추정법

정책처방의 인과효과를 추정하고자 할 때 정책수혜에 대한 내생성을 통제하기 위해서 가장 오랜 기간 동안 사용되어 온 방법이 도구변수 추정법이다. 도구변수 추정법은 외생적인, 즉 추정모형의 오차항( $u$ )과는 상관이 없는 도구변수를 활용하는 방법으로, 구체적인 추정방법은 2단계 최소자승(Two Stage Least Squares : 2SLS 또는 TSLS) 추정법이 활용된다. 이제 우리의 10만 명 자료에 외생적인 변수  $z$ 가 주어졌다고 하자. 그렇다면 Stata에서 ivregress 명령어를 활용하여 [그림 9-5]와 같이 2SLS 추정치를 구할 수 있다.

도구변수를 활용한 2SLS 추정 결과(그림 9-5)를 살펴보면, OLS 및 PSM의 결과와는 달리 정책처방( $d$ )이 개인의 성과( $y$ )에 미치는 영향의 크기가 약 0.01 수준으로 크게 감소한 것은 물론이거니와 통계적으로도 유의하지 않은 것을 확인할 수 있다. 다시 말해서, 정책의 수혜 여부가 개인의 성과에는 영향을 미치지 않는다는 것이다. 이것은 OLS 및 PSM의 결과와는 상충되는 결과로, 앞서 제기한 바와 같이 OLS와 PSM을 통해서는 여전히 내생성을 제대로 통제할 수 없다는 문제를 보여주는 단적인 예라 할 수 있다.

실증적으로 도구변수 추정법을 사용하기 위해서는 두 가지 충족되어야 하는 조건이 있다. 하나는 도구변수의 외생성(exogeneity)이고, 다른 하나는 도구변수의 적절성(relevance)이다. 이 중에서 도구변수의 적절성은 도구변수가 추정모형 내 내생변수와 상당히 높은 상관관계가 유지되어야 한다는 조건으로서, 이 조건이 충족되지 않을 경우에 약한 도구변수 문제(weak IV problem)가 발생하여 2SLS 추정치에 편의(bias)가 발생하게 된다. 도구변수의 적절성은 2SLS 추정과정에서 검정이 가능하다. 1단계 최소자승(OLS) 추정에서 도구변수의 유의성(0과 다른지 여부)에 대한 F-검정 통계량의 크기를 통



[그림 9-6] Weak IV 검정 Stata 프로그램 및 결과

```
. regress d x1 x2 z
```

Source	SS	df	MS	
Model	5153.50542	3	1717.83514	
Residual	19846.2666	99996	.198470605	
Total	24999.772	99999	.25000022	

Number of obs = 100000  
 F( 3, 99996) = 8655.36  
 Prob > F = 0.0000  
 R-squared = 0.2061  
 Adj R-squared = 0.2061  
 Root MSE = .4455

d	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	-.0552313	.0016337	-33.81	0.000	-.0584333 -.0520294
x2	.0979284	.0016336	59.95	0.000	.0947265 .1011302
<b>z</b>	<b>-.210322</b>	<b>.0014071</b>	<b>-149.47</b>	<b>0.000</b>	<b>-.2130799 -.2075642</b>
_cons	.5007425	.0014088	355.44	0.000	.4979813 .5035038

```
. test z=0
```

(1) z = 0

**F( 1, 99996) =22342.66**  
 Prob > F = 0.0000

[그림 9-6]에 제시된 1단계 추정에서 도구변수( $z$ )의 유의성에 대한 F-검정 통계량을 살펴보면, 무려 약 22,343 수준으로 일반적으로 받아들여지는 임계치(critical value)인 10보다 훨씬 크다는 것을 확인할 수 있다. 이로부터 우리에게 주어진 도구변수  $z$ 는 적절성 측면에서는 2SLS 추정을 수행하는 데 큰 문제가 없다는 결론을 내릴 수 있다. Stata에서 2SLS 추정치를 구할 때, ivregress 명령어를 사용할 경우에는 weak IV 검정을 위해서 위와 같이 1단계 OLS 추정을 별도로 수행해야 하는 번거로움이 있다. 이러한 불편함을 보완하고자 Stata는 ivreg2 명령어를 제공하고 있는데, 앞선 PSM 명령어들과 마찬가지로 ivreg2 명령어는 Stata에 내재되어 있지 않기 때문에 findit 또는 search 명령어를 통해 검색·설치한 후 사용해야 한다. ivreg2 명령어를 통해서 2SLS 추정을 수행하면, 다음 [그림 9-7]과 같은 추정 결과가 도출되면서 도구변수의 적절성에 대한 weak IV 검정 결과가 함께 제시되는 장점이 있다.

[그림 9-7] IV Stata 프로그램 및 결과(ivreg2.ado)

```
. ivreg2 y (d = z) x1 x2
```

---

IV (2SLS) estimation

---

Estimates efficient for homoskedasticity only  
Statistics consistent for homoskedasticity only

	Number of obs =	100000
	F( 3, 99996) =	13406.98
	Prob > F =	0.0000
Total (centered) SS =	279796.7979	Centered R2 = 0.2885
Total (uncentered) SS =	3881003.531	Uncentered R2 = 0.9487
Residual SS =	199088.1357	Root MSE = 1.411

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
d	.0100367	.0211888	0.47	0.636	-.0314926	.0515661
x1	1.010956	.005304	190.60	0.000	1.00056	1.021351
x2	-.3065302	.0055761	-54.97	0.000	-.3174592	-.2956013
_cons	-6.004055	.0115178	-521.28	0.000	-6.02663	-5.981481

Underidentification test (Anderson canon. corr. LM statistic): 1.8e+04  
Chi-sq(1) P-val = 0.0000

---

**Weak identification test (Cragg-Donald Wald F statistic): 2.2e+04**

Stock-Yogo weak ID test critical values:	10% maximal IV size	16.38
	15% maximal IV size	8.96
	20% maximal IV size	6.66
	25% maximal IV size	5.53

Source: Stock-Yogo (2005). Reproduced by permission.

---

**Sargan statistic (overidentification test of all instruments): 0.000**  
**(equation exactly identified)**

---

Instrumented: d  
Included instruments: x1 x2  
Excluded instruments: z

---

도구변수의 외생성을 확인하는 방법은 지금까지 논의한 도구변수의 적절성 검정과는 다소 차이를 보인다. 계량경제학 기법으로는 과다 식별검정 (overidentification test)을 통해서 도구변수가 외생적인 변수인지를 검정할

수 있으나, 이것은 추정에 사용된 도구변수가 내생변수보다 많아야 사용할 수 있다. 다시 말해서, 우리의 모형과 같이 내생변수( $d$ )와 도구변수( $z$ )의 개수가 동일할 경우에는 도구변수의 외생성을 통계적으로 검정할 수 있는 방법이 없다. ivreg2 명령어의 추정 결과에서는 weak IV 검정 결과와 더불어 과다 식별검정의 결과도 함께 제시하고 있는데, 앞선 추정 결과와 같이 모형이 과다 식별되지 않고 정확하게 식별된(exactly identified) 경우에는 검정통계량을 제시하지 못하고 있음을 확인할 수 있다.

일반적으로 순수하게 외생적인 도구변수를 실증분석 과정에서 찾는 것조차 쉽지 않은데, 그나마도 도구변수의 수가 내생변수보다 많아야 도구변수의 외생성을 통계적으로 검정할 수 있다. 따라서 도구변수 추정법은 순수하게 외생적인 도구변수를 현실에서는 구하는 것이 쉽지 않다는 한계를 지니고 있다.

## 제4절 회귀단절모형

정책처방의 인과효과를 도구변수 추정법을 활용하여 추정하려 할 경우에는 반드시 외생적인 도구변수가 필요한데, 이러한 변수를 찾아내는 일은 일반적으로 상당히 어려운 것으로 알려져 있다. 횡단면 자료가 주어져 있을 때 순수하게 외생적인 도구변수를 찾는 어려움 때문에 도구변수 추정법을 적용하여 정책의 인과효과를 추정하는 것에는 한계가 있다.

회귀단절모형(RDD)은 정책처방 결정의 규칙이 알려져 있을 때 경계값 근처에서 이루어지는 불연속성을 활용하여 정책처방의 인과효과를 추정하는 방법이다. 우리에게 주어진 데이터에서 정책처방 변수( $d$ )가 다음과 같은 규칙에 의해서 결정된다고 하자.

$$d_i = d_i(x_3) = 1(x_{3i} > 0)$$

여기서  $1(\cdot)$ 은 괄호 안의 조건이 만족되면 1, 아니면 0의 값을 갖는 지표 함수(indicator function)를 의미한다. 따라서 우리에게 주어진 자료는 임계치 ( $x_{3i}=0$ )에서 정확하게 절단이 이루어진 Sharp 회귀단절에 해당한다.<sup>6)</sup> 이와 같이 정책 수혜 여부 결정과정에서 발생하는 절단의 특징을 인과효과 추정에 활용하기 위해서 회귀단절은 정책처방의 임계치인  $x_{3i}=0$  근처의 개인들에 게만 관심을 가진다. 왜냐하면, 임계치 주변의 개인들은 다양한 인적 특성의 측면에서 상당히 유사할 가능성이 많기 때문이다. 따라서 이들 사이에서는 정책처방이 무작위하게 배정되었다고 볼 수 있다는 것이 회귀단절 추정의 기본 생각이다.

이러한 회귀단절의 기본 아이디어하에서 다음과 같은 추정모형을 설정할 수 있다.

$$y_i = \beta_0 + \beta_1 d_i + f(x_{3i}) + u_i$$

회귀단절을 추정하는 데 있어서는 규칙변수(forcing variable)의 함수  $f(x_{3i})$ 의 형태를 어떻게 가정하고 추정하느냐에 따라서 모수적 방법(parametric method)과 비모수적 방법(non-parametric method)으로 구분한다.

우선 모수적 방법을 Stata에서는 어떻게 적용할 수 있는지 살펴보자. 모수적 방법을 활용할 경우에는 비교적 쉽게 인과효과의 추정치를 구할 수 있다. 규칙변수의 함수를  $f(x_{3i}) = \gamma_1 x_{3i} + \gamma_2 x_{3i}^2 + \gamma_3 x_{3i}^3 + v_i$ 와 같이 모수적으로 확정하고, 표본을 임계치 주변으로 한정한 후, OLS 추정을 통해서 구한  $\hat{\beta}_1$ 이 우리가 관심 있는 회귀단절 추정치다. Stata 추정 결과는 다음과 같다.

6) 정책처방에 따른 절단이 어떤 한 값에서 정확하게 이루어지지 않고, 그 값 주변의 여러 값에서 이루어져 임계치가 일정 구간을 갖는 경우를 Fuzzy 회귀단절이라 한다.

[그림 9-8] RDD Stata 프로그램 및 결과 : 모수적 방법

```

.gen x3squ=x3^2
.gen x3cub=x3^3

.regress y d x3 x3squ x3cub if x3>=0.5 & x3<0.5

```

Source	SS	df	MS			
Model	1316.72	4	329.18	Number of obs = 39351		
Residual	88296.8606	39346	2.24411276	F( 4, 39346) = 146.69		
				Prob > F = 0.0000		
				R-squared = 0.0147		
				Adj R-squared = 0.0146		
Total	89613.5806	39350	2.27734639	Root MSE = 1.498		

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d	-.0145842	.0396396	-0.37	0.713	-.0922789	.0631105
x3	.7057318	.1302178	5.42	0.000	.4505017	.9609619
x3squ	.0331378	.1021882	0.32	0.746	-.1671535	.2334291
x3cub	-.2691739	.535966	-0.50	0.616	-1.31968	.7813325
_cons	-6.003104	.0226998	-264.46	0.000	-6.047596	-5.958612

```

.regress y d x3 x3squ x3cub if x3>=1 & x3<=1

```

Source	SS	df	MS			
Model	9405.08055	4	2351.27014	Number of obs = 69264		
Residual	145938.179	69259	2.10713668	F( 4, 69259) = 1115.86		
				Prob > F = 0.0000		
				R-squared = 0.0605		
				Adj R-squared = 0.0605		
Total	155343.26	69263	2.24280294	Root MSE = 1.4516		

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d	-.0089336	.0276522	-0.32	0.747	-.063132	.0452647
x3	.6633844	.0470751	14.09	0.000	.5711173	.7556515
x3squ	.0030745	.0196594	0.16	0.876	-.0354579	.0416069
x3cub	.0561257	.0507912	1.11	0.269	-.043425	.1556764
_cons	-6.005236	.0159065	-377.53	0.000	-6.036412	-5.974059

```

.regress y d x3 x3squ x3cub if x3>=2 & x3<=2

```

Source	SS	df	MS			
Model	41167.0964	4	10291.7741	Number of obs = 95090		
Residual	209729.831	95085	2.2057089	F( 4, 95085) = 4665.97		
				Prob > F = 0.0000		
				R-squared = 0.1641		
				Adj R-squared = 0.1640		
Total	250896.927	95089	2.63854838	Root MSE = 1.4852		

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d	-.0399261	.0213247	-1.87	0.061	-.0817224	.0018701
x3	.7154412	.0206927	34.57	0.000	.6748838	.7559986
x3squ	.0073413	.0053692	1.37	0.172	-.0031822	.0178649
x3cub	.0336603	.006619	5.09	0.000	.0206871	.0466336
_cons	-5.991594	.0123551	-484.95	0.000	-6.01581	-5.967378



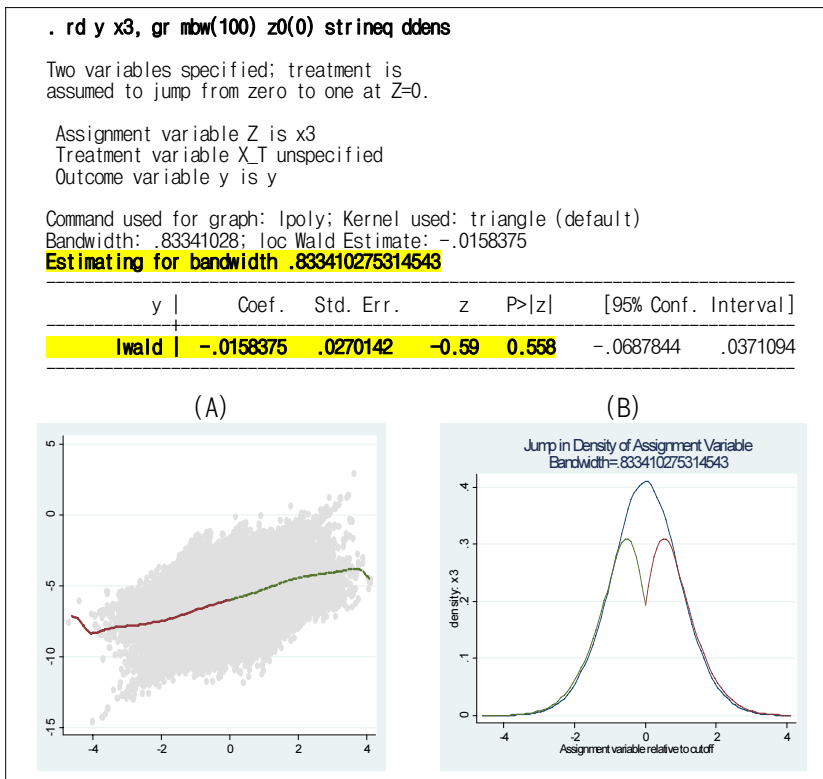
이상의 모수적 방법에 의한 회귀단절 결과를 살펴보면, 정책처방 결정의 임계치인  $x_{3i} = 0$ 에 근접해 있는 개인들로 표본을 제한하면 정책의 효과가 유의하지 않은 것으로 추정된다. 예를 들어,  $x_{3i} \in (-0.5, 0.5)$  또는  $x_{3i} \in (-1, 1)$ 의 범위로 표본을 제한한 후 계산한 회귀단절 추정치는 각각 -0.015, -0.009 수준이며 통계적으로 유의하지 않은 것으로 나타났다. 이는 앞서 도구변수 추정을 통해서 도출하였던 정책효과의 결과와 유사한 결과다. 따라서 IV 추정법과 회귀단절 추정법의 결과를 바탕으로, 내생성이 통제될 경우에 정책 수혜 여부( $d$ )가 개인의 성과( $y$ )에 미치는 인과효과는 없다는 결론을 유추할 수 있다.

하지만  $x_{3i} \in (-2, 2)$ 인 표본에 대해서는  $\hat{\beta}_1 = -0.040$  수준으로 10% 유의수준에서 통계적으로 유의한 값이 된다. 이렇듯 회귀단절 추정 내에서도 표본을 확장하는 기준에 따라서 결과가 상이하게 나타난다. 왜 이러한 결과가 나타난 것일까? 위의 Stata 결과에서도 확인할 수 있듯이,  $x_{3i} \in (-2, 2)$ 인 표본은 95,090명으로 전체 10만 명의 약 95%에 해당하는 표본이다. 임계치로부터 멀리 떨어져 있는 개인들까지 포함하는 표본으로 회귀단절 분석으로 수행할 경우, 임계치 주변의 개인들에게는 정책처방이 무작위적으로 배정되었을 것이라는 회귀단절의 아이디어가 퇴색되어 정책처방 변수( $d$ )가 가지는 내생성을 회귀단절을 통해서도 통제하지 못하게 될 수도 있다. 그렇기 때문에 앞선 두 회귀단절 분석과는 다른 결과가 나타났을 가능성이 있다. 결국 회귀단절 추정에서 표본을 임계치로부터 어느 수준까지 확장해야 하는지가 분석의 핵심이 된다. 임계치로부터 좁게 표본을 설정할 경우에는 정책처방의 무작위성은 확보할 수 있겠으나 표본의 크기가 작아지는 단점을 갖는 반면, 표본을 넓게 확장하면 표본의 크기는 확보할 수 있으나 정책처방의 무작위성이 담보될 수 없는 단점이 있다. 즉, 회귀단절 추정에 있어서는 표본의 크기와 정책처방의 무작위성(외생성) 사이의 trade-off가 존재한다.

이제 비모수적 방법으로 회귀단절 추정을 Stata에서는 어떻게 수행할 수 있는지 살펴보도록 한다. 앞에서도 언급했듯이 회귀단절 추정의 비모수적 방법

은 회귀단절 추정식에서 규칙변수(forcing variable)의 함수  $f(x_{3i})$ 의 형태를 모수적으로 설정하지 않고 비모수적으로 추정하는 방법을 의미한다. Stata에서는 rd 명령어를 통해서 비모수적 방법을 활용한 회귀단절 추정을 수행할 수 있는데, rd 명령어는 개인이 작성하여 배포한 명령어로 Stata 내에서 findit 또는 search 명령어로도 검색할 수 없으며, 인터넷 검색을 통해 해당 프로그램을 다운로드하여 사용해야 한다.<sup>7)</sup> rd 명령어를 사용한 회귀단절 분석 결과는 다음과 같다.

[그림 9-9] RDD Stata 프로그램 및 결과:비모수적 방법(rd.ado)



7) rd 명령어는 인터넷 사이트(<http://ideas.repec.org/c/boc/bocode/s456888.html>)에서 다운로드 받을 수 있으며, 다운로드 받은 파일들은 c:\wado\personal 폴더에 저장하여 사용할 수 있다.

이상의 Stata 추정 결과를 살펴보면, 회귀단절 추정치는 -0.158 수준이며 통계적으로 유의하지 않은 것으로 나타났다. 이는 그림을 통해서도 확인할 수 있는데, 그림 (A)를 보면 임계치  $x_3 = 0$ 을 기준으로 좌측( $x_3 < 0$ )에서 접근하는 개인의 성과( $y_0$ )와 우측( $x_3 < 0$ )에서 접근하는 성과( $y_1$ ) 사이의 절단이 크게 이루어지지 않았음을 확인할 수 있다. 이러한 결과로부터 앞선 IV 추정 결과 또는 모수적 회귀단절 추정 결과와 마찬가지로, 정책처방의 내생성을 통제하면 정책 수혜 여부가 개인의 성과에 미치는 인과효과는 없다는 결론을 도출할 수 있다.

비모수적 방법인 rd 명령어 추정 결과에서는 회귀단절 추정치 이외에도 관심을 가질 만한 결과를 몇 개 제시하고 있다. 그 중 하나는 앞에서도 언급한 바 있는 표본의 확정과 관련된 구간 폭(bandwidth)을 제시한다는 점이다. 위 결과를 살펴보면, 구간 폭을 약 0.833 수준으로 하여 분석을 수행했다는 것을 확인할 수 있다. 따라서 앞선 모수적 방법 추정 결과에서  $x_3 \in (-2, 2)$ 인 표본은 지나치게 크게 설정된 것으로, 이로부터 도출된 회귀단절 추정치에는 편의(bias)가 발생했을 가능성이 많다는 것을 유추할 수 있다. 아울러, rd 명령어는 그림 (B)와 같이 임계치를 기준으로 정책 수혜 집단( $x_3 > 0$ )과 정책 비수혜 집단( $x_3 < 0$ )의 분포를 보여준다. 회귀단절의 기본 아이디어는 임계치 주변의 개인들에게는 정책처방이 무작위적으로 배정되었을 것이라는 점이다. 이를 위해 충족되어야 하는 가장 중요한 조건 중 하나가 임계치 주변의 분포가 유사해야 한다는 것인데, 그림 (B)는 이를 확인할 수 있는 근거를 제시한다. 그림 (B)를 살펴보면, 우리의 예에서는 임계치를 기준으로 정책 수혜 집단과 비수혜 집단의 분포가 서로 매우 유사하다는 것을 확인할 수 있으며, 이로부터 우리에게 주어진 자료는 임계치  $x_3 = 0$  주변에서 비교적 무작위적으로 정책처방( $d$ )이 결정되었을 가능성이 높다는 결론을 유추할 수 있다.

## 제5절 이중차분법(Difference in Differences : DID)

지금까지 횡단면 자료가 주어져 있을 때 정책처방이 성과에 미치는 인과효과를 추정하는 두 가지 방법, 도구변수모형과 회귀단절모형에 대해서 논의하였다. 만약에 우리에게 주어진 데이터가 두 시점에서 수집된 지속관측 자료(longitudinal data)라면, 정책처방의 내생성이 통제된 인과효과를 추정하기 위해서 이중차분법(DID)을 활용할 수 있다.

우리에게 주어진 10만 명의 자료를 생각해 보자. 이는 정책이 시행된 이후( $t = 1$ )에 정책 수혜 여부( $d_{i1}$ )와 개인의 성과( $y_{i1}$ )가 주어진 것이다. 따라서 이 자료를 정책 수혜 여부에 따라 정책 수혜 집단( $d_{i1} = 1 \equiv g_i$ )과 비수혜 집단( $d_{i1} = 0 \equiv g_i$ )으로 구분할 수 있을 것이다. 이제 이러한 두 집단의 정책 시행 이전( $t = 0$ ) 10만 명의 자료가 함께 있다고 하자.<sup>8)</sup> 즉, 0期の 성과( $y_{i0}$ )가 주어져 있다는 것을 의미하며, 0期에는 모든 개인이 정책처방을 받지 않았으므로  $d_{i0} = 0$ 이 된다. 이로부터 시점변수( $t$ )와 집단변수( $g$ ), 그리고 정책처방 변수( $d$ ) 사이에  $d = (t \times g)$ 의 관계가 성립함을 알 수 있다.

이중차분법에 의한 정책의 인과효과 추정치는 정책 시행 이전의 두 집단 간 평균 성과 차이를 정책 시행 이후의 두 집단 간 평균 성과 차이에서 빼주는 것으로, 다음과 같이 수식으로 표현할 수 있다.

$$ATE = (E[y_{i1} | g_i = 1] - E[y_{i1} | g_i = 0]) - (E[y_{i0} | g_i = 1] - E[y_{i0} | g_i = 0])$$

8) 지속관측 자료(longitudinal data)가 반드시 10만 명의 개인 패널자료(panel data)일 필요는 없다.

위에서 제시된 정책의 평균처방효과(ATE)에 대한 이중차분(DID) 추정치의 관계로부터 다음과 같은 회귀모형을 설정할 수 있으며, 이 모형에서 추정된  $\hat{\beta}_1$ 이 곧 ATE의 추정치이다.

$$y = \beta_0 + \beta_1 d + \beta_2 t + \beta_3 g + u = \beta_0 + \beta_1 (t \times g) + \beta_2 t + \beta_3 g + u$$

이상에서 제시된 DID 회귀모형은 OLS를 통해 간단하게 추정할 수 있으며, Stata 추정 결과는 [그림 9-10]과 같다.

[그림 9-10] DID Stata 프로그램 및 결과

<b>. regress y d t g</b>						
Source	SS	df	MS		Number of obs = 200000	
Model	67145.5287	3	22381.8429		F( 3,199996) = 9078.99	
Residual	493036.907199996	2.46523384			Prob > F = 0.0000	
					R-squared = 0.1199	
					Adj R-squared = 0.1199	
Total	560182.435199999	2.80092618			Root MSE = 1.5701	
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d	<b>-.0103156</b>	<b>.0140435</b>	<b>-0.73</b>	<b>0.463</b>	-.0378406	.0172093
t	.0010748	.0099453	0.11	0.914	-.0184176	.0205673
g	1.163985	.0099303	117.22	0.000	1.144521	1.183448
_cons	-6.580657	.0070324	-935.77	0.000	-6.59444	-6.566874
<b>. areg y d t, absorb(g)</b>						
Linear regression, absorbing indicators					Number of obs = 200000	
					F( 2,199996) = 0.44	
					Prob > F = 0.6440	
					R-squared = 0.1199	
					Adj R-squared = 0.1199	
					Root MSE = 1.5701	
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d	<b>-.0103156</b>	<b>.0140435</b>	<b>-0.73</b>	<b>0.463</b>	-.0378406	.0172093
t	.0010748	.0099453	0.11	0.914	-.0184176	.0205673
_cons	-5.996907	.0049651	-1207.81	0.000	-6.006638	-5.987175
g	F(1, 199996) = 13739.545			0.000	(2 categories)	

위의 Stata 결과에는 areg 명령어를 이용하여 추정한 결과도 함께 제시되어 있다. areg 명령어는 회귀모형에 큰 규모의 두 집단이 존재하는 경우에 선행 회귀 추정치를 구하는 데 활용할 수 있는 명령어로, 우리의 경우와 같이 정책 처방이 주어진 집단과 아닌 집단으로 구분될 때 DID 추정을 위해 활용될 수 있다. 이상의 추정 결과를 살펴보면, 정책의 효과가 약 -0.010 수준으로 나타나지만 통계적으로 유의하지 않게 나타나고 있다. 이 역시 앞선 도구변수모형 및 회귀단절모형의 결과와 동일한 결과로서, 내생성을 통제하면 정책 수혜 여부가 개인의 성과에 미치는 인과효과는 없다는 결론을 유추할 수 있다.

지금까지의 논의를 바탕으로 DID 추정을 수행함에 있어서는 정책 시행 이후의 두 집단(정책 수혜 및 비수혜 집단)에 대한 정책 시행 이전의 성과 자료가 반드시 필요하다는 것을 알 수 있다. 하지만 이것이 개인의 패널자료(panel data)가 구축되어야 한다는 것을 바로 의미하는 것은 아니다. 다시 말해서, DID 추정에서는 두 집단에 대한 개인성과가 정책 시행 이전과 이후에 존재하기만 하면 되는 것이지, 그것이 반드시 개인별 패널자료로 구축될 필요는 없다는 것이다.

이제 우리에게 패널자료가 주어졌다고 하자. 즉, 정책 시행 이후에 관찰된 10만 명 각 개인에 대해서 정책 시행 이전의 성과자료가 주어져 있는 상황이다. 이러한 패널자료가 주어진 경우에는 다음과 같이 회귀모형을 설정하여 보다 쉽게 DID 추정치를 구할 수 있다.

$$y_{it} = \beta_0 + \beta_1 d_{it} + \beta_2 t + u_{it} ; u_{it} = \alpha_i + \varepsilon_{it}$$

위에서 제시된 패널모형으로부터, within 추정치  $\hat{\beta}_1^w$ 를 구하면 ATE를 추정할 수 있다. Stata 추정 결과는 [그림 9-11]과 같다.

패널모형에 의한 within 추정 결과에서도 정책의 수혜 여부가 개인의 성과에 미치는 인과효과는 없는 것으로 나타났다. 아울러, OLS 및 areg 명령어, 그리고 within 추정을 통해서 구해진  $\hat{\beta}_1$ 이 모두 -0.0103156으로 동일하다는 것을 발견할 수 있다. 따라서 이러한 세 가지 방법이 모두 DID 추정치를 구하는 동일한 방법이라는 것을 확인할 수 있다.

[그림 9-11] DID Stata 프로그램 및 결과(패널분석)

```

. tsset id t
    panel variable: id (strongly balanced)
    time variable: t, 0 to 1
    delta: 1 unit

. xtreg y d t, fe

Fixed-effects (within) regression              Number of obs   =   200000
Group variable: id                            Number of groups =   100000

R-sq:  within = 0.0000                        Obs per group:  min =    2
        between = 0.1765                      avg   =    2.0
        overall  = 0.0449                      max   =    2

corr(u_i, Xb) = -0.2599                      F(2,99998)      =    0.60
                                                Prob > F        =    0.5468

```

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
d	<b>-.0103156</b>	<b>.0119907</b>	<b>-0.86</b>	<b>0.390</b>	-.0338172 .0131859
t	.0010748	.0084915	0.13	0.899	-.0155683 .017718
_cons	-5.996907	.0042393	-1414.59	0.000	-6.005216 -5.988598

sigma_u	1.3803406
sigma_e	1.3405918
rho	.51460541 (fraction of variance due to u_i)

F test that all u_i=0:	F(99999, 99998) =	1.93	Prob > F = 0.0000
------------------------	-------------------	------	-------------------

## 제6절 소 결

지금까지 정책처방( $d$ )이 개인의 성과( $y$ )에 미치는 인과효과(causal effect)를 추정하기 위해서 다양한 방법을 활용한 Stata 추정 결과를 살펴보았다. 추정 결과를 비교하면, 최소자승법(OLS) 및 성향점수 짝짓기(PSM)를 통해서는 정책이 개인의 성과에 양(+)의 매우 유의한 영향을 미치는 것으로 추정된 반면, 도구변수모형(IV) 및 회귀단절모형(RDD), 그리고 이중차분법(DID)을 통해서는 정책의 수혜 여부가 개인의 성과에는 영향을 미치지 못한다는 결론을 도출할 수 있었다. 그렇다면 우리는 어떠한 결과를 받아들여야 할 것인가? 앞서에서도 언급했다시피, PSM의 아이디어는 OLS에서 통제변수(control variable)

를 추가하는 아이디어와 다를 바가 없기 때문에, 여전히 내생성에 대한 위험이 존재한다. 반면 IV, 회귀단절, DID는 완전히 무작위적인 정책처방이 주어지지 않은 현실에서 내생성을 제대로 통제할 수 있는 추정방법으로 현대 실증경제학에서 각광을 받고 있다. 따라서 우리 역시 주어진 자료로부터 정책처방( $d$ )이 개인의 성과( $y$ )에는 아무 영향을 미치지 못했다는 결론을 도출하는 것이 바람직한 것이다.

그렇다면 왜 이렇게 다른 결과가 나타난 것일까? 당연히 정책처방( $d$ )이 갖는 내생성 때문일 것이다. 사실 지금까지 우리가 분석한 자료는 실제로 존재하는 통계자료가 아니라 시뮬레이션을 위해 생성된 자료다. 지금까지 언급된 변수들을 살펴보면, 개인의 성과변수(outcome variable:  $y$ ), 정책처방변수(treatment variable:  $d$ ), 정책처방의 결정에 대한 규칙변수(forcing variable:  $x_3$ ), 개인의 인적 특성( $x_1, x_2$ ), 도구변수(instrumental variable:  $z$ ) 등이 사용되었으며, 오차항( $u$ )은 관찰되지 않는 개인의 특수효과(unobservable individual specific effect:  $\alpha$ )와 백색잡음(white noise)인  $\varepsilon$ 으로 구성되어 있다. 이러한 변수들을 다음과 같은 다변량 정규분포(multi-variate normal distribution)를 갖도록 임의 생성(random generation)하였다.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ z \\ \alpha \\ \varepsilon \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & .5 & 0 & 0 & 0 & 0 \\ .5 & 1 & .2 & 0 & 0 & 0 \\ 0 & .2 & 1 & -.5 & .8 & 0 \\ 0 & 0 & -.5 & 1 & 0 & 0 \\ 0 & 0 & .8 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right)$$

이렇게 생성된 변수들을 바탕으로, 종속변수인 개인의 성과( $y$ )를 다음의 관계식에 기반하여 생성하였다.

$$y = -6 + 0 \times d + 1 \times x - 0.3 \times x + u ; u = \alpha + \varepsilon$$

따라서 실제모형(true model)에서 정책처방( $d$ )의 효과는 영(0)이다. 따라



서 제대로 추정이 이루어진다면, 실제 모형에서와 같이 정책 수혜 여부가 개인의 성과에 미치는 인과효과는 없다는 결론을 도출할 수 있어야 한다.

그런데 변수들의 분포로부터  $Cov(x_3, \alpha) = 0.8$ 임을 확인할 수 있으므로, 정책처방변수  $d = 1(x_3 > 0)$ 와 오차항  $u = (\alpha + \varepsilon)$  사이에  $Cov(d, u) \neq 0$ 이 되어 정책처방이 내생적인 변수가 된다. 따라서 이러한 자료를 OLS로 추정하면 당연히 편의(bias)가 발생하게 된다. 앞선 OLS 추정 결과를 상기해 보면, 정책의 인과효과 추정치  $\hat{\beta}_1$ 이 크기 측면에서도 실제 모형과는 많이 다른 1.30 수준인데, 그뿐만 아니라 통계적으로도 아주 유의한 것으로 추정될 만큼 내생성으로 인한 편의의 수준은 상당하다는 것을 확인할 수 있다. 아울러, 정책처방의 인과효과 추정에서 종종 사용된 PSM 역시 내생성의 문제를 완전히 해결하지는 못한다는 것을 함께 확인할 수 있었다.

반면, 현대 실증경제학에서 각광을 받고 있는 IV, 회귀단절, DID의 추정 결과는 모두 실제 모형을 잘 반영하여 정책의 인과효과가 없는 것으로 나타났다. 즉, 이 세 가지 추정방법을 활용할 경우에는 정책처방변수가 갖는 내생성을 통제하여 실제 모형과 같은 (또는 유사한) 추정 결과를 얻을 수 있다.

이상의 논의로부터 정책처방에 대한 내생성이 존재할 경우에 진정한 인과효과 추정을 위해서는 OLS 및 PSM이 아닌 IV, 회귀단절, DID 등을 활용하는 것이 바람직하다는 잠정적인 결론을 얻을 수 있다. 하지만 앞에서 언급했다시피, 이러한 세 가지 추정방법을 적용함에 있어서 주의를 기울여야 할 몇 가지 사항들이 있다. IV에서 요구되는 외생적인 도구변수, 회귀단절 분석을 위해 필요한 규칙변수(forcing variable), DID를 위한 지속관측 자료(longitudinal data) 등이 그것이다. 현실적으로 이러한 것들이 충족되지도 못하는 정책평가 환경이 조성되는 경우도 존재할 수 있다. 하지만 정책처방에 대한 내생성이 존재하는 경우에는 단순한 OLS 추정만으로는 내생성으로 인한 편의를 제거할 수 없으므로, 가능한 모든 수단을 동원하여 IV, 회귀단절, DID 중 한 가지 이상의 추정방법이 적용될 수 있는 환경을 조성함으로써, 내생성이 제거된 순수한 정책의 인과효과를 추정하려 노력해야 한다.

[부도 9-1] 시뮬레이션 데이터 생성을 위한 Stata 프로그램(Do 파일)

```

/***** Random Variables Generate (endogenous treatment) *****/
*** Define Covariance Matrix
mat P=(1,0,5,0,0,0,0,0 \ 0,1,0,5,0,0,0,0 \ .5,0,1,0,2,0,0,0 \ 0,.5,0,1,2,0,0,0 \ ///
0,0,2,2,1,-5,-8,0,0 \ 0,0,0,0,-5,1,0,0,0 \ 0,0,0,0,-8,0,1,0,0 \ 0,0,0,0,0,0,1,0 \ 0,0,0,0,0,0,0,1)
mat A=cholesky(P)

*** Define Data Size: # of observations
set obs 100000

*** Variables Generating Process
local k=rowsof(P)
forvalue i=1/^k' {
set seed 10`i'
gen c`i'=invnorm(uniform())
}

forvalue i=1/^k' {
gen v`i'=0

forvalue j=1/^k' {
replace v`i'=v`i'+A[`i',`j']*c`j'
}
}

drop c*

rename v1 x10
rename v2 x11
rename v3 x20
rename v4 x21
rename v5 x3
rename v6 z
rename v7 a
rename v8 e0
rename v9 e1

/**** Define Treatment Variable ****/
gen d0=0
gen d1=(x3>0)

/**** Construct Panel Data ****/
gen id=_n
gen g=(d1==1)
reshape long x1 x2 d e, i(id) j(t)

/**** Define Outcome Variable ****/
gen u = a + e
gen y = -6 + 1*x1 - 0.3*x2 + 0*d + u

/**** SAVE Dataset ****/
** Panel Data
saveold exdata_panel, replace
** Cross Sectional Data
keep if t==1
saveold exdata_cross, replace

```

```
/** Load Cross Sectional Data */
use exdata_cross, clear

/** OLS Regression */
regress y d
regress y d x1 x2

/** PSM(Propensity Score Matching) */
attnd y d x1 x2
attr y d x1 x2

/** IV(Instrumental Variable) Estimation */
ivregress 2sls y (d = z) x1 x2, first
ivreg2 y (d = z) x1 x2

/** RDD(Regression Discontinuity Design) */
gen x3squ=x3^2
gen x3cub=x3^3
*** parametric method
regress y d x3 x3squ x3cub if x3>-0.5 & x3<0.5
regress y d x3 x3squ x3cub if x3>-1 & x3<1
regress y d x3 x3squ x3cub if x3>-2 & x3<2
*** non-parametric method
rd y x3, gr mbw(100) z0(0) strineq ddens

/** Load Panel Data */
use exdata_panel, clear

/** DID(Difference-in-Differences) */
** OLS Estimator
regress y d t g
areg y d t, absorb(g)
** Within Estimator
tsset id t
xtreg y d t, fe
```



**【제 2 부】**

**산업별 고용효과 분석을 위한  
다부문 거시산업모형 DB 구축**



1990년대 후반 외환위기 이후 한국 경제는 5%대 이하의 경제성장률과 함께 경제성장의 고용유발효과 또한 지속적으로 낮아지는 추세를 보이고 있다. 특히, 몇 년 전부터 계속되고 있는 세계 금융위기의 여파로 경기가 위축되고 국내 노동시장 상황이 악화되면서 고용창출에 대한 사회적 관심이 높아지고 있다. 따라서 매년 대규모로 이루어지고 있는 정부의 재정을 통한 일자리 지원사업의 고용창출 효과에 대한 지속적이고 체계적인 분석의 필요성이 대두되고 있다.

그럼에도 불구하고, 그동안 재정을 통한 일자리 지원사업이 고용창출에 직·간접적으로 미치는 영향에 대한 체계적인 논의와 연구는 많지 않은 실정이다. 더욱이 정부의 다양한 일자리 대책들이 직접적인 고용창출 이외에 경기 활성화에 기여하여 파급시키는 간접적인 고용효과에 대한 분석을 수행할 수 있는 모형의 개발 및 구축은 상당히 미흡한 상태라 할 수 있다.

경제는 다양한 부문(또는 산업)으로 구성되어 있으며, 또한 각 부문들은 상호 연계되어 있다. 그렇기 때문에 외부적 환경이 변화하거나 새로운 정책이 시행됨에 따라서 발생하는 고용 및 경제적 성과 등의 변화 역시 부문(또는 산업)별로 상이할 것이다. 따라서 재정을 통한 일자리 지원사업을 시행함에 따라 발생하는 고용효과를 부문(또는 산업)별로 분석하고, 또한 부문 간 상호연계로 인해 유발되는 파급효과 분석을 위해서 경제 전체를 대상으로 하는 모형의 구축과 함께 다양한 부문·산업별 정책효과와 차이와 이들 간의 연관

관계를 고려하는 경제균형모형의 구축이 필요하다.

일반적으로 정부 정책이 경제에 미치는 영향을 분석함에 있어서 산업(또는 부문)간 연관효과를 고려하는 경제균형모형으로, 산업연관분석(Input-Output Analysis : IO), 연산가능 일반균형모형(Computable General Equilibrium: CGE), 다부문 거시산업모형 등이 많이 사용된다. 본 연구에서는 정부의 재정을 통한 일자리 지원사업의 고용창출효과를 분석하기 위해 경제 전체를 대상으로 하는 모형의 구축과 함께 경제를 구성하는 다양한 부문(또는 산업)별 정책효과의 차이 및 이들 간의 연관효과를 고려하는 다부문 거시산업모형 구축을 시도한다.<sup>9)</sup> 구체적으로는 기존에 다루어진 다부문 거시산업모형을 활용하여 노동시장 사정을 잘 반영할 수 있는 일자리창출 분석·평가 모형을 구축하고자 한다. 다시 말해서, 기존의 다부문 거시산업모형에 한국 노동시장의 중요한 특징들을 반영할 수 있는, 즉 노동부문이 확장된 다부문 거시산업모형을 개발하여 일자리창출 효과 등의 각종 정책효과를 평가할 수 있도록 유도하고자 한다.

---

9) 다부문 거시산업모형을 개발·구축하는 작업은 상당히 오랜 시간을 필요로 한다. 그렇기 때문에 다부문 거시산업모형 개발을 위한 DB구축 과정을 본 연구의 내용으로 한정하고, 구축된 DB를 활용하여 모형을 개발하고 분석을 수행하는 작업은 추후 연구과제로 남겨두기로 한다.



제1절 모형의 이론적 성격<sup>10)</sup>

경제 전체를 대상으로 균형을 가정하는 경제균형모형(Economic model)은 경제 주체들의 경제 행위가 상호 연관성을 유지하면서 반복적으로 순환하는 과정을 소득의 발생과 처분이라는 순환 과정으로 파악하는 거시경제균형모형(Macroeconomic equilibrium model)과 미시경제적 최적화에 기초하고, 기업 혹은 산업 간 상호관계에 의한 소득 순환까지를 포괄하는 산업부문과 거시경제를 연계하는 일반균형모형(General equilibrium model) 두 가지 측면에서 발전해 왔다. 이들 경제균형모형은 각 부문 간 상호관계를 연립방정식 체계에 입각하여 분석한다는 점에서는 동일하나, 거시경제균형모형은 거시경제이론에 기초하여 집계변수들 상호 간의 관계를 추정하고 예측하여 중단기적 경기 순환과 추세적 분석을 행하는 데 주로 이용되는 반면, 일반균형모형은 미시경제이론에 입각하여 정책효과가 장기적으로 경제구조에 미치는 영향을 분석하는 데 주로 활용된다. 일반균형모형은 여러 개의 산업을 포함할 뿐만 아니라 산업별 소비, 투자, 수출입 등 경제부문도 동시에 고려한다는 점에서 다부문모형(multisectoral model)이라고도 한다.

10) 본절의 내용은 이진면 외(2007)의 제2장 내용을 참고하여 정리한 것이다.

이러한 다부문모형은 균형해(均衡解)의 산출 방식과 파라미터의 추정 여부에 따라 전통적인 산업연관모형(Input-output model), 거시·산업모형(interindustry macroeconomic model)과 연산가능 일반균형모형(computable general equilibrium model)으로 구분된다. 일반균형모형은 초기에는 산업연관표의 기술계수에 입각하여 경제정책의 산업별 파급효과를 정태적인 측면에서 분석하는 것이 일반적이었으나, 미시경제적 최적화에 기초한 Warlas의 일반균형이론을 실증모형으로 구축하는 시도가 이루어지면서 산업 간 및 거시경제를 연계하는 모형으로 발전하고 있다. 거시·산업모형은 거시경제모형의 연계 방식에 따라 하향식 동태산업·거시모형과 상향식 동태산업·거시모형으로 구별된다. 연산가능 일반균형모형은 정태적 모형에서 동태적 요인을 부분적으로 고려하고 있는 동태 기간 간 일반균형모형으로 발전하고 있다.

이렇게 일반균형모형이 다양한 형태로 발전하고 있는 것은 거시경제적 소득 순환과 미시경제적 산업연관 관계는 서로 유기적이고 보완적인 관계를 갖고 있으므로 이러한 경제 현상의 구조적 효과와 순환적 효과를 동시에 파악하는 것이 바람직하며, 여기에 시간 변화를 고려하는 동태성을 부여할 경우 현실 경제에 대한 계량적 접근이 매우 완벽해질 수 있기 때문이다. 즉, 정책 변화에 따른 총량적 집계변수의 변화는 개별 경제 주체가 직면하는 제약조건에 영향을 미치고, 개별 경제 주체의 변화는 궁극적으로 집계적 행동에 영향을 미치게 되며, 또한 집계변수의 변화는 개별 경제 주체의 경제 행위의 범위를 규정하는 제약조건에 영향을 미친다. 이 경우 미시경제적 현상과 거시경제적 현상을 구분하기가 용이하지 않게 된다. 즉 정책변수나 경제적 충격이 경제구조(미시경제적 현상)와 경기순환(거시경제적 현상)에 미치는 영향은 서로 분리할 수 없으며, 이 두 영향을 동시에 고려하려는 경우 거시경제 계량모형이나 산업연관모형 중 어느 한쪽만 가지고는 불충분하다는 것이다. 더욱이 산업연관모형이나 연산가능 일반균형모형의 경우 대부분 시간 변화에 대해서 투입-산출 관계가 고정된다는 가정에 입각하고 있기 때문에 동태성을 고려하지 못한다는 단점도 가지고 있다.

다부문 거시산업모형은 Warlas의 일반균형이론에 입각한 동태적 산업연관

모형에 거시경제계량모형을 연계시킨 모형으로 거시경제적으로는 소득 순환의 동태성과 미시경제적으로는 산업 간 투입-산출 계수의 동태적 분석을 동시에 행할 수 있다. 투입-산출 계수의 내생화를 통해서 얻을 수 있는 분석상의 이점으로는 외부 충격에 대하여 중·장기적으로 산업연관효과가 파생되어 가는 과정을 분석할 수 있으며, 생산성의 변화에 따른 투입-산출 계수의 변화를 고려함으로써 기술변화에 따른 산업구조의 변화과정을 분석할 수 있다는 것이다. 또한 각종 산업정책에 대한 단순한 경제성장률, 국제수지 등의 거시경제적 분석을 포함하여 산업구조, 투입-산출 계수에 미치는 영향까지를 분석함으로써 정책 결정에 더 많은 정보를 제공할 수 있다.

## 제2절 모형의 체계와 하부 모형의 구조<sup>11)</sup>

### 1. 모형의 체계

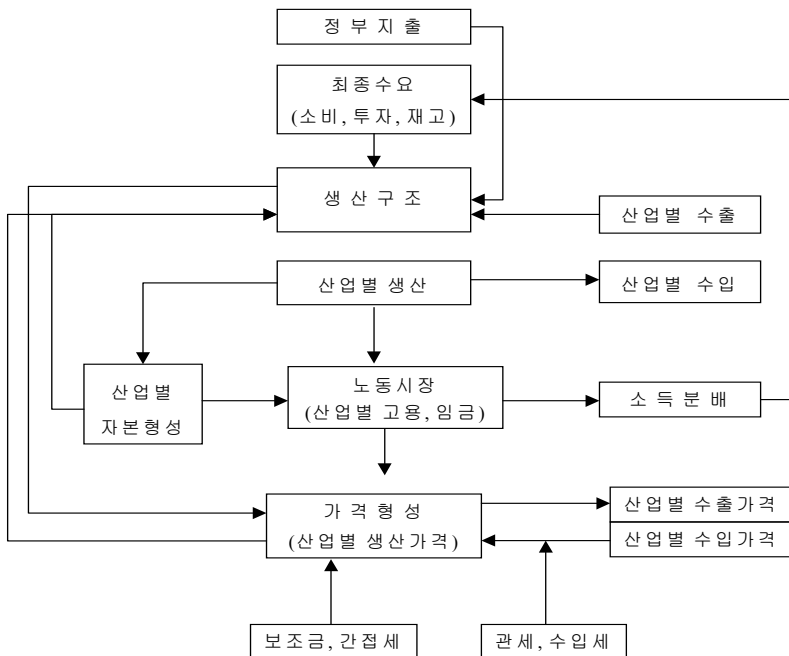
다부문 거시산업모형은 거시경제변수들과 미시적 산업변수들 간에 정합성을 확보하고, 국민경제의 상호 의존관계를 명시적으로 취급하는 일반균형적 계량경제모형이다. 거시경제의 총량 부문과 미시경제의 산업별 생산과 수요 간의 정합성을 보증하는 산업연관자료를 시계열화하여 개별 방정식을 추정하고, 경제이론을 바탕으로 경제부문별 인과관계를 설정한다. 그리고 국내외 환경변화와 개별 산업의 장기 전망에 관한 다양한 정보를 이용하여 산업별 생산성과 요소가격의 변동에 의한 산업경쟁력 추이를 분석하여 산업별 중장기 성장 전망이나 다양한 정책에 대한 효과, 특히 산업별 고용효과를 평가할 수 있다.

이러한 다부문 거시산업모형의 기본 구조는 [그림 11-1] 과 같이 나타낼 수

11) 본절의 내용은 이진면 외(2001)의 제2장과 이진면 외(2007)의 제3장을 참고하여 정리한 것이다.

있는데, 우선 외생적으로 주어진 정부지출 및 수출과 내생적으로 설명되는 소비 및 투자로부터 최종수요가 결정되고, 이와 함께 산업연관 분석의 산업 간 상호의존관계를 반영하여 산업별 생산이 결정된다. 다만, 이때 가격의 결정은 투입구조를 반영하는 비용구조에 의해 결정된다. 이러한 것들이 결정되면 분배국민소득이 산업별로 구해질 수 있으며, 산업별 분배국민소득으로부터 총소득을 다시 얻을 수 있다. 마지막으로 이 소득을 기초로 하여 최종수요의 크기가 결정되고 이것이 다시 생산에 파급되어 간다. 이와 같이 다부문 거시산업모형은 거시경제적인 측면에서 최종수요의 각 지출 항목에 대한 산업별 생산·고용·투자·가격 등의 움직임을 시간적으로 설명하는 데 중점을 둔다. 이러한 분석을 위해 다부문모형은 가격·기술모형, 수요·공급모형, 총량모형 등 3개의 하위 모형으로 구성되며, 이들 하위 모형 간의 연결을 통해서 정책적 시뮬레이션이 가능하도록 되어 있다.

[그림 11-1] 다부문 거시산업모형의 체계



## 2. 하부 모형의 구조

### 가. 가격·기술모형

가격·기술모형에서는 산업별 투입계수, 산업별 국내생산가격, 산업별 요소가격, 산업별 자본가격, 수입계수가 결정된다. 이들은 산업별 수입재가격, 산업별 감가상각률, 기준연도의 투입계수, 총생산성 증가율 등의 외생변수와 총량모형에서 결정되는 전 산업 평균임금, 이자율, 환율 및 수요·공급모형에서 결정되는 산업별 초과수요 등의 함수로 결정된다. 가격·기술모형에서 결정되는 산업별 요소가격, 총수요가격, 수입계수는 수요·공급모형에 설명변수로 투입된다.

우선 산업별 중간투입계수 및 자본·노동계수의 변화는 상대가격 변화에 의한 대체효과와 기술변화에 의한 효율성을 동시에 고려하여 추정된다.

$$\begin{aligned} \langle \text{중간투입계수} \rangle a_{ij} &= p_i^\gamma a_{ij}^0 \lambda_j & \langle \text{노동계수} \rangle l_j &= p_j^\gamma l_j^0 \lambda_j \\ & & \langle \text{자본계수} \rangle k_j &= p_j^\gamma k_j^0 \lambda_j \end{aligned}$$

- $a_{ij}^0, l_j^0, k_j^0$  : 초기년 산업별 중간투입계수, 노동계수, 자본계수
- $a_{ij}, l_j, k_j$  : 비교년 산업별 중간투입계수, 노동계수, 자본계수
- $p_i^r = (p_i/p_i^0)/(p/p^0)$  : 대체효과를 측정하는 산업별 상대가격 변화
- $p_i$  : 산업별 중간재가격,  $p$  : 전 산업 중간재 평균가격
- $\lambda_j = [a_{ij}^0 p_i^0 + l_j^0 w_j^0 + k_j^0 c_j^0] / [\bar{a}_{ij} p_i^0 + \bar{l}_j w_j^0 + \bar{k}_j c_j^0]$  : 효율성을 측정하는 산업별 총생산성지수
- $w_j$  : 산업별 임금,  $c_j$  : 산업별 고정자본 사용자 비용
- $\bar{a}_{ij}, \bar{l}_j, \bar{k}_j$  : 기준연도 중간투입계수, 노동계수, 자본계수

한편 산업별 국내생산가격의 결정은 풀 코스트(full cost) 원리에 기초하여 함수를 설정하여 추정한다.

〈산업별 국내생산가격〉  $p_i^x = f(p_i^z, p_i^m, \rho_i)$

- $p_j^z = \sum a_{ij}p_j + l_jw_j + k_jc_j$  : 산업별 생산비용
- $p_i = (1 - m_i)p_i^x + m_i p_i^m$  : 산업별 중간재가격
- $p_i^m$  : 산업별 수입재가격
- $\rho_i = (Y_j - Y_j^*)$  : 산업별 초과수요 혹은 국내총생산 겹
- $Y_j$  : 산업별 국내총생산,  $Y_j^*$  : 산업별 생산능력

그리고 산업별 임금은 전 산업 평균 임금을 기초로 각 산업의 상대노동생산성에 의해 조정되며, 산업별 고정자산의 사용자 비용(user cost)은 이자율, 산업별 감가상각률 및 고정자본 가격의 함수로 근사 추정한다.

〈산업별 임금〉  $w_i = f(w, \eta_i^r)$

〈산업별 고정자산 사용자 비용〉  $c_i = f(r + d_i - \dot{p}_j^k)$

- $\eta$  : 전 산업 평균 노동생산성
- $w$  : 전 산업 평균 임금(총량모형에서 결정)
- $\eta_i^r$  : 산업의 상대노동생산성( $\eta_i/\eta$ ),  $r$  : 이자율,  $d_i$  : 산업별 감가상각률
- $\dot{p}_j^k$  : 고정자산의 가격 상승에 의한 자본이득

산업별 총수요( $D_i$ )에서 수입( $M_i$ )이 차지하는 비율을 나타내는 수입계수는 생산물시장에서 국내가격과 수입가격의 상대가격의 함수로 추정한다.

〈산업별 수입계수〉  $m_i = f(p_i^m/p_i^x, \epsilon)$

- $p_i^x$  : 산업별 생산가격,  $p_i^m$  : 산업별 수입재가격,  $\epsilon$  : 환율(원/달러)로 단위 조정함

## 나. 수요 · 공급모형

수요 · 공급모형에서는 산업별 최종수요, 산업별 중간수요, 산업별 요소수요, 산업별 국내총생산, 산업별 총산출, 산업별 수입 및 초과수요가 결정된다. 즉, 산업별 정부소비지출, 산업별 재고 증감, 산업별 수출 등이 외생적으로 주어지고, 산업별 최종수요에서는 산업별 민간소비지출, 자본재별 투자수요가 내생적으로 결정되며, 산업별 요소수요에서는 산업별 고정자본스톡과 취업자가 결정된다.

산업별 민간소비지출은 비목별 민간소비지출에 산업별 · 비목별 민간소비지출 행렬을 이용하여 추정하며, 비목별 민간소비지출은 소비자의 소득 제약 조건하에서 효용극대화 행동에 입각하여 유도되는 수요함수체계(Demand System)에 의해서 결정된다.

$$\langle \text{산업별 민간소비지출} \rangle \quad C_i = \sum_{h=1}^n \theta_{ih} C_h$$

$$\langle \text{비목별 민간소비지출} \rangle \quad C_h = f(p_1^c, p_2^c, \dots, p_n^c, C)$$

$$\langle \text{비목별 소비재가격} \rangle \quad p_h^c = \sum_{i=1}^m \theta_{ih}^w p_i$$

- $C_h$  :  $h$  비목의 민간소비지출,  $C$  : 총 민간소비지출(총량모형에서 결정)
- $\theta_{ih}$  = 민간소비지출의 산업별( $i$ ) · 비목별( $h$ ) 변환 행렬
- $\theta_{ih}^w$  : 민간소비지출의 산업별( $i$ ) · 비목별( $h$ ) 가중치 행렬

자본재별 투자수요는 자본재 수요산업의 투자함수와 자본재별 · 수요산업별 투자지출 행렬을 이용하여 추정한다. 그리고 자본재 수요산업의 투자지출은 기업의 자금제약하에서 생산 극대화를 위한 자본투입 및 노동투입 간의 장기 최적화 행동에 입각한 투자수요함수에 의해 추정한다. 즉, 기업은 기대 산출량( $X_i^*$ )과 기대요소가격( $w_i^*, c_i^*$ )하에서 요소투입량( $K_i^*, L_i^*$ )을 결정하는데, 이때  $K_i^*$ 가 결정되면 기업은 전기의 자본스톡( $K_{i-1}$ )과의 차이만큼

신규투자를 투입할 것이며, 신규투자는 자금제약하에서  $K_{i,t}^*$  과  $K_{i,-1}$  에 의해 결정된다고 가정한다. 여기서 기대에 대한 통계가 용이하지 않으므로 기대변수는 자기상관모형(autoregressive model)을 이용한다. 그리고 자본재 수요산업의 자본스톡은 감가상각률과 투자지출에 의한 항등식으로 결정한다.

$$\langle \text{자본재별 투자수요} \rangle I_i = \sum_{j=1}^m \Phi_{ij} I_j$$

$$\langle \text{자본재 수요산업별 투자지출} \rangle I_j = f(c_j^*/w_j^*, X_j^*, K_{j,-1})$$

$$\langle \text{자본재 수요산업별 자본스톡} \rangle K_j = (1 - d_j) K_{j,-1} + I_j$$

- $I_i$  :  $i$  자본재의 투자수요,  $\Phi_{ij}$  = 자본재별( $i$ ) · 수요산업별( $j$ ) 변환행렬
- $c_j^*$  : 기대자본비용,  $w_j^*$  : 기대임금,  $X_j^*$  : 기대산출량,  $K_{j,-1}$  : 전기의 자본스톡

산업별 노동수요는 투자지출과 마찬가지로 기업의 장기적 최적화 행동의 구조에 입각하는데, 실제 노동투입량( $L_i$ )은 최적의 노동투입량으로부터 조정과정에 있다고 가정하여 조정식을 도입하여 산출량, 상대가격, 전기의 노동수요에 의해서 결정되는 함수를 추정한다.

$$\langle \text{산업별 최적 노동투입량} \rangle L_i^* = f(X_i, w_i/c_i)$$

$$\langle \text{실제와 최적 투입량 조정식} \rangle \Delta L_i^* = f(L_i^* - L_{i,-1})$$

$$\langle \text{실제 노동투입량} \rangle L_i = f(X_i, w_i/c_i, L_{i,-1})$$

산업별 최종수요는 산업별 민간소비지출, 자본재 투자수요와 외생적으로 결정되는 산업별 정부소비지출, 재고 변동 및 수출의 합으로 결정되며, 산업별 총산출은 산업연관표상의 수급균형식에 기초하여 최종수요에 레온티에프의 역행렬(Leontief inverse matrix)을 곱하여 산출, 즉 산업연관모형의 수급균형식을 이용하여 산업별 총산출을 구한다.



〈산업별 최종수요〉  $F_i = C_i + G_i + I_i + H_i + E_i$

〈산업별 총산출〉  $X_i = (II - A + \hat{M})^{-1} F_i$

- $G_i$ : 산업별 정부소비지출,  $H_i$ : 산업별 재고변동,  $E_i$ : 산업별 수출
- $II$ : 항등 행렬
- $A = [a_{ij}]$ : 중간재 투입계수 행렬,  $\hat{M}$ : 수입계수의 대각 행렬

산업별 중간수요는 산업별 총산출에 중간재 투입계수를 곱하여 산출하고, 산업별 수입은 산업별 총수요에 수입계수를 곱하여 구하며, 산업별 국내총생산은 산업별 최종수요에서 수입을 감하여 구한다.

〈산업별 중간수요〉  $U_i = \sum_{j=1}^m a_{ij} X_j$

〈산업별 수입〉  $M_i = m_i D_i$

- 산업별 총수요:  $D_i = U_i + F_i$ , 산업별 국내총생산:  $Y_i = F_i - M_i$
- 전 산업 국내총생산:  $Y = \sum_{i=1}^m Y_i$  (총량모형에서 설명변수로 투입)

산업별 생산능력은 노동투입량과 자본투입량 및 기술진보를 의미하는 총생산성의 변화를 고려하는 생산함수를 통해 추정하며, 여기서 산업별 생산능력과 실제 생산 간 갭(gap)이 수급조정 요인으로 가격형성, 고용, 자본형성에 영향을 미치는 것으로 고려한다.

〈산업별 생산능력〉  $Y_i^* / \lambda_i = f(L_i, K_i)$

- 산업별 초과수요 혹은 갭:  $Y_i - Y_i^*$

## 다. 총량모형

총량모형은 산업구조 변화와 거시경제변수 변화 간에 존재하는 상호 의존

관계 분석을 위한 것으로서 케인지언의 수요체계에 기초한다. 부연하면, 총량 모형은 소득, 통화, 이자율, 환율, 물가, 국제수지 등의 거시경제변수의 변화가 산업구조에 파급되는 과정과 산업구조의 변화가 소득, 실업, 물가 등의 거시경제변수에 미치는 영향을 분석하는 것이 목적이다. 이러한 총량모형에서 우선 국민총소득은 수요·공급모형에서 결정되는 국내총생산에 외생적으로 주어지는 국외순수취요소소득을 더하여 산출한다. 그리고 조세수입은 국민총소득과 내국세율을 설명변수로 추정하며, 국민가처분소득은 국민총소득에서 세금과 감가상각액을 차감하여 산출한다.

$$\langle \text{국민총소득} \rangle \quad Y^t = Y + Y^f$$

$$\langle \text{조세수입} \rangle \quad T = f(Y^t, t^d)$$

$$\langle \text{국민가처분소득} \rangle \quad Y^n = Y^t - T - K^c$$

■  $Y$  : 국내총생산(GDP),  $Y^f$  : 국외순수취요소소득,  $t^d$  : 내국세율

■  $K^c = \sum_{j=1}^m (1 - d_j) K_{j,-1}$  : 전 산업 감가상각액

민간소비지출총액은 항상소득가설(permanent income hypothesis)에 습관형성가설(habit formation hypothesis)을 포함하는 형태의 함수로 추정하며, 소비자물가의 상승률은 임금상승률, 통화공급량 증가 및 수입물가 증가율로 설명되는 함수로 추정한다.

$$\langle \text{민간소비지출총액} \rangle \quad C = f(Y^n, C_{-1})$$

$$\langle \text{소비자물가상승률} \rangle \quad \dot{p}^c = f(\dot{w}, \dot{M}^s, \dot{p}^m)$$

■  $C_{-1}$  : 전년도 민간소비지출총액

■  $\dot{w}$  : 전 산업 평균 임금상승률,  $\dot{p}^m$  : 수입물가변화율,  $\dot{M}^s$  : 통화공급변화율

전 산업 평균 임금의 상승률은 소비자물가상승률, 생산성증가율 및 실업률

로 설명되는 함수로, 이자율은 국내총생산 증가율과 통화공급 변화율의 함수로, 그리고 환율은 무역수지와 국내외 가격차 및 국내외 금리차로 설명하는 함수로 설정하여 추정한다.

$$\langle \text{임금상승률} \rangle \dot{w} = f(\dot{p}^c, \mu, \dot{\eta})$$

$$\langle \text{이자율} \rangle r = f(\dot{Y}, \dot{M}^s)$$

$$\langle \text{환율} \rangle \epsilon = f(B^T, p^w/p, r^w/r)$$

■  $\mu$  = 실업률,  $\dot{\eta}$  = 전 산업 노동생산성 변화율

■  $B^T = E - M$  : 무역수지,  $E = \sum_{i=1}^m E_i$  : 총수출,  $M = \sum_{i=1}^m M_i$  : 총수입

■  $p^w$  : 해외물가,  $r^w$  : 국제금리

## 다부문 거시산업모형 DB 구축

### 제1절 DB의 체계

다부문 거시산업모형은 앞장에서 살펴본 바와 같이 거시적 측면에서는 경제변수 간 인과관계, 미시적 측면에서는 산업별 수요와 공급의 균형조건에 의해서 경제 전체 및 산업별 균형을 유도하는 모형이다. 이러한 모형을 구축하기 위한 데이터베이스(data base)는 경제에 있어서 산업 및 소득의 순환과정을 통계로 정비한 산업연관표에 기초한다. 국민경제의 순환과정은 소득의 발생과 처분으로 이루어지는 흐름을 소득 순환과 기업 상호 간에 이루어지는 중간재의 매매 거래인 산업 간 순환으로 대별되는데, 산업연관표는 산업 간 순환을 주축으로 하여 소득 순환까지 포함한 통계표라고 볼 수 있다.<sup>12)</sup> 다부문 거시산업모형의 DB는 이러한 산업연관표를 수집 및 추정한 것으로서 농업, 제조업, 서비스업 등의 산업별 중간수요, 소비, 투자, 수출입, 생산, 임금, 고용 등의 경제부문별 행렬이 연도별로 정리된 통계의 집합으로 구성된다. 또한 경제이론에서 경제변수 간의 관계는 가격요인이 제거된 물량 기준(불변가격)으로 형성되어 있기 때문에 DB의 구축도 명목가격과 불변가격 기준의 통계가 동시에 요구된다.

12) 한국은행(2007)의 p.20에서 인용.

[그림 12-1] 다부문 거시산업모형의 DB체계



따라서 [그림 12-1]에서 보는 바와 같이 산업별 총산출, 부가가치, 중간수요, 중간투입, 최종수요(소비, 투자, 수출, 수입) 등의 경상가격, 불변가격 기준 시계열, 산업별 자본스톡, 노동량, 임금, 사용자 비용 등 부가가치의 세분화, 각 항목별 산업별 디플레이터, 투입·산출계수, 노동계수, 자본계수 등에 대한 통계를 한국은행의 기존 산업연관표, 국민계정 및 통계청과 관세청의 통계를 이용하여 정비 및 추정 과정을 거쳐 DB화하게 된다.

## 제2절 산업분류

### 1. 중요성

다부문 거시산업모형은 특성상 산업분류가 세분화될수록 많은 산업에 대한 다양한 정보 및 정책 실험을 수행할 수 있는 장점이 있는 반면, 통계자료의

양과 추정방정식의 수가 기하급수적으로 증대하여 모형을 구축하는 데 막대한 시간과 노력이 요구된다는 단점도 있다. 더욱이 산업분류가 세분화되어 모형을 구성하는 방정식의 수가 증가할수록 모형의 안정성은 낮아질 가능성이 높아 산업분류의 세분화와 모형의 안정성 사이에는 상충(trade off) 관계가 존재한다고 볼 수 있다. 따라서 구축을 위해서는 산업분류의 수준을 사전에 결정할 필요가 있으며, 이를 위해서는 관심의 대상이나 핵심적인 산업을 중심으로 세분화하는 방안을 사전에 검토할 필요가 있다. 즉, 중점적인 연구대상 산업군에 대해서는 비교적 세분화하고 이외의 산업군에 대해서는 대분류 수준으로 통합하는 것이 현실적이라고 판단된다.

하나의 대안으로는 통계 DB와 모형 구축을 별도의 산업분류로 이원화하는 방안도 고려해 볼 필요가 있다. 즉, 통계 DB의 구축은 가능한 세분화하고, 모형 내의 산업분류는 모형의 활용성과 국내 산업구조의 변화를 감안하여 유사 성격 혹은 관심 대상이 아닌 산업은 그룹으로 통합하는 것이다. 이러한 산업분류의 이원화가 가지는 장점은 산업부문과 거시경제 간 정합성이 보증하는 통계 DB의 구축에는 많은 시간과 비용이 초래되지만, 한 번 구축해 놓으면 다양한 산업분류로 모형 개발이 가능하며, 모형 개발 이외에도 다른 목적으로 활용이 가능하다는 것이다. 그리고 세부산업별 개별 모형의 구축이나 다양한 계수의 산출을 통해 본 모형에 의해 통합된 산업별 분석 결과를 세부산업별로 배분하는 축차적인 분석도 가능하다. 반면, 유지관리 측면에서는 많은 시간과 비용이 초래되는 단점은 있다.

## 2. 고려사항

### 가. 정합적 통계의 존재 여부

산업분류 시 우선적으로 고려해야 할 점은 산업부문과 거시경제부문 간의 정합성을 확보할 수 있는 기초통계가 어느 정도 이용 가능한가이다. 즉, 산업분류를 결정하는 핵심 요소는 기초통계의 활용 가능성으로서 기본적으로 한

국은행의 산업연관표, 국민계정 그리고 통계청 및 관세청 통계 간의 정합성을 확보할 수 있는 정도를 사전에 검토해야 할 필요가 있다. 구체적으로는 한국은행 통계의 경우 산업연관표의 산업별 중간투입, 최종수요 구조, 취업계수 등과 국민계정의 최종수요별 총량 및 전 산업 총산출, 부가가치 등을 고려해야 한다. 그리고 통계청 통계는 제조업의 생산·고용구조 및 유형 고정자산을 추정하기 위한 광업·제조업 통계, 산업별 자본스톡 추정을 위한 국부통계의 산업별 자산, 산업·비목별 행렬 추정을 위한 가계지출 통계, 산업별 노동통계의 정합성을 위한 경제활동인구 통계, 사업체 기초통계 등을 검토할 필요가 있다. 그리고 관세청 통계는 산업별 수출입의 시계열 추정을 위해 산업연관표의 수출입과 연계 가능성을 검토해야 한다.

기초통계를 검토함에 있어서 가장 중요시 고려해야 할 사항은 기초통계 간 표준분류 연계 및 정합성 확보의 가능성이다. 기초통계의 사용이 가능할지라도 산업분류와 품목분류 등 표준분류 간 연계성 및 정합성 확보가 가능해야 한다. 예를 들면, 통계청의 경제활동인구 통계는 장기 시계열의 이용이 가능하나 산업분류가 기간별로 다르고, 대분류에 한해서 이용이 가능하므로 산업세분화 시 이용에 한계가 있다. 또한 관세청의 수출입 통계는 세부품목별 이용이 가능하나 HS 코드분류를 산업분류와 연계 시 문제가 발생할 가능성이 존재한다.

## 나. 산업연관표의 이용 가능성

거시산업모형의 핵심 통계인 산업연관표는 산업별 부문별 정합성이 확보된 국내 유일의 통계다. 즉, 거시경제부문과 산업별 총산출, 부가가치, 중간수요 및 중간투입, 민간소비, 정부소비, 투자, 수출, 수입 등과 정합성이 유지된 공식적인 통계는 한국은행이 작성해 오고 있는 산업연관표가 유일하다. 그러나 불연속적인 시계열이라는 점과 더불어 연도별로 산업분류 및 각 부문의 개념이나 정의, 포괄 범위, 산출액의 추계방법 등을 다소 조정해 왔기 때문에 시계열 분석을 위한 통계로서는 한계가 있다. 더욱이 경제모형은 대부분 물량

변수 간의 인과관계를 중심으로 설정하기 때문에 경사가격 기준의 통계보다는 불변화된 변수를 이용하고 있다는 점에서 경사 산업연관표의 이용상 한계를 가지고 있다.

이와 같은 문제점을 해결하기 위해서는 1차적으로 한국은행의 접속불변 산업연관표를 이용하는 것이다. 이것은 시간 경과에 따른 산업구조의 변화가 가격의 변동에 의한 것인지, 물적 투입량의 변동에 의한 것인지를 구분하여 파악하기 위해 산업연관표의 실측표를 대상으로 부문 분류와 각 부문의 개념, 정의, 범위, 추계방법 등에 대한 기준을 조정하고, 기준연도 가격으로 불변화한 산업연관표다. 그러나 이러한 불변접속 산업연관표도 10년간 3개년을 작성하는 관계로 기준연도의 변화와 불연속 시계열이라는 문제를 안고 있다.

〈표 12-1〉 1980년 이후 산업연관표의 작성 현황

경사표			접속불변표		
대상연도	기본부문	표종류	접속연도	기본부문	기준연도
1980	396	실측표	1980-85-90	168	1990
1983	396	연장표	1985-90-95	396	1995
1985	402	실측표	1990-95-00	396	2000
1986	161	연장표	1995-00-05	350	2005
1987	161	연장표	2006	350	2005
1988	161	연장표	2007	350	2005
1990	405	실측표	2008	350	2005
1993	405	연장표	2009	350	2005
1995	402	실측표			
1998	405	연장표			
2000	404	실측표			
2003	404	실측표			
2005	403	실측표			
2006	403	연장표			
2007	403	연장표			
2008	403	연장표			
2009	403	연장표			

자료: 한국은행.



#### 다. 국민계정 및 기타 통계의 활용 가능성

한편 산업연관표의 불연속 시계열 문제는 산업분류에 있어서 제한적이기는 하나 국민계정으로 보완이 가능하다. 즉, 산업연관표의 불연속적인 시계열이라는 단점을 국민계정의 부문별 통계 및 보간법(RAS)으로 보완이 가능한데, 국민계정 통계는 현재 2005년 기준으로 부가가치와 총산출이 경상과 불변가격으로 81개 산업분류로는 1995~2010년 기간, 27개 산업분류로는 1970~2010년까지 연속적인 시계열로 발표하고 있다. 그러나 이러한 국민계정 통계도 최종수요 부문이 산업별로 세분화되어 있지 않고, 소비는 비목별로, 투자는 16개 산업분류 그리고 수출과 수입은 산업별 구분되지 않는 문제가 존재한다.

따라서 산업분류는 국민계정으로부터 얻을 수 있는 통계는 비목별 민간소비, 산업별 투자수요, 산업별 총산출 및 부가가치로 한정할 수밖에 없으므로 통계청 및 관세청 통계는 거시경제 혹은 산업의 일부 항목에 대해 보완적으로 활용해야 한다. 즉, 산업별 노동통계의 보완은 사업체 기초통계, 경제활동인구 통계, 광업·제조업 통계, 서비스업 조사통계 등을 이용해야 하며, 산업별 자본스톡의 추정은 국부통계, 광업·제조업 통계 등을 통해 보완이 가능하며, 산업별 소비통계는 가계지출통계, 산업별 수출입은 관세청의 품목별 수출입통계를 이용하여 보완할 필요가 있다.

〈표 12-2〉 국민계정의 통계제공 산업분류

81개 기본분류		27개 산업분류		16개 산업분류	
1	작물	1	농림어업	1	농림어업
2	축산물				
3	임산물				
4	수산물				
5	농림어업서비스업				
6	석탄	2	광업	2	광업
7	원유천연가스				
8	금속광석				
9	비금속광석				

〈표 12-2〉의 계속

81개 기본분류		27개 산업분류		16개 산업분류	
10	식료품	3	음식료품담배	3	제조업
11	음료품				
12	담배				
13	섬유사직물	4	섬유가죽제품		
14	의복섬유제품				
15	가죽제품				
16	목재목제품	5	목재종이인쇄복제업		
17	펄프종이제품				
18	인쇄복제				
19	석유석탄제품	6	석유석탄화학제품		
20	기초화학제품				
21	합성수지합성고무				
22	화학섬유				
23	비료농약				
24	의약품화장품				
25	기타화학제품				
26	고무플라스틱제품				
27	유리유리제품	7	비금속광물제품		
28	도자기점토제품				
29	기타비금속광물제품				
30	철강	8	금속제품		
31	비철금속제품				
32	금속제품				
33	일반산업기계	9	일반기계		
34	특수산업기계				
35	전기기계장치	10	전기전자기기		
36	반도체전자부품				
37	영상음향통신기기				
38	컴퓨터사무기기				
39	가정용전기기기				
40	정밀기기	11	정밀기기		
41	자동차	12	운송장비		
42	선박				
43	기타수송장비				
44	가구기타제품	13	가구기타제품		
45	전력	14	전력가스수도사업	4	전력가스수도사업
46	가스증기온수공급업				
47	수도				
48	건설	15	건설	5	건설업
49	도소매업	16	도소매업	6	도소매음식 숙박업
50	음식점숙박업	17	음식점숙박업		
51	운수보관	18	운수보관	7	운수보관업
52	금융보험	19	금융보험업	8	금융보험업

〈표 12-2〉의 계속

81개 기본분류		27개 산업분류		16개 산업분류	
53	주거서비스	20	부동산임대업	9	부동산임대업
54	부동산서비스				
55	기계장비용품임대				
56	출판서비스	21	정보통신업	10	정보통신업
57	영화				
58	방송				
59	통신				
60	정보서비스	22	사업서비스	11	사업서비스
61	광고				
62	연구기관				
63	사업관련전문서비스				
64	사업지원서비스	23	공공행정국방	12	공공행정국방
65	공공행정국방				
66	교육〈국공립〉	24	교육서비스업	13	교육서비스업
67	교육〈비영리〉				
68	교육〈산업〉				
69	의료보건〈국공립〉	25	보건사회복지서비스업	14	보건사회복지서비스업
70	의료보건〈비영리〉				
71	의료보건〈산업〉				
72	사회복지사업〈국공립〉				
73	사회복지사업〈비영리〉	26	문화오락서비스업	15	문화오락서비스업
74	문화서비스				
75	오락서비스	27	기타서비스업	16	기타서비스업
76	위생서비스〈국공립〉				
77	위생서비스〈산업〉				
78	대개인서비스				
79	수리서비스				
80	사회단체				
81	가사서비스				

자료: 한국은행.

### 3. 한국노동연구원 모형의 분류

한국노동연구원의 다부문 거시산업모형은 기본적으로 정부의 재정지출을 통한 일자리 지원사업의 고용창출 효과를 분석하기 위한 것이므로 모형 개발 및 DB 구축을 위한 산업분류도 이와 관련하여 정부 지원사업의 산업적 범위와 이에 의해서 직·간접적으로 영향을 받는 산업, 즉 평가대상의 산업적 범위를 기본적으로 고려해야 한다. 그리고 추가적으로 산업별 정합성 및 일관성이 확보된 고용통계의 이용 가능성도 동시에 감안해야 한다. 우선 정부재

정에 의한 일자리 지원사업은 세분류된 산업 범위에서 특정 산업을 대상으로 추진되기보다는 중앙정부의 부처별 혹은 지방정부에 의해서 포괄적인 산업 범위를 위주로 추진되고 있기 때문에 국민계정의 16개 산업분류 혹은 27개 산업분류의 범주에서 분류가 가능하다. 그리고 사업 추진에 따른 평가대상의 산업 또한 아주 세분화된 산업을 대상으로 평가할 필요성은 그리 크지 않다. 다만, 산업별 고용유발계수와 같은 평가지표 등에서 현저한 차이를 보일 가능성이 높은 산업은 좀더 세분화하여 평가의 객관성과 정밀성을 확보할 필요는 있을 것이다. 이러한 점을 고려할 때 제조업을 하나의 산업으로 분류한 16개 산업분류보다는 27개 산업분류 또는 81개 산업분류가 더 적합하다고 볼 수 있다. 그러나 산업을 세분화할수록 산업별 정합성 및 일관성이 확보된 고용통계의 이용 가능성은 낮아진다.

현재 우리나라의 고용통계 가운데 전 산업을 포괄하면서 정합성과 일관성이 유지되고 있는 것은 가구를 대상으로 한 경제활동인구조사의 산업별 취업자, 기업이나 사업체를 대상으로 하는 산업연관표의 부속표로 제공되는 고용표의 취업자 및 피용자, 사업체를 대상으로 하는 전국사업체조사의 종사자 3종만이 가능하다. 이 중 경제활동인구의 산업별 취업자는 표준산업분류 중분류(약 75개 산업)에서 통계 작성이 이루어지고 있으나, 그동안 표준산업분류의 개정, 산업분류의 조정 및 통계조사 기준의 변경 등으로 인해 20년 정도의 장기 시계열로 일관성 있는 산업별 취업자 통계의 입수는 대분류 혹은 중분류를 다소 통합한 수준에서만 가능하다. 산업연관표의 고용표는 168개 산업분류에서 취업자와 피용자의 통계 이용이 가능하나 불연속이라는 단점이 있고, 전국사업체조사의 종사자는 세세분류까지의 통계 이용이 가능하나, 시계열이 1993년 이후부터나 이용할 수 있으며, 농림어업 부문이 사업체만을 대상으로 하기 때문에 일부분만을 포함하고 있다는 단점이 있다.

이상과 같은 정책적 측면과 고용통계의 이용 가능성, 그리고 앞에서 언급한 국민계정 및 산업연관표와 고용통계 간 정합성 등을 고려할 경우 한국노동연구원의 산업분류는 다음 표에서 제시한 국민계정의 27개 대분류 기준이 비교적 적절한 것으로 판단된다.

〈표 12-3〉 한국노동연구원의 산업분류

다부문모형 27부문		산업연관표(2009년 기준) 168부문 코드	한국표준산업분류 9차 개정코드
1	농림어업	001-011	01-03
2	광업	012-017	05-08
3	음식료품담배	018-033	10-12
4	섬유가죽	034-045	13-15
5	목재종이인쇄	046-051	16-18
6	석유석탄화학	052-071	19-22
7	비금속광물	072-077	23
8	금속제품	078-089	24-25
9	일반기계	090-097	29
10	전기전자	098-107	26, 28
11	정밀기기	108-110	27
12	운송장비	111-117	30-31
13	가구기타제조업	118-120	32-33
14	전기가스수도	121-124	35-36
15	건설업	125-130	41-42
16	도소매업	131	45-47
17	음식숙박업	132-133	55-56
18	운수보관업	134-141	49-52
19	금융보험업	144-146	64-66
20	부동산임대업	147	68-69
21	정보통신업	142-143, 160	58-63
22	사업서비스	148-154	70-75
23	공공행정국방	155	84
24	교육서비스업	156	85
25	보건사회복지	157-159	86-87, 37-39
26	문화오락	161-162	90-91
27	기타서비스	163-168	94-99

## 제3절 산업별 중간수요 · 투입 및 최종수요 · 공급 DB

### 1. 산업연관표와 국민계정의 연계

DB 구축의 기본 원칙은 산업 전체 혹은 국민경제 전체의 총량변수(혹은 집계변수, aggregated variables)는 국민계정의 통계값을 사용하며, 산업별 구

성비는 산업연관표를 통해 산출하여 국민계정과 산업연관표를 연계하였다. 현재 국민계정은 2005년 기준으로 작성 원칙과 평가 기준을 통일하여 1970년부터 최근까지 통계를 제공하지만, 산업연관표는 작성 연도의 산업구조와 기술구조를 보다 정확하게 파악하기 위하여 매년 산업분류와 작성 원칙을 조금씩 개편해 왔으며, 이러한 비일관성을 조정하기 위하여 기존 산업연관표를 10년 기간의 3개 시점에 대해서 접속하는 불변접속산업연관표를 재작성하여 공표하고 있다. 따라서 산업연관표의 통계는 값 자체보다는 산업별 구성비나 노동계수, 부가가치를 등과 같이 산업구조를 나타내는 지표에 유의적이라고 볼 수 있다. 이러한 점을 감안하여 국민계정의 27개 산업으로 산업연관표의 연도별 실측표와 불변접속산업연관표의 산업분류를 연계하고 총량은 국민계정, 산업별 구성이나 계수는 산업연관표에서 산출된 값을 이용함으로써 통계적 정합성을 확보하였다. 즉, 총산출, 부가가치, 수출, 수입 등 부문별 총량은 국민계정의 값을 이용하며, 개별 산업은 산업연관표의 산업별 비중으로 할당하였다.

한편 국민계정의 통계값과 산업연관표의 산업별 구성비만을 이용하여 양자의 통계를 연계할 수 없는 ‘산업별 중간투입과 중간수요의 관계인 중간투입행렬’, ‘민간소비의 산업별·비목별 전환행렬’, ‘투자재의 산업별 수요와 공급을 나타내는 고정자본형성행렬’은 양비례법인 RAS 방식을 이용해 추정하였다.

우선 중간투입행렬의 경우 산업연관표가 발표되지 않는 연도의 산업별 중간수요 합계는 산업연관표가 존재하는 연도 간의 비중 변화율을 기준으로 보정하며, 산업별 중간투입 합계는 산업연관표가 존재하는 연도의 중간투입계수를 가중 평균하여 초기값을 산출한 다음, 국민계정의 산업별 중간투입 합계와 상기에서 추정한 산업별 중간수요 합계가 일치하도록 RAS방법을 통해서 조정하는 과정을 거쳤다.

민간소비의 산업별·비목별 전환행렬은 국민계정의 비목별 통계를 산업연관표의 산업별 통계와 연계한 후 RAS방법으로 정합성을 확보하였다. 즉, 국민계정에서는 가계와 비민간 영리단체의 목적별 최종소비지출의 비목분류 기준으로, 산업연관표에서는 산업분류 기준이 작성되고 있어 양자 간에 정합성

을 유지하면서 상호 전환이 가능하도록 조정한 것이다. 구체적으로는 1단계로 국민계정에서 12대 비목별로 작성된 가계소비액과 비민간 영리단체의 최종소비지출액을 11대 비목별로 통합하였다.<sup>13)</sup> 2단계로 국민계정 기준의 11대 비목별에 대한 세부 품목을 통계청의 가계소비지출 세부항목을 통해 파악하였다. 3단계로는 이를 산업연관표의 27개 산업과 매칭하여 국민계정의 11대 비목이 산업연관표의 27개 산업으로 배분되는 1차 추정작업을 수행하였다. 4단계로 양자의 비중이 일치하도록 RAS방법을 적용하여 민간소비의 비목별·산업별 전환행렬을 추정하였다.

한편 산업별 투자는 고정자본 형성행렬을 통해 수요산업과 공급산업을 연계하였다. 이러한 이유는 산업별 투자(고정자본형성)는 국민계정의 경우 13개 분류<sup>14)</sup> 산업별 투자수요 개념으로, 산업연관표에서는 산업별 자본재의 공급 개념으로 작성하고 있으나 경제이론 및 모형에서는 투자수요함수가 중심이 되고 있기 때문이다. 양자 간의 정합성을 유지하면서 상호 전환이 가능하도록 산업연관표의 부속표로 제공되는 고정자본 형성행렬을 이용하여 전환행렬(27×13)을 작성하여 국민계정의 13개 산업별 투자수요와 연계하는 것이다. 고정자본 형성행렬이 작성되지 않는 연도는 고정자본 형성행렬의 비중을 보간법으로 추정하여 전환행렬을 추정하였다.

이상과 같은 산업연관표와 국민계정의 연계에 의한 DB 구축은 경제가 전반적으로 안정화되기 시작한 1990년에서 가장 최근의 산업연관표가 발표된 2009년까지 20년 기간을 대상으로 하였으며, 연계 과정에서 활용한 RAS 방법은 다음 절에서 상세히 기술할 것이다.

13) 국민계정의 소비지출액은 1. 식료품 및 비주류음료품, 2. 주류 및 담배, 3. 의류 및 신발, 4. 임료 및 수도광열, 5. 가계시설 및 운영, 6. 의료 보건, 7. 교통, 8. 통신, 9. 오락문화, 10. 교육, 11. 음식 숙박, 12. 기타 등 12개 비목으로 분류하고 있으나, 본 연구에서는 통상 음식료품으로 분류될 수 있는 1. 식료품 및 비주류음료품과 2. 주류 및 담배를 통합하여 11개 비목으로 분류하였다.

14) 1. 농림어업, 2. 광업, 3. 제조업, 4. 전기가스수도업, 5. 건설업, 6. 도소매음식숙박, 7. 운수통신, 8. 금융보험업, 9. 사업서비스, 10. 공공행정국방, 11. 교육서비스, 12. 의료서비스, 13. 기타서비스

〈표 12-4〉 소비지출항목의 연계

	국민계정의 소비지출 11개 항목										
	식료품 및 담배	의류 및 신발	임료 및 수도광열	가계시설 및 가구·조명·장 식서비스	의료·보건 사회복지	교통	통신	오락문화	교육	음식 숙박	기타
통계청 가계소비지출 세부항목	곡물·가공 곡물·가공 빵·떡류 육류·가공 품 신선농산물 염건농산물 가공농산물 유제품 과일·과일 채소·채소 가공품 해조·해조 가공품 당류·과자 류 조미식품 기타식품 커피·차 주스·기타 음료 주류 담배	직물·외의 내의 기타의복 의복관련서 비스 신발 신발서비스	실제주거비 주택유지·수 선 상하수도·폐 기물처리 기타주거관련 서비스 연료비	가구·조명 실내장식 가구·조명·장 식서비스 가정용섬유 가전·가정용 기기 가전관련서비 스 가사용품 가정용공구· 기타 가사소모품 가사서비스	의약품 의료용소모품 보건·의료용품 외래의료서비 스 치과서비스 기타의료서비 스 입원서비스 사회복지	자동차구입 기타운송기구 구입 운송기구유자 수리 운송기구연료 비 기타개인교통 서비스 철도운송 육상운송 기타운송 기타교통관련 서비스	우편서비스 통신장비 통신서비스	영상음향기기 사진광학장비 정보처리장치 기록매체 영상음향·정보 기기수리 오락문화내구 재 악기기구 오락문화내구 재유지·수리 장난감·취미용 품 캠핑·운동관련 용품 화훼관련용품 애완동물관련 물품 화훼·애완동물 서비스 운동·오락서비 스 문화서비스 복권 서적 기타인쇄물 문구 단체여행비	정규교육 초·중·고 고·중·고 학원·보습교육 학생학원 교육 성인학원 교육 기타교육	식사비 숙박비	이미용서비스 이미용기기 위생·이미용용 품 시계·장신구 기타개인용품 보험 기타금융 기타서비스

자료: 한국은행, 통계청.



〈표 12-5〉 민간소비의 비목별 · 산업별 전환행렬 추정 결과(2005)

27개 산업	산업별 비중	비목별 비중										
		식료품 및 담배	의류 및 신발	임료 및 수도광열	가계시설 및 운영	의료보건 사회복지	교통	통신	오락문화	교육	음식 숙박	기타
1 농림어업	0.0285	0.0285	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2 광업	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3 음식료품담배	0.0901	0.0901	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4 섬유가죽	0.0403	0.0000	0.0403	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5 목재종이인쇄	0.0016	0.0000	0.0000	0.0000	-0.0005	0.0000	0.0000	0.0000	0.0021	0.0000	0.0000	0.0000
6 석유석탄화학	0.0453	0.0000	0.0000	0.0280	0.0000	0.0138	0.0000	0.0000	0.0035	0.0000	0.0000	0.0000
7 비금속광물	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0000	0.0000	0.0000
8 금속제품	0.0018	0.0000	0.0000	0.0000	0.0018	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9 일반기계	0.0030	0.0000	0.0000	0.0000	0.0030	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10 전기전자	0.0322	0.0000	0.0000	0.0000	0.0091	0.0000	0.0000	0.0115	0.0116	0.0000	0.0000	0.0000
11 정밀기기	0.0041	0.0000	0.0000	0.0000	0.0000	0.0035	0.0000	0.0000	0.0000	0.0000	0.0000	0.0006
12 운송장비	0.0251	0.0000	0.0000	0.0000	0.0000	0.0000	0.0251	0.0000	0.0000	0.0000	0.0000	0.0000
13 가구기타제조업	0.0089	0.0000	0.0000	0.0000	0.0048	0.0000	0.0000	0.0000	0.0022	0.0000	0.0000	0.0019
14 전기가스수도	0.0262	0.0000	0.0000	0.0262	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15 건설업	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
16 도소매업	0.0814	0.0105	0.0036	0.0167	0.0016	0.0066	0.0057	0.0050	0.0060	0.0065	0.0072	0.0120
17 음식숙박업	0.0811	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0811	0.0000
18 운수보관업	0.0395	0.0000	0.0000	0.0000	0.0000	0.0000	0.0395	0.0000	0.0000	0.0000	0.0000	0.0000
19 금융보험업	0.0711	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0711
20 부동산임대업	0.1321	0.0000	0.0000	0.1321	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
21 정보통신업	0.0523	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0452	0.0070	0.0000	0.0000	0.0000
22 사업서비스	0.0113	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0113
23 공공행정국방	0.0012	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0012
24 교육서비스업	0.0739	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0739	0.0000	0.0000
25 보건사회복지	0.0595	0.0000	0.0000	0.0018	0.0000	0.0577	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
26 문화오락	0.0399	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0399	0.0000	0.0000	0.0000
27 기타서비스	0.0490	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0490
중간투입 계	1.0000	0.1291	0.0438	0.2047	0.0196	0.0816	0.0703	0.0618	0.0733	0.0804	0.0883	0.1470
국민계정 비중	1.0000	0.1499	0.0527	0.1760	0.0346	0.0571	0.1101	0.0497	0.0792	0.0738	0.0814	0.1355

〈표 12-6〉 고정자본 형성행렬의 추정 결과(2005)

27개 산업	산업별 비중	비목별 비중															
		농림 어업	광업	제조업	전기 가스 수도	건설업	도소매 음식숙 박	운수 보관업	금융보 험업	부동산 임대업	정보통 신업	사업서 비스	공공행 정국방	교육서 비스업	보건 사회복 지서비 스업	문화오 락서비 스업	기타서 비스업
1 농림어업	0.0023	0.0013	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0001	0.0000	0.0002	0.0000
2 광업	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3 음식료품담배	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4 섬유가죽	0.0009	0.0002	0.0000	0.0003	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0001	0.0000
5 목재종이인쇄	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6 석유석탄화학	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7 비금속광물	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8 금속제품	0.0113	0.0002	0.0000	0.0078	0.0005	0.0003	0.0002	0.0006	0.0000	0.0000	0.0002	0.0001	0.0004	0.0001	0.0002	0.0006	0.0000
9 일반기계	0.2291	0.0091	0.0002	0.1879	0.0062	0.0050	0.0028	0.0015	0.0004	0.0012	0.0005	0.0046	0.0031	0.0016	0.0010	0.0031	0.0009
10 전기전자	0.1049	0.0010	0.0001	0.0368	0.0050	0.0005	0.0065	0.0011	0.0039	0.0006	0.0215	0.0067	0.0072	0.0063	0.0018	0.0050	0.0011
11 정밀기기	0.0471	0.0002	0.0000	0.0211	0.0017	0.0002	0.0006	0.0005	0.0001	0.0001	0.0021	0.0079	0.0026	0.0011	0.0083	0.0007	0.0001
12 운송장비	0.1115	0.0037	0.0002	0.0209	0.0007	0.0041	0.0073	0.0458	0.0006	0.0014	0.0026	0.0095	0.0048	0.0033	0.0015	0.0042	0.0008
13 가구기타제조업	0.0149	0.0001	0.0000	0.0036	0.0001	0.0002	0.0019	0.0007	0.0007	0.0002	0.0004	0.0015	0.0016	0.0017	0.0005	0.0013	0.0004
14 전기가스수도	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15 건설업	0.3389	0.0024	0.0001	0.0566	0.0033	0.0037	0.0736	0.0089	0.0036	0.0387	0.0089	0.0169	0.0775	0.0187	0.0094	0.0084	0.0081
16 도소매업	0.0380	0.0018	0.0000	0.0167	0.0018	0.0005	0.0022	0.0014	0.0009	0.0003	0.0031	0.0032	0.0009	0.0015	0.0020	0.0014	0.0004
17 음식숙박업	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
18 운수보관업	0.0020	0.0001	0.0000	0.0009	0.0000	0.0000	0.0001	0.0002	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000
19 금융보험업	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20 부동산임대업	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
21 정보통신업	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
22 사업서비스	0.0992	0.0001	0.0005	0.0306	0.0028	0.0024	0.0088	0.0013	0.0131	0.0018	0.0086	0.0102	0.0019	0.0098	0.0035	0.0032	0.0005
23 공공행정국방	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
24 교육서비스업	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
25 보건사회복지	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
26 문화오락	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
27 기타서비스	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
중간투입 계	1.0000	0.0201	0.0012	0.3833	0.0221	0.0169	0.1040	0.0621	0.0234	0.0444	0.0481	0.0607	0.1007	0.0444	0.0283	0.0281	0.0122
국민계정 비중	1.0000	0.0168	0.0008	0.2612	0.0414	0.0157	0.0407	0.0558	0.0225	0.2054	0.0596	0.0253	0.1494	0.0435	0.0180	0.0167	0.0271

## 2. 시계열화

### 가. 불연속 계열의 보정방법

다부문 거시산업모형의 DB 구축은 앞에서 기술한 바와 같이 산업연관표와 국민계정 통계가 근간이 되는데, 산업연관표가 불연속 시계열로 작성되는 관계로 인하여 이를 연속적 시계열로 보정하는 것이 매우 중요하다. 많은 경우 양비례법의 일종인 RAS에 의해서 보정하지만, 불연속 시계열의 유형에 따라 다른 방법을 사용하는 경우도 발생하므로 본절에서는 불연속 시계열의 유형과 이에 대응하는 보정방법에 대해서 기술하고자 한다.

불연속 시계열은 크게 독립적인 단일변수, 종속적인 변수집단, 행렬변수집단의 세 가지 유형으로 대별할 수 있다.

〈표 12-7〉 불연속 시계열의 유형

독립적 단일변수		T	T+1	T+2	T+3	T+4	T+5	T+6
	K 변수	240			242			280
	S 변수	200			202			240
	W 변수	170			178			200

종속적인 변수집단		T	T+1	T+2	T+3	T+4	T+5	T+6
	산업 1	240			242			280
	산업 2	200			202			240
	산업 3	170			178			200
	전 체	610	615	618	622	660	695	720

〈행렬변수집단〉

T년	산업1	산업2	산업3	전 체	T+1	산업1	산업2	산업3	전 체	T+2	산업1	산업2	산업3	전 체
산업1	80	120	40	240	산업1				242	산업1	87	155	38	280
산업2	120	60	20	200	산업2				202	산업2	138	82	20	240
산업3	80	30	60	170	산업3				178	산업3	95	43	62	200
전 체	280	210	120	610	전 체	286	214	122	622	전 체	320	280	120	720

우선 독립적인 단일변수의 불연속은 자본스톡과 같이 10년을 주기로 조사하는 관계로 인하여 국가 전체의 총량적 자본스톡이나 개별 산업별 자본스톡이 독립적으로 연속적 시계열을 확보할 수 없는 경우다. 종속적인 변수집단은 집계변수(총량변수) 연속계열이 존재하나 이를 구성하는 변수들이 불연속인 경우다. 예를 들면 세부산업별 취업자의 변화를 파악하려 할 경우 전 산업의 취업자 수는 경제활동인구를 통해서 파악이 가능하나 세부산업별 취업자는 산업연관표의 고용표에서 불연속적이므로 중간 연도에 대한 현황을 파악하기 어렵다. 마지막으로 행렬변수집단의 형태는 행과 열의 총량변수는 연속이나 인자(element)들이 불연속인 경우로서 앞에서 언급한 산업연관표, 고정자본 형성행렬, 민간소비의 산업별·비목별 전환행렬 등이 이에 속한다.

이러한 불연속 계열을 보정하는 방법은 해당 변수의 추세선이나 다항식을 이용하는 방법 또는 밀접하게 관련된 변수를 설명변수로 하여 회귀분석을 이용하는 방법으로 구별되는데, 불연속의 유형에 따라 적합한 방법을 사용할 필요가 있다. 우선 독립적 단일변수의 불연속 시계열을 보정하는 방법으로는 변수의 추세에 따라 선형, 2차 혹은 3차 등의 추세식 혹은 다항식이나 회귀방법을 이용할 수 있는데, 이때 수준, 증가율, 증분 등의 추세선을 목적할 필요가 있으며, 회귀방법의 경우에는 설명변수의 선택에 매우 유의할 필요가 있다. 종속적 집단변수의 경우에는 개별변수는 독립적 단일변수의 방법과 동일하나 수준, 증가율, 증분 이외에 비중의 추세도 고려할 수 있는데, 이때에는 개별변수의 보정값이 총량변수와 일치해야 하는 정합성의 확보가 필요충분조건이라는 점을 유의해야 한다.

한편 행렬변수의 불연속 시계열을 보정하는 방법으로는 한계투입계수법, 라그랑지 미정계수법, 평균 증가배율법, 2차형 계획법, 선형계획법, RAS 추정법 등이 이용되고 있다. 여기서는 이들에 대해서 간단하게 설명하고, 이 중 가장 활용도가 높은 RAS 추정법에 대해서는 다음 절에서 자세하게 기술하고자 한다. 우선 한계투입계수법(Marginal input coefficient method)은 다음의 수식에서처럼 두 시점과 변화분을 단순히 연장하는 방법이다.

-  $t$  연도 평균 투입계수:  $A_t = [a_{ij}(t)] = \frac{X_{ij}(t)}{X_j(t)}$

$X_{ij}(t) = t$  연도에 있어서  $i$  산업  $\rightarrow j$  산업 중간재 투입

$X_j(t) = t$  연도에 있어서  $j$  산업 총산출

-  $r$  기간 동안 한계투입계수 :

$$\hat{a}_{ij}(t-r) = \frac{X_{ij}(t) - X_{ij}(t-r)}{X_j(t) - X_j(t-r)} = \frac{\Delta X_{ij}}{\Delta X_j}$$

-  $t+r$  연도 중간투입 :

$$X_{ij}(t+r) = X_{ij}(t) + \Delta X_{ij} = X_{ij}(t) + \hat{a}_{ij}(t-r) \Delta X_j$$

라그랑지 미정계수법(Lagrangian multiplier method)은 기준연도와 예측연도의 투입계수 간 편차 자승합을 최소화하는 방법이다. 기준연도의 실측된 투입계수, 예측연도의 중간수요 합계, 중간투입 합계 및 예측연도 생산액이 주어졌을 때 이를 제약조건으로 하여 기준연도 투입계수와 예측연도 투입계수와의 편차자승합이 최소가 되도록 예측연도 투입계수를 구하는 방법이다.

- 제약조건 :  $\sum_j^m X_{ij}^* = U_i$  (중간수요 합계 조건)

$$\sum_i^{m-1} X_{ij}^* = V_j \text{ (중간수요 합계 조건)}$$

$$\sum_{ij \in t}^n X_{ij}^* = C_t \text{ (행정원소제약조건, 예측연도 중간거래 관련 외생적 정보)}$$

$$X_{ij}^* \geq 0$$

- LM:  $Min. = \left[ \frac{\sum (X_{ij}^* - X_{ij})^2}{W_{ij}} \right]$

$X_{ij}$  = 기준연도 중간거래 행렬 요소

$X_{ij}^*$  = 예측연도 중간거래 행렬 요소

$W_{ij}$  = 기초 추계치 변동계수에 비례하는 가중치 요소

이러한 라그랑지 미정계수법은 정밀도가 높고, 최소자승법이라는 이론적 체계가 분명하다는 장점을 가지고 있으나, 부등식 제약을 포함하기 어려워 부의(-) 값이 발생할 가능성이 존재한다는 단점이 있다.

평균 증가배율법은 기준연도의 투입계수와 예측연도의 산출액, 중간수요 합계 및 중간투입 합계의 평균치를 적용하여 예측연도의 투입계수를 추정하는 방법이다. 그리고 2차형 계획법(quadratic programming method)은 Non-negativity 제약 부여 가능 등 신축적 제약조건 부여로 모형 자체가 탄력적이다.

$$Mn.Z = \left[ \frac{\sum (X_{ij}^* - X_{ij})^2}{W_{ij}} \right] + \left[ \frac{\sum (V_j^* - V_j)^2}{W_j} \right] + \left[ \frac{\sum (U_j^* - U_j)^2}{U_j} \right] + \left[ \frac{\sum (C_i^* - C_i)^2}{W_i} \right]$$

그리고 선형계획법(linear programming method)은 기준연도에 실측된 각 원소와 그에 대응하는 연장 추정 원소와의 편차의 최대 절대치를 최소화하는 것으로 2차형 계획법에 비해 계산이 간편하다는 장점이 있다.

그러나 이들 방법도 행과 열 간에 정합성을 보장하기 어렵다는 단점이 있어 제한적인 범위 내에서만 활용되고 있다.

#### 나. 양비례법(RAS)을 활용한 시계열화

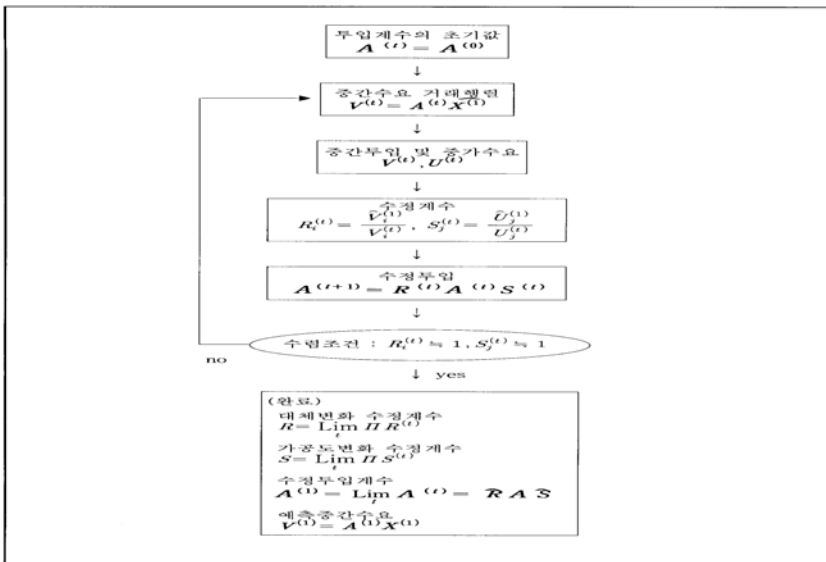
RAS 추정법(biproportional adjustment method)은 기준연도의 투입계수와 예측연도의 중간수요 합계, 중간투입 합계 및 총산출액이 주어지면 대체변화 계수(행의 R계수), 가공도변화계수(열의 S계수)를 측정하여 예측연도의 중간수요 합계와 중간투입 합계가 같아질 때까지 반복 계산하는 것이다. 이 방법은 R. Stone이 개발한 것으로 추정된 투입계수의 Non-negativity 및 안정성을 보장, 이용 가능한 모든 통계를 최대한 활용 가능하며, 투입-산출모형의 기본 원리를 충족한다는 장점으로 가장 보편적으로 이용되고 있다. RAS 추정법은 R. Stone이 1961년에 개발한 단순 방법과 이를 수정한 방법 둘로 대별되고 있다.

우선 단순 RAS(simple RAS, Stone, 1961) 방법은 기준연도와 비교연도 사이에 실현된 구조변화를 나타내는 대체변화계수(R)와 가공도변화계수(S)를 추정하여 연장하는 것이다. 다만, 추정하려는 시점의 부분적인 정보, 예를 들면 산업연관표가 부재한 시점의 산업별 총산출액, 중간수요 합계, 중간투입 합계 존재한다고 가정한다.

- 기본식:  $A^{(1)} = \hat{R} A \hat{S}$
- $A^{(1)}$  = 변화 후의 투입계수행렬
- $A$  = 변화 전의 투입계수행렬
- $\hat{R}$  = 대체변화를 수정하는 계수의 대각행렬
- $\hat{S}$  = 가공도를 수정하는 계수의 대각행렬

한편 수정 RAS(Modified RAS)는 투입계수 가운데 비교적 안정적인 추세를 갖는 요소들은 고정하고 나머지만 RAS 방법을 적용하는 방식으로 1964년 J. R. C. Lecomber에 의해서 제안되었다. 그러나 이 방법은 고정시킬 투입계수의 결정에 있어 자의성 문제가 초래될 수 있다는 단점이 있다.

[그림 12-2] 수정 RAS 방법의 절차



## 제4절 산업별 노동과 임금 DB

### 1. 산업별 노동통계

노동과 관련된 통계는 현재 고용노동부나 통계청 혹은 한국은행 등의 국가 통계작성기관에서 다양하게 조사하고 있으나, 전 산업을 포괄하는 통계로는 앞에서도 기술한 바와 같이 통계청의 가구대상 경제활동인구조사와 사업체 대상의 전국사업체조사, 한국은행의 산업연관표 부속표로서 고용표 3종이 있다. 우선 경제활동인구조사의 취업자는 매월 일정 기간 동안 수입을 목적으로 취업한 사람의 수를 조사한 것으로서 1963년부터 이용 가능하지만, 그동안 표준산업분류의 개정, 산업분류의 조정 및 통계조사 기준의 변경 등으로 인해 20년 정도의 장기 시계열로서 일관성 있는 산업별 취업자 통계의 입수는 대분류 혹은 중분류를 다소 통합한 수준에서만 가능하다. 사업체기초통계는 세세분류로 사업체에 종사한 월평균 종업원 수를 각각 조사한 것으로 1993년부터 이용 가능하지만, 농림어업 부문의 사업체만을 대상으로 하기 때문에 일부분만을 포함하고 있다는 단점이 있다. 산업연관표의 고용표에서 노동통계는 1년 동안 각 산업에 투입된 노동량을 실제 종사한 개월 수로 가중 평균한 연인원(man-year)으로 추정한 것으로서 노동량을 취업자와 피용자로 구분하고, 생산활동에 종사한 내외국인을 모두 포함하며, 1970년부터 100인 이상의 산업별로 이용이 가능하지만, 불연속적이라는 단점이 있다.

이상과 같이 개별 노동통계가 있어서 노동의 개념, 산업분류, 시계열 확보에서 차이가 있어 어느 하나의 통계를 전적으로 이용하는 것이 불가능한 상황이다. 따라서 본 연구에서는 장기 시계열 통계의 입수가 가능하면서 경기변동을 잘 반영한다고 볼 수 있는 경제활동인구의 통계를 대분류 산업별 취업자의 총량으로 사용하고, 중분류 산업에 대해서는 사업체기초통계와 산업연관표의 고용표를 통해 산업별 비중을 산출하고 이를 경제활동인구조사의



대분류 취업자에 적용하여 개별 산업별 취업자 수를 추정하는 방식을 채택하였다. 이 과정에서 고용표가 갖는 불연속 시계열의 단점을 보완하기 위해 사업체기초통계 및 산업연관표의 고용표를 3, 4, 5개년 연평균 비중 변화의 단순평균을 이용하여 해당 대분류별 경제활동인구조사의 취업자 수 기준으로 추정하였다.

## 2. 산업별 임금통계

또한 대부분 거시산업모형에서 임금은 노동생산성, 노동과 자본의 대체성을 통해 노동수요와 더불어 가격·기술부문, 수요·공급부문 및 총량부문을 연결하는 중요한 매개 역할을 하고 있다. 따라서 임금은 근로자의 소득이라는 개념보다는 노동생산성의 측정, 물가에 대한 반영도, 자본과 노동 간의 대체성을 정밀하게 반영될 수 있는 통계로 구축될 필요가 있다. 현재 국내의 산업별 임금통계는 고용노동부의 사업체노동력조사(구 매월노동통계조사), 임금구조기본통계조사, 기업체노동비용조사가 대표적이나 농림어업 등 일부 산업이 제외되어 있거나 상용근로자만을 대상으로 하고 있으며, 표준산업분류의 변경이나 조사방법 혹은 조사대상의 변경 등으로 일관성 있는 장기 시

〈표 12-8〉 노동량의 측정단위

측정단위	측정방법
연인원	한 사람이 1년 동안 일하는 작업량의 단위로 1년 동안 각 사업체에서 투입한 실제 노동량(cf. total man-days : 어떤 일에 동원된 인원 수와 일수를 계산하여 그 일을 하루에 끝내는 것으로 가정하고 환산한 총인원 수와는 다른 개념) * 1년 중 6개월만 영업한 사업체에 근무한 근로자 1명과 1년 동안 영업한 사업체에 6개월만 근무한 근로자 1명은 0.5명으로 산정, 1명의 근로자가 두 개의 직업을 가지고 두 개의 사업체에서 각각 1명의 근로자로 1년 동안 계속 일한 경우에는 2명으로 산정)
월평균 인원	해당 사업체의 영업기간 중 월별 총 종사자수를 영업월수로 나눈 것 (일반 사업체 조사)
전업 환산	전체 취업자의 총 근로시간을 전업 직무의 연평균 근로시간으로 나누어 산출

자료: 한국은행(2007), 「산업연관분석 해설」.

계열 통계를 확보하기 어려운 측면이 있다. 더욱이 이러한 조사통계는 국민 계정이나 산업연관표와 같이 경제의 다른 부문과 연계성을 고려하지 않는 조사통계라는 점에서 계량모형에의 활용에 한계가 있다.

이에 본 연구에서는 이상과 같은 현재 임금조사통계의 한계점과 모형 내에서 임금의 다양한 매개 역할을 고려하여 산업별 임금은 노동계수 혹은 노동 생산성의 변화나 모형 내 다른 변수와 일관성 및 적합성 측면에서 유용하도록 단위당 노동비용의 개념에 입각하여 추정하였다. 구체적으로는 산업연관표와 국민계정을 연계하는 과정에서 추정되는 산업별 피용자보수를 앞에서 추정한 산업별 노동으로 나누어 추정하였다. 그리고 모형 내에서 산업별 임금의 상대적 격차나 물가에 대한 영향 등이 정밀하게 추정되도록 2005년을 100으로 지수화하였다.

한편 노동부의 임금관련 산업별 조사통계는 임금정책 측면에서 유용한 정보를 제공하므로 상기의 임금지수와 연계하는 방정식을 추정하여 임금의 단기적 전망이나 정책효과에 대한 평가에 활용할 필요가 있다.

## 제5절 산업별 자본스톡과 수출입 DB

### 1. 자본스톡의 개념과 추정방법

자본스톡은 어떠한 시점에 생산자가 보유한 모든 자산을 그 시점에서 다시 구입하는 데 소요되는 비용으로 평가한 총자본스톡과 어떤 시점의 총자본스톡에서 그 시점까지 누적된 총고정자본 소모를 차감한 순자본스톡으로 구분된다. 전자의 총자본스톡은 생산자가 소유한 자본스톡과 임대한 자본스톡 그리고 생산과정에 투입되지 않는 유헴자본스톡도 포함하는 개념이며, 순자본스톡은 어떤 시점의 경제 전체 고정자산의 시장가치를 의미하는 것으로서 이들간의 관계는 다음과 같이 표현된다.

- 순자본스톡 = 총고정자본 + 총고정자본형성 - 고정자본소모 + 자본스톡을  
 변화시키는 다른 요인에 의한 순변동의 총가치 + 자산가격  
 변동으로 인한 보유수익 변화에 따른 순변동의 총가치

여기서 고정자산 소모란, 회계 기간 동안 물리적 감소, 정상적 진부화 (obsolescence), 또는 정상적 사고·훼손 등의 결과로 생산자에 의해 사용된 고정자본스톡의 경상가치가 감소한 것을 의미한다. 여기에 포함되지 않는 것은 전쟁 또는 발생 빈도가 낮은 자연재해 등과 같은 예외적인 사건으로 인한 고정자본스톡의 감소이며, 이러한 경우는 자산규모의 기타 변화 항목에 기록된다. 진부화란 생산성이 향상된 동종의 자산의 도입으로 구 자산의 가치가 하락하는 것을 의미하는 것으로서 물리적 마모(physical deterioration)를 포함하지 않는 개념이다. 그리고 자본스톡의 경상가치 감소는 인플레이션을 감안한 가격, 즉 실질가격으로 계산하는데 생산자에 의해 사용됨은 생산활동에 이용되지 않았더라도 사용자가 구입한 자본스톡을 포함한다. 한편 물리적 마모란 수리해서 고치지 못한 마모·훼손 또는 마모된 부속의 교체를 의미하는 것으로서 관리를 잘하는 경우 물리적 마모는 상대적으로 적게 발생하게 된다. 정상적인 사고·훼손이란 보험에 의해서 보상이 가능한 사고를 말하는데, 교통수단의 경우 사고로 인하여 조기에 못 쓰게 되는 수도 있다. 그리고 특별한 언급이 없으면 비정상적 진부화는 고정자본 소모에 포함되지 않는데, 이것은 예상치 못한 기술발전, 요소가격의 갑작스런 상대가격 변화 등에 의해 발생할 수 있으며, 1960년대 전자계산기의 도입, 1974년 오일쇼크 등이 대표적인 예라고 할 수 있다.

이러한 자본스톡은 원래 조사를 통해 파악해야 하나, 조사비용에 대한 부담 등으로 인해 최근에는 대부분 간접적인 방법으로 추정하고 있다. 우리나라는 1968년부터 10년을 주기로 1977년, 1987년, 1997년 4년을 대상으로 국부통계조사를 실시하였으나 시의성, 대규모 예산 소요, 응답 부담 등으로 1997년을 마지막으로 조사가 중단되고 대신 가공 통계인 국가자산통계로 대체하고 있다. 국가자산통계는 자산을 생산자산, 비생산자산으로 구분하고, 생

산자산에 대해서는 영구재고법, 비생산자산에 대해서는 물량가격법(토지, 임목), 수익환원법(지하자산)으로 추정하고 있으며, 산업분류는 제9차 한국표준산업분류(KSIC)의 대분류를 적용(16개 산업분류)하고 있다.

국가자산통계에서 이용하는 영구재고법(perpetual inventory method)은 특정 연도 해당 자산의 총자본스톡은 내용연수 기간이 지나 폐기된 자산을 제외한 투자자산의 합으로 추정하는 것이다.

$$\text{- 총자본스톡 } GK_t = \sum_{i=1}^L w_i I_{it}, \text{ 순자본스톡 } NK_t = \sum_{i=1}^L w_i I_{it} d_i$$

$L$  = 자산의 내용연수,  $w_i$  = 잔존율,  $I_{it}$  =  $t$ 시점 해당 자산에 대한 투자  
 $d_i = 1 - \delta_i$  ( $\delta_i$ =감가상각)

이러한 영구재고법을 이용하여 자본스톡을 추계하기 위해서는 해당 자산에 대한 기초자료가 충분히 확보되어야 한다. 즉, 해당 자산의 경제적 내용연수와 시간이 감에 따라 해당 자산의 폐기 형태에 대한 기초자료가 확보된 후에야 추정이 가능하다. 그러나 자산별 폐기 분포와 경제적 내용연수의 추계는 그 자체가 많은 비용과 노력을 필요로 한다는 단점이 있다.

한편 자본스톡을 추정하는 또 다른 방법으로는 두 개의 기준연도 자본스톡 자료에다 투자시계열 자료를 접속하여 기준연도 사이에 있는 각 연도의 자본스톡을 추계하는 기준연도접속법(Benchmark year estimation method)이 있다.

- 총자본스톡

$$\begin{aligned} GK_t = & I_t + (1-r_t)I_{t-1} + (1-r_t)(1-r_{t-1})I_{t-2} + \dots \\ & + (1-r_t)\dots(1-r_{t-s-2})I_{t-s+1} + (1-r_t)\dots(1-r_{t-s+1})GK_{t-s} \\ & r_t : \text{폐기율} \end{aligned}$$

- 순자본스톡(변동감가상각률)

$$NK_t = I_t + (1 - \delta_t) I_{t-1} + (1 - \delta_t)(1 - \delta_{t-1}) I_{t-2} + \dots \\ + (1 - \delta_t) \dots (1 - \delta_{t-s+2}) I_{t-s+1} + (1 - \delta_t) \dots (1 - \delta_{t-s+1}) NK_{t-s}$$

- 순자본스톡(고정감가상각률)

$$NK_t = I_t + (1 - \delta) I_{t-1} + (1 - \delta)^2 I_{t-2} + \dots \\ + (1 - \delta)^{s-1} I_{t-s+1} + (1 - \delta)^s NK_{t-s}$$

이러한 기준연도접속법은 주기적으로 일정한 기간에 따라 실시한 자본스톡 자료를 중심으로 간접적으로 폐기율과 감가상각률을 추정하여 연도별 자본스톡을 추계하는 방법이다.

본 연구에서는 산업별 자본스톡을 추정함에 있어서 국부통계와 기준연도접속법 및 영구재고법을 혼용하였다. 즉, 산업별 자본스톡은 1987년과 1997년 국부통계조사의 산업별 자본스톡과 국민계정의 산업별 고정자본형성 시계열을 이용하여 추정하였는데, 1987~97년 기간은 기준연도접속법으로, 이외의 연도는 영구재고법을 적용하였다. 그리고 제조업 세부산업별 고정자본형성 시계열은 광공업통계조사에서 제공되는 고정자본형성 연말 잔액의 산업별 구성비를 이용하여 추정하였다.

## 2. 수출입통계의 구축

### 가. 수출입통계의 종류

우리나라의 수출입 관련 통계는 관세청의 무역통계, 한국은행의 국제수지, 국민계정 및 산업연관표 등 4종류가 공표되고 있다. 우선 관세청의 무역통계는 보통 통관기준 수출입으로 불리어지는데, 이는 UN에서 권고하고 있는 국제무역통계의 개념과 원칙에 입각하여 한국 자원량의 증감을 유발시키는 경

제선을 통과하는 물품의 이동을 모두 포함한다. 수출입 품목은 관세의 기준이 되는 HSK 10단위를 기준에 의거하여 분류하며, 금액 기준으로 수입은 운임보험료포함조건(CIF, cost, insurance and freight), 수출은 본선인도조건(FOB, free on board) 기준으로 평가한다.<sup>15)</sup> 그리고 화폐 단위의 변환은 관세청장이 매주 고시하는 환율에 의하여 각국의 화폐를 원화로 환산한 후 원화를 다시 US\$로 환산하고 환산된 US\$가 수출입 신고 시 적용한다.

한편 국제수지의 수출입 통계는 일정 기간 동안, 국가 간에 발생한 모든 경제적 거래를 체계적으로 기록한 국제수지표를 통해서 작성된다. 국제수지에 포함되는 거래는 경제활동을 영위하는 주 소재지가 국민경제 영역 내에 있는 경제 주체(거주자)와 영역 밖에 있는 여타 경제 주체(비거주자) 간의 거래인데 여기서 거주자와 비거주자의 구분은 법률상의 국적보다는 경제활동에 있어 이익의 중심(acerter of economic interest)이 어디에 있느냐를 기준으로 구분한다. 그리고 거래 유형에 있어서는 모든 경제적 거래가 대상이 되는데 재화 및 서비스, 소득, 자본 간의 교환거래뿐만 아니라 국제 간의 증여 등 대가가 없이 이루어지는 이전거래도 포함한다.<sup>16)</sup> 이러한 국제수지표에서는 경제적 가치가 생성·변화·교환·이전 또는 소멸될 때를 거래의 계상 시점으로 하는 발생주의 회계원칙이 적용된다. 즉, 채권과 채무는 소유권이 변동되었을 때 발생하는 것으로 인식하며 여기서 소유권 변동이란 통제 및 소유에 관련된 법적인 변동이나 경제적인 변동이 있을 때를 의미한다. 한국은행은 국제수지표 작성을 위해 관세청의 통관통계, 외환수급통계, 외국환은행의 외화 B/S와 같은 공식적인 통계 이외에도 정부와 금융감독원, 기업, 협회 및 금융기관 등으로부터 입수한 여러 기초자료를 이용하고 있다.

15) 운임보험료포함조건은 선적항에서 본선의 난간을 통과시켜 물품을 인도하되 목적항까지의 운임과 보험료를 매도인이 부담하는 조건이며, 본선인도조건은 선적항에서 본선의 갑판상에 물품을 인도하는 조건으로서, 수출자가 수출품을 항구 또는 공항에서 하역할 때까지의 운임, 기타 제비용을 부담하는 조건을 의미한다.

16) 복식부기(double-entry system)의 원리에 의해 상품, 서비스, 소득, 이전거래 등 실물거래와 금융거래를 체계적으로 구분, 국가 간에 비교가 가능하도록 IMF가 정해 놓은 국제수지 통계의 포괄 범위, 분류, 평가 등에 관한 국제기준(IMF BOP Manual)에 의해 기록.

한편 국민계정의 수출입통계는 국외거래계정에서 경상거래로 구분하여 국내총생산에 대한 지출로 계상하는 과정에서 작성된다. 즉, 국내에서 생산된 재화와 서비스는 중간소비, 최종소비, 총자본형성으로 사용되는데, 이 중 국외로 수출되어 외국에서 사용되는 것을 ‘재화와 서비스의 수출’, 외국에서 생산된 재화와 서비스가 국내에 수입되어 사용되는 것을 ‘재화와 서비스의 수입’으로 작성하게 된다. 수출입의 평가 시점은 재화의 수출입의 경우 국내거래와 마찬가지로 법적 소유권이 이전된 시점에서 기록하며, 재화는 일반상품, 가공용 재화, 재화 수리, 운수조달재화 및 비화폐용 금(金)을 포함한다. 그리고 재화수출은 수출국 세관 국경의 본선인도(FOB) 가격으로 평가하며, 재화의 가액뿐만 아니라 그 지점까지의 유통서비스 가액을 포함하며, 필요한 경우 계속적인 수송을 위해 운송수단에 적재하는 비용도 포함한다. 그리고 재화수입도 본선인도가격에 의해 평가되는데 보통 수입가격은 수입국 국경에서의 구매자가격(CIF)으로 작성되므로 이 경우에는 CIF 가격에서 운송비용과 보험료를 차감하여 기록한다. 한편 서비스 수출입은 서비스를 제공한 시점에서 기록하며, 재화에 포함되지 않은 가공 및 수선의 부가가치는 그 활동이 수행된 시점에서 기록하며, 여기에는 운수 및 통신서비스, 보험서비스, 기타서비스, 비거주자 가계의 국내소비지출(거주자가계의 국외소비지출), 외국기관의 국내소비지출(정부의 국외소비지출)을 포함된다.

그리고 산업연관표의 수출입통계는 재화 수출입의 경우 무역통계 자료를 부문 및 품목 분류에 따라 재분류하여 추계하며, 서비스 수출입은 외환수급통계 및 국제수지통계 등을 이용하여 추계한다. 수입품은 CIF 가격에 관세 및 수입상품세를 더한 가격으로, 수출품은 FOB 가격으로 평가된다. 그리고 상품수출입에 대해서는 일평균 환율을, 서비스 수출입에 대해서는 월평균 환율을 적용하여 추계하는데, 수입거래표는 개별 수입품의 수요처에 대한 특별조사 결과 등을 이용하여 작성한다.

다부문 거시산업모형의 DB 구축을 위한 산업별 수출입통계는 실물거래인 재화 및 서비스 수출입을 대상으로 하는데, 국내 생산활동을 대상으로 하므로 생산요소의 국제 간 이동에서 발생하는 해외요소소득의 지급이나 수취는 제

외하였다. 그리고 총수출입의 총량 규모는 국민계정 재화 및 서비스 수출입, 산업별 및 수요형태별 비중은 산업연관표의 산업별 구조에 입각하였다. 즉, 산업별 수출입은 국민계정의 총수출입을 산업연관표의 산업별 비중으로 배분하며, 산업연관표 미작성 연도에 대해서는 관세청의 통관기준 수출입, 국제수지의 서비스 수출입을 산업분류에 연계하여 비중 변화를 추정하였다. 그리고 수출입의 실질화는 산업별 수출은 산업별 산출디플레이터, 수입은 수입물가지수를 이용하였다.

여기서 한 가지 유의해야 할 점은 산업연관표와 국민소득계정은 개념적으로 서로 밀접하지만, 목적이 상이하어 추계에 이용되는 통계자료나 추계방법상의 차이 등으로 동일 변수에 대해 계수가 완전히 일치하지 않는다는 것이다. 산업연관표에서는 수입품과 국산품을 동일한 가격 기준으로 평가하고 있으나, 국민소득계정에서는 수입을 CIF 가격으로 평가하고 관세와 수입상품세를 별도의 부가가치 항목으로 설정하고 있으므로 산업연관표의 부가가치 총액은 관세와 수입상품세만큼 국민소득계정의 국내총생산(GDP)과 차이가 발생한다. 따라서 산업연관표의 부가가치에 관세와 수입상품세를 더하고 해외순수취요소소득을 더하면 국민총소득(GNI)의 개념과 일치하게 된다는 것이다.

## 제6절 DB의 불변가격(실질변수)화 작업

산업연관표를 국민계정과 연계하여 산업연관표의 미작성 연도에 대해서는 국민계정의 총량변수와 기존 산업연관표의 투입-산출 구조를 이용하여 추정함으로써 경상가격 기준 산업연관표의 시계열화가 완료되면 이를 특정 연도의 가격으로 고정시키기 위해 실질화(혹은 불변화)를 수행해야 한다. 앞에서도 언급한 바와 같이 경제 분석에서는 보통 물량 개념을 이용하므로 산업별 경상가격 자료는 불변가격으로 전환, 즉 실질화가 필요하다. 이를 위해 경상



가격 기준의 시계열 산업연관표를 국민계정의 산업별 산출디플레이터를 이용하여 실질화하였다. 즉, 연도별·산업별·부문별 경상가격 기준의 규모는 국민계정의 산업별 산출디플레이터를 이용하여 2005년 기준으로 불변화한 다음 다시 국민계정의 각 부문별 불변가격 총규모와 일치하도록 조정한 것이다. 그리고 산업별 수출입은 관세청의 무역통계를 산업연관표의 27개 산업을 연계하고, 이들 각 산업에 대해서 금액을 중량으로 나누어 산업별 수출입 단가 지수를 산출한 다음, 이를 이용하여 2005년 기준으로 불변화한 산업별 규모를 가중치로 국민계정의 불변가격 재화 수출과 재화 수입의 총 규모와 일치하도록 조정하였다. 이러한 과정에서 산업별로 디플레이터를 개별적으로 적용하여 합산된 산업별 실질 중간투입에 산업별 부가가치 디플레이터를 적용한 산업별 실질 부가가치를 더한 것과 전 산업 총산출에 해당 산출디플레이터를 적용하여 환산된 실질 총산출액 간에는 차이가 발생하게 된다. 중간재와 부가가치에 적용한 디플레이터와 총산출에 적용한 디플레이터 간의 차이에 의해서 발생하는 소위 이중디플레이터(double deflator) 문제를 해소하기 위해 국민계정의 실질 총산출을 기준으로 총규모를 결정하고, 최종적으로는 여기에 산업별 부가가치율을 적용하여 중간투입 비중과 부가가치 비중을 산출하여 적용하였다.

## 다부문 거시산업모형 DB를 활용한 기초 분석

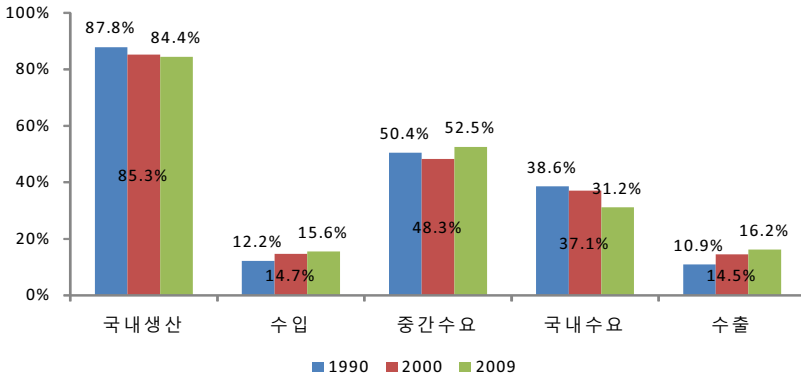
### 제1절 산업별 수요 및 공급구조 분석

#### 1. 산업의 수요와 공급구조

우리 경제의 공급구조는 전반적으로 수출입의 비중이 높아지는 추세로서 대외경제에 대한 의존도가 심화되는 구조로 변화하고 있다. 수요 측면에서는 1990~2009년 기간 동안 해외수요인 수출의 비중이 10.9%에서 16.2%로 높아진 반면, 내수의 비중은 38.6%에서 31.2%로 낮아졌다. 공급 측면에서는 국내 생산이 87.8%에서 84.4%로 낮아진 반면에 수입 비중은 12.2%에서 15.6%로 확대되었다.

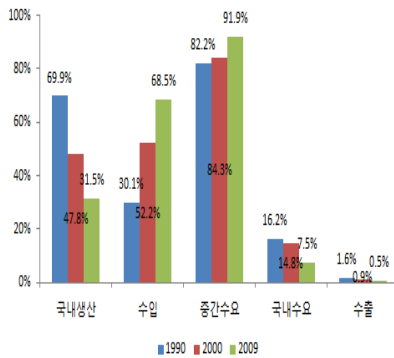
대분류 산업별로 공급구조를 살펴보면 1990~2009년의 20년 동안 농림어업의 경우 수요 측면에서는 산업적 수요인 중간수요가 82.2%에서 91.9%로 확대되고, 공급에서는 수입 비중이 30.1%에서 68.5%로 2배 이상 확대되었다. 제조업 역시 중간수요가 절반 이상을 차지하는 산업이지만, 그 비중은 57.0%에서 60.0%로 다소 확대된 반면, 수출 비중이 17.1%에서 27.7%로 크게 확대되었고, 공급구조는 큰 변화 없이 국내생산과 수입 비중이 각각 80%, 20%대를 유지하고 있다. 전기수도가스건설업과 서비스업은 내수 위주의 산업으로

[그림 13-1] 전 산업의 수요 및 공급구조

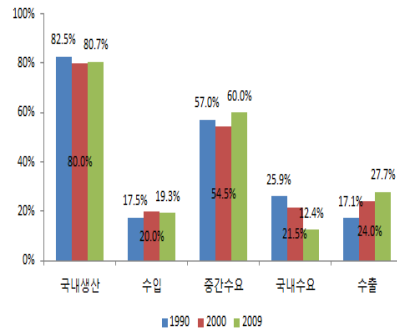


[그림 13-2] 전 산업의 수요 및 공급구조

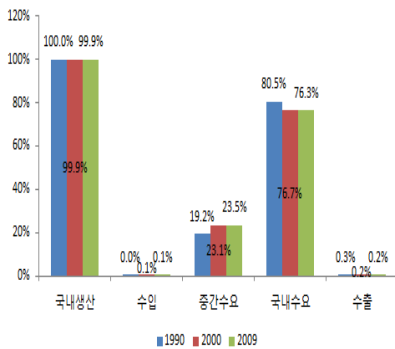
<농림어업>



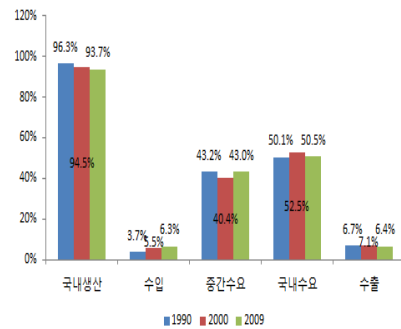
<제조업>



<전기수도가스건설업>



<서비스업>



수출입의 비중이 낮으며, 수급구조가 비교적 안정된 추세를 보이고 있다. 다만, 전기수도가스건설업에서는 중간수요의 비중이 다소 높아지는 경향을 보이고 있고, 서비스업에서는 수입의 비중이 확대되는 추세다.

〈표 13-1〉 27개 산업의 수요와 공급구조(1990)

(단위 : %)

	총공급		총수요		
	국내생산	수입	중간수요	국내수요	해외수요
농림어업	87.2	12.8	74.6	23.4	2.0
광업	23.7	76.3	102.4	-3.1	0.8
제조업	82.5	17.5	57.0	25.9%	17.1
음식료품담배	91.3	8.7	40.0	56.1	3.9
섬유가죽	91.5	8.5	41.5	15.1	43.5
목재종이인쇄	82.7	17.3	94.4	0.6	4.9
석유석탄화학	78.4	21.6	78.0	12.1	9.8
비금속광물	91.5	8.5	96.6	-1.4	4.8
금속제품	83.3	16.7	87.3	0.8	11.9
일반기계	62.4	37.6	41.3	52.7	6.0
전기전자	77.4	22.6	43.0	24.7	32.3
정밀기기	47.8	52.2	41.2	43.2	15.6
운송장비	90.8	9.2	33.8	50.2	16.0
기타제조업	91.6	8.4	25.7	42.2	32.1
전기가스수도	99.8	0.2	79.3	20.2	0.4
건설업	100.0	0.0	9.2	90.5	0.2
서비스업	96.3	3.7	43.2	50.1	6.7
도소매업	98.6	1.4	41.9	45.8	12.3
음식숙박업	92.9	7.1	28.8	61.6	9.5
운수보관업	90.4	9.6	41.9	34.3	23.8
금융보험업	97.8	2.2	80.7	17.0	2.3
부동산임대업	99.4	0.6	30.8	69.0	0.3
정보통신업	95.8	4.2	66.3	29.2	4.5
사업서비스	91.0	9.0	85.5	12.2	2.3
공공행정국방	100.0	0.0	0.2	99.4	0.4
교육서비스업	95.1	4.9	1.9	98.1	0.0
보건사회복지	99.8	0.2	7.0	92.9	0.0
문화오락	94.8	5.2	9.7	83.2	7.1
기타서비스	98.3	1.7	66.7	31.1	2.2
전 산업	87.8	12.2	50.4	38.6	10.9

〈표 13-2〉 27개 산업의 수요와 공급구조(2000)

(단위 : %)

	총공급		총수요		
	국내생산	수입	중간수요	국내수요	해외수요
농림어업	87.6	12.4	70.0	28.4	1.6
광업	6.2	93.8	99.3	0.5	0.2
제조업	80.0	20.0	54.5	21.5	24.0
음식료품담배	86.9	13.1	40.0	55.5	4.5
섬유가죽	85.4	14.6	36.8	21.2	42.0
목재종이인쇄	82.6	17.4	87.8	2.8	9.4
석유석탄화학	82.4	17.6	70.5	10.9	18.6
비금속광물	90.6	9.4	93.4	1.1	5.5
금속제품	82.7	17.3	86.0	0.6	13.5
일반기계	68.9	31.1	41.9	42.5	15.5
전기전자	73.4	26.6	41.9	20.8	37.3
정밀기기	43.5	56.5	35.8	50.1	14.1
운송장비	90.3	9.7	35.4	27.4	37.2
기타제조업	81.6	18.4	30.8	45.0	24.2
전기가스수도	99.8	0.2	67.2	32.6	0.2
건설업	99.9	0.1	9.1	90.7	0.2
서비스업	94.5	5.5	40.4	52.5	7.1
도소매업	98.1	1.9	41.8	47.5	10.7
음식숙박업	87.2	12.8	34.9	56.3	8.8
운수보관업	85.1	14.9	37.4	26.1	36.5
금융보험업	97.7	2.3	61.1	36.2	2.7
부동산임대업	98.8	1.2	29.9	69.9	0.2
정보통신업	96.7	3.3	58.4	40.0	1.6
사업서비스	86.9	13.1	73.7	21.1	5.2
공공행정국방	100.0	0.0	0.0	100.0	0.0
교육서비스업	95.5	4.5	2.9	96.9	0.2
보건사회복지	99.8	0.2	8.8	91.1	0.1
문화오락	90.7	9.3	20.2	70.7	9.0
기타서비스	96.5	3.5	62.8	32.8	4.4
전 산업	85.3	14.7	48.3	37.1	14.5

〈표 13-3〉 27개 산업의 수요와 공급구조(2009)

(단위 : %)

	총공급		총수요		
	국내생산	수입	중간수요	국내수요	해외수요
농림어업	83.2	16.8	73.6	25.1	1.3
광업	3.3	96.7	101.9	-2.0	0.1
제조업	80.7	19.3	60.0	12.4	27.7
음식료품담배	84.5	15.5	49.6	44.9	5.5
섬유가죽	75.0	25.0	34.7	38.3	26.9
목재종이인쇄	85.1	14.9	91.7	-0.2	8.5
석유석탄화학	79.9	20.1	72.8	2.9	24.3
비금속광물	84.1	15.9	95.9	0.1	4.0
금속제품	83.3	16.7	88.6	-5.2	16.6
일반기계	73.0	27.0	46.3	29.8	23.9
전기전자	77.7	22.3	44.7	11.0	44.3
정밀기기	55.7	44.3	49.4	32.3	18.3
운송장비	90.6	9.4	32.6	19.9	47.4
기타제조업	78.7	21.3	50.5	34.5	14.9
전기가스수도	99.7	0.3	75.3	24.6	0.1
건설업	100.0	0.0	5.6	94.2	0.2
서비스업	93.7	6.3	43.0	50.5	6.4
도소매업	98.0	2.0	51.7	37.9	10.4
음식숙박업	90.4	9.6	35.1	59.2	5.7
운수보관업	85.1	14.9	51.8	18.5	29.7
금융보험업	96.3	3.7	61.4	34.1	4.5
부동산임대업	98.4	1.6	27.3	72.2	0.5
정보통신업	94.7	5.3	59.0	38.7	2.3
사업서비스	82.2	17.8	75.1	16.7	8.2
공공행정국방	99.6	0.4	1.5	98.5	0.0
교육서비스업	95.4	4.6	1.6	98.2	0.2
보건사회복지	99.7	0.3	13.3	86.5	0.1
문화오락	95.5	4.5	24.8	71.5	3.7
기타서비스	98.9	1.1	66.1	33.4	0.5
전 산업	84.4	15.6	52.5	31.2	16.2

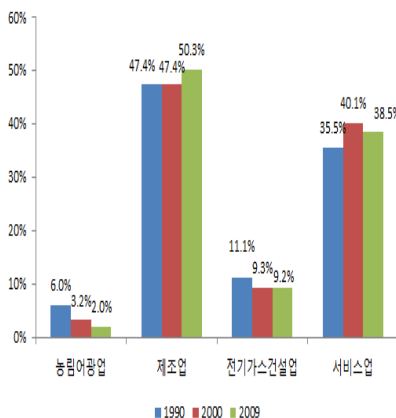
## 2. 산업생산구조의 변화

우리나라의 생산구조는 중간재를 포함한 총산출 기준에 입각할 경우 제조업이 전체 생산의 절반 이상을 차지하나 부가가치를 기준할 때는 서비스업이 60% 이상을 차지한다. 1990년 이후 생산구조는 총산출 및 부가가치 양면에서 농림어업업과 전기가스수도건설업의 비중이 낮아지는 추세다. 그러나 제조업의 비중은 총산출 기준에서는 높아지는 추세이나 부가가치 기준으로는 2000년 이후 낮아지고 있으며, 서비스업의 비중은 그 반대의 추세를 보이고 있다. 이것은 제조업의 생산과정에서 원료, 부품, 반제품 등의 투입이 상대적으로 많으며, 서비스업의 생산은 대부분 노동투입에 의존하므로 부가가치의 비중이 높은 산업적 특성을 반영하는 것이다. 따라서 산업생산구조의 현황이나 변화를 분석할 때에는 양자 중 어떤 지표를 사용하느냐에 따라 결과에 차이가 있으므로 분석지표에 대한 명시가 필요하다는 점에 유의해야 한다.

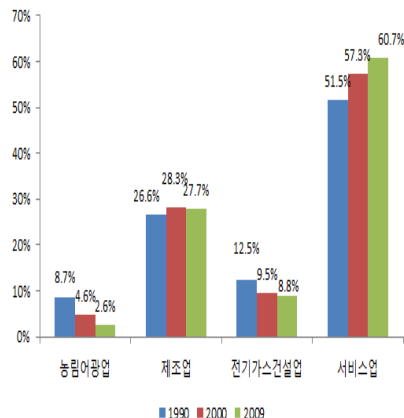
한편 제조업 내에서는 총산출과 부가가치 양면에서 음식료, 섬유가죽 등 경공업의 성장이 둔화 내지 감소하는 반면 기계, 운수장비, 전기전자, 화학 등의 중화학공업은 대부분 높은 성장을 유지하면서 생산 비중이 확대되는 추

[그림 13-3] 산업생산구조 변화 추이

<경상 총산출액 비중>



<경상 부가가치액 비중>



세다. 그리고 서비스업에서는 보건·사회복지, 금융보험의 성장 가속으로 견실한 성장세를 이어가면서 비중이 확대되는 추세다.

〈표 13-4〉 27개 산업별 생산구조 변화 추이(총산출 기준)

	경상비중(%)					실질증가율(%)			
	1990	1995	2000	2005	2009	91~95	96~00	01~05	06~09
농림어업	5.49	3.97	2.97	2.19	1.90	3.58	1.10	0.54	2.63
광업	0.52	0.31	0.19	0.17	0.14	-2.75	-2.22	0.91	-1.89
제조업	47.37	45.23	47.36	47.94	50.31	8.09	6.71	6.96	4.34
음식료품담배	6.22	4.59	4.14	3.21	3.23	5.92	3.14	0.28	2.11
섬유가죽	6.14	3.99	3.23	2.06	1.62	-3.66	1.58	0.06	-1.73
목재종이인쇄	1.97	1.84	1.58	1.33	1.19	6.80	1.47	3.54	-0.26
석유석탄화학	7.97	8.09	9.83	10.17	11.09	11.26	7.95	3.45	3.00
비금속광물	1.90	1.88	1.34	1.23	1.20	12.33	0.03	5.04	3.57
금속제품	6.90	6.97	6.28	8.23	8.78	7.81	3.81	5.91	0.13
일반기계	2.69	3.29	2.97	3.43	3.44	12.03	4.66	9.82	3.22
전기전자	6.77	7.68	10.85	9.97	10.24	8.19	23.31	15.08	10.86
정밀기기	0.45	0.51	0.50	0.58	0.59	6.22	6.71	12.81	9.11
운송장비	5.06	5.43	5.82	7.07	8.34	16.79	6.98	11.12	4.61
기타제조업	1.29	0.95	0.81	0.65	0.59	-1.03	3.39	1.44	2.36
전기가스수도	1.59	1.65	2.24	2.25	2.31	13.78	8.65	6.66	2.74
건설업	9.52	9.93	7.09	7.32	6.87	7.35	-1.52	3.70	1.24
서비스업	35.51	38.92	40.15	40.13	38.48	8.50	5.38	4.66	4.05
도소매업	7.33	6.84	6.35	5.52	5.14	5.71	4.78	1.95	2.58
음식숙박업	3.53	3.43	2.91	2.68	2.59	5.75	4.38	3.03	3.45
운수보관업	3.64	3.85	4.03	4.14	3.97	8.10	6.95	5.33	2.62
금융보험업	3.44	4.33	4.43	4.46	4.55	15.61	4.16	5.17	8.44
부동산임대업	3.80	5.00	5.12	4.45	3.79	9.07	4.17	2.37	1.82
정보통신업	2.04	2.63	4.07	4.10	3.58	17.20	17.24	8.35	3.38
사업서비스	2.17	2.70	2.71	3.06	2.92	9.80	4.07	6.50	2.93
공공행정국방	3.70	3.50	3.39	3.58	3.60	5.46	2.91	3.55	4.38
교육서비스업	2.28	2.57	2.61	2.96	3.01	5.50	3.09	4.83	3.78
보건사회복지	1.39	1.52	2.07	2.62	2.94	9.69	7.31	7.95	8.06
문화오락	0.64	0.80	0.75	0.89	0.86	15.42	5.88	8.32	4.98
기타서비스	1.55	1.75	1.70	1.66	1.53	8.65	4.74	4.91	2.89
전 산업	100.00	100.00	100.00	100.00	100.00	8.16	6.18	5.85	3.79



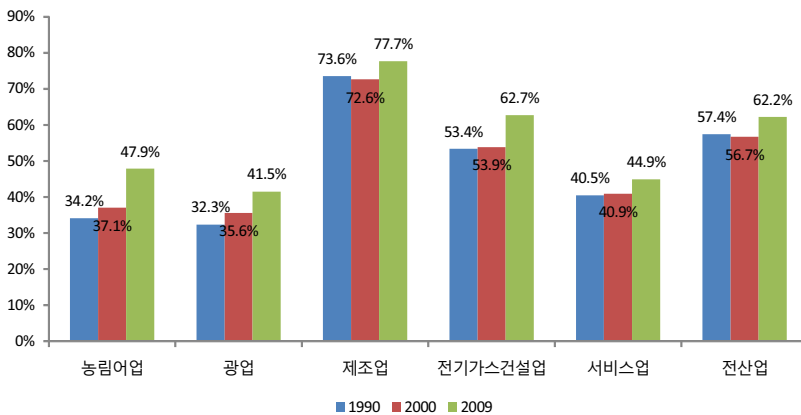
〈표 13-5〉 27개 산업별 생산구조 변화 추이(부가가치 기준)

	경상비중(%)					실질증가율(%)			
	1990	1995	2000	2005	2009	91-95	96-00	01-05	2009
농림어업	8.69	6.16	4.63	3.33	2.60	2.57	1.23	0.77	3.18
광업	0.76	0.48	0.31	0.26	0.23	-0.40	-2.21	0.40	-1.26
제조업	26.62	26.68	28.26	27.54	27.74	7.20	7.03	6.14	4.32
음식료품담배	1.96	1.75	1.79	1.45	1.43	2.87	2.45	0.60	0.60
섬유가죽	3.59	2.36	2.34	1.41	1.16	-5.74	3.73	-1.68	-1.55
목재종이인쇄	1.35	1.39	1.21	0.99	0.94	5.38	2.55	0.30	-0.52
석유석탄화학	3.87	4.42	4.33	4.74	4.35	10.64	7.25	5.20	2.28
비금속광물	1.77	1.40	1.17	0.98	1.06	7.94	0.56	2.92	1.86
금속제품	3.71	3.76	3.61	4.59	3.99	8.31	4.71	5.29	-3.60
일반기계	1.78	2.07	2.24	2.28	2.37	16.65	4.95	7.17	3.27
전기전자	4.16	4.96	7.01	6.49	6.65	13.43	23.17	13.65	10.84
정밀기기	0.39	0.44	0.39	0.42	0.48	16.14	6.81	6.27	7.40
운송장비	3.14	3.45	3.51	3.75	4.91	14.13	7.56	7.81	8.68
기타제조업	0.87	0.68	0.66	0.45	0.41	-3.76	3.42	-3.61	0.38
전기가스수도	2.07	1.99	2.54	2.27	1.82	11.54	9.03	6.64	4.72
건설업	10.40	10.11	6.95	7.64	6.94	5.90	-2.95	3.85	1.02
서비스업	51.45	54.57	57.31	58.96	60.68	7.90	4.36	3.85	3.30
도소매업	11.76	9.78	9.64	8.27	8.59	6.45	4.78	1.44	2.66
음식숙박업	2.43	2.65	2.76	2.36	2.40	8.86	4.58	2.20	1.86
운수보관업	4.74	4.57	4.51	4.55	4.40	6.47	6.89	3.67	2.68
금융보험업	5.23	6.08	5.78	6.88	6.92	14.36	1.01	7.08	5.89
부동산임대업	6.52	8.42	9.16	8.15	7.76	8.49	4.09	2.26	1.32
정보통신업	3.04	3.40	4.46	4.67	4.19	17.39	16.41	9.79	3.29
사업서비스	2.95	3.98	4.36	4.88	5.19	10.40	5.03	4.22	2.65
공공행정국방	5.27	5.41	5.66	6.21	6.72	4.40	2.85	2.18	3.21
교육서비스업	4.67	4.94	5.11	5.99	6.62	3.99	2.40	4.08	2.94
보건사회복지	2.02	2.12	2.75	3.68	4.50	7.53	2.94	3.99	6.84
문화오락	1.00	1.14	1.08	1.30	1.37	14.43	7.13	6.79	4.49
기타서비스	1.82	2.07	2.04	2.01	2.02	8.25	4.61	3.77	2.82
전 산업	100.00	100.00	100.00	100.00	100.00	7.67	5.06	4.55	3.34

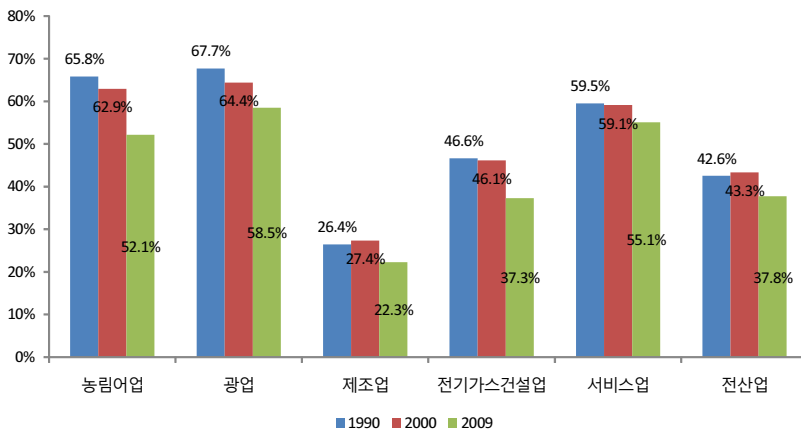
### 3. 산업별 중간투입계수와 부가가치계수 구조 변화

산업별 투입구조를 나타내는 중간투입계수와 부가가치계수의 변화를 보면 전반적으로 중간투입의 비중이 높아지고 부가가치의 비중이 낮아지는 추세다. 전 산업을 기준할 때 중간투입계수는 1990년에 57.4%에서 62.2%로 상승한 반면 부가가치계수는 42.6%에서 37.8%로 낮아졌다. 특히 농림어업과 광업에서 부가가치계수가 크게 낮아졌다.

[그림 13-4] 산업별 중간투입계수 변화 추이



[그림 13-5] 산업별 부가가치계수 변화 추이



〈표 13-6〉 27개 산업별 중간투입계수 변화 추이(경상 중간투입액/총산출액)

	1990	1995	2000	2005	2009
농림어업	34.16	33.74	37.05	42.03	47.85
광업	32.34	29.28	35.59	41.81	41.49
제조업	73.56	68.95	72.65	74.63	77.71
음식료품담배	75.89	71.71	72.92	71.28	74.57
섬유가죽	76.27	69.71	70.24	69.66	70.18
목재종이인쇄	73.36	67.89	71.32	68.66	69.76
석유석탄화학	75.88	68.34	72.28	74.45	79.85
비금속광물	61.47	62.72	65.87	68.29	69.63
금속제품	77.47	74.34	75.86	77.46	81.85
일반기계	67.11	65.70	69.18	72.21	73.79
전기전자	73.84	66.18	72.61	75.14	78.95
정밀기기	69.62	66.88	74.18	71.57	74.48
운송장비	68.45	69.92	75.92	78.44	76.24
기타제조업	65.78	62.52	65.87	70.22	71.83
전기가스수도사업	47.29	51.80	54.37	59.15	70.99
건설업	54.38	56.54	53.69	54.75	59.90
서비스업	40.47	39.83	40.88	42.95	44.90
도소매업	31.62	34.27	35.97	40.44	41.57
음식숙박업	56.00	61.49	59.43	59.88	62.05
운수보관업	48.53	47.83	57.45	57.17	62.51
금융보험업	32.16	28.49	31.11	37.30	44.51
부동산임대업	23.77	22.62	21.90	24.83	24.02
정보통신업	32.36	32.80	46.45	53.84	57.09
사업서비스	41.99	37.76	35.88	38.75	39.02
공공행정국방	45.17	39.80	31.64	31.07	32.79
교육서비스업	10.95	14.03	17.30	19.46	22.09
보건사회복지	45.49	40.36	48.13	43.47	44.45
문화오락	41.56	45.08	44.93	40.18	39.95
기타서비스	79.43	82.03	80.78	81.02	81.00
전 산업	57.44	54.84	56.67	58.82	62.24

〈표 13-7〉 27개 산업별 부가가치계수 변화 추이(경상 부가가치액/총투입액)

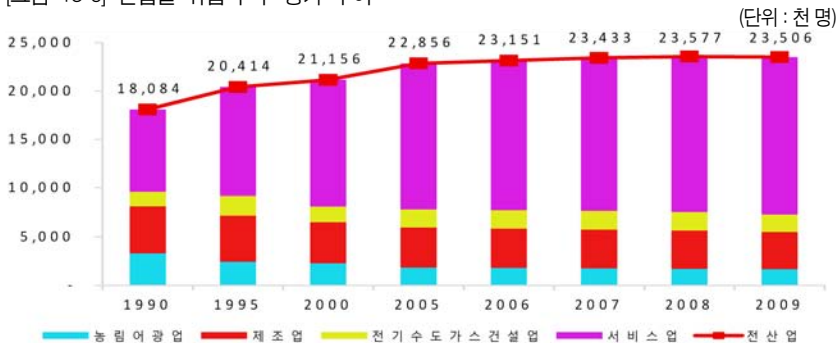
	1990	1995	2000	2005	2009
농림어업	65.84	66.26	62.95	57.97	52.15
광업	67.66	70.72	64.41	58.19	58.51
제조업	26.44	31.05	27.35	25.37	22.29
음식료품담배	24.11	28.29	27.08	28.72	25.43
섬유가죽	23.73	30.29	29.76	30.34	29.82
목재종이인쇄	26.64	32.11	28.68	31.34	30.24
석유석탄화학	24.12	31.66	27.72	25.55	20.15
비금속광물	38.53	37.28	34.13	31.71	30.37
금속제품	22.53	25.66	24.14	22.54	18.15
일반기계	32.89	34.30	30.82	27.79	26.21
전기전자	26.16	33.82	27.39	24.86	21.05
정밀기기	30.38	33.12	25.82	28.43	25.52
운송장비	31.55	30.08	24.08	21.56	23.76
기타제조업	34.22	37.48	34.13	29.78	28.17
전기가사수도	52.71	48.20	45.63	40.85	29.01
건설업	45.62	43.46	46.31	45.25	40.10
서비스업	59.53	60.17	59.12	57.05	55.10
도소매업	68.38	65.73	64.03	59.56	58.43
음식숙박업	44.00	38.51	40.57	40.12	37.95
운수보관업	51.47	52.17	42.55	42.83	37.49
금융보험업	67.84	71.51	68.89	62.70	55.49
부동산임대업	76.23	77.38	78.10	75.17	75.98
정보통신업	67.64	67.20	53.55	46.16	42.91
사업서비스	58.01	62.24	64.12	61.25	60.98
공공행정국방	54.83	60.20	68.36	68.93	67.21
교육서비스업	89.05	85.97	82.70	80.54	77.91
보건사회복지	54.51	59.64	51.87	56.53	55.55
문화오락	58.44	54.92	55.07	59.82	60.05
기타서비스	20.57	17.97	19.22	18.98	19.00
전 산업	42.56	45.16	43.33	41.18	37.76

## 제2절 산업별 고용구조

우리나라의 취업자 수 규모는 증가하고 있지만 증가율 자체는 계속 둔화되는 추세에 있다. 2009년 현재 2,351만 명으로 1990년의 1,808만 명 대비 1.3배 증가하였지만 증가율은 1990년대 연평균 3.6%에서 1980년대 2.8%, 1990년대 1.6%에서 2000년대 1.2%로 둔화되었다.

산업별로는 농림어업과 제조업의 비중이 계속 축소되는 반면, 서비스업의 비중은 확대되는 추세다. 농림어업은 절대적인 생산 수준이 증가하고 있지만, 기계화와 대규모화 등의 영향으로 취업자 수가 지속적으로 감소하면서 취업 비중이 1990년 18.2%에서 2009년에 7.1%로 축소되었다. 제조업은 경공업의 생산 감소, 중화학공업에 있어서 자동화 등 고용대체 생산설비의 지속적 확대로 고용의 절대 수준이 감소하면서 전 산업 대비 비중이 1990년 26.9%에서 2009년 16.35로 크게 축소되었다. 특히 음식료품, 섬유가죽, 목재 종이, 가구 등의 경공업은 감소세를 보이고 있다. 전기·가스·건설업의 취업 비중은 전반적으로 낮아지는 추세이지만 경기변동이 민감한 산업적 특성으로 인하여 축소와 확대를 반복하고 있다. 반면에 서비스업은 경제서비스화의 진전으로 생산 비중이 증가하면서 취업 비중도 1990년 47.1%에서 2009년에 68.8%로 크게 확대되었다. 특히, 부동산·사업서비스, 보건·사회복지의 취업자 비중이 크게 확대되는 추세다.

[그림 13-6] 산업별 취업자 수 증가 추이



〈표 13-8〉 27개 산업별 취업자 비중과 연평균 증가율 추이

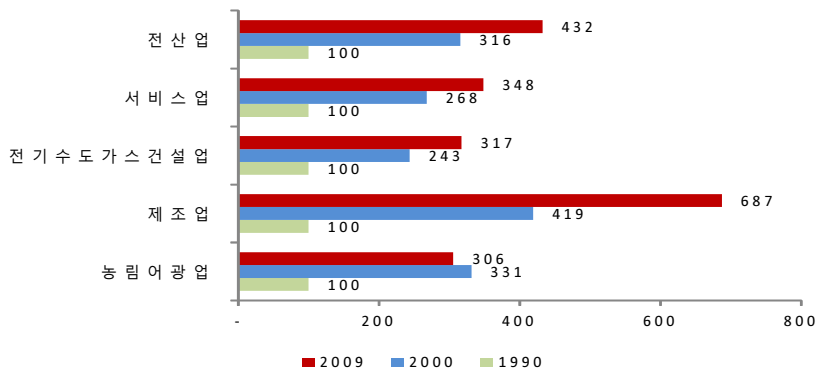
	취업자 비중					연평균 증가율			
	1990	1995	2000	2005	2009	91-95	96-00	01-05	06-09
농림어업	17.8	11.8	10.6	7.9	7.0	-5.7	-1.4	-4.2	-2.4
광업	0.4	0.1	0.1	0.1	0.1	-19.8	-8.1	0.0	7.8
제조업	26.9	23.4	20.1	18.1	16.3	-0.4	-2.3	-0.6	-1.8
음식료품담배	2.1	2.0	1.8	1.6	1.4	1.7	-2.0	-0.6	-2.0
섬유가죽	6.8	4.6	3.4	2.0	1.8	-5.1	-5.4	-8.2	-2.0
목재종이인쇄	1.9	1.4	1.1	1.0	0.9	-2.6	-4.5	-0.3	-2.9
석유석탄화학	2.4	2.2	2.0	2.1	1.9	1.4	-1.3	2.1	-1.9
비금속광물	1.6	1.0	0.7	0.6	0.5	-5.4	-7.3	-2.2	-1.9
금속제품	2.7	2.7	2.5	2.4	2.2	2.7	-1.0	1.1	-1.3
일반기계	2.1	2.1	1.9	1.8	1.6	2.5	-1.9	0.5	-1.4
전기전자	3.3	3.4	3.5	3.4	3.1	2.9	1.2	1.4	-2.3
정밀기기	0.4	0.4	0.4	0.4	0.4	3.1	0.7	1.3	-1.5
운송장비	2.4	2.3	2.0	2.0	1.9	1.1	-2.1	1.6	-0.9
기타제조업	1.3	1.1	0.9	0.8	0.7	-0.9	-3.3	-2.0	-2.5
전기가스수도	0.4	0.3	0.3	0.3	0.4	0.1	-1.8	2.1	7.9
건설업	7.4	9.4	7.5	7.9	7.3	7.4	-3.7	2.8	-1.3
서비스업	47.1	55.0	61.5	65.7	68.8	5.7	3.0	2.9	1.9
도소매업	14.9	17.3	18.1	16.4	15.3	5.5	1.7	-0.5	-1.0
음식숙박업	5.7	7.9	9.1	9.0	8.2	9.6	3.5	1.4	-1.5
운수보관업	4.5	4.7	5.1	5.1	5.5	3.2	2.1	1.7	2.7
금융보험업	2.7	3.6	3.6	3.3	3.3	8.5	0.6	-0.2	0.7
부동산임대업	1.0	1.7	2.1	2.4	2.4	14.1	4.9	4.5	0.6
정보통신업	0.7	0.8	1.2	1.6	1.7	4.8	9.3	7.9	2.3
사업서비스	1.8	3.2	4.7	6.9	7.8	15.1	8.7	9.8	3.8
공공행정국방	2.7	3.2	3.6	3.5	4.4	5.5	3.1	0.8	6.9
교육서비스업	4.8	5.0	5.6	7.0	7.6	3.6	3.0	5.9	2.9
보건사회복지	1.3	1.5	2.0	2.8	4.2	5.1	6.8	8.6	11.5
문화오락	0.9	1.0	1.4	1.8	1.9	5.8	7.8	6.7	1.6
기타서비스	6.1	5.0	5.0	5.9	6.4	-1.3	0.4	5.0	3.1
전 산업	100.0	100.0	100.0	100.0	100.0	2.5	0.7	1.6	0.6

### 제3절 산업별 임금구조

피용자 1인당 보수로 측정한 임금수준은 1990~2009년의 20년 동안 전 산업 평균으로 4.3배가 상승하였고, 대분류 산업별에서는 제조업의 임금이 6.9배 가장 빠르게 상승한 반면, 농림어광업의 임금이 3배 정도로 가장 느리게 상승하였다. 산업별 상대적 임금수준은 농림어광업이 1990년에 전 산업 대비 83%에서 2009년에는 59%로 하락한 반면, 제조업은 같은 기간 72%에서 114%로 상승하였다. 전기수도가스건설업과 서비스업은 1990년에 전 산업 대비 120%로 높았으나, 2009년에는 91%와 98%로 낮아졌다.

제조업 내부에서는 목재종이인쇄, 금속제품, 전기전자, 석유석탄화학, 섬유가죽 등은 1990년에 비해 2009년에 6배 이상의 상대적으로 빠른 임금상승을 보인 반면, 음식료품담배, 가구기타제조업, 비금속광물, 운송장비, 일반기계, 정밀기기 등은 상대적으로 느린 임금상승을 보였다. 한편 서비스업 내에서는 공공국방행정인 지난 20년 동안 10배로 가장 높은 임금상승을 보였으며, 다음으로는 부동산임대업, 문화오락, 사업서비스 등의 순으로 4배 이상 상승하였고, 도소매업, 교육서비스, 보건사회복지, 금융보험업, 정보통신업 등은 3배 이하의 상대적으로 낮은 임금상승을 보인 것으로 나타났다.

[그림 13-7] 산업별 임금변화 추이(1990=100)



〈표 13-9〉 27개 산업별 임금수준 및 연평균 증가율 추이

	상대적 임금수준(전 산업=100.0)					연평균증가율(%)			
	1990	1995	2000	2005	2009	91-95	96-00	01-05	06-09
농림어업	79.7	71.1	82.9	48.5	51.0	13.0	12.2	-5.8	3.3
광업	101.2	115.1	134.5	139.0	155.1	18.7	12.3	5.5	4.8
제조업	71.8	82.4	95.3	109.6	114.0	18.9	12.0	7.8	3.0
음식료품담배	62.5	69.1	86.8	84.4	85.5	18.0	13.9	4.3	2.3
섬유가죽	49.4	56.1	69.5	73.7	72.0	18.6	13.6	6.1	1.4
목재종이인쇄	45.0	91.0	84.1	92.9	91.5	33.1	7.1	6.9	1.6
석유석탄화학	88.1	98.0	119.9	127.8	132.2	18.1	13.3	6.2	2.9
비금속광물	88.5	96.3	107.2	111.0	113.8	17.6	11.2	5.6	2.6
금속제품	81.4	102.7	109.9	127.4	130.8	21.1	10.3	8.0	2.7
일반기계	96.1	93.2	93.4	110.5	120.3	14.9	8.9	8.4	4.2
전기전자	74.4	72.6	93.8	102.6	113.0	15.1	14.5	6.7	4.5
정밀기기	69.7	72.8	67.6	76.0	77.9	16.7	7.2	7.3	2.6
운송장비	105.7	110.1	117.5	144.7	135.3	16.6	10.2	9.3	0.3
기타제조업	73.8	58.7	70.6	91.0	98.5	10.5	12.9	10.3	4.0
전기가스수도	147.0	141.0	133.7	187.2	249.0	14.7	7.7	12.1	9.5
건설업	123.1	98.7	93.5	82.7	84.1	10.7	7.6	2.3	2.4
서비스업	121.6	112.6	103.2	100.2	97.9	13.9	6.9	4.2	1.4
도소매업	73.9	86.1	62.0	69.3	63.0	19.3	1.9	7.2	-0.4
음식숙박업	72.0	57.1	65.0	51.7	51.9	10.4	11.6	0.2	2.1
운수보관업	103.8	103.9	89.8	85.5	93.1	15.7	5.7	3.8	4.2
금융보험업	192.8	290.8	263.1	151.1	125.5	25.6	6.6	-6.2	-2.6
부동산임대업	93.8	155.9	200.5	142.8	117.7	28.0	14.4	-2.0	-2.8
정보통신업	253.7	49.4	59.7	63.9	51.3	-16.6	13.0	6.3	-3.5
사업서비스	170.3	183.6	180.6	180.5	178.0	17.4	8.4	4.8	1.6
공공행정국방	148.6	612.4	602.3	486.2	348.3	53.5	8.4	0.4	-6.2
교육서비스업	166.4	125.2	118.2	122.1	126.9	9.3	7.6	5.5	3.0
보건사회복지	80.7	47.6	40.7	49.7	55.3	4.1	5.5	9.1	4.8
문화오락	80.0	94.3	95.3	97.9	99.5	19.5	9.0	5.4	2.4
기타서비스	79.0	50.1	61.9	59.8	60.8	5.6	13.5	4.1	2.4
전 산업	100.0	100.0	100.0	100.0	100.0	15.6	8.8	4.8	2.0



## 결론 및 향후 발전방향

### 제1절 노동 및 임금통계의 확대

정부정책의 노동영향평가를 위해서는 고용창출의 양적 규모뿐만 아니라 창출되는 노동의 질적 측면도 중요한 평가 대상으로 삼아야 할 것이다. 주요 선진국의 노동모형 개발에서도 보듯이, 산업별 노동뿐만 아니라 직업별·성별·학력별 노동까지 고려하는 것이 최근의 추세다. 특히 정부가 추진하는 재정지출사업은 유형이 다양하여 노동의 양적 규모만으로 평가하기 어려운 측면이 있으며, 평가 과정에서 사업별 형평성의 문제도 제기될 수 있다. 따라서 이러한 점들을 고려할 때 향후 모형의 개발에서는 고용의 유형별 분석이 가능하도록 노동 및 임금통계의 양적·질적 확대가 필요하다.

고용의 유형별 분석을 위한 대안으로는 노동수요 측면에서 산업별·직업별·성별 등으로 노동과 임금관련 통계를 질적·양적으로 확대 및 세분화하는 방법이다. 이렇게 확대된 통계에 기초하여 노동 및 임금함수가 다양하게 추정될 경우 일자리 지원사업에 대한 평가가 좀더 정밀하게 분석될 것이며, 다양한 이해집단의 정책효과에 대한 이의제기도 조정이 가능할 것으로 판단된다. 한편 기초통계의 부족으로 장기 시계열 데이터의 구축이 어려운 경우에는 가능한 연도만을 대상으로 DB를 구축하고, 자유도의 부족으로 함수를

추정하는 것이 어려울 경우에는 미국의 BLS 등에서 활용하고 있는 산업별·직업별 행렬을 작성하여 산업별 노동수요를 직업별로 배분하는 방식을 취하는 것도 하나의 대안이 될 수 있다.

또 하나의 개선안으로는 노동수요함수의 도출에 기초가 되는 생산함수의 설정에서 노동의 질적 측면을 고려하기 위하여 물적자본(physical stock)인 자본스톡을 유형고정자본과 무형고정자본으로 분리하는 것과 마찬가지로 인적자본(human capital)으로서 노동을 숙련도에 따라 구분하여 생산함수의 설명변수로 추가하여 노동수요에서 질적인 측면이 고려되도록 하는 방안이 있을 수 있다.

## 제2절 모형과 DB 구축의 이원화

다부문 거시산업모형은 산업별·부문별을 동시에 고려하고 또한 동태적 연립방정식의 형태를 취하고 있기 때문에 모형의 안정성을 확보하고, 이에 기초하여 균형해를 도출하여 정책적 시뮬레이션을 수행하는 것은 그리 쉬운 작업이 아니다. 앞서서도 언급한 바와 같이 산업을 세분화하면 다양한 산업에 대한 정책적 영향 분석이 가능하다는 장점이 있으나, 다부문 거시산업모형의 특성상 산업분류의 세분화 정도에 따라 구축해야 하는 DB 크기와 추정해야 할 방정식 및 파라미터의 수가 기하급수적으로 증가하게 되므로 노력과 시간이 소요되며, 모형의 안정성도 낮아질 가능성이 높다.

따라서 본 연구에서는 27개 산업으로 분류하여 DB를 구축하였으나, 경우에 따라서는 정부 정책, 특히 일자리관련 사업의 효과를 세부적인 산업 수준에서 파악해야 할 가능성도 배제하기 어렵다. 또한 통계 DB는 특정한 목적보다는 다양한 측면에서 활용성을 높일 필요가 있다. 즉, 통계 DB의 가치는 활용의 빈도와 비례한다. 본 연구에서 구축된 DB는 일자리창출 지원사업의 효과 측면을 위한 다부문 거시산업모형을 개발하는 용도로 구축되었지만, 다른

용도로도 활용이 가능하다면 DB의 가치는 더욱 높아질 것이다.

따라서 통계 DB의 구축과 모형 내의 산업분류를 이원화하는 방안을 검토할 필요도 있다고 판단된다. 통계 DB는 모형을 구축하기 위한 기초적인 인프라이지만, 모형으로 분석하기에는 어려운 부분은 기초통계에 입각하여 특정 계수나 비중 등을 이용하여 축차적 분석을 할 수 있다. 즉, 산업부문과 거시경제 간 정합성이 보증되는 통계 DB의 구축에는 많은 시간과 비용이 초래되지만 한 번 구축해 놓으면 모형 개발 이외에도 다양한 목적으로 활용될 수 있다. 특히, 세부산업별 개별 모형의 구축이나 다양한 계수의 산출을 통해 산업거시모형에 의해 통합된 산업별 분석 결과를 세부산업별로 배분하는 축차적인 분석도 가능하다. 이러한 점을 고려하여 DB 구축은 가능한 세분화하고, 모형 내의 산업분류는 분석의 목적과 국내 산업구조의 변화를 감안하여 산업들은 성격이 유사한 산업 그룹으로 통합할 필요가 있다.

### 제3절 정부 정책의 유형화 및 모형 내 연계방안 마련

정부의 재정지출사업은 SOC 투자, 낙후지역 개발, 취약계층 지원, 산업단지 조성, 연구개발투자, 인력양성 등 사업 내용이 다양하며, 또한 지원 형태에 따라서도 보조, 출연, 융자 등으로 구분되고 있다. 이러한 다양한 형태의 재정지출사업이 고용에 미치는 영향을 정밀하게 평가하기 위해서는 무엇보다도 사업의 유형 분류가 매우 중요하다. 즉, 사업별로 인력양성 등과 같이 고용창출이 기본 목적인 사업도 있으나, 사업의 추진 과정에서 고용창출이 간접적으로 유발되는 사업도 있다. 따라서 이를 고용창출을 기준으로 혹은 관련 산업을 중심으로 유형 분류를 할 필요가 있다. 그리고 여기에 지원 형태를 고려해야 한다. 즉, 융자사업은 지원을 받는 대상이 다시 이를 되돌려주어야 하므로 보조금 사업과는 성격과 효과도 다르며, 평가 시 이를 고려해야 한다.

유형 분류 시 고려해야 할 또 하나의 사항은 모형 내에 이를 어떻게 연계

하는가이다. SOC 관련 사업의 경우는 통상적으로 건설업에 대한 수요 증가로 가정하여 건설업에 대한 정부지출 증가로 분석이 어느 정도 가능하며, 이와 유사하게 지역개발사업도 지출 내역에 따라 관련 산업에 대한 지출 증가로 분석할 수 있다.

그러나 연구개발 등은 그 효과가 복잡하므로 다양한 방안을 강구할 필요가 있다. 연구개발에 대한 정부투자는 기술혁신뿐만 아니라 산학연 협력을 통해 연구자 간의 네트워크 확대 등을 초래한다. 또한 연구개발은 그 효과가 장기에 걸쳐서 나타나는 특성이 있다. 이러한 효과를 모형 내에서 반영할 수 있는 방안이 강구되어야 한다. 이론적 모형에서는 이를 연구개발 부분에서 산출되는 생산물을 사용하는 산업에서 중간투입으로 발생하는 고정비용의 감소나 무형의 자본비용 감소 등을 유발하므로 이를 모형 내에 반영하도록 추천하고 있고, 생산함수에서 통상적으로는 외생적으로 취급하였던 기술진보에 의한 생산성 증대를 내생화하는 방안도 고려할 수 있는 하나의 방안이 될 수 있다.

또한 교육훈련이나 사회복지 등에 인적자본과 관련되는 정부지출사업은 분석에 많은 어려움을 초래한다. 기존 연구에서는 이처럼 비생산적 정부지출이면서 가계부문으로 직접 이전되는 경향이 있는 정부 사업은 노동의 숙련도 향상에 의한 생산성 증대로 분석할 수 있는 방안을 제시하고 있다. 사람에 대한 투자는 개인들의 학력이나 숙련도의 향상을 의미하며, 이는 인적자본의 축적이 생산 증대를 초래하므로 이러한 것을 모형 내에 반영하는 방안을 강구할 필요가 있다.

## 참고문헌

---

### [제1부]

- 강창희·유경준(2009), 「고용보험의 사업주 직업능력개발 지원사업이 기업의 훈련 투자결정에 미치는 영향」, 『한국경제의 분석』 15 (3), pp.209~253.
- 최강식(2007), 「고용영향 분석평가 방법론 연구」, 『직업능력개발원』 10 (3), pp. 181~202.

- Angrist, J.(1998), "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data On Military Applicants," *Econometrica* 66, pp.249~288.
- Angrist, J. and A. Krueger(1991), "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics* 106(4), pp. 979~1014.
- Angrist, J. and W. Evans(1998), "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review* 88(3), pp.450~477.
- Angrist, J. and Victor Lavy(1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics* 114(2), pp.533~575.
- Becker, S. O. and L. Woessmann(2009), "Was Weber Wrong? A Human Capital Theory of Protestant Economic History," *The Quarterly Journal of Economics* 124(2), pp.531~596.
- Bertrand, M. and S. Mullainathan(2004), "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *The American Economic Review* 94(4), pp.991~1013.
- Black, S. E.(1999), "Do Better Schools Matter? Parental Valuation Of Elementary Education," *Quarterly Journal of Economics* 114(2), pp.577~599.

- Card, David and Alan B. Krueger(1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review* 84(4), pp.772~793.
- Cohen-Zada, D. and W. Sander(2011), "Religious Participation versus Shopping: What Makes People Happier?," *Journal of Law and Economics* (forthcoming).
- DiNardo, J. and D. S. Lee(2004), "Economic Impacts of New Unionization on Private Sector Employers: 1984-2001," *Quarterly Journal of Economics* 119(4), pp.1383~1441.
- Drago, F., R. Galbiati, and P. Vertova(2009), "The Deterrent Effects of Prison: Evidence from a Natural Experiment," *Journal of Political Economy* 117(2), pp.257~280.
- Gould, E., V. Lavy, and D. M. Paserman(2004), "Immigrating to Opportunity: Estimating the Effect of School Quality Using a Natural Experiment on Ethiopians in Israel," *The Quarterly Journal of Economics* 119(2), pp. 489~526.
- Gruber, Jonathan and Daniel M. Hungerman(2008), "The Church versus the Mall: What Happens When Religion Faces Increased Secular Competition?," *The Quarterly Journal of Economics* 123(2), pp.831~862.
- Hanushek, E. A. and L. Woessmann(2006), "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries," *Economic Journal* 116(510), pp.C63~C76.
- Heckman, J., H. Ichimura and P. Todd(1997), "Matching as An Econometric Evaluation Estimator: Evidence From Evaluating A Job Training Programme," *Review of Economic Studies* 65, pp.605~654.
- Imbens, G. and J. Wooldridge(2009), "Recent Development in the Econometrics of Program Evaluation," *Journal of Economic Literature* 47(1), pp.5~86.
- Jensen, Robert(2007), "The Digital Divide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector," *The Quarterly Journal of Economics* 122(3), pp.879~924.
- Kang, Changhui(2011), "Family Size and Educational Investments in Children:

- Evidence from Private Tutoring Expenditures in South Korea,” *Oxford Bulletin of Economics and Statistics* 73(1), pp.59~78.
- Kang, Changhui, Cheolsung Park and Myoung-Jae Lee(2007), “Effects of Ability Mixing in High School on Adulthood Earnings: Quasi-Experimental Evidence from South Korea,” *Journal of Population Economics* 20(2), pp. 269~297.
- Kling, J. R., J. B. Liebman, and L. F. Katz(2007), “Experimental Analysis of Neighborhood Effects,” *Econometrica* 75(1), pp.83~119.
- Lee, David S.(2008), “Randomized Experiments from Non-random Selection in U.S. House Elections,” *Journal of Econometrics* 142(2), pp.675~697.
- Lee, S. and C. Kang(2011), “Labor Market Effects of School Ties: Evidence from Graduates of Leveled High Schools in South Korea”, Mimeo.
- Sacerdote, B.(2001), “Peer Effects With Random Assignment: Results For Dartmouth Roommates,” *The Quarterly Journal of Economics* 116(2), pp. 681~704.
- Stinebrickner, R. and T. R. Stinebrickner(2008), “The Causal Effect of Studying on Academic Performance,” The B.E. *Journal of Economic Analysis & Policy* 8(1), Berkeley Electronic Press, p.14.
- Stock, J. H., J. Wright and M. Yogo(2002), “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business & Economic Statistics* 20(4), pp.518~529.
- Wooldridge, J.(2008), *Introductory Econometrics : A Modern Approach*, South Western College.

## [제2부]

- 이진면 · 김동석 · 김민수(2001), 『다부문모형에 의한 산업구조 변화의 장기전망』, 한국개발연구원.
- 이진면 · 변창욱 · 최용재 · 김진웅 · 이상호(2007), 『KIET 산업경제계량모형』, 산업연구원.
- 한국은행(2007), 『산업연관분석 해설』, 한국은행.





◆ 執筆陣

- 윤윤규(한국노동연구원 연구위원)
- 홍민기(한국노동연구원 연구위원)
- 강창희(중앙대학교 교수)
- 이진면(산업연구원 연구위원)
- 이혜정(한국노동연구원 책임연구원)
- 고영우(한국노동연구원 책임연구원)

노동시장정책 평가방법론 및 다부문 거시산업모형 DB 구축

- |           |   |
|-----------|---|
| ▪ 발행연월일   | 2012년 2월 24일 인쇄<br>2012년 2월 29일 발행  |
| ▪ 발 행 인   | 김 승 태 원장직무대행  |
| ▪ 발 행 처   | <b>한국노동연구원</b><br>11510-0110 서울특별시 영등포구<br>은행로 30<br>☎ 대표 (02) 782-0141 Fax (02) 786-1862 |
| ▪ 조판 · 인쇄 | 창보문화사 (02) 2272-6997  |
| ▪ 등 록 일 자 | 1988년 9월 13일  |
| ▪ 등 록 번 호 | 제13-155호  |

© 한국노동연구원      정가 8,000원

ISBN 978-89-7356-901-4